Official Title: An Open-label, Multicohort, Phase II Study of Atezolizumab in Advanced Solid Tumors

NCT Number: NCT02458638

Document Date: SAP Version 4: 24-February-2020

STATISTICAL ANALYSIS PLAN

TITLE: **AN OPEN-LABEL, MULTICOHORT, PHASE II STUDY OF ATEZOLIZUMAB IN ADVANCED SOLID TUMORS**

PROTOCOL NUMBER: MO29518 (Version 8, 23-Oct-2019)

STUDY DRUG: Atezolizumab (RO5541267)

VERSION NUMBER: 4.0

IND NUMBER: 111271

EUDRACT NUMBER: 2015-000269-30

SPONSOR: F. Hoffmann-La Roche Ltd

AUTHORS: [redacted], Accovion (Version 1)
[redacted],
Clinipace/Accovion (Version 2)
[redacted], Roche (Version 3)
[redacted], Roche; [redacted], Cytel (Version 4)

DATE FINAL: 12 February 2020

SIGNATURE PAGE

Signature: [redacted] *Electronically signed by: [redacted] Reason: Approve Date: 2020-02-24 11:23:48+01:00*

Biostatistician [redacted] **Email:** [redacted]

F. Hoffmann-La Roche Ltd

Signature

24-Feb-2020

Date

Signature: [redacted] *Electronically signed by: [redacted] Reason: Approve Date: 2020-02-14 15:07:40+01:00*

[redacted] **Email:** [redacted]

F. Hoffmann-La Roche Ltd

[redacted]

Signature

14-Feb-2020

Date

STATISTICAL ANALYSIS PLAN HISTORY

SAP Version / Date	Description
Version 1 / 12 November 2015	Definition of Stage I analyses
Version 2 / 10 July 2017	• Overview of planned cohorts and sub-cohorts per protocol amendment 4 (section 2) • Rules added to impute, where possible, NPR endpoint for patients with no tumor assessment in the appropriate assessment window (section 4.4.1 for NPR at 18 weeks and section 4.4.2 for NPR at 24 weeks) • Description of Stage I analyses expanded • Expanded and clarified definition of eligibility and evaluability for all efficacy analysis populations (section 4.1.5) • Analyses for Stage II, and Final analysis added.
Version 3 / 20 April 2018	• Clarified exclusion from this SAP of antibody analysis, pharmacokinetic analysis as well as exploratory biomarker analysis • Removed modified RECIST analysis • Rescheduled Stage II analysis to be part of combined Month 12 analysis • Clarified analysis set definitions • Added waterfall plot • Clarified DOR will only be analyzed if >3 patients with CR/PR in the cohort • Simplified summary of study drug exposure (removed dose intensity and detailed summaries of delayed infusions) • Specified AESIs defined by Roche current standard (superseding eCRF tickbox) • Removed most listings (exceptions include deaths, SAEs and AESIs)
Version 4.0 / 12 February 2020	• Added more information about calculation of missed and delayed doses • Allowed for minor departures from 12 patients at Stage I analysis time • Removed Koyama and Chen adjustment of Stage II analysis of NPR (not directly applicable in case of departure from intended Stage I sample size) • Updated extended time window for NPR at 18 weeks (section 4.4.1): extended window to define imputed CR/PR/SD increased up to Day 209 (as for NPR at 24 weeks) • Analysis of modified RECIST added (BOR only) • PK and ADA analyses added • Biomarker (PD-L1) analysis added. • Clarifications added on the analysis of all cohorts combined

TABLE OF CONTENTS

1. BACKGROUND

This study will assess the efficacy and safety of the anti-PDL1 therapeutic antibody Atezolizumab for the treatment of various advanced solid tumors. The tumors included in this study are listed in Section 2 below. A brief description of their epidemiology and prognoses is presented in Table 1 of the Protocol (Section 1.1).

2. STUDY DESIGN

This is an open-label, multicenter, multinational, multi-cohort, phase II study. For each patient, the study will consist of a Screening Period (Day –35 to −1), a Treatment Period, and a Treatment Discontinuation Visit occurring $\leqslant$ 30 days after the last dose of study medication. Each cohort will be followed for survival until 2 years after the first dose for the final patient in the cohort. For further illustration please see Figure 1 on study schema in the protocol.

The study will include cohorts of patients with the following solid tumors (see Protocol version 8):

1) Cervical cancer
2) Nasopharyngeal carcinoma
3) Known high microsatellite instability (MSI-H) or mismatch repair (MMR) deficient colorectal cancer
4) Known BRCA 1/2 mutated cancers
    a) BRCA mutated ovarian cancer
    b) BRCA mutated breast cancer and that are not characterized as triple negative breast cancer
5) Soft tissue and visceral sarcoma
    a) Liposarcoma
    b) Leiomyosarcoma
    c) Gastrointestinal stromal tumor (GIST)
    d) Undifferentiated pleomorphic sarcoma and
    e) Known translocation related sarcomas
    f) Radiation induced sarcoma
    g) Osteosarcoma
    h) Chondrosarcoma
6) Mesothelioma
    a) Pleural mesothelioma

b) Peritoneal mesothelioma

7) Cholangiocarcinoma/cancer of the biliary tract
8) Thyroid cancer
    a) Anaplastic thyroid cancer
    b) Follicular or papillary thyroid cancer
    c) Medullary thyroid cancer and mixed medullary and follicular or papillary thyroid cancer
9) Gastric adenocarcinoma/adenocarcinoma of the gastro-esophageal junction (GEJ)
10) Other solid tumors – closed to recruitment from Protocol Version 4 onwards
11) Malignant germ cell tumors
12) ER+/HER2- metastatic breast cancer with known high mutation load (>100 mutations) by local test
13) Thymoma and thymic cancer
    a) Thymoma
    b) Thymic cancer
14) Gastroenteropancreatic (GEP) and lung neuroendocrine tumors (NETs)
    a) Low and intermediate grades (typical or atypical carcinoid)
    b) Poorly differentiated grade (excluding SCLC)
15) Known HPV induced squamous cell carcinoma
    a) Head and Neck
    b) Penile cancer
    c) Vaginal cancer/ vulvar cancer
    d) Anal cancer
16) Known MSI high or MMR deficient tumors (excluding colorectal and gastric cancers)

Enrollment in this study will be based on Simon's optimal two-stage design (Simon, 1989). The primary objective of this study is to evaluate investigator-determined non-progression rate (NPR) at 18 weeks in the individual cohorts.

2.1 OBJECTIVES

Efficacy Objectives

The primary efficacy objective for this study is as follows:

To evaluate non-progression rate (NPR) at 18 weeks in patients with advanced solid tumors treated with Atezolizumab, defined as the percentage of patients with complete response (CR) partial response (PR) or stable disease (SD) as assessed by the Investigator[1]. The secondary efficacy objectives for this study are as follows:

- To evaluate NPR at 24 weeks, overall response rate (ORR), best overall response (BOR), clinical benefit rate (CBR), duration of response (DOR), time to tumor progression (TTP) and progression-free survival (PFS), as assessed by the Investigator[1].

- To evaluate NPR at 18 and 24 weeks, ORR, BOR, CBR, DOR, TTP and PFS, as assessed by the Investigator using *modified RECIST*[2]

- To evaluate overall survival (OS)

Safety Objectives

The safety objectives for this study are as follows:

- To evaluate the safety and tolerability of Atezolizumab in patients with advanced solid tumors

- To characterize the immunogenic potential of Atezolizumab by measuring anti-Atezolizumab antibodies and to explore the potential relationship of the immunogenicity response with safety and efficacy

Pharmacokinetic Objectives

The pharmacokinetic (PK) objectives for this study are as follows:

---

[1] That is, using disease progression status as recorded in the clinical database. Investigator assessments are carried out according to Response Evaluation Criteria In Solid Tumors, Version 1.1 (RECIST, v1.1) (except for prostate cancer and malignant pleural mesothelioma) and disease-specific criteria for patients with prostate cancer (see Protocol Appendix 6) and malignant pleural mesothelioma (see Protocol Appendix 7).

[2] That is, using disease progression status as recorded in the clinical database, for Investigator assessments carried out according to *modified* RECIST criteria as defined in Protocol Appendix 3.

- To characterize the pharmacokinetics of Atezolizumab

Exploratory Objectives

The exploratory objectives for this study are as follows:

- To evaluate the relationship between tumor tissue PD-L1 expression and measures of efficacy, including NPR at 18 weeks and 24 weeks, ORR, BOR, CBR, DOR, TTP, PFS and OS

- To assess predictive and prognostic exploratory biomarkers(e.g. but not limited to protein and genetic markers on DNA and RNA) in archival and/or fresh tumor tissue and plasma and their association with disease status and/or response to study treatment

- To evaluate exploratory pharmacodynamic (PD) biomarkers (e.g., genetic markers, T, B and NK cell enumeration, T cell subpopulations like CD8+ T, effector/memory T cells, regulatory T cells, changes in expression of CD25 or human leukocyte antigen-DR [HLA-DR], interferon [IFN]-gamma production,IL-2 and other exploratory biomarkers) in tumor tissue, tumor microenvironment and plasma and their association with disease status, and/or response to study treatment, tumor immunobiology or tumor type

2.2 DETERMINATION OF SAMPLE SIZE

The sample size for this phase II study is based on Simon's optimal two-stage design (Simon, 1989).

In this Phase II study, an NPR of 20% is a level of activity that is not of interest for further clinical development, whereas an NPR of 40% is of clinical interest. The type I error will be 10% and the study will have 80% power to reject the null hypothesis when the true NPR is 40%.

The Simon design in this study requires 12 fully evaluable patients for first stage. If at the end of the first stage there are only 0, 1 or 2 patients with non-progressive disease at Week 18, the enrollment into this cohort will be terminated (or the cohort may be expanded as described below). Otherwise, if more than 2 patients with non-progressive disease are observed at the end of Stage I, an additional 13 fully-evaluable patients will be enrolled into Stage II. The study drug will be considered of clinical interest in this cohort if, at the end of the second stage, there are 8 or more patients with non-progressive disease out of 25 total fully-evaluable patients.

It can be assumed that some patients will not be evaluable for NPR. A patient will be considered evaluable for NPR if they received study drug, have a baseline tumor assessment and at least one tumor assessment post-baseline (per protocol mandated

timelines). Thus, more than 25 patients may be needed to be enrolled to obtain 25 fully-evaluable patients.

The hypotheses and clinical assumptions will be applied to all cohorts individually.

2.2.1 Definition of Cohorts and Extensions/Expansions

Statistical analyses will be based on patients with the same individual tumor type in the protocol-defined cohorts stated in section 2 above. The cohort for each patient will be identified from the data recorded on the “Cohort Assignment” CRF record.

If, during the conduct of the study, a specific subgroup of patients in a cohort is identified, the cohort may be modified to include only that specific subgroup, and then statistical analysis may be performed based on this subgroup of patients. Subgroups of cohorts will be identified by first selecting patients in the appropriate cohort(s), and then filtering for patients meeting the required subgroup criteria.

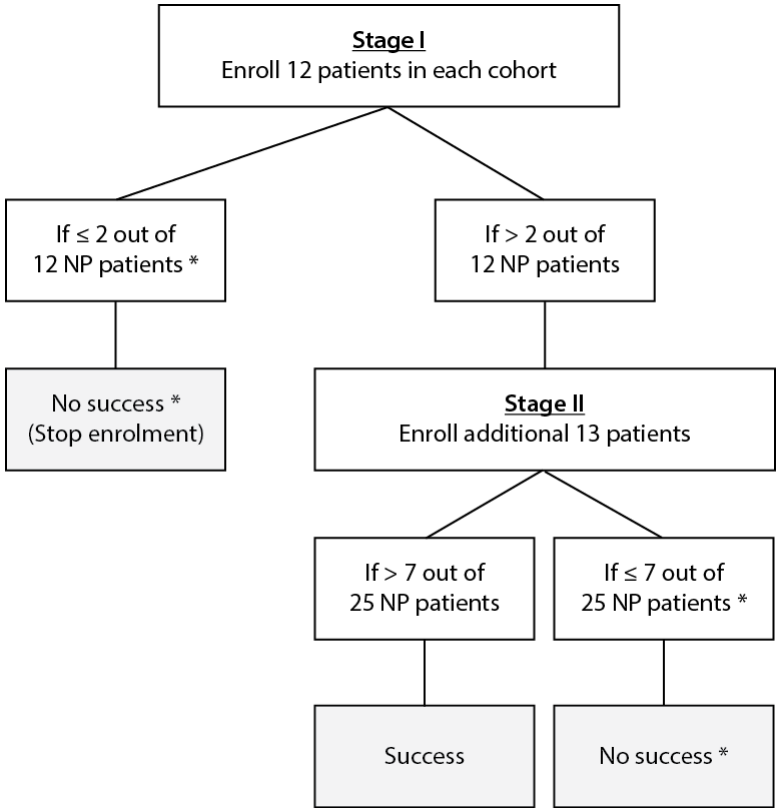
Additionally, a subgroup that applies across multiple tumor types (e.g. a biomarker positive subgroup) may be identified to be of interest.

Cohorts of these latter types will be analyzed using the same methods as for the protocol-defined cohorts.

Statistical analyses will be performed as well on all patients from all cohorts combined. The “All combined” cohort will be analyzed using the same methods as for the protocol-defined cohorts, but won’t include the following analyses: demographics, by-patient plots, time-to-event analyses, PK, immunogenicity and biomarkers. Some of the analyses will be performed only for “All combined” (as specified throughout the SAP).

The enrollment pattern across Stage I and Stage II is shown in Figure 1 below.

Figure 1 Sample Size for Each Cohort in Stage I/II



NP = non-progressors; NPR = non-progression rate.

Assumptions: H_0 = NPR 20%; H_1 = NPR 40%; power = 80%; alpha = 10%.

The sample size was estimated using Simon's optimal two-stage design (Simon 1989).

* And no additional clinical benefit in a subgroup.

2.3 ANALYSIS TIMING

Each cohort will be analyzed independently at the following time points:

1. Stage I: when the first 12 eligible and fully evaluable patients of the cohort have passed the week 18 assessment. This analysis will focus on the primary efficacy endpoint (NPR) and a targeted subset of Safety analyses (as agreed with the Steering Committee as being sufficient for the purpose of making a decision whether or not to continue to Stage 2).

2. Month 12: for cohorts that progress to Stage 2, there may be an analysis of the majority of outcomes at Month 12 (when all patients of the cohort have been on the study for at least 12 months, have prematurely discontinued from the study or died). Such analyses are considered informative for potential future trials and early publication of individual cohort first results, but are not considered final.

3. Final analysis: the definitive Final Analysis, on which the CSR will be based, will take place following final database lock, which will be when all patients satisfy at least one of the following conditions: died, withdrawn consent or lost to follow up, or has been followed for survival for a minimum of 24 months after the last patient was enrolled, whichever occurs first. In the Final Analysis, all cohorts will be analysed both separately and combined.

3. STUDY CONDUCT

Stopping and continuation rules for Stage I:

Enrolment into the cohort will stop at the end of Stage I if the number of patients with non-progressive disease (confirmed or unconfirmed) is 2, 1 or 0 out of 12 patients. However, if the overall patient population does not meet requirements at Stage I, but a clinical benefit is observed in a specific subgroup (e.g. biomarker positive subgroup) of patients, then only this subgroup may be added for further evaluation at Stage I and Stage II.

Decisions at the end of Stage I (termination, continuation or continuation in a subgroup) will be made by the Sponsor in discussion with the study Steering Committee members.

Rules for Stage II:

A study treatment will be considered to be promising for a given cohort if:

- there is no unacceptable toxicity , and
- at 18 weeks, there are 8 or more patients with non-progressive disease out of 25 fully-evaluable patients.
- if for logistical reasons there are not exactly 25 fully-evaluable patients in a cohort, then a p-value will be calculated allowing for the departure from the planned Simon's design, using the method of Koyama and Chen (2008). By analogy with the original design being based on alpha=0.1, a treatment will be considered to be promising if $p < 0.1$. This p-value can be mapped to a new cutoff point in the manner of the original design calculation. Some values for the cutoff as a function of sample size are: 7 or more out of 21 to 24; 8 or more out of 25; 9 or more out of 26 to 29; 10 or more out of 30.

If the overall patient population does not meet the Stage II criteria, but a clinical benefit is observed in a specific subgroup (e.g. biomarker positive subgroup) of patients, then

Stage II may be expanded to include additional patients from this specific subgroup in order to re-evaluate the Stage II criteria in this patient subgroup.

In all such cases where a “new” cohort is defined based on partial results from the planned cohorts, the assumptions for Simon’s design and formal hypothesis testing will not hold. Therefore, only the standard summary statistics and confidence intervals will be presented. As an informal guideline, we will consider 8 or more NPR out of 25 patients as a promising result, but to be interpreted with suitable caution.

4. STATISTICAL METHODS

Categorical data (such as discontinuation reasons or adverse events) will be summarized using frequencies and percentages (including a category for missing, if appropriate). Continuous endpoints (such as durations or concentrations) will be summarized using descriptive statistics (N, mean, standard deviation, median, lower quartile, upper quartile, minimum and maximum). Other analysis methods will be specified (below) where applicable.

Study Day 1 will be defined as the first day a patient receives study medication. Baseline values are determined by the last measurement up to and including date of treatment start (Study Day 1). For statistical evaluations (e.g. descriptive summaries) of post-baseline data summarised by visit, or by visit and timepoint, only measurements at scheduled visits or timepoints are considered. If a measurement is repeated at a scheduled visit or time point, then only the latest retest result will be used in calculation of the statistical summary. All data – including from unscheduled measurements – will be included in the listings, and used in the calculation of minimum or maximum value over the course of the study, and in the potential identification of adverse events.

Any cohort that does not reach the Stage I analysis (i.e. does not reach 12 eligible and evaluable patients), will not be presented for statistical analysis per se. (This does not preclude that some or all patients within such cohorts may be included in cross-cohort analyses.)

4.1 ANALYSIS SETS

4.1.1 Intention-to-treat Set

For each cohort, the intention-to-treat (ITT) set will include all patients enrolled in the study irrespective of whether they have received study medication or not. In practice during the course of this study, no patient has been enrolled and not received treatment. Therefore ITT and Safety sets are the same, and we label this as Safety Set in all outputs.

4.1.2 Safety Set

For each cohort, the Safety Set will include all patients who have received at least one dose of study medication. All tables (demographics, efficacy and safety) will include a summary by the Safety Set.

4.1.3 Analysis sets for efficacy analyses

The Protocol mandates performing the efficacy analysis on "fully evaluable" patients. We clarify below how this will be interpreted in detail, and subsequently refer to this Efficacy Set as "eligible and evaluable".

4.1.3.1 Eligibility and evaluability

All patients will be considered **eligible** for the cohort identified from the data recorded on the "Cohort Assignment" CRF unless they have a protocol deviation that compromises their efficacy endpoint, as identified in the current version of the Protocol Deviation Plan at the time of the analysis.

A patient will be considered **evaluable** for NPR if they received study drug, have a baseline tumor assessment and at least one tumor assessment post-baseline (per protocol mandated timelines). Specifically, if a single assessment has been performed, this assessment must be performed between days 18 and 66 (inclusive) post Day 1 of study treatment. (This latter restriction does not apply in case there are two or more assessments.)

4.1.3.2 Stage I Efficacy Set

For each cohort, the Stage I Efficacy analysis set is by intention 12 eligible and evaluable patients having reached their Week 18 visit (or to have had their final study data collection before Week 18, e.g. due to death or other reason for study withdrawal). In some cases, due to late reclassification of protocol deviations, there may in practice be a small number fewer or greater than 12 for Stage 1 analysis.

4.1.3.3 Stage II (and Subsequent) Efficacy Set

For each cohort, the Stage II Efficacy analysis set will include all of the eligible and evaluable patients in the corresponding safety population, once all patients in the cohort have reached Week 18 (or terminated the study). Patients will be considered eligible and evaluable for Stage II, independently of eligibility and evaluability assessed at the time of Stage I analysis.

The same analysis set will be used for all subsequent efficacy analyses.

4.2 SUMMARIES OF STUDY CONDUCT

Enrollment, reasons for treatment discontinuation, reasons for study discontinuation and duration of follow up will be summarized for both Safety and Efficacy Sets. For the analysis of "All combined", a summary of enrollment by protocol-defined cohort will be provided.

Major protocol deviations, as identified in the current version of the Protocol Deviation Plan at the time of the analysis, will be listed (and summarized for “All combined” only).

4.3 SUMMARIES OF TREATMENT GROUP COMPARABILITY

This is a single arm study and, consequently, no formal treatment group comparisons will be conducted. Demographics (age, sex, ethnicity, and race) and disease history will be summarized. Medical history, prior and concomitant therapies data will be summarized for “All combined” only.

4.4 EFFICACY ANALYSIS

The majority of efficacy analyses will be performed on both Efficacy and Safety Sets. Exceptions are: waterfall plots, PFS, DOR and TTP analyses, which are only presented for efficacy analysis sets by convention.

4.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint for this study is non progression rate (NPR) at 18 weeks.

Week 18 occurs at Study Day 127. Allowing 15 days either side of the exact Week 18 study day, assessments that occur between Study Day 112 and Study Day 142 (inclusive) will be considered as a Week 18 assessment.

Individual patient response at 18 weeks, based on RECIST v1.1, will be categorized as “CR”, “CR (imputed)”, “PR”, “PR (imputed)”, “SD”, “SD (imputed)”, “PD”, ”Death in absence of PD”, “Discontinuation in absence of PD” or ”Missing” as follows:

1. any patient who recorded PD (as determined by the Investigator) at any post-baseline tumor assessment up to Study Day 142 inclusive, will be classified as “PD”,

2. any remaining patient who died at or prior to Study Day 142 will be classified as “Death in the absence of PD”,

3. any remaining patient will be classified as “CR”, “PR”, or “SD” according to their Investigator-reported tumor assessment data that falls within the window of Study Day 112 – 142 inclusive (taking the worst case if there are two conflicting assessments within the window)

4. for a patient who does not meet any of the above criteria, but who has a nearby later assessment (up to Day 209), the patient will be classified as “CR (imputed)”, “PR (imputed)” or “SD (imputed)” according to the first tumor assessment in the extended window.

5. any remaining patient will be classified as “Discontinuation in absence of PD” if they have an End of Study date prior to Study Day 142 or, otherwise, as “Missing/Out of window”.

Non progression rate (NPR) at 18 weeks is defined as the percentage of patients with "CR", "CR (imputed)", "PR", "PR (imputed)", "SD" or "SD (imputed)" at 18 weeks. The denominator is all patients in the analysis (thus in this calculation, Death or Discontinuation in the absence of PD, and Missing/Out of window, are, conservatively, effectively treated as PD).

4.4.2 Analysis of the Primary Efficacy Endpoint

The number and percentage of progression-free patients (NPR) at 18 weeks (Stage II) will be presented along with standard 95% (exact binomial, two-sided) confidence intervals for proportions.

4.4.3 Secondary Efficacy Endpoints

Non progression rate (NPR) at 24 weeks is defined as the percentage of patients with "CR", "CR (imputed)", "PR", "PR (imputed)", "SD" or "SD (imputed)" at 24 weeks.

Week 24 occurs at Study Day 169. Allowing 20 days either side of the exact Week 24 study day, assessments that occur between Study Day 149 and Study Day 189 (inclusive) will be considered as a Week 24 assessment.

Individual patient response at 24 weeks, based on RECIST v1.1, will be categorized as "CR", "CR (imputed)", "PR", "PR (imputed)", "SD", "SD (imputed)", "PD", "Death in absence of PD", "Discontinuation in absence of PD" or "Missing" as follows:

1. any patient who recorded PD at any post-baseline tumor assessment up to Study Day 189 inclusive, will be classified as "PD",

2. any remaining patient who died at or prior to Study Day 189 will be classified as "Death in the absence of PD",

3. any remaining patient will be classified as "CR", "PR", or "SD" according to their Investigator-reported tumor assessment data that falls within the window of Study Day 149 – 189 inclusive (taking the worst case if there are two conflicting assessments within the window),

4. for a patient who does not meet any of the above criteria, but who has a nearby later assessment (up to Day 209), the patient will be classified as "CR (imputed)", "PR (imputed)" or "SD (imputed)" according to the first tumor assessment in the extended window.

5. any remaining patient will be classified as “Discontinuation in absence of PD” if they have an End of Study date prior to Study Day 189 or, otherwise, as “Missing/Out of window”.

Best overall response (BOR), based on RECIST v1.1, for an individual patient, is defined as the best overall response obtained until progressive disease, death or lost to follow up, whichever is the first, as follows:

1. CR for a patient with an overall tumour response assessment of CR at two consecutive visits at least 28 days apart. If applicable, any tumour assessment within the 28 day period must also record CR.

2. PR for a patient with an overall tumour response assessment of PR or CR at two consecutive visits at least 28 days apart without being a CR. If applicable, any tumour assessment within the 28 day period must also record PR or CR.

3. SD, for a patient with an overall tumour response assessment of SD, PR, or CR at one or more visits at least 42 days after start of study treatment, but are not a confirmed CR or PR.

4. PD, for a patient with an overall tumour response assessment of PD at any visit, and does not meet the criteria for a BOR of CR, PR or SD.

5. Missing, for a patient with an assessment of SD, PR or CR in the first 42 days after start of study treatment and no further tumour assessments thereafter.

Best overall response will summarize the proportion of patients with each best overall response.

Overall response rate (ORR) is defined as the proportion of patients with a BOR of CR or PR.

Clinical benefit rate (CBR) is defined as the proportion of patients with a BOR of CR, PR or SD.

Duration of response (DOR), based on RECIST v1.1, is defined as the time from the first occurrence of a documented objective response (CR or PR) to the time of progression or death from any cause, whichever occurs first. For patients who do not die or experience disease progression before the end of the study or who are lost to follow-up, duration of objective response will be censored at the day of the last tumor

assessment. Note: DOR will not be analysed if there are less than 4 patients available for the analysis.

Progression-free survival (PFS), based on RECIST v1.1, is defined as the time from the first day of study treatment to the first occurrence of disease progression or death from any cause, whichever occurs first. A patient without a PFS event will be censored at the time of the last evaluable tumor assessment or, for patients with no tumor assessment after the baseline visit, at the time of the first day of study treatment plus 1 day.

Time to progression (TTP), based on RECIST v1.1, is defined as time from the first day of study treatment to the first occurrence of progressive disease or death due to disease progression, whichever occurs first. Patients who have not progressed or died due to disease progression at the time of study completion or who are lost to follow-up will be censored at the date of the last evaluable tumor assessment or, for patients with no tumor assessment after the baseline visit, at the time of the first day of study treatment plus 1 day.

Overall survival (OS) is defined as the time from the first day of study treatment to death from any cause. Patients who are still alive at the time of analysis (clinical cut-off) and patients who are lost to follow-up will be censored at the last date they were known to be alive.

Modified best overall response (mBOR), based on modified RECIST, for an individual patient, is defined as the best overall response obtained until death or lost to follow up, whichever is the first. The mBOR categories are obtained in the same manner as for the BOR based on RECIST v1.1, as described above.

Modified best overall response will summarize the proportion of patients with each modified best overall response.

4.4.4 Analysis of the Secondary Efficacy Endpoints

4.4.4.1 Analysis of NPR, BOR, ORR and CBR

The secondary endpoints NPR at 24 weeks, BOR, ORR and CBR based on RECIST v1.1, as well as mBOR based on modified RECIST, will be analyzed using standard 95% (exact binomial, two-sided) confidence intervals for proportions.

4.4.4.2 Response and lesion assessments

A swimmer plot will be provided to illustrate the response assessments over time, and a spider plot will be presented to show the change in sum of target tumor diameters (SLD) over time. A waterfall plot will show best SLD change from baseline.

4.4.4.3 Time to event analyses

The time-to-event variables PFS, OS, TTP and DOR will be presented graphically using the Kaplan-Meier (KM) approach. The median survival time and its 95% confidence interval will be presented. Additionally, the estimated survival rate together with its associated 95% confidence interval at 24 weeks, at 6 and 12 months, and at 24 months (for OS only) will be presented.

Duration of response will be evaluated only for patients who achieve CR or PR (and only analysed if there are more than 3 such patients in a cohort).

4.5 PHARMACOKINETIC (PK) ANALYSIS

The PK analysis will be performed on the safety population. Atezolizumab serum concentration data will be summarized descriptively, for each cycle and timepoint at which PK is measured. Descriptive statistics will include the numbers of patients below the limit of quantification (BLQ), mean, SD, median, minimum and maximum.

4.6 IMMUNOGENICITY ANALYSIS

The immunogenicity analysis will be performed on the safety population.

The anti-atezolizumab antibodies (ADAs) status at baseline will be summarized.

In addition, the incidence of treatment-emergent ADAs will be summarized overall, for all patients with an ADA result from at least one post-baseline sample.

The overall ADAs status will be reported in 4 categories:

- Negative
  - patient with negative or missing ADA result at baseline and always negative post-baseline

- Negative (Treatment unaffected)
  - patient with positive ADA at baseline and always negative post-baseline
  - patient with positive ADA at baseline and treatment unaffected, i.e. with all post-baseline titer results <0.6 compared compared to baseline titer

- Positive (Treatment enhanced)
  - patient with positive ADA and treatment enhanced, defined as at least one post-baseline titer result ≥0.6 compared to patient's baseline titer

- Positive (Treatment induced)
  - patient with negative or missing ADA result at baseline and at least one positive post-baseline

A patient is considered positive if a titer value > 0 is reported.

4.7 SAFETY ANALYSES

Safety will be assessed through summaries of AEs, exposure to Atezolizumab, changes in laboratory test results as well as changes in vital signs and ECGs. The safety variables will be summarized for the safety population. Unless stated otherwise, all safety variables will be summarized for each cohort separately.

4.7.1 Exposure to Study Medication

Number of infusions will be summarized by descriptive statistics. Time on treatment (last dose date minus first dose date, plus 1, expressed in weeks), and total dose received will be summarized.
Missed treatment cycles are not recorded as such in the eCRF (sequential visits are labelled with sequential cycle number regardless of length of time between visits). Therefore to capture "missed doses" and delayed dose irregularities we categorize

- *Patients with one or more missed cycle* as being patients who have >=42 days between two consecutively labelled cycles (target cycle time being 21 days)

- *Patients with at least one cycle delayed by 10 days or more* (i.e. >=31 days between two consecutively labelled cycles). Patients with one or more missed cycle as just defined are also included in this category.

We also present total patient years of exposure to Atezolizumab (per cohort, and across the study) calculated as the sum of invididual time on treatment (last day of treatment minus first day of treatment plus one day), expressed in years (i.e. days/365.25).

4.7.2 Adverse Events

Adverse events (AE) will be coded using MedDRA.

AEs will be graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE), Version 4.0. For events of varying severity, the most severe grade as documented on the eCRF will be used in the summaries.

The analysis will focus on treatment-emergent AEs – AEs which start on or after the date of first dose of study drug. (For the assignment of 'treatment-emergent' in case of incomplete dates being recorded on the CRF, see Section 4.8.)

AEs of Special Interest are defined according to current Roche global guidelines, and are identified in the clinical database. (This classification overrides the flag set in the eCRF by the Investigator.)

An overview that summarizes the number of patients with treatment-emergent AEs (and the number of events) will be provided including all AEs, related AEs, AEs of Special Interest (AESI), Serious AEs (SAEs), related SAEs, SAEs leading to death, related SAEs leading to death and AEs leading to treatment discontinuation. The number of patients experiencing AEs at each grade (by highest grade) and related AEs at each grade (by highest grade) will also be included.

The number and frequency of patients with treatment-emergent AEs will be presented by primary System Organ Class (SOC) and Preferred Term (PT) within SOC, both overall and for highest grade within patient, for the following:

- All AEs

- Related AEs [“All combined” only]
- AESIs (by AESI class and PT rather than by SOC and PT) [“All combined” only]
- SAEs [“All combined” only]
- Related SAEs [“All combined” only]
- SAEs leading to death [“All combined” only]
- Related SAEs leading to death [“All combined” only]

In addition, a summary of non-serious AEs occurring in 5% or more of patients will be provided (for regulatory disclosure purposes), for “All combined” only.

Patient listings will be provided for:

- AEs leading to treatment discontinuation or interruption
- AESIs
- SAEs
- All deaths.

4.7.3 Laboratory Data

Laboratory data will be analysis for “All combined” only. Hematology and chemistry laboratory data will be summarized in tables presenting the shift from baseline grade to worst NCI CTCAE grade recorded on treatment. Laboratory data recorded at an unscheduled visit (if on treatment) will be included in the calculation of worst grade for the shift tables.

4.7.4 Vital Signs and ECG

Vital signs and ECG data will be summarized by visit (for vital signs, by timepoint within visit) and presented together with change from baseline, for “All combined” only.

4.8 BIOMARKER ANALYSIS

Status of PD-L1 immune cells (IC) and tumor cells (TC) expression will be summarized at baseline, on both efficacy and safety populations. The following categories will be used:

- IC
  - IC0: <1%
  - IC1: 1%-<5%
  - IC2: 5%-<10%
  - IC3: ⩾10%
- TC
  - TC0: <1%
  - TC1: 1%-<5%

- TC2: 5%-<50%
- TC3: ⩾50%

In addition, box plots displaying the baseline PD-L1 continuous IC and TC values will be provided by NPR status at Week 18 (Yes/No), for the efficacy population.

4.9 MISSING DATA

Missing data in the non-progression rate, for the eligible-and-evaluable efficacy analyses, will be minimized by extending recruitment to allow for patients who do not meet the definitions of eligible and evaluable. Further imputation rules are described in section 4.4.1 for NPR at Week 18 and section 4.4.3 for NPR at Week 24.

Time to event endpoints will implement the censoring rules described in Section 4.4.3.

For the flagging of an adverse event or concomitant medication with partial start and/or end date as being on treatment: the missing date is not imputed per se but it is calculated whether the event or medication could *possibly* have started or been taken while on treatment, in which case it is taken as such.

5. REFERENCES

Simon, R. Optimal two-stage designs for phase II clinical trials. Control Clin Trials 1989; 10:1-10.