NOVARTIS

Clinical Development

ACZ885/Canakinumab

CACZ885U2301 / NCT03631199

A randomized, double-blind, placebo-controlled, Phase III study evaluating the efficacy and safety of pembrolizumab plus platinum-based doublet chemotherapy with or without canakinumab as first line therapy for locally advanced or metastatic non-squamous and squamous non-small cell lung cancer subjects (CANOPY-1)

Statistical Analysis Plan (SAP)

Author:	Trial statistician, [redacted]

Document type:	SAP Amendment 1.0
Document status:	Final
Release date:	03-Jul-2020
Number of pages:	87

Document History

Date/Version	Section	Changes
18-Dec-2018/ Final		N/A - First version
3-Jul-2020/ Amendment 1	Section 1	To update to latest protocol amendment version 3.0 prior to primary PFS analysis.
	Section 1.1.2	Updated wording "final analysis" to "primary analysis" for PFS as PFS can be re-tested with alpha 0.025 if not statistically significant at primary analysis given OS is statistically significant at one of the interim or final analyses. This update applied throughout the document Removed approximate timing of primary PFS analysis as it was no longer accurate given faster than expected enrollment rate.
	Section 2.1.1	To clarify data collected after withdrawal of consent will not be used: data collected after patients' withdrawal of informed consent for further participation in the study will not be reported (except for death date, which might be obtained from public records).
	Section 2.1.3.1	Updated wording: "paclitaxel, nab-paclitaxel" to "paclitaxel or nab-paclitaxel"; "pembrolizumab with induction chemotherapy" to "pembrolizumab and induction chemotherapy".
	Section 2.1.3.6	To clarify the dosing schedule and definition of last date of exposure to study drug for nab-paclitaxel.
	Section 2.1.3.10	Updated wording: Table 2-1 footnote [a]: remove first "for PRO".
	Section 2.2.3.1	CCI [redacted] clarified randomization factors will be both IRT and CRF; [redacted]
	Section 2.2.3.2	To update smoking history subgroup from former/current vs. never to former vs. current vs. never for safety subgroup analyses.
	Section 2.2.3.3	To support Japan submission: removed the requirement of minimum number of subjects in Japan for conduct of Japanese subgroup analyses; clarified an un-stratified Cox model will be used if number of Japanese subject is <20.
	Section 2.3.1.8	In light of COVID-19 pandemic, additional analysis of protocol deviations related to COVID-19 pandemic were added. COVID-19 related PD vs. corresponding relationship summary and COVID-19 related outcomes summary were added based on health authority feedback
	Section 2.4.1.1	Updated wording: "pemetrexed being given" to "Pemetrexed will be given".

Date/Version	Section	Changes
	Section 2.4.1.1.1	Hyperlink correction: updated “as outlined in section 2.1.2” to “as outlined in section 2.1.3”.
	Section 2.4.1.1.2	Updated wording: “canakinumab (placebo) and pembrolizuamb” to “canakinumab/placebo and pembrolizumab”.
	Section 2.4.1.1.3	To update the methods used for the calculation of actual dose intensity: New calculation can avoid the impact of early drop–out or death, which can lead to higher ADI and RDI.
	Section 2.4.1.1.4	To clarify duration of interruption will be calculated for each interruption if a subject has multiple interruptions by adding wording “for each observed interruption”.
	Section 2.5.1.2	Typo correction: “statisfies” to “satisfies”.
	Section 2.5.2.1	To update the estimand language and align with the ICH E9 addendum: Treatment was added as new attribute and handling of intercurrent event of start of new ANP was incorporated into treatment attribute. Missing tumor assessment was removed from intercurrent event and moved to missing data section. Clarified that intercurrent events due to COVID-19 will be handled by treatment policy strategy. Discontinuation of study due to lost to follow-up or withdrawal of consent was also removed from the intercurrent event for OS.
	Section 2.5.2.3	Handling of missing tumor assessment moved to this section as this was not an intercurrent event but missing data problem: Handling of missing tumor assessment was added in this section.
	Section 2.5.2.5	Clarify for PFS BIRC analysis, no p value will be provided given BIRC was audit based.
	Section 2.5.2.5.1 and 2.5.2.5.2	Estimand languages were refined for sensitivity and supplementary analyses. In addition, second sensitivity analysis estimand of covariate adjusted cox model was moved from sensitivity estimand section 2.5.2.5.1 to supplementary estimand section 2.5.2.5.2, as it answered a different question and implied a different estimand; fifth and sixth supplementary analyses estimands of actual events and backdating approach were moved from supplementary estimand section 2.5.2.5.2 to sensitivity estimand section 2.5.2.5.1 as missing tumor assessment is a missing data problem and associates with underlying model imputation assumption; per-protocol set (PPS) supplementary analysis was removed, as PPS is not recommended under estimand framework. Additional supplementary analyses were added to assess the impact of COVID-19 on PFS and OS. Hypothetical strategy was used to handle intercurrent events due to COVID-19. Definition of new anticancer therapy was clarified for special exceptions for PFS analysis. The same definition was also aligned and updated for BOR (Section 2.6.1.1), time to event for AE and

Date/Version	Section	Changes
		lab (Section 2.8.1.2; Section 2.8.1.4), and PRO (Section 2.12) analyses.
	Section 2.5.2.5.3	Unknown (UNK) response updated to Not evaluable (NE) based on latest RECIST macro. This update applied throughout the document
	Section 2.5.2.5.4	China subgroup was added for primary PFS/OS analyses to support China submission.
	Section 2.5.2.5.3	Correction on the hyperlink to Section 5.4.2.
	[redacted]	[redacted]
	Section 2.8.1.1	Removed one redundant "having at least one"; clarified in overview of adverse events that AE leading to each component of study treatment will also be summarized.
	Section 2.8.1.2	AESI term 'DILI (Hepatic transaminase and bilirubin elevations)' was updated to 'Abnormal liver parameters' based on latest eCRS. These updates were also made in any other applicable sections.
	Section 2.8.1.4	To clarify in the by visit analysis, for a subject with multiple assessments in a time window, the average value will be used; Analyses for liver function parameters were updated based on the newly released internal hepatic values guidance dated 28-Feb-2020.
	Section 2.8.1.5.2	Clarified the clinical notable criteria for weight and added for respiratory rate in Table 2.8.
	Section 2.9.1	PK parameter analyses description added for randomized part (Detailed contents was moved from SRI part. SRI part then referred to randomized part).
	Section 2.11	Clarified if 2 assessments within a time window are equidistant from the target date, the assessment obtained prior to visit will be considered.
	Section 2.12	Sensitivity analyses of time to definitive deterioration with 5 points were removed as they were the same as 10 points analyses for the defined symptom endpoints.
	[redacted]	[redacted]

Date/Version	Section	Changes
	[redacted]	[redacted]
	Section 5.1.3	Typo correction: “january” to “January”.
	Section 6	Four new references added.

Table of contents

List of abbreviations

ADA	Anti-drug antibodies
AE	Adverse event
AESI	Adverse events of special interest
ATC	Anatomical Therapeutic Classification
AUC	Area Under the Curve
BLRM	Bayesian Logistic Regression Model
BIRC	Blinded Independent Review Committee
BMI	Body mass index
BSA	Body surface area
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DCR	Disease control rate
DDS	Dose determining set
DI	Dose intensity
DLT	Dose Limiting Toxicity
DLRT	Dose Level Review Team
DMC	Data Monitoring Committee
DOR	Duration of response
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
FAS	Full Analysis Set
IA	Interim analysis
[redacted]	[redacted]
MedDRA	Medical Dictionary for Drug Regulatory Affairs
NCI	National Cancer Institute
NE	Not evaluable
NSCLC	Non-small cell lung cancer
ORR	Overall response rate
OS	Overall Survival
PAS	Pharmacokinetic analysis set
PDI	Planned dose intensity
PFS	Progression-Free Survival
PK	Pharmacokinetics
PRO	Patient-reported Outcomes
PT	Preferred term
QoL	Quality of Life
RDI	Relative dose intensity
RECIST	Response Evaluation Criteria in Solid Tumors
RP3R	Recommended Phase 3 dose Regimen
SAP	Statistical Analysis Plan
SOC	System Organ Class
TA	Tumor assessment

TFLs	Tables, Figures, Listings
TTR	Time to response
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes the planned analyses for the primary Clinical Study Report (CSR) of the study CACZ885U2301, a randomized, double-blind, placebo-controlled, phase III study evaluating the efficacy and safety of pembrolizumab plus platinum-based doublet chemotherapy with or without canakinumab as first line therapy for locally advanced or metastatic non-squamous and squamous non-small cell lung cancer subjects.

As specified in the section 12-7 of the study protocol, the primary analysis of PFS will be performed when approximately 253 PFS events have been observed. The 1st interim OS analysis will be done at that time, where 120 OS events are expected. In case PFS and/or OS is declared statistically significant at the time of the primary PFS analysis, the primary PFS analysis and 1st IA OS analysis will constitute the basis of the primary CSR. In case none of these primary endpoints are found statistically significant at time of the primary PFS, the next OS analysis (at 2nd IA or final OS), which meet statistically significance, will be considered as the primary analysis and will constitute the basis of the primary CSR. In the primary CSR, the safety run-in data will also be reported.

The content of this SAP is based on protocol amendment CACZ885U2301 version v03. All decisions regarding the analysis, as defined in the SAP document, have been made prior to database lock and unblinding of the study data.

1.1 Study design

This is a double-blind, randomized, multicenter Phase III study evaluating the efficacy and safety of canakinumab vs. canakinumab matching-placebo in combination with pembrolizumab in addition to 4 cycles of platinum-based doublet induction chemotherapy, followed by maintenance therapy in subjects with AJCC v. 8 stage IIIB/IIIC (not eligible for definite chemoradiation therapy) or stage IV (metastatic) NSCLC regardless of PD-L1 levels and histology (squamous and non-squamous). Subjects who completed 4 cycles of induction treatment and fulfill eligibility criteria as defined in Section 6.1.3 of the protocol, will enter the maintenance treatment phase.

This study has two parts: Safety run-in part and randomized part, which are defined below.

1.1.1 Safety run-in part (part 1)

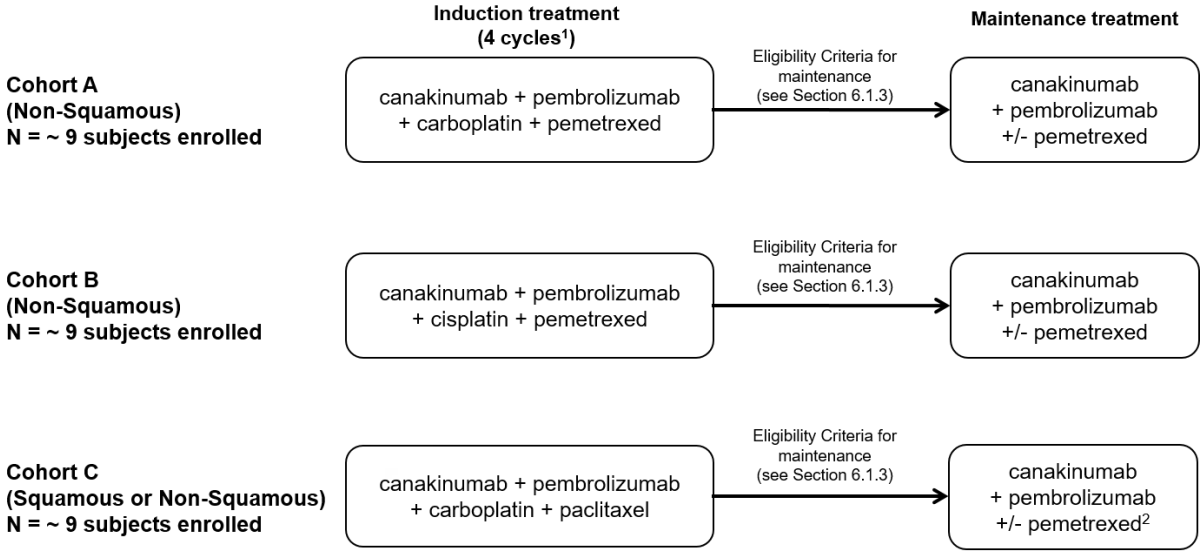
The safety run-in part will determine the recommended regimen of canakinumab (recommended Phase 3 dose regimen, RP3R) in combination with pembrolizumab and platinum-based doublet chemotherapy and will comprise of 3 cohorts. Approximately 9 subjects will be enrolled in each cohort, defined by the different platinum-based doublet chemotherapy, to get at least 6 evaluable subject in each cohort, as described in below Figure 1-1. The starting dose regimen for canakinumab in all 3 cohorts is the following: canakinumab 200 mg Q3W (DL1). If this regimen is not considered to be safe, additional subjects are foreseen to be enrolled to assess the next lowest dose level (DL-1) with canakinumab 200 mg Q6W. Of note, all other drugs (pembrolizumab and platinum-based

double chemotherapy) remain at fixed doses, as mentioned in the protocol.

The determination of RP3R will be guided by a Bayesian analysis of DLT data for each cohort for the first 42 days (6 weeks) during which subjects receive the combination of canakinumab, pembrolizumab and platinum-based doublet chemotherapy. The dose-toxicity relationship of canakinumab in combination with pembrolizumab and platinum-based doublet chemotherapy will be modeled by a 5-parameter Bayesian logistic regression model (BLRM) for each dose regimen. In addition to the recommendations from the BLRM, toxicity information (including adverse events and laboratory abnormalities that are not DLTs) and PK information will be evaluated by the Dose Level Review Team (DLRT), comprising of study Investigators and Novartis study personnel: the study physician and statistician.

Figure 1-1 safety run-in design provides an overview on the safety run-in part. If judged necessary, additional subjects might be enrolled in the Dose Level 1 (DL1), or a de-escalation to Dose level minus 1 (DL-1) might also be considered. There will be no dose de-escalation beyond DL-1.

Figure 1-1 Study design: Safety run-in (part 1)



[1] 1 cycle = 3 weeks, [2] for Non-Squamous subjects only

1.1.2 Double-blind randomized part (part 2)

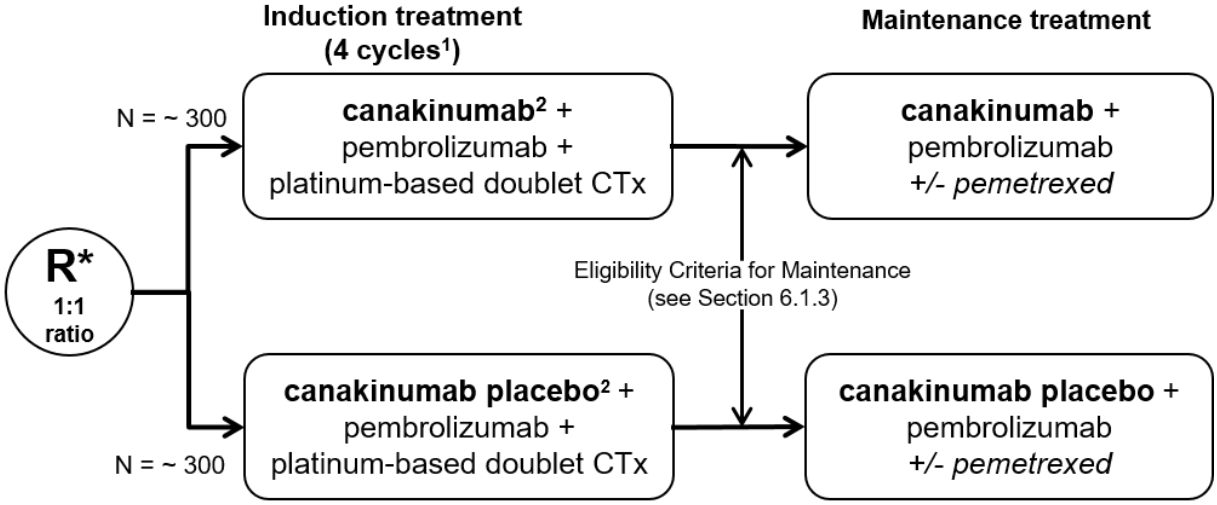
Once the RP3R for canakinumab in combination with pembrolizumab and platinum-based doublet chemotherapy is confirmed in the safety run-in part, the double-blind, randomized, placebo-controlled part of the study will open.

The randomization will be stratified based on PD-L1 status (Tumor Proportion Score (TPS) <1% vs. ≥1%), geographic region (East Asia vs. North America + Western Europe vs. Rest of the world) and histology (squamous vs. non-squamous). PD-L1 unevaluable subjects will be

included with the TPS <1% group. Approximately 600 subjects will be randomized in a 1:1 ratio to two treatment arms as described in below Figure 1-2.

No cross-over treatment from canakinumab matching-placebo treatment arm to canakinumab treatment arm will be allowed.

Figure 1-2 Study design: double-blind randomized part (part 2)



*Stratification factors:
- PD-L1 expression: (<1% vs. ≥1%) -> by Novartis designated laboratory
- Histology: squamous vs. non-squamous
- Geographic region: East Asia vs North America + Western Europe vs RoW

[1] cycle = 3 weeks; [2] canakinumab/matching-placebo at RP3R; CTx = chemotherapy; R = Randomization

An independent Data Monitoring Committee (DMC) will monitor semi-blinded safety and efficacy data during the trial. A separate DMC SAP will specify the analyses to be performed for the DMC reviews.

There is no interim analysis (IA) for PFS. The primary PFS analysis, is planned after approximately 253 PFS events have been observed. The 1st IA for OS will be performed at the same time of primary PFS analysis when approximately 120 among the 304 (39%) deaths are expected to be observed. The intent of this analysis is to assess superior efficacy with either PFS and/or OS results. There is no intent to assess OS futility at the time of the 1st interim OS analysis. At the time of the primary PFS analysis, both PFS and 1st IA for OS will be performed by an independent statistician and reviewed by a data monitoring committee (DMC). If the primary PFS and 1st IA for OS are not statistically significant, the 2nd IA for OS will also be done by an independent statistician. If the primary PFS analysis is statistically significant, the sponsor's clinical team will be unblinded and will perform the second OS interim and/or final OS analyses. Further details on the group sequential design are provided in section 2.16.

1.2 Study objectives and endpoints

Table 1-1 Study objectives and endpoints

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• **Safety run-in part:** To determine the RP3R of canakinumab in combination with pembrolizumab plus platinum-based doublet chemotherapy	• Incidence of dose limiting toxicities in the first 42 days of study treatment
• **Double-blind, randomized, placebo-controlled part:** To compare PFS by local investigator assessment as per RECIST1.1 and OS between the two treatment arms.	• PFS based on local investigator assessment as per RECIST 1.1 and OS are multiple primary endpoints
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**
• **Safety run-in part:**	
○ To assess the preliminary clinical anti-tumor activity (overall response rate, disease control rate and duration of response) of canakinumab in combination with pembrolizumab plus platinum-based doublet chemotherapy.	• ORR, DCR, DOR by local investigator's assessment according to RECIST 1.1
○ To characterize the safety and tolerability of canakinumab in combination with pembrolizumab plus platinum-based doublet chemotherapy.	• Type, frequency and severity of adverse events and reactions, changes in laboratory values, vital signs, ECGs
○ To characterize the pharmacokinetics of canakinumab in combination with pembrolizumab plus platinum-based doublet chemotherapy.	• Concentration and PK parameters of canakinumab, pembrolizumab and chemotherapy
○ To characterize the immunogenicity (anti-drug antibodies) of canakinumab and pembrolizumab	• Anti-drug antibodies (ADA) prevalence at baseline and ADA incidence on-treatment of canakinumab and pembrolizumab
• **Double-blind, randomized, placebo-controlled part:**	
○ To evaluate overall response rate (ORR), disease control rate (DCR), time to response (TTR) and duration of response (DOR) by local investigator assessment per RECIST 1.1 in the treatment arms.	• ORR, DCR, TTR and DOR based on local investigator assessment as per RECIST 1.1
○ To characterize the safety profile of the two treatment arms.	• Frequency of adverse events, serious adverse events, AEs leading to treatment discontinuation, proportion of patients

Objective(s)	Endpoint(s)
	with laboratory abnormalities, ECG, and vital signs.
○ To characterize the pharmacokinetics of canakinumab, pembrolizumab and chemotherapy.	● Concentration and PK parameters of canakinumab, pembrolizumab and chemotherapy
○ To characterize the immunogenicity (anti-drug antibodies, ADA) of canakinumab and pembrolizumab.	● Anti-drug antibodies (ADA) prevalence at baseline and ADA incidence on-treatment of canakinumab and pembrolizumab
○ To assess PROs (EORTC QLQ-C30 with the QLQ-LC13 lung cancer module and EQ-5D-5L) including symptoms, physical functioning and health-related quality of life in the two treatment arms.	● Time to definitive 10-point deterioration symptom scores for chest pain, cough and dyspnea per QLQ-LC13 questionnaire as three primary PRO variables of interest and time to definitive deterioration in global health status/QoL, shortness of breath and pain per QLQ-C30 as secondary PRO variables of interest.

2 Statistical methods

2.1 Data analysis general information

The primary PFS analysis and first interim OS analyses will be performed by an independent statistician external to Novartis. Depending on the results, the following will apply for the primary CSR:

- If PFS and/or OS are statistically significant at the time of the primary PFS analysis, these analyses will be the basis of the primary CSR and the study team will be unblinded. If OS is not statistically significant at the time of the primary PFS analysis, then the next OS analyses (i.e. 2nd IA and final OS analysis if OS is not statistically significant at time of 2nd IA OS) will be performed by Novartis team
- If PFS and OS are not statistically significant at time of the primary PFS analysis, the team will remain blinded to the study treatment. The next OS analysis (2nd IA OS) will be performed by an independent statistician. If it is statistically significant, it will be considered as the basis for the primary CSR. If not, the final OS analysis, performed by Novartis team, will constitute the primary CSR.

Novartis will perform the analysis specified in this SAP for the primary CSR. SAS version 9.4 or later will be used to perform all data analyses and to generate tables, figures and listings. R version 2.3.2 will be used to perform the BLRM analysis.

2.1.1 Data included in the analysis

Data from all patients who signed main informed consent in centers that participate in this study will be used in the analysis. Data collected after patients' withdrawal of informed consent for further participation in the study will not be reported (except for death date, if it is obtained from public records).

Safety run-in part (part 1)

The analysis cut-off date for **the primary analysis** of study data for safety run-in part will be established for each of the cohort after at least 6 evaluable subjects have been treated and observed for 42 days in each cohort. The primary analysis will comprise the DLT data to determine the RP3R (after having assessed the safety in each cohort) using the BLRM.

The analysis cut-off date for **the final analysis** of study data for safety run-in part will be established at the time of the primary analysis for randomized part, when PFS and/or OS reach statistical significance at primary PFS or at 2nd IA OS analysis. Otherwise, the cut-off date will correspond to the final OS analysis cut-off (regardless of the significance of the endpoints), when approximately 304 deaths are observed. At this time, data for all subjects from part 1 will be included and will be analyzed by canakinumab dosing regimen and cohort: DL1 and DL-1 for each cohort of treatment (A, B, C).

The final analysis of the safety run-in part will be included in the primary CSR along with BLRM and DLT results from the primary analysis.

Double-blind randomized part (part 2)

There is one primary PFS analysis and 2 interim OS analyses planned for the primary efficacy endpoints. A unique **cut-off date** will be established after the targeted number of events for the planned primary PFS/interim OS analyses and final analyses has been documented. For each of the analyses, all statistical analyses will be performed using all data collected in the database up to the data cut-off date. All data with an assessment date or event start date (e.g. vital sign assessment date or start date of an adverse event) prior to or on the cut-off date will be included in the analysis. Any data collected beyond the cut-off date will not be included in the analysis and will not be used for any derivations.

All events with start date before or on the cut-off date and end date after the cut-off date will be reported as 'ongoing'. The same rule will be applied to events starting before or on the cut-off date and not having documented end date. This approach applies, in particular, to adverse event and concomitant medication reports. For these cases, the end date will not be imputed and therefore will not appear in the listings.

2.1.2 General analysis conventions

Pooling of centers: Unless specified otherwise, data from all study centers will be pooled for the analysis. Due to expected small number of subjects enrolled at centers, no center effect will be assessed.

Qualitative data (e.g., gender, race, etc.) will be summarized by means of contingency tables by treatment arm; a missing category will be included as applicable. Percentages will be calculated using the number of subjects in the relevant population or subgroup as the denominator.

Quantitative data (e.g., age, body weight, etc.) will be summarized by appropriate descriptive statistics (e.g. mean, standard deviation, median, percentiles, minimum, and maximum) by treatment arm.

2.1.3 General definitions

Definitions below apply for each study part of the study, a subject included in the safety run-in will not enter in the double-blind randomized phase. Nab-paclitaxel and canakinumab matching placebo drug are only applicable to the double-blind randomized part of the study. Canakinumab matching placebo will be referred to as "placebo" in the next sections of this document.

2.1.3.1 Study drug and study treatment

Study drug refers to canakinumab, placebo, pembrolizumab, pemetrexed, carboplatin, cisplatin, paclitaxel, or nab-paclitaxel

Study treatment will refer to the combination of canakinumab with pembrolizumab and induction chemotherapy (+/- pemetrexed in maintenance) or placebo with pembrolizumab and

induction chemotherapy (+/-pemetrexed in maintenance). One of the following platinum-based doublet chemotherapy regimen will be administered during the induction period of the two parts of the study:

- Pemetrexed + carboplatin
- Pemetrexed + cisplatin
- Carboplatin + paclitaxel
- Carboplatin + nab-paclitaxel (only applicable to double-blind randomized part)

2.1.3.2 Date of first administration of study drug

The date of first administration of study drug is defined as the first date when a non-zero dose of study drug is administered and recorded on the study treatment (e)CRF page.

2.1.3.3 Date of last administration of study drug

The date of last administration of the study drug is defined as the last date when a non-zero dose of the respective study drug is administered and recorded on the study treatment (e)CRF page. This date will also be referred as ***last date of study drug***. Last date of study drug is defined for each drug which is part of the study treatment.

Last date of study drug exposure may not be the same as the last date of study drug (see section 2.1.3.6)

2.1.3.4 Date of first administration of study treatment

The ***date of first administration of study treatment*** is defined as the first date when a non-zero dose of any component of the study treatment (Canakinumab/Placebo, pembrolizumab, chemotherapy drugs, or pemetrexed) is administered.

For the sake of simplicity, the date of first administration of study treatment will also be referred as ***start date of study treatment.***

2.1.3.5 Date of last administration of study treatment

The ***date of last administration of study treatment*** is defined as the last date when the last non-zero dose of any last component of the study treatment (Canakinumab/Placebo, pembrolizumab, chemotherapy drugs, or pemetrexed) is administered.

For example, if the last dose of canakinumab/placebo, pembrolizumab, pemetrexed, carboplatin, is taken on 13Oct2019, 25Dec2019, 13Oct2019, 13Sep2019, respectively, then the date of last administration of study treatment is on 25Dec2019.

2.1.3.6 Last date of exposure to study drug/treatment

Pembrolizumab, chemotherapy drugs (except for nab-paclitaxel) and pemetrexed in maintenance are planned to be administered on day 1 every 3 weeks, on a cycle duration of 21 days. Nab-paclitaxel will be given three doses on day 1, day 8 and day 15 for each cycle.

Canakinumab/placebo is also administered on day 1 every 3 weeks (i.e. cycle duration of 21 days), unless there was a toxicity leading to a dosing interval increase or if the recommended phase III dose (RP3R) is Q6W regimen.

The ***date of last exposure*** to canakinumab/ placebo is then calculated as:

- Min ([(last date of administration of study drug)+ (length of cycle duration – 1 = 21-1 = 20)], date of death, last contact date in case subject is lost to follow-up), if dosing regimen of the last canakinumab/placebo injection is Q3W
- Min ([(last date of administration of study drug)+ (length of cycle duration – 1 = 42-1 = 41)], date of death, last contact date in case subject is lost to follow-up), if dosing regimen of the last canakinumab/placebo injection is Q6W
- Min ([(last date of administration of study drug)+ (length of cycle duration – 1 = 63-1 = 62)], date of death, last contact date in case subject is lost to follow-up), if dosing regimen of the last canakinumab/placebo injection is Q9W

The ***date of last exposure*** to pembrolizumab, chemotherapy drugs (except for nab-paclitaxel) and pemetrexed is calculated as:

- Min ([(last date of administration of study drug)+ (length of cycle duration – 1 = 21-1 = 20)], date of death, last contact date in case subject is lost to follow-up)

The ***date of last exposure*** to nab-paclitaxel is calculated as:

- Min ([(last date of administration of study drug)+ (length of each dosing – 1 = 7-1 = 6)], date of death, last contact date in case subject is lost to follow-up)

"Last date of administration of study drug" and "date of last contact" are defined in sections 2.1.3.3 and 2.1.3.11 respectively.

The ***last date of exposure to study treatment*** is defined as the latest date among the last date of exposure to canakinumab/placebo, pembrolizumab, chemotherapy drugs and pemetrexed.

If the derived last date of exposure to study drug/study treatment goes beyond the data cutoff date, it should be truncated to the date of data cutoff.

2.1.3.7 Study day

The study day, describes the day of the event or assessment date, relative to the reference start date (randomization date or start date of study treatment).

The study day is calculated as follows:

- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date + 1 if event is on or after the reference start date;
- The date of the event (visit date, onset date of an event, assessment date etc.) − reference start date if event precedes the reference start date.

In Safety run-in part, the reference start date for all assessments is the start of study treatment.

In the double-blind randomized part, the reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose adjustments,

PK etc.) is the start of study treatment. The reference start date for all other non-safety assessments (i.e., tumor assessment, survival, EQ-5D, ECOG performance status, and patient reported outcomes (PRO)) is the date of randomization.

The study day will be displayed in the data listings. If an event starts before the reference start date, the study day displayed on the listing will be negative.

Time unit

A year length is defined as 365.25 days. A month length is 30.4375 days (365.25/12). If duration is reported in months, duration in days will be divided by 30.4375. If duration is reported in years, duration in days will be divided by 365.25.

2.1.3.8 Baseline

In the **Safety run-in part,** the last available assessment on or before the date of start of study treatment is defined as "baseline" assessment for both safety and efficacy.

In the **double-blind randomized part**, for efficacy evaluations, the last non-missing assessment, including unscheduled assessments on or before the date of randomization is taken as "baseline" value or "baseline" assessment. In the context of baseline definition, the efficacy evaluations also include [redacted] PRO and ECOG performance status. For PRO, if the randomization date is not the cycle 1 day1, PRO assessment on cycle 1 day 1 will be defined as "baseline". For safety evaluations, the last available assessment on or before the date of start of study treatment is taken as ''baseline" assessment.

If subjects have no value as defined above, the baseline result will be missing.

In rare cases where multiple measurements meet the baseline definition, with no further flag or label that can identify the chronological order, then the following rule should be applied: If values are from central and local laboratories, the value from central assessment should be considered as baseline. If multiple values are from the same laboratory (local or central) or collected for ECGs or vital signs, then the last value should be considered as baseline.

2.1.3.9 On-treatment assessment/event and observation periods

For adverse event reporting, the overall observation period will be divided into three mutually exclusive segments:

1. ***pre-treatment period***: from day of subject's informed consent to the day before first administration of study treatment
2. ***on-treatment period*** (including the lower and upper limits of the time interval) : from date of first administration of study treatment up to 130 days (approximately five terminal half-lives of canakinumab) after date of last administration of study treatment
3. ***post-treatment period***: starting at day 131 after last administration of study treatment.

Safety summaries (tables, figures) and summaries of on-treatment death include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries). In addition, all deaths which

occurred during the study (i.e. during the on-treatment and post-treatment periods) will be summarized.

An ***on-treatment adverse event*** (or ***treatment-emergent*** AEs) is defined as any adverse event reported in the on-treatment period.

An ***on-treatment assessment*** is defined as any assessment performed after the date of first administration of study treatment i.e. assessments performed in the following time interval (including the lower and upper limits): from date of first administration of study treatment +1 up to 130 days after the date of last administration of study treatment.

In case at time of the analysis, the date of last administration of study treatment is missing, on-treatment adverse event/assessment include any adverse event/assessment recorded in the database and which occur after the start date of the study treatment.

Data listings will include all assessments/events, flagging those which are not on-treatment assessment/event (i.e. pre-treatment and post-treatment period).

2.1.3.10 Windows for multiple assessments

Time windows will be defined for descriptive summary of PRO and ECOG performance status data by visit and longitudinal data analysis. If more than one assessment is available in the same time window, the assessment closest to the planned date will be considered. If two assessments are obtained with the same time difference compared to the scheduled visit day, the assessment obtained prior to visit will be considered. Data obtained at the end of treatment will be classified as other assessment in the corresponding time window.

Table 2-1 Time windows for PRO and ECOG PS assessments

Assessment	Target day of assessment	Time Window Definition
Baseline	On or before Study Day 1[a]	≤ Study Day 1
During treatment phase		
Week 3	Study Day 22	Study Days 2 to 32
Week 6	Study Day 43	Study Days 33 to 53
Week k	Study Day = 21*(k/3)+1	Study Day
(with k=9, 12, 15, 18, etc...)		21*(k/3)-9 to 21*(k/3)+11 For last cycle of dosing : from 21*(k/3)-9 to end of treatment visit date +7 "Note: EOT data will be included if obtained within 7 days of permanent discontinuation of study treatment"

Assessment	Target day of assessment	Time Window Definition
After treatment discontinuation		
Safety follow-up 1*	26 days after treatment discontinuation	[end of treatment visit date+8; end of treatment visit date + 52]
Safety follow-up 3*	78 days after treatment discontinuation	[end of treatment visit +53; end of treatment visit + 104]
Safety follow-up 5*	130 days after treatment discontinuation	[end of treatment visit+105; end of treatment visit + 156]
Post treatment follow-up** (Efficacy follow up x)	Every 6, 9 or 12 weeks depending on the stage of the study	For the first time window : [upper bound of the last previous time windows with assessment + 1; TA date + TA interval at next visit /2] Otherwise : [TA date – TA interval at this visit /2; TA date+TA interval at next visit/2]
7 days post disease progression***	Within 7 days post disease progression	[disease progression date + 1; disease progression date +21]
28 days post disease progression***	28 days post disease progression	[disease progression date+22; disease progression date +42]

[a] Study Day 1 = randomization date and assessment on Cycle 1 Day1 will be baseline for PRO if randomization date is not Cycle 1 Day1.

* visits planned only for ECOG assessment

** visits planned for the subjects who discontinued the study treatment other than RECIST PD.

*** visits planned only for PRO assessments

2.1.3.11 Last contact date

The last contact date will be derived for subjects not known to have died at the analysis cut-off using the last complete date among the following:

Table 2-2 Last contact date data sources

Source data	Conditions
Date of Randomization	No condition
Last date subject was known to be alive from Survival Follow-up page	Subject status is reported to be alive, unknown.
Start/End dates from further antineoplastic therapy	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose.
Imaging assessment date	Imaging marked as done
Laboratory/PK collection dates	Sample collection marked as 'done'.
Vital signs date	At least one non-missing parameter value
Performance Status date	Non-missing performance status
Start/End dates of AE	Non-missing verbatim term

The last contact date is defined as the latest complete date from the above list on or before the data cut-off date. The cut-off date will not be used for last contact date, unless the subject was seen or contacted on that date. No date post cut-off date will be used. Completely imputed dates (e.g. the analysis cut-off date programmatically imputed to replace the missing end date of a dose administration record) will not be used to derive the last contact date. Partial date imputation is allowed for event (death)/censoring if coming from 'Survival information' eCRF (further details in section 5.1.3).

The last contact date will be used for censoring of subjects in the analysis of overall survival.

2.2 Analysis sets

2.2.1 Safety run-in part (part 1)

Full analysis set and safety set

The full analysis set (FAS) and safety set are defined in the same way and comprise all subjects to whom study treatment has been assigned and who received at least one dose of the study treatment (i.e. at least one dose of any component of the study treatment that is canakinumab or pembrolizumab or platinum-based doublet chemotherapy (including incomplete infusion) or pemetrexed). Subjects will be analyzed according to the dose regimen they have been assigned to and by cohort (A, B, and C) as defined in Figure 1-1.

Dose-determining set

The Dose-Determining Set (DDS) includes all subjects from the safety set who meet the minimum exposure criterion and have sufficient safety evaluations, or experienced a dose limiting toxicity (DLT) during the first 42 days (6 weeks) of dosing.

A subject meets the minimum exposure criterion if the subject received

- At least 2 doses of canakinumab (for Q3W dosing regimen cohort) or at least 1 dose of canakinumab (for Q6W dosing regimen cohort in case the canakinumab dose is de-escalated to Q6W) and 2 infusions of pembrolizumab (full dose) within the first 42 days
- Takes at least 75% of planned doses of the overall platinum-based doublet chemotherapy within the first 42 days (6 weeks) of study treatment

The following examples describe situations how to check the 2nd bullet point criteria of the minimum exposure. Subject is assumed to have already met the first criteria of the minimum exposure for canakinumab and pembrolizumab.

Example 1: *Subject receives pemetrexed and cisplatin as the induction chemotherapy regimen. At C1D1, subject receives 500mg/m^2 of pemetrexed and 75 mg/ m^2 of cisplatin. At cycle 2 (C2D1), subject has toxicity which prevents him to receive the chemotherapy drugs on day 22. The infusions are delayed 23 days later, still at full dose of each drug. Summary of exposure for this subject over the first 2 cycles is:*

- *For pemetrexed, cumulative dose = 500+0 = 500 mg/m2; Planned dose over the 2 cycles = 500+500 = 1000 mg/m2, subject received (500/1000)*100=50% of the planned pemetrexed dose*
- *For cisplatin, cumulative dose = 75+0=75 mg/m2; Planned dose over the 2 cycles = 75+75 = 150 mg/m2, subject received (75/150)*100=50% of the planned cisplatin dose*

⇨ *Subject considered <u>not evaluable</u> and not included in the DDS*

Example 2: *Subject receives pemetrexed and cisplatin as the induction chemotherapy regimen. At C1D1, subject receives 500 mg/m2 of pemetrexed and 75 mg/m2 of cisplatin. At Cycle 2 day 1 (day 22), subject experienced a severe hematological toxicity at cycle 1 which led to a dose reduction of pemetrexed from 500 mg/m2 to 250 mg/m2 and cisplatin remains at full dose of 75 mg/m2. Summary of exposure for this subject over the first 2 cycles will be as follows:*

- *For pemetrexed, cumulative dose = 500+250 = 750 mg/m2; Planned dose over the 2 cycles = 500+500 = 1000 mg/m2, subject received (750/1000)*100=75% of the planned permetrexed dose*
- *For cisplatin, cumulative dose = 75+75=150 mg/m2; Planned dose over the 2 cycles = 75+75 = 150 mg/m2, subject received (150/150)*100=100% of the planned cisplatin dose*

⇨ *Subject considered <u>evaluable</u> and included in the DDS*

Example 3: *Subject receives pemetrexed and cisplatin as the induction chemotherapy regimen. At C1D1, subject was assigned to a lower dose than the planned protocol dose by dosing error 375 mg/m2 of pemetrexed and 75 mg/m2 of cisplatin. At Cycle 2 day 1 (day 22), subject experienced a toxicity at cycle 1 which led to a dose reduction of pemetrexed from 375 mg/m2 to 281 mg/m2 and cisplatin remains at full dose of 75 mg/m2. Summary of exposure for this subject over the first 2 cycles will be as follows:*

- *For pemetrexed, cumulative dose = 375+281 = 656 mg/m2; Planned dose over the 2 cycles = 500+500 = 1000 mg/m2, subject received (656/1000)*100=65.6% of the planned permetrexed dose*
- *For cisplatin, cumulative dose = 75+75=150 mg/m2; Planned dose over the 2 cycles = 75+75 = 150 mg/m2, subject received (150/150)*100=100% of the planned cisplatin dose*

⇨ *Subject considered <u>not evaluable</u> and not included in the DDS*

Example 4: *Subject receives pemetrexed and carboplatin as the induction chemotherapy regimen. At C1D1, subject receives 500 mg/m2 of pemetrexed and AUC 5 for carboplatin. At Cycle 2 day 1 (day 22), subject experienced a toxicity at cycle 1 which lead to a dose reduction of pemetrexed from 500 mg/m2 to 375 mg/m2 and carboplatin from AUC5 to AUC 3.75. Summary of exposure for this subject over the first 2 cycles will be as follows:*

- *For pemetrexed, cumulative dose = 500+375 = 875 mg/m2; Planned dose over the 2 cycles = 500+500 = 1000 mg/m2, subject received (875/1000)*100=87.5% of the planned permetrexed dose*
- *For carboplatin, cumulative dose = 5+3.75=AUC 8.75; Planned dose over the 2 cycles = 5+5 = AUC 10, subject received then (8.75/10)*100=87.5% of the planned carboplatin dose*

⇨ *Subject considered <u>evaluable</u> and included in the DDS*

Pharmacokinetic set

The Pharmacokinetic Analysis Set (PAS) consists of all subjects who received at least one dose of study drug and have at least one evaluable pharmacokinetic (PK) sample for one of the study drug. PAS will be defined for canakinumab, pembrolizumab, carboplatin, cisplatin, paclitaxel and pemetrexed separately.

The **PAS-canakinumab** includes all subjects who received at least one dose of canakinumab and have at least one evaluable PK sample for canakinumab.

A PK sample for canakinumab is considered evaluable if:

- Subject receive at least one dose of canakinumab prior to sampling except C1D1 pre-dose sample
- Have pre-dose samples drawn prior to the next dose of canakinumab
- Subject receive 200 mg of canakinumab prior to post-dose PK sampling
- It is collected (except for cycle 1 day 1) between 16 days (384 hours) (37 days for Q6W) and 26 days (684 hours) (47 days for Q6W) after the last 200 mg canakinumab dose administration

The **PAS-pembrolizumab** includes all subjects who received at least one dose of pembrolizumab and have at least one evaluable PK sample for pembrolizumab.

A PK sample for pembrolizumab is evaluable if:

- For pre-dose sample:
  - It is collected before the next dose administration
  - It is collected (except for cycle 1 day 1) between 16 days (384 hours) and 26 days (624 hours) after the last 200 mg pembrolizumab dose administration
- For post-dose sample:
  - Subject received 200 mg of pembrolizumab prior to post-dose PK sampling
  - For end-of-infusion samples, have the sample collected within 2 hours post end of infusion

The pharmacokinetic analysis set of the platinum-doublet chemotherapy agents (PAS for each agent: PAS-cisplatin, PAS-pemetrexed, PAS-carboplatin, PAS-paclitaxel) will include all subjects who provide at least one evaluable PK concentration for the respective agent. For a concentration to be evaluable, subjects are required to:

- receive a planned dose of platinum-doublet chemotherapy agent prior to sampling,
- for pre-dose samples, have the sample collected before the next dose administration
- for samples scheduled to be taken prior to EOI, have the samples collected prior to EOI

for post-EOI samples, have the samples collected post EOI

Subject Classification

Subjects may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific subject classification rules defined in Table 2-3.

Table 2-3 Subject classification based on protocol deviations and non-PD criteria

Analysis set	Protocol deviations leading to exclusion	Non protocol deviation leading to exclusion
FAS	No written informed consent	No dose of any component of study treatment
Safety Set	No written informed consent	No dose of any component of study treatment
DDS	No written informed consent	See definition of DDS
PAS-canakinumab	No written informed consent	See definition of PAS-canakinumab
PAS-pembrolizumab, -pemetrexed, -carboplatin, -cisplatin, -paclitaxel	No written informed consent	See definition of PAS-pembrolizumab, -pemetrexed, -carboplatin, -cisplatin, -paclitaxel

2.2.2 Double-blind, randomized part (part 2)

Full analysis set

The FAS comprises of all subjects to whom study treatment has been assigned by randomization. According to the intent to treat principle, subjects will be analyzed according to the treatment and strata which they have been assigned during the randomization procedure.

Safety set

The safety set includes all subjects who received at least one dose of any component of the study treatment. Subjects will be analyzed according to the study treatment they received, either canakinumab plus pembrolizumab plus platinum-based doublet chemotherapy or canakinumab matching-placebo plus pembrolizumab plus platinum-based doublet chemotherapy. The treatment received is defined as the randomized treatment if the subject took at least one dose of that treatment (i.e. at least one dose of canakinumab or placebo) or the first treatment received if the randomized treatment was never received.

Pharmacokinetic set

The Pharmacokinetic set in the study part 2 will be defined as in part 1. PAS-nab-paclitaxel for nab-paclitaxel applicable to only randomized part will be defined as for the other chemotherapy agents.

Subject Classification

Subjects may be excluded from the analysis populations defined above based on the protocol deviations entered in the database and/or on specific subject classification rules as defined in

Table 2-3. PAS-nab-paclitaxel for the randomized part will be handled based on the PAS definition detailed above and protocol deviation rules described in Table 2-3.

2.2.3 Subgroup of interest

In **Safety run-in part**, no subgroup analysis will be performed.

2.2.3.1 Efficacy

If the primary endpoint analyses for PFS and/or OS are statistically significant, subgroup analyses to assess the homogeneity of the treatment effect across demographic and baseline disease characteristics will be performed for the subgroups as shown below.

The subgroups are:

- Gender (male vs. female)
- Age (<65 vs. ≥65 years)
- Race (White vs. Black vs. Asian vs. Others (includes "Native Hawaiian or other pacific Islander" and "American Indian or Alaska Native"))
- ECOG performance status (0 vs. 1)
- Smoking history (Former/Current vs. Never)
- PD-L1 baseline expression level (<1% vs. ≥1%, <50% vs. ≥50%, <1% vs. 1-49% vs. ≥ 50%, in case of non-evaluable PD-L1 status the subjects will be excluded from the subgroup analyses) (both IRT and eCRF)
- Histology (squamous vs. non-squamous) (both IRT and eCRF)
- Disease stage (locally advanced (stage III) vs. metastatic (stage IV))
- Presence of brain metastasis (yes vs. no)
- Geographic region (East Asia vs. Western Europe and North America vs. Rest of the world) (both IRT and eCRF)
- Chemotherapy regimen with the four following subgroups :
    - Canakinumab + pembrolizumab +pemetrexed + carboplatin vs. placebo + pembrolizumab + pemetrexed + carboplatin
    - Canakinumab + pembrolizumab + pemetrexed + cisplatin vs. placebo + pembrolizumab + pemetrexed + cisplatin
    - Canakinumab + pembrolizumab + carboplatin + paclitaxel vs. placebo + pembrolizumab + carboplatin + paclitaxel

- Canakinumab + pembrolizumab + carboplatin + nab- paclitaxel vs. placebo + pembrolizumab + carboplatin + nab- paclitaxel

- Tumor mutation burden (TMB) status at baseline (<10 mutations/megabase vs. ≥10 mutations/megabase, <16 mutations/megabase vs. ≥16 mutations/megabase)
- [redacted] CCI
- hsCRP at baseline (<2mg/L vs. ≥2mg/L, <10mg/L vs. ≥10mg/L, <50 mg/L vs. ≥50 mg/L)

No formal statistical test of hypotheses will be performed for the subgroups. [redacted] Kaplan-Meier method will be used for summarizing medians and the corresponding 95% CI. Hazard ratio (HR) for the treatment effect and 95% CIs will be summarized based on cox regression model stratified by the randomization stratification factors using forest plot.

2.2.3.2 Safety

Key safety analyses, including AEs and AESIs will be repeated on safety set in the following subgroups:

- Gender (male vs. female)
- Race (White vs. Black vs. Asian vs. Others (includes "Native Hawaiian or other pacific Islander" and "American Indian or Alaska Native"))
- Chemotherapy regimen (pemetrexed+cisplatin vs. pemetrexed+carboplatin vs. carboplatin+paclitaxel vs. carboplatin+nab-paclitaxel, as described for the efficacy subgroups in the above section 2.2.3.1)
- ECOG PS status (0 vs. ≥ 1)
- Smoking history (Former vs. Current vs. Never)
- Age (<65 years vs. ≥65 years)

The objective for carrying out these subgroup analyses is to identify potential safety issues that may be limited to a subgroup of subjects, or safety issues that are more commonly observed in a subgroup of subjects.

The following safety summaries will be performed by subgroup:

- AEs, regardless of study treatment, by preferred term and maximum CTC grade
- AEs with suspected relationship to study drugs by preferred term and maximum CTC grade
- Adverse Event of Special Interest, irrespective of causality, by grouping, preferred term, maximum CTC grade

2.2.3.3 Japan-specific subgroup analyses

Key efficacy and safety outputs, including baseline characteristics, will be repeated for the subjects randomized in sites from Japan. An un-stratified Cox model will be used for primary estimands if number of Japanese subjects is less than 20. These will be specified in the TFL shells.

2.3 Subject disposition, demographics and other baseline characteristics

2.3.1 Double-blind randomized part (part 2)

2.3.1.1 Enrollment status

The following summaries will be provided in each treatment group for the FAS:

- Number (%) of subjects who were randomized
- Number (%) of subjects who received at least one dose of study treatment after randomization

Number (%) of subjects screened will be summarized by country and center. In addition, the number (%) of subjects randomized will be summarized by country, center and treatment group.

For subjects who are screen failures, the reasons for not completing screening will be summarized based on “Screening Phase Disposition” eCRF.

2.3.1.2 Basic demographic and background data

Demographic and other baseline data including disease characteristics will be listed and summarized descriptively by treatment group using the FAS and safety set for the double-blind randomized part. Categorical data will be presented as frequencies and percentages. For continuous data, mean, standard deviation, median, minimum, and maximum will be presented.

BMI (kg/m2) will be calculated as weight[kg] / (height[m]^2) using weight and height at baseline.

2.3.1.3 Baseline stratification factors

The number (%) of subjects in each stratum (histology, PD-L1 status and geographic region) based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS only for the double-blind randomized part. Discordances between the stratum recorded in IRT at the time of randomization and the actual stratum recorded in the clinical database through the data collected on eCRF will be cross-tabulated and listed.

2.3.1.4 Diagnosis and extent of cancer

Summary statistics will be tabulated for diagnosis and extent of cancer using FAS. This analysis will include the following: primary site of cancer, details of tumor histology/cytology, stage at

initial diagnosis with time since initial diagnosis, stage at time of study entry, number and type of metastatic sites. Metastatic sites will be based on diagnosis page.

2.3.1.5 Medical history

Relevant medical histories and current medical conditions at baseline will be summarized and listed by treatment group in the double-blind randomized part using the FAS. The summaries will be presented by primary system organ class (SOC) and preferred term (PT), by treatment group in the double-blind randomized part. Medical history and current medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The MedDRA version used for reporting will be specified in the CSR and as a footnote in the applicable tables/listings.

2.3.1.6 Other

All data collected at baseline (e.g. treatment beyond progression informed consent) will be listed.

2.3.1.7 Subject disposition

The number (%) of subjects treated/randomized subjects for the double-blind randomized part, included in the FAS, will be presented by treatment group in the double-blind randomized part.

The number (%) of screened subjects and not randomized in the double-blind randomized part and the reasons for screening failure will also be displayed. The number (%) of subjects in the FAS who are still on treatment, who discontinued the study phases and the reason for treatment/post-treatment follow-up discontinuation will be presented by treatment group in the double-blind randomized part.

2.3.1.8 Protocol deviations

The number (%) of subjects in the FAS with any protocol deviation will be tabulated by deviation category (as specified in the edit checks specification) overall and by treatment arm for the FAS. Protocol deviations leading to exclusion from analysis sets will be tabulated separately overall and by treatment arm as specified in Section 2.2.2. All protocol deviations will be listed.

In addition to the pre-defined standard PD terms, Novartis has also defined 6 new protocol deviations and the corresponding relationship (health status related vs. site lockdown, patient concerns, etc.) to the COVID-19 pandemic in line with “FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency” (March 2020) and “Guidance on the management of clinical trials during the COVID-19 (coronavirus) pandemic” (April 2020) from EMA as listed below. One additional study specific PD is also defined to capture treatment delay/interruption due to COVID-19. The following deviations related to the COVID-19 pandemic will be summarized.

- Missing visits
- Changes in procedures and assessments
- Planned visits not done at sites

- Changes in drug supply method
- Treatment not given
- Patient discontinuation due to COVID-19 situation
- Treatment delayed/interrupted (CANOPY-1 study specific)

A cross-tabulation of COVID-19 related PD vs. corresponding relationship will also be produced by treatment arm. In addition, COVID-19 related outcomes (e.g., COVID-19 AEs, discontinuation due to COVID-19, death due to COVID-19) will be descriptively summarized by country, site and treatment arm.

2.3.1.9 Analysis sets

The number (%) of subjects in each analysis set will be summarized by treatment arm and randomization stratum.

2.3.2 Safety run-in part (part 1)

Summaries and listings will be reported by canakinumab dose regimen and by cohort for all subjects. The following analyses will be repeated for Safety run-in part using the analysis details specified for the randomized part in section 2.3.1:

- Number (%) of subjects treated will be summarized by country, center, dose regimen of canakinumab and by cohort
- Basic demographic and background data
- Diagnosis and extent of cancer
- Medical history will be listed
- Subject disposition and screening disposition
- Protocol deviations will be listed
- Analysis sets

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

2.4.1 Double-blind randomized part (part 2)

2.4.1.1 Study treatment / compliance

The safety set will be used for the analyses below. The exposure related analyses will be presented by treatment group.

The duration of exposure for study treatment and for each study drug (canakinumab (/placebo), pembrolizumab, platinum-based doublet chemotherapy drugs and pemetrexed) will be presented by treatment group as well as by chemotherapy regimen using the following groups:

- Canakinumab + pembrolizumab +pemetrexed + carboplatin vs. placebo + pembrolizumab + pemetrexed + carboplatin
- Canakinumab + pembrolizumab + pemetrexed + cisplatin vs. placebo + pembrolizumab + pemetrexed + cisplatin

- Canakinumab + pembrolizumab + carboplatin + paclitaxel vs. placebo + pembrolizumab + carboplatin + paclitaxel
- Canakinumab + pembrolizumab + carboplatin + nab-paclitaxel vs. placebo + pembrolizumab + carboplatin + nab- paclitaxel

Pemetrexed will be given in maintenance as per local approved label and local practice after induction treatment for non-squamous NSCLC subjects.

2.4.1.1.1 Duration of exposure to study drug

The duration of exposure for a study drug is defined according to the dosing regimen of each study drug as outline in section 2.1.3.

Duration of exposure (days) = (last date of exposure for study drug – date of first administration of study drug + 1)

The duration of exposure to study treatment is defined considering the duration of exposure of each study drug as:

Duration of exposure (days) = (last date of exposure to study treatment – date of first administration of study treatment + 1)

The duration of exposure includes the periods of temporary interruption except for the last one, a dose interruption occurring after the last date of exposure to study drug won't be considered.

Duration of exposure to study drug/treatment will be categorized into time intervals in months (<1 month, at least 1 month, at least 2 months etc.). In addition summary statistics will be displayed in months.

2.4.1.1.2 Cumulative dose

The cumulative dose is defined as the total dose given during the study drug exposure and will be summarized for each of the study drug separately by chemotherapy regimen as defined above for the duration of exposure.

For subjects who did not take any drug the cumulative dose is by definition equal to zero.

The cumulative dose is calculated using the information from the study treatment eCRF pages and is expressed in mg for pembrolizumab and canakinumab/placebo, in mg/m2 for chemotherapy drugs other than carboplatin, and in AUC for carboplatin.

For canakinumab/placebo and pembrolizumab, the cumulative dose in mg is the sum of "dose administered" from the eCRF of all cycles during the exposure of canakinumab/placebo, pembrolizumab respectively.

For chemotherapy drugs:

- Cisplatin, paclitaxel, nab-paclitaxel, pemetrexed

  The dose in mg/m^2 at cycle N is equal to the dose administered (mg) at cycle N divided by the body surface area (BSA) at the beginning of the cycle N using the weight measured before the infusion. For nab-paclitaxel, the dose in mg/m^2 at cycle N will be the sum of the 3 doses planned to be received at day 1, day 8 and day 15 of cycle N. The

dose in mg/m^2 will be calculated using the BSA at the beginning of the cycle (i.e. day 1).

BSA (m^2) at cycle N = $\sqrt{Wt\ (kg) * Ht(cm)/3600}$ (Mosteller formula)

Where Wt=weight at the beginning of cycle N (Usually this is the weight taken before the infusion of cycle N, or last weight available if this is missing)

Ht=Height at the beginning of the study

- Carboplatin

  The dose in AUC at cycle N will be derived using the Calvert Cockroft-Gault formula as follows:

  Total dose (AUC) = total carboplatin dose (mg) / [GFR (ml/min)+25]

  Where GFR will be substituted with creatinine clearance.

  Creatinine clearance will be calculated using the Cockroft-Gault formula as follows:

  - Male GFR (mL/min) = [140 - age (years)] x weight (kg) x 1.23 / serum creatinine (μ mol/L)
  - Female GFR (mL/min) = [140 - age (years)] x weight (kg) x 1.04 / serum creatinine (μ mol/L)

    OR

  - Male GFR (mL/min) = [140 - age (years)] x weight (kg) / [72 x serum creatinine (mg/dL)]
  - Female GFR (mL/min) = [140 - age (years)] x weight (kg) x 0.85 / [72 x serum creatinine (mg/dL)]

The cumulative dose of the respective chemotherapy drugs will correspond to the sum of the dose as defined above across all the visits.

2.4.1.1.3 Dose intensity and relative dose intensity

The actual dose intensity and the relative dose intensity (RDI) will be summarized for each study drug component nested within the groups as defined for the cumulative dose above by descriptive statistics. In addition, categorical summary of RDI for each study drug will be presented. The number of administrations (or treatment cycles) of each study drug component of the study treatment nested within the treatment groups as defined for the cumulative dose will be summarized by frequencies and descriptive statistics.

Actual dose intensity (DI) is defined for subjects with non-zero duration of exposure. For subjects who did not take the drug, the DI is by definition equal to zero.

PDI (dosing unit/unit of time) = planned dose intensity

PDI is the assigned dose by unit of time planned to be given to subjects as per protocol in the same dose unit and unit of time as that of the dose intensity.

DI, PDI and relative dose intensity (RDI) is defined as:

For canakinumab (/placebo):

- DI (mg/cycle) = [cumulative dose (mg) / (last date of administration of study drug - first date of administration of study drug + length of dose frequency)]*21,

  Where length of dose frequency = 21 if last dose is given on Q3W, and 42 if last dose is given on Q6W and 63 if last dose is given on Q9W
- PDI is the planned dose as per protocol in mg (i.e. 200 mg Q3W)
- RDI (%) = DI (mg/cycle) / PDI (mg/cycle) * 100

For pembrolizumab:

- DI (mg/cycle) = [cumulative dose (mg) / (last date of administration of study drug - first date of administration of study drug + 21)]*21,
- PDI is the planned dose as per protocol in mg (i.e. 200 mg Q3W)
- RDI (%) = DI (mg/cycle) / PDI (mg/cycle) * 100

For chemotherapy:

- Cisplatin, paclitaxel, pemetrexed
  - DI (mg/m^2/cycle) = [cumulative dose (mg/m^2) / (last date of administration of study drug - first date of administration of study drug + 21)]*21
  - PDI is the planned dose as per protocol in mg/m^2 (i.e. for pemetrexed 500 mg/m^2 Q3W, for paclitaxel 200 mg/m^2 Q3W, for cisplatin 75 mg/m^2 Q3W)
  - RDI (%) = DI (mg/m^2/cycle) / PDI (mg/m^2/cycle) * 100
- Carboplatin
  - DI (AUC/cycle) = [cumulative dose (AUC) / (last date of administration of study drug - first date of administration of study drug + 21)]*21
  - PDI is the planned dose as per protocol in AUC (i.e. when combined with pemetrexed it's 5 mg/ml*min Q3W, when combined with paclitaxel/nab-paclitaxel it's 6 mg/ml*min Q3W)
  - RDI (%) = DI (AUC/cycle) / PDI (AUC/cycle) * 100
- Nab-paclitaxel
  - DI (mg/m^2/cycle) = [cumulative dose (mg/m^2) / (last date of administration of study drug - first date of administration of study drug + i)]*21

    Where i = 7 if last dose is given on CxD15, i = 14 if last dose is given on CxD8 and i = 21 if last dose is given on CxD1
  - PDI is the planned dose as per protocol in mg/m^2 (i.e. 300 mg/m^2/cycle)
  - RDI (%) = DI (mg/m^2/cycle) / PDI (mg/m^2/cycle) * 100

2.4.1.1.4 Dose reduction, interruption, frequency regimen change and permanent discontinuation

The number of subjects (%) with dose changes, dose interruptions and dose permanent discontinuations, and associated reasons, will be summarized separately for each study drug, as well as overall by treatment group. All dosing data will be listed.

'Dose interrupted', 'dose change', and 'dose permanently discontinued' option from the type of change field of the study treatment eCRF pages will be used to determine the dose interruptions, change (dose reduction for chemotherapy drugs, dosing interval increase for canakinumab) and permanent discontinuations, respectively.

The corresponding fields 'Reasons for dose change' from the study treatment eCRF pages, will be used to summarize the reasons.

The duration of the interruption will be summarized by time intervals in weeks for each observed interruption: <1week, [1-2 weeks), [2-3 weeks), … as well as by descriptive statistics. The time intervals may be adjusted depending on the observed data.

Of note, a dose interruption reported in the eCRF after the last dose of study drug, won't be considered in the analysis as a dose interruption.

A dose change is defined as a reduction of the study chemotherapy dose from the protocol planned starting dose or a decrease from the previous non-zero dose, even if this decrease has been directly preceded by an interruption. For canakinumab, dose change is defined as any dosing interval increase from Q3W to Q6W, or from Q6W to Q9W.

There is no planned dose reduction or dosing interval increase for pembrolizumab, only dose interruptions are allowed.

2.4.1.1.5 Treatment beyond RECIST progression

The number of subjects who continue treatment beyond RECIST1.1 progression according to local investigator assessment based on protocol specified criteria will be summarized using FAS. It includes all subjects who received any study treatment (i.e. at least one dose of canakinumab/placebo and pembrolizumab) after RECIST1.1 progression assessed by local investigator assessment. Those subjects will be identified using the eCRF 'Verification for treatment beyond RECIST 1.1 PD' page. The reasons for subjects who did not go beyond RECIST 1.1 while eligible will be presented based on the same eCRF page.

2.4.1.2 Prior, concomitant and post therapies

2.4.1.2.1 Prior anti-cancer therapy

The number and percentage of subjects who received any prior anti-neoplastic medications, prior anti-neoplastic radiotherapy or prior anti-neoplastic surgery will be summarized by treatment arm. Prior anti-neoplastic medications will be summarized by therapy type for the last treatment (e.g. chemotherapy, etc.), setting for the last treatment (e.g. adjuvant, neo-adjuvant) and also by ATC class, preferred term and treatment arm. Summaries will include total number of regimens. The medication therapy type of any combination therapy will be counted in each

category. For example, a subject receiving a combination therapy of chemotherapy and targeted therapy will be counted under ‘chemotherapy’ and ‘targeted therapy’.

For radiotherapy, time since last radiotherapy, locations and setting of last therapy will be summarized by treatment arm. For prior surgery, time since last surgery, procedure and residual disease of last therapy will be summarized by treatment arm.

Separate listings will be produced for prior anti-neoplastic medications, radiotherapy, and surgery.

Anti-neoplastic medications will be coded using the WHO Drug Dictionary (WHO-DD); anti-neoplastic surgery will be coded using MedDRA. Details regarding MedDRA and WHO-DD version will be included in the footnote in the tables/listings.

The above analyses will be performed using the FAS.

2.4.1.2.2 Post treatment anti-cancer therapy

Anti-neoplastic therapies since discontinuation of study treatment will be listed and summarized by ATC class, preferred term, overall and by treatment arm by means of frequency counts and percentages using FAS. In addition a summary of the first anti-neoplastic therapy received after treatment discontinuation will be summarized by treatment regimen (i.e. in case of combined treatments) overall and by treatment arm.

2.4.1.2.3 Concomitant medications

Concomitant therapy is defined as all interventions (therapeutic treatments and procedures) other than the study treatment administered to a subject coinciding with the study treatment period. Concomitant therapy include medications (other than study treatment) starting on or after the start date of study treatment or medications starting prior to the start date of study treatment and continuing after the start date of study treatment.

Concomitant medications will be coded using the World Health Organization (WHO) Drug Reference Listing (DRL) dictionary that employs the WHO Anatomical Therapeutic Chemical (ATC) classification system and summarized by lowest ATC class and preferred term using frequency counts and percentages. These summaries will include:

- Medications starting on or after the start of study treatment but no later than 130 days after the last date of administration of study treatment and
- Medications starting prior to start of study treatment and continuing after the start of study treatment

All concomitant therapies will be listed. Any concomitant therapies starting and ending prior to the start of study treatment or starting more than 130 days after the last date of study treatment will be flagged in the listing. The safety set will be used for all concomitant medication tables and listings. In addition, surgical and medical procedures will be coded using MedDRA and summarized by SOC, preferred term and treatment arm in the safety set.

2.4.2 Safety run-in part (part 1)

The analyses for study treatment / compliance in Section 2.4.1.1 will be repeated for safety run-in part by canakinumab dose regimen and cohort (instead of treatment group). All prior, concomitant and post therapies in safety run-in part will be listed by canakinumab dose regimen and by cohort.

Of note treatment cohort corresponds to either cohort A, B or C. Canakinumab dose regimen refers to either Q3W or Q6W.

2.5 Analysis of the primary objective

2.5.1 Safety run-in part (part 1)

2.5.1.1 Primary endpoint

The primary objective is to determine the recommended phase 3 dose regimen (RP3R) of canakinumab in combination with pembrolizumab, pemetrexed and carboplatin (cohort A) in non-squamous NSCLC cohort, canakinumab in combination with pembrolizumab, pemetrexed and cisplatin (cohort B) in non-squamous NSCLC cohort and canakinumab in combination with pembrolizumab, carboplatin and paclitaxel in squamous or non-squamous NSCLC cohort (cohort C).

The primary endpoint is the incidence of dose limiting toxicities (DLT) in the first 42 days (6 weeks) of the study treatment.

Summaries of the posterior distribution of DLT rates for each canakinumab dose regimen for each cohort will be provided based on DDS.

An AE reported as a DLT by the investigator will be identified on the AE eCRF page with the question ‘AE meets definition of DLT’.

The number (%) of subjects reporting any DLT during the first 42 days (i.e. start date of event ≤ date of first study treatment+41) will be summarized by primary system organ class, preferred term, worst grade (using CTCAE v5.0) and by cohort using the DDS. DLTs will be listed and subjects with no DLT will be indicated as such by ‘no DLT’.

2.5.1.2 Statistical hypothesis, model, and method of analysis

Identification of recommended regimen

Determination of the RP3R will be based upon the estimation of the probability of DLT for the first 42 days (6 weeks) following the first dose for subjects in the dose-determining set (DDS). A lower recommended regimen may be identified based on other safety and PK data from the current study.

The RP3R is identified when the following conditions are met **in each cohort**:

- At least 6 evaluable subjects have been treated at this dose and observed for 42 days **and**

- This dose satisfies EWOC criteria **and**
- The selected dose regimen is recommended either per the model or by review of all clinical data by Dose Level Review Team (DLRT) in a dose decision meeting.

Bayesian adaptive approach

The determination of RP3R will be guided by a Bayesian analysis of DLT data for each cohort for the first 42 days (6 weeks) during which subjects receive the combination of canakinumab, pembrolizumab and platinum-based doublet chemotherapy. The relationship between dose and the probability of DLT is modeled using logistic regression. Details of the Bayesian logistic regression model (BLRM) are given in Appendix Section 16.3 of the protocol.

The dose limiting toxicity (DLT) relationship of canakinumab in combination with pembrolizumab and platinum-based doublet chemotherapy is modeled by a BLRM for each dose regimen that comprises single agent toxicity parts and interaction part. Single agent toxicity is modeled using logistic regression for the probability of a subject experiencing a DLT against log-dose. The odds of a DLT for each dose regimen are then calculated under no interaction for the single agent toxicities, and interaction is accounted for by adjusting these odds with additional model parameter (odds multipliers).

Starting dose

The starting dosing regimen are the following:

- Cohort A: canakinumab 200 mg Q3W in combination with pembrolizumab 200 mg, carboplatin AUC 5 mg/mL*min and pemetrexed 500 mg/m^2 Q3W
- Cohort B: canakinumab 200 mg Q3W in combination with pembrolizumab 200 mg, cisplatin 75 mg/m^2 and pemetrexed 500 mg/m^2 Q3W
- Cohort C: canakinumab 200 mg Q3W in combination with pembrolizumab 200 mg, carboplatin AUC 6 mg/mL*min and paclitaxel 200 mg/m^2 Q3W.

For this starting dose level of canakinumab (i.e. 200 mg Q3W), the prior risk of excessive toxicity (i.e. risk of DLT within [33%-100%]) is 15%, 10% and 22% for cohorts A, B and C, respectively, which satisfies the EWOC criterion.

Dose recommendation

Dose recommendations will be based on summaries of the posterior distribution of DLT rates for each dose level of the respective combination therapy. After each cohort of subjects, the posterior distribution for the risk of DLT for new subjects at combination doses of interest will be evaluated. The posterior distributions will be summarized to provide the posterior probability that the risk of DLT for each dose regimen lies within the following intervals:

- Under-dosing: [0, 16%]
- Targeted toxicity: [16%, 33%]
- Excessive toxicity: [33%, 100%]

Dosing regimen decisions are guided by the escalation with overdose control (EWOC) principle (Rogatko et al 2007). A dosing regimen may only be used for newly enrolled subjects if the risk of excessive toxicity at that dosing regimen is less than 25%.

2.5.1.3 Handling of missing data

Subjects who are ineligible for the DDS will be excluded from the primary analysis (assessment of RP3R using incidence of DLT during first 42 days of treatment with canakinumab either in combination with pembrolizumab, carboplatin and pemetrexed or in combination with pembrolizumab, cisplatin and pemetrexed or in combination with pembrolizumab, carboplatin and paclitaxel), although their data will be used for all remaining analyses.

Other missing data will simply be noted as missing on appropriate tables/listings.

2.5.2 Double-blind, randomized part (part 2)

2.5.2.1 Primary endpoints

The primary objectives of the study are to compare progression-free survival (PFS) by local investigator assessment as per RECIST 1.1 and overall survival (OS) between the two study treatment arms.

Progression-free survival (PFS)

PFS based on local investigator assessment as per RECIST 1.1 will be one of the primary efficacy endpoints. The primary analysis will be based on the FAS and will include all data observed up-to the cut-off date. Censoring conventions are defined in section 2.5.2.5.5.

The scientific objective guiding the primary estimand is to estimate the treatment effect based on the primary endpoint of PFS for the combination of canakinumab with pembrolizumab +/- pemetrexed and induction chemotherapy compared to the combination of placebo with pembrolizumab +/- pemetrexed and induction chemotherapy, for the target population irrespective of any post-treatment anti-neoplastic therapy received. The primary estimand will be described by the following five attributes:

- The **target population** is defined by all subjects randomized in the study (FAS).
- The **primary variable** is PFS defined as the time from the date of randomization to the date of the first documented disease progression based on local investigator assessment as per RECIST 1.1 or date of death due to any cause, whichever occurs first.
- The **treatment of interest** is the randomized treatment (canakinumab arm or the matching placebo arm) with or without any new anti-neoplastic therapy post randomization as needed.
- The remaining **intercurrent event** describes how events that may occur after randomization are considered when assessing the treatment effect.
  - **Discontinuation of study treatment**: PFS will take into account all PFS events irrespective of the study treatment discontinuation reasons (treatment policy)
  - Any unforeseen intercurrent events due to COVID-19 pandemic will be handled by treatment policy strategy.
- The **summary measure** is the hazard ratio (HR) for PFS between the two treatment arms. It will be estimated using Cox proportional hazard model stratified by the

randomization stratification factors. PFS will be tested using the log-rank test stratified by randomization stratification factors.

The analysis of PFS will be based on the local radiological assessments done until the cut-off date defined in Section 2.1.1. The analysis will be performed on the FAS and will use the default censoring and event date options from table 16-5 of the appendix 16.1 of the protocol based on options A(1), B(1), C1(1), C2(1), D(1), E(1), and F(1). In particular, PFS will not be censored if a new antineoplastic therapy is started; instead, an Intent to treat (ITT) approach will be used and this new antineoplastic therapy will be ignored for the purposes of PFS derivation (and tumor assessments will continue), i.e. option F (1) in Table 16-5 of the protocol appendix 16.1 will be used. Discontinuation of study treatment (for any reason) will not be considered as a reason for censoring.

Overall survival (OS)

OS is the other primary efficacy endpoint. The primary analysis will be based on the FAS and will include all data observed up-to the cut-off date (i.e. primary PFS cut-off date, 2nd IA OS or final OS analysis in case not significant before at any interim analysis). Censoring conventions are defined in section 2.5.2.5.5.

The scientific objective guiding the primary estimand is to estimate the treatment effect based on the primary endpoint of OS for the combination of canakinumab with pembrolizumab +/- pemetrexed and induction chemotherapy compared to the combination of placebo with pembrolizumab +/- pemetrexed and induction chemotherapy, for the target population irrespective of any post-treatment anti-neoplastic therapy received. The primary estimand will be described by the following five attributes:

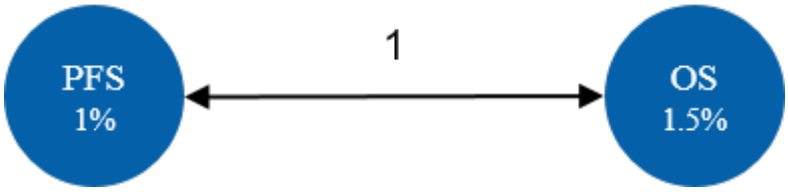
- The **target population** is defined by all subjects randomized in the study (FAS).
- The **primary variable** is OS defined as the time from the date of randomization to the date of death due to any cause.
- The **treatment of interest** is the randomized treatment (canakinumab arm or the matching placebo arm) with or without any new anti-neoplastic therapy post randomization as needed.
- The remaining **intercurrent event** describes how events that may occur after randomization are considered when assessing the treatment effect.
    - **Discontinuation of study treatment**: OS will take into account all deaths irrespective of the study treatment discontinuation reasons (treatment policy)
    - Any unforeseen intercurrent events due to COVID-19 pandemic will be handled by treatment policy strategy.
- The **summary measure** is the hazard ratio (HR) for OS between the two treatment arms. It will be estimated using Cox proportional hazard model stratified by the randomization stratification factors. The OS will be tested using the log-rank test stratified by randomization stratification factors.

2.5.2.2 Statistical hypothesis, model, and method of analysis

2.5.2.2.1 Gate-keeping procedure

In order to conserve the overall type-1 error (one-sided level of significance α=0.025) in testing the primary endpoints of PFS and OS, an alpha split with a graphical gate-keeping approach will be implemented as shown in the figure below (Bretz et al 2009, Bretz et al 2011).

Figure 2-1 Graphical gate-keeping procedure to test primary endpoints in order to control overall type 1 error



2.5.2.2.2 Progression free survival (PFS)

Assuming a proportional hazards model for PFS, the null hypothesis will be tested at one-sided 1% level of significance. If OS is statistically significant at either interim or final analysis, then the 1.5 % alpha assigned to OS will be transferred to PFS and PFS will be tested at one-sided 2.5% level of significance.

H_{01} (null hypothesis): $\Theta 1 \geq 0$ vs. H_{a1} (alternative hypothesis): $\Theta 1 < 0$

Where $\Theta 1$ is the log hazard ratio of PFS in the investigational arm vs. control arm.

In the primary analysis, PFS will be tested using the log-rank test stratified by randomization stratification factors when approximately 253 PFS events are observed. The statistical basis for a claim of efficacy will be the statistical significance (at the 1% one-sided level of significance if OS is not statistically significant at interim and final analysis or 2.5% one-sided level of significance if OS is statistically significant at either interim or final analysis) for PFS in favor of the canakinumab arm.

The distribution of PFS will be estimated using the Kaplan-Meier method. The median PFS and PFS rate at different timepoints (i.e. 3, 6, 9, 12, 15 mos) along with 95% confidence intervals (CIs) will be presented by treatment arms. A Cox regression model stratified by randomization stratification factors will be used to estimate the hazard ratio (HR) of PFS, along with 95% CI based on the Wald test.

2.5.2.2.3 Overall survival (OS)

Assuming a proportional hazards model for OS, the null hypothesis will be tested at one-sided 1.5% level of significance. If PFS by investigator assessment per RECIST1.1 is statistically significant at 1% level of significance, then the 1 % alpha assigned to PFS will be transferred to OS and OS will be tested at one-sided 2.5% level of significance.

H_{02} (null hypothesis): $\Theta_2 \geq 0$ vs. H_{a2} (alternative hypothesis): $\Theta_2 < 0$

Where Θ_2 is the log hazard ratio of OS in the investigational arm vs. control arm.

In the primary analysis, OS will be tested using the log-rank test stratified by randomization stratification factors. The statistical basis for a claim of efficacy will be the statistical significance (at the 1.5% one-sided level of significance if PFS is not statistically significant at 1% one-sided level of significance or 2.5% one-sided level of significance if PFS is statistically significant at 1% one-sided level of significance) for OS in favor of the canakinumab arm.

The distribution of OS will be estimated using the Kaplan-Meier method. The median OS and OS rate at different timepoints (i.e., 6, 12, 18, 24 mos) along with 95% confidence intervals (CIs) will be presented by treatment arms. A Cox regression model stratified by randomization stratification factors will be used to estimate the hazard ratio (HR) of OS, along with 95% CI based on the Wald test.

2.5.2.3 Handling of missing values/censoring/discontinuations

A subject whose disease has not progressed or died by the date of the analysis cut-off will have their PFS censored at the time of the last adequate tumor evaluation performed on or before the cut-off date. Clinical deterioration will not be considered as documented disease progression. PFS events will be included in the analysis if it occurs after one missing assessment. PFS will be censored at the last adequate tumor assessment if a subject didn't have an event or the event occurred after two or more consecutive missing tumor assessments. Censoring rules for PFS follow the RECIST 1.1 guidelines and details can be found in section 2.5.2.5.5.

A subject who has not died by the date of the analysis cut-off date will have their OS censored at the last known date the subject is alive. Details on the censoring rules are provided on section 2.5.2.5.5.

2.5.2.4 Checking proportional hazard assumption

Visual checks of the proportional hazard assumption will be performed based on the plots (SURVIVAL LOGSURV LOGLOGS) generated by LIFETEST procedure in SAS (Section 5.4.1). No formal analysis will be generated. If proportional hazard assumption doesn't hold good, additional methods may be explored to assess the treatment effect. Further details will be provided in a separate post-hoc analysis plan.

2.5.2.5 Supportive analyses

For all sensitivity and supplementary estimand analyses, nominal p-values will be presented without any multiplicity adjustments (except for PFS BIRC analysis). So, no statistical significance conclusion should be performed. The following summaries will be provided: Kaplan-Meier estimates, estimate of the median PFS/OS along with 95% confidence interval, and hazard ratio obtained using the Cox proportional hazards model.

2.5.2.5.1 Sensitivity estimand analysis related to primary endpoints

Progression-free survival

First sensitivity analysis estimand: the target population, the primary variable, the treatment of interest, intercurrent events and the summary measure of this endpoint are the same as for the primary estimand. To assess the impact of stratification, the two treatment arms will be

compared using the unstratified log-rank test. The HR together with the associated 95% confidence interval obtained using the unstratified Cox regression model will also be presented. In the summary tables, this approach is referred as 'unstratified PFS sensitivity analysis'.

Second sensitivity analysis estimand: the target population, the primary variable, the treatment of interest, intercurrent events and the summary measure of this endpoint are the same as for the primary estimand. The two treatment arms will be compared using the stratified log-rank test with stratification factors derived from the clinical database, in case at least 5% of the subjects have discrepancies between strata at randomization and strata derived from the CRF data. The HR together with the associated 95% confidence interval obtained using the stratified Cox regression model will also be presented. In the summary tables, this approach is referred as 'PFS sensitivity analysis with stratification factors from eCRF'.

Third sensitivity analysis estimand: the target population, the primary variable, intercurrent events, treatment of interest and the summary measure of this endpoint are the same as for the primary estimand. PFS event observed after 1, 2 or more missing assessment will result in considering the PFS event based on the actual date of the event in the analysis. In the summary tables, this approach is referred as 'actual event PFS sensitivity analysis'.

Fourth sensitivity analysis estimand: the target population, the primary variable, intercurrent events, treatment of interest and the summary measure of this endpoint are the same as for the primary estimand. PFS event observed after 1, 2 or more missing assessment will result in backdating the PFS event at the date of the next scheduled tumor assessment. In the summary tables, this approach is referred as 'backdate event PFS sensitivity analysis'.

Overall survival

The first and second sensitivity analyses estimand for PFS will be conducted for OS. They will be named as 'unstratified OS sensitivity analysis' and 'OS sensitivity analysis with stratification factors from eCRF', respectively.

2.5.2.5.2 Supplementary estimand analysis related to primary endpoints

Progression-free survival

First supplementary analysis estimand: the target population, the primary variable, the treatment of interest and intercurrent events are the same as for the primary estimand. A Cox regression model stratified by randomization stratification factors will be fitted to evaluate the effect of other baseline demographic and disease characteristics on the estimated hazard ratio. The fitted model adjusting the treatment difference for key baseline and prognostic factors will include as covariates the following: age category (<65 years vs. ≥65 years), gender (male vs. female), ECOG status (0 vs. 1), smoking history (former/current vs. never). In the summary tables, this approach is referred as 'PFS supplementary analysis adjusted for baseline covariates'.

Second supplementary analysis estimand: PFS assessed by Blinded Independent Review Committee (BIRC) will serve as supportive evidence of the primary endpoint. The treatment of interest, the intercurrent events and the summary measure of this endpoint are the same as for the primary estimand. The primary variable is PFS based on BIRC assessments and the target population is defined as the randomized subjects selected for the BIRC assessment.

Third supplementary analysis estimand: the target population, the primary variable, intercurrent events and the summary measure of this endpoint are the same as for the primary estimand. The treatment of interest is the randomized treatment (canakinumab arm or the matching placebo arm without any antineoplastic therapy post randomization). An additional intercurrent event of antineoplastic therapy post randomization prior to PFS event will be included and handled using hypothetical estimand strategy i.e., PFS will be censored at the date of the last adequate assessment prior to the start of new anticancer therapy if no PFS event is observed prior to the start of new antineoplastic therapy). In the summary tables, this approach is referred as 'new anticancer therapy leading to censoring PFS supplementary analysis'.

Fourth supplementary analysis estimand: the target population, intercurrent events and the summary measure of this endpoint are the same as for the primary estimand. The treatment of interest is the randomized treatment (canakinumab arm or the matching placebo arm without any antineoplastic therapy post randomization). The primary variable is defined as the time from the date of randomization to the date of the first documented disease progression based on local investigator assessment as per RECIST 1.1 or date of death due to any cause or date of new antineoplastic therapy, whichever occurs first. Any new antineoplastic therapy post randomization prior to progressive disease/death will result in PFS event (composite strategy) (consider as PFS event at the start of the new antineoplastic therapy if no PFS event occurred before). In the summary tables, this approach is referred as 'new anticancer therapy leading to PFS event supplementary analysis'.

In the third and fourth supplementary analysis estimands, a new anticancer therapy is defined as any systemic anticancer therapy or any radiotherapy other than palliative for bone pain received during or post study treatment. Definitive cancer related surgery to either debulk tumor or remove an isolated metastasis would be considered as a new anticancer therapy. Surgery that is not related to the underlying NSCLC or done for alleviation of symptoms or adverse clinical consequences from the tumor such as, but not limited, to relieving an impending spinal cord compression, fracture from a lytic lesion or pain relief will not be considered as a new anticancer therapy. Continuation of pembrolizumab monotherapy or in combination with pemetrexed as 1st new anti-neoplastic therapy after end of treatment without prior PD and collected in the "Post antineoplastic therapy –medications" eCRF, will not be considered as a new anticancer therapy.

Additional supplementary estimand analysis to assess COVID-19 pandemic

Fifth supplementary analysis estimand:

In light of COVID-19 pandemic and its potential impact on treatment effect, an important scientific question of interest to address is to estimate the treatment effect based on PFS had COVID-19 pandemic not occurred. The target population, treatment of interest and the summary measure of this endpoint are the same as for the primary estimand. The primary variable is defined as the time from the date of randomization to the date of the first documented disease progression based on local investigator assessment as per RECIST 1.1 or date of death due to non-COVID-19 pandemic reasons, whichever occurs first. The remaining intercurrent events will be handled as follows:

- **Discontinuation of study treatment due to any non-COVID-19 pandemic reasons**: PFS will take into account all PFS events irrespective of the study treatment discontinuation reasons (treatment policy)

- **Discontinuation of study treatment due to COVID-19 pandemic reasons**: PFS will be censored at the date of the last adequate assessment prior to discontinuation of treatment due to COVID-19 pandemic (hypothetical strategy). The discontinuation reason due to COVID-19 pandemic will be identified from the defined COVID-19 protocol deviations.
- **Medications used for treating suspected or confirmed COVID-19 cases:** PFS will be censored at the date of the last adequate assessment prior to the administration of COVID-19 medication (hypothetical strategy). The medications will be identified by drug class of aminoquinolines, protease inhibitors, glucocorticoids, macrolides, antiviral or antiretroviral OR drug names that contain key word of "hydroxychloroquine" or "chloroquine".
- **Death due to COVID-19**: PFS will be censored at the date of the last adequate assessment prior to death due to COVID-19 (hypothetical strategy)

In the summary tables, this approach is referred as 'COVID-19 PFS supplementary analysis'.

A summary of the censoring reasons of COVID-19 PFS supplementary analysis will be provided by treatment arm. Additional analyses to assess the impact of intercurrent event of treatment interruption/delay due to COVID-19 pandemic may be explored as post hoc analyses.

Overall survival

The first supplementary analysis estimand for PFS will be conducted on OS and will be named as 'OS supplementary analysis adjusted for baseline covariates'.

In addition, analysis to assess the treatment effect based on OS had COVID-19 pandemic not occurred will also be conducted. The target population, treatment of interest and the summary measure of this endpoint are the same as for the primary estimand. The primary variable is defined as the time from the date of randomization to the date of death due to non-COVID-19 pandemic reasons. The remaining intercurrent events will be handled as follows:

- **Discontinuation of study treatment due to any non-COVID-19 pandemic reasons**: OS will take into account all deaths irrespective of the study treatment discontinuation reasons (treatment policy)
- **Discontinuation of study treatment due to COVID-19 pandemic reasons**: OS will be censored on the date of discontinuation of treatment due to COVID-19 pandemic (hypothetical strategy). The discontinuation reason due to COVID-19 pandemic will be identified from the defined COVID-19 protocol deviations.
- **Medications used for treating suspected or confirmed COVID-19 cases:** OS will be censored on the date of administration of COVID-19 medication (hypothetical strategy). The medications will be identified by drug class of aminoquinolines, protease inhibitors, glucocorticoids, macrolides, antiviral or antiretroviral OR drug names that contain key word of "hydroxychloroquine" or "chloroquine".
- **Death due to COVID-19**: OS will be censored on the date of death due to COVID-19 (hypothetical strategy)

In the summary tables, this approach is referred as 'COVID-19 OS supplementary analysis'.

A summary of the censoring reasons of COVID-19 OS supplementary analysis will be provided by treatment arm.

2.5.2.5.3 Other supportive analyses

Data analysis from BIRC assessment

For studies with PFS based on local radiology assessment as the primary endpoint, PFS assessment done centrally has generally been used as a secondary or supplementary estimand analysis of the treatment effect observed in the primary efficacy analysis. Although 100% central review of scans has been performed in many trials, there is a growing body of evidence that an audit based approach for central evaluation is sufficient (Zhang et al, 2012, FDA ODAC 2012). An audit (sample) based approach will therefore be implemented for the BIRC assessment of PFS, whereby all assessments for a randomly selected subset of randomized subjects will be assessed by BIRC. An independent random sampling process, implemented by the third party IRT vendor, will select approximately 40% of randomized subjects. This random allocation will be stratified by randomized treatment arm and the strata used for the randomization of subjects to treatment arms.

Two additional methods will be used to summarize the data from the BIRC assessment in order to decide whether a 100% BIRC review should be conducted.

1. The NCI (National Cancer Institute) method (Dodd et al. 2011), uses an auxiliary variable estimator of the log-hazard ratio that combines information from subject-level investigator assessment from all subjects and the BIRC assessment of these subjects randomly selected for central review (see Section 5.4.2 for methodological details). This estimate and its one-sided 95% CI will be provided. The NCI method will be used for audit sample size determination (see Section 3.2.3) and summary of treatment effect (HR, 95% confidence intervals) based on the supportive BIRC assessment.
2. The data from the BIRC assessment generated following the sampling scheme as described above will also be summarized using the method proposed by Amit et al. 2011, referred to as the PhRMA (Pharmaceutical Research and Manufacturers in America) method, based on the early discrepancy rate (EDR) and late discrepancy rate (LDR).The EDR quantifies the frequency with which the investigator declares progression early relative to BIRC within each arm as a proportion of the total number of investigator assessed PDs. The LDR quantifies the frequency that the investigator declares progression later than BIRC as a proportion of the total number of discrepancies within the arm. If the distribution of discrepancies is similar between the arms this suggests the absence of evaluation bias favoring a particular arm (see Section 5.4.2 for details on the calculations). With this approach, the differential discordance (DD) of the early discrepancy rate (EDR) and late discrepancy rate (LDR) between the two arms will be estimated as the difference between canakinumab+pembrolizumab+chemotherapy arm and placebo+pembrolizumab+chemotherapy arm. The EDR and LDR results will also be summarized by treatment arm.

The following thresholds based on the NCI and PhRMA methods will be used to define the trigger for a full BIRC review:

- If the upper-bound of the one-sided 95% confidence interval for BIRC-based loghazard ratio exceeds zero (i.e. HR>1) based on the NCI method

  and/or

- If ≥ 15% differential discordance is observed in EDR or LDR based on the PhRMA method (i.e. ≤ -15% differential discordance for EDR or ≥ 15% differential discordance for LDR between treatment arms)

Concordance analysis of PFS

Cross-tabulation of 'PFS by central radiology' vs. 'PFS by investigator' by PFS event type (i.e. 'death', 'PD', 'censor' for each of the two sources) and by treatment will be constructed to investigate discordance between the two sources (in subjects selected for BIRC assessment). The discrepancy rate between central radiology and investigator will be summarized by treatment arm.

A cross-tabulation will be produced displaying the PFS timings for the local investigators' assessment compared to the BIRC assessment (in subjects selected for BIRC assessment). For progression assessments, the frequency and percent of subjects with complete agreement [occurring on the same date plus or minus 7 days of each other], progression later, progression earlier, and cases where progression was called by one method and censored by the other will be displayed. Similarly, if censoring was recorded, the frequency and percent of subjects with complete agreement, censoring called later, censoring called earlier, and cases where censoring was called by one method and progression was called by the other method will be displayed.

Missing tumor assessments

The number of subjects with at least one missing/not evaluable (NE) TA based on local assessment will be presented together with the following breakdown categories: number of subjects with 1, 2, 3, 4, 5, >5 missing/NE TAs. The purpose of this analysis is to gain an insight as to whether the TAs have been carried out in accordance with the protocol and to understand if any meaningful discrepancies exist between the pattern of missing assessments by treatment arms.

Since the planned tumor assessments are every 6 weeks in the first 12 weeks, then every 9 weeks through week 75 and every 12 weeks thereafter, the following time windows (in weeks) will be constructed for each subject:

(note the open parenthesis such as (9, 16.5] indicates that week 9 doesn't belong to this interval and week 9+1 day belongs to that interval)

- Until ~12 weeks post randomization [0, 9], (9, 16.5]
- After ~12 weeks and until week 75 post randomization (16.5, 25.5], (25.5, 34.5], (34.5, 43.5], (43.5, 52.5], (52.5, 61.5], (61.5, 70.5], (70.5, 81]

- After week 75 post randomization, (81, 93], (93, 105], (105, 117], …

where '0' is the subject's date of randomization. Every time-window (with the exception of the initials, broader ones) is centered at the scheduled time of a TA, i.e., around week 12, week 30 for second and fourth window respectively, etc. A subject will be considered 'at risk' of missing his/her TA for any one of these time-windows if he/she either:

- is 'on study' for at least the first 3 weeks of the time-window for the first 12 weeks (6 weeks for the first time window i.e. [0, 9]), or at least the first 4.5 weeks of time-window thereafter until week 75, or at least the first 6 weeks of time-window thereafter, i.e., if the subject is ongoing at the time of the scheduled TA, or
- discontinued treatment due to documented disease progression within the specific time window.

For example, if a subject discontinued due to documented disease progression during Week 24, then he/she would have been 'at risk' of a missing/NE TA for the (16.5, 25.5] week time window.

For the purpose of this analysis, 'NE' TAs (i.e., evaluations with an overall lesion response of 'NE') will be considered to be missing. However, a clear distinction between 'truly missing' and 'present but NE' needs to be made in the derived dataset to allow for both a combined analysis, i.e. missing and NE treated the same, and separate analyses.

TAs performed after a documented disease progression will not be considered. In other words, the final time-window for which a subject would be at risk of a missing/NE scan would be that during which the documented progression occurred.

For subjects without documented progression, all TAs are considered up to the earliest of the following dates: death, the analysis cut-off, disease progression, withdrawal of consent or loss to follow-up.

The disease progression mentioned in the above analyses is the first PD observed as per RECIST 1.1.

[redacted]

2.5.2.5.4 Subgroup analyses for the primary endpoints

If the primary analysis for PFS and/or OS are statistically significant, subgroup analyses to assess the homogeneity of the treatment effect across demographic and baseline disease characteristics will be performed for the subgroups specified in Section 2.2.3 and China subgroup, defined as subjects randomized in sites from China across the endpoints and using the same conventions as for the primary analysis. If sample size of China subgroup is small, then an un-stratified Cox model will be used.

2.5.2.5.5 Censoring pattern

Progression-free survival

Determination of missing adequate assessments

The term 'missing adequate tumor assessment' is defined as a tumor assessment not done or tumor assessment with overall lesion response 'NE'. For the sake of simplicity, a 'missing adequate tumor assessment' will also be referred to as a 'missing assessment'.

As described in Table 16-5 in the Appendix 16.1 of the study protocol, the PFS censoring and event date options depend on the presence and the number of missing tumor assessments (TAs). In the primary analysis of PFS, an event occurring after two or more missing assessments is censored at the last adequate tumor assessment, i.e. the last assessment preceding the missing assessment(s).

An exact rule to determine whether there is no, one or two missing TAs is therefore needed. This rule is based:

- on the time interval between the last adequate tumor assessment (LATA) date and the event date
- on the time interval between the last adequate tumor assessment (LATA) date and the cut-off date

In this study, the protocol defined schedule of assessments is

- every 6 weeks during the first 12 weeks (i.e. at weeks 6, 12),
- every 9 weeks through week 75 (i.e. at weeks 21, 30, 39, 48, 57, 66, 75)
- and every 12 weeks thereafter (i.e. at weeks 87, 99, 111, 123, 135 etc…) up to the primary PFS analysis

The scheduled date of tumor assessments (in weeks from randomization), protocol specified windows for tumor assessments, and the thresholds for LATA to belong to a visit can be found in the following table.

Table 2-4 Schedule for tumor assessment and time windows

Assessment schedule		Scheduled date – 1 week	Scheduled date (weeks from randomization)	Scheduled date +1 week	Threshold (weeks)*
	Baseline	0	0^	1	n/a

Assessment schedule		Scheduled date – 1 week	Scheduled date (weeks from randomization)	Scheduled date +1 week	Threshold (weeks)*
Every 6 weeks for the first 12 weeks	C3D1	5	6	7	9
	C5D1	11	12	13	16.5
Every 9 weeks through week 75	C8D1	20	21	22	25.5
	C11D1	29	30	31	34.5
	C14D1	38	39	40	43.5
	C17D1	47	48	49	52.5
	C20D1	56	57	58	61.5
	C23D1	65	66	67	70.5
	C26D1	74	75	76	81
Every 12 weeks thereafter	C30D1	86	87	88	93
	C34D1	98	99	100	105
	…				

* The threshold correspond to the mid-point between current and next visit (except for baseline) and to the upper limit for LATA to be matched to a certain scheduled assessment. e.g. if LATA is at week 10, this is after threshold for C3D1 and before that for C5D1, so the matching scheduled assessment is C5D1.

^ Day of randomization is taken as 0.

To calculate the number of missing tumor assessments, the LATA before an event is matched with a scheduled tumor assessment using the time window in the table above (essentially whichever scheduled assessment it is closest to).

D1 and D2 are calculated for that scheduled assessment based on the protocol-specified schedule and windows.

- The threshold D1 is defined as the protocol-specified time interval between the TAs plus 2x the protocol-allowed time window around the assessments (i.e. = 2*1 week=2 weeks).
- The threshold D2 is defined as twice the protocol-specified time interval between the TAs plus 2x the protocol-allowed time window around the assessments (except when the matched scheduled tumor assessment is C3D1 and C23D1, in which case D2 is defined in Rule 1 and 3 below).

Since there is a change of schedule for tumor assessments after first 12 weeks and after week 75, D1 and D2 are defined differently depending on when LATA occurs. Here are the rules based on when LATA occurs:

- Rule 1: if LATA happens within 9 weeks of randomization (≤ study day 63), the matched scheduled tumor assessment is C3D1, D1=6+2=8 weeks (=56 days) and D2=6+9+2=17 weeks (119 days).
- Rule 2: if LATA happens after 9 weeks from randomization (≥ study day 64), the matched scheduled tumor assessment is C5D1 or after, but within 61.5 weeks (≤study day 430) (the matched scheduled tumor assessment is C20D1), D1=9+2=11 weeks (=77 days) and D2=9*2+2=20 weeks (=140 days).
- Rule 3: if LATA happens after 61.5 weeks from randomization (≥ study day 431) but within 70.5 weeks (≤ study day 493) (the matched scheduled tumor assessment is C23D1), D1=9+2=11 weeks (=77 days) and D2=9+12+2=23 weeks (=161 days).
- Rule 4: if LATA happens after 70.5 weeks (≥ study day 494) from randomization (the matched scheduled tumor assessment is C26D1 or later), D1=12+2=14 weeks (=98 days) and D2=12*2+2=26 weeks (=182 days).

The number of missing events is defined as:

- An event after LATA+D1 weeks will be considered as having ≥1 missing assessment
- An event after LATA+D2 weeks will be considered as having ≥2 missing assessments

The same definition of D2 will be used to determine the PFS censoring reason. If there is no post-baseline adequate tumor assessment available (before an event or a censoring reason occurred), the randomization date will be used to compute the interval.

If the time interval between the last adequate TA date and the earliest of the following dates:

- Analysis cut-off date
- Date of consent withdrawal
- Date of study treatment discontinuation due to lost to follow-up or end of post-treatment follow-up discontinuation due to lost to follow-up.

is smaller or equal to D2 days, then the PFS censoring reason will be respectively:

- ‘Ongoing without event’
- ‘Withdrew consent’
- ‘Lost to follow-up’

However if the time interval is larger than D2 days with no event then the PFS censoring reason will always default to ‘Adequate assessment no longer available’. If the time interval between the last adequate tumor assessment date and the PFS event date is larger than D2 then the subject will be censored and the censoring reason will be 'Event documented after two or more missing tumor assessments'.

No baseline tumor assessment

As described in Table 16.5 in the Appendix 16.1 of the study protocol, since the timing of disease progression cannot be determined for subjects with missing baseline tumor assessment, these subjects are censored in the PFS analysis at the date of randomization. This rule, however, only applies to the ‘progressive disease’ component of the PFS assessment. Subjects without

any baseline tumor assessment who die within D2 time interval from date of randomization will be counted as having an event in the analysis of PFS at the date of death.

A summary of the censoring reasons of the primary PFS analysis will be provided by treatment arm.

Overall survival

Number of subjects censored for the OS analysis will be summarized. In addition, a summary of reasons for OS censoring will be provided by treatment arm.

The pattern of censored data will be examined between the treatment arms: reasons for censoring ('Alive' or 'Lost to follow-up') will be summarized by treatment arm. Subjects not known to have died will be censored for 'Lost to follow-up' if the time between their last contact date and the analysis cut-off date is longer than 12 + 2 weeks = 14 weeks = 98 days (i.e. the planned interval between two OS follow-up visits plus the 1 week window on either side). Otherwise, subjects will be censored for 'Alive'.

2.6 Analysis of secondary efficacy objective(s)

2.6.1 Double-blind randomized part (part 2)

The secondary efficacy endpoints will be assessed using the FAS. The following analyses will be performed based on local investigator assessment unless otherwise specified.

2.6.1.1 Overall response rate

Overall response rate (ORR) is defined as the proportion of subjects with a best overall response (BOR) of complete response (CR) or partial response (PR) as per local investigator assessment. ORR will be evaluated according to RECIST 1.1. ORR based on RECIST1.1 will be calculated based on the FAS and according to the intent-to-treat (ITT) principle. ORR and its 95% exact confidence interval (Clopper 1934) will be presented by treatment group. Waterfall plot will also be generated (section 2.6.1.2).

The BOR will be determined from response assessments undertaken while on treatment. In addition, only tumor assessments performed before the start of any further anti-neoplastic therapies (i.e. same definition of new anticancer therapy as in section 2.5.2.5.2) will be considered in the assessment of BOR.

BOR for each subject is determined from the sequence of overall (lesion) responses according to the following rules:

- CR = at least two determinations of CR at least 4 weeks apart before progression.
- PR = at least two determinations of PR or better at least 4 weeks apart before progression (and not qualifying for a CR).
- SD = at least one SD assessment (or better) > 5 weeks after randomization (and not qualifying for CR or PR).

- PD = progression ≤ 13 weeks after randomization (and not qualifying for CR, PR or SD)
- NE = all other cases (i.e. not qualifying for confirmed CR or PR and without SD after more than 5 weeks or early progression within the first 13 weeks)

Complete and partial responses must be confirmed by repeat assessments that should be performed not less than 4 weeks after the criteria for response are first met.

Subjects with 'NE' BOR will be summarized by reason for having NE status. The following reasons will be used:

- No valid post-baseline assessment
- All post-baseline assessments have overall lesion response NE
- New anti-neoplastic therapy started before first post-baseline assessment
- SD too early
- PD too late

Note 1: A SD is considered as "SD too early"' if the SD is documented within first 5 weeks after randomization date.

Note 2: A PD is considered as "PD too late" if the first documentation of PD is recorded more than 13 weeks after randomization date with no qualifying CR, PR or SD in between.

Note 3: Special (and rare) cases where BOR is "NE" due to both too early SD and too late PD will be classified as "SD too early".

2.6.1.2 Construction of waterfall graphs

Waterfall graphs will be used to depict the anti-tumor activity based on local investigator assessment. These plots will display the best percentage change from baseline in the sum of diameters of all target lesions for each subject. Only subjects with measurable disease at baseline will be included in the waterfall graphs.

Special consideration is needed for assessments where the target lesion response is CR, PR or SD, but the appearance of a new lesion or a worsening of non-target lesions results in an overall lesion response of PD. As a conservative approach, such assessments will not be considered for display as bars in the graph, since the percentage change in the sum of diameters of target lesions reflects the non-PD target lesion response, but the overall lesion response is PD. A subject with only such assessments will be represented by a special symbol (e.g.✶) in the waterfall graph.

Assessments with "NE" target lesion response and assessments with NE overall response will be excluded from the waterfall plots. Subjects without any valid assessments will be completely excluded from the graphs.

The total number of subjects displayed in the graph will be shown and this number will be used as the denominator for calculating the percentages of subjects with tumor shrinkage and tumor growth. Footnote will explain the reason for excluding some subjects (due to absence of any valid assessment).

All possible assessment scenarios are described in Table 2-5.

Table 2-5 Assessments considered for calculation of best percentage change for waterfall graphs

Case	Target response	Overall lesion response	Calculate % change from baseline in sum of diameters?
1	NE	Any	No, exclude assessment
2	Any	NE	No, exclude assessment
3	CR/PR/SD	PD	No, flag assessment with ✶
4	PD	PD	Yes
5	CR/PR/SD	CR/PR/SD	Yes

Based on the above considerations, the following algorithm will be used to construct the graph:

1. Select "valid" post-baseline assessments to be included, i.e. for each subject and each assessment repeat the following four steps.
   1.1. Check the target lesion response and overall lesion response. If at least one of them is NE then exclude the whole assessment. Otherwise, go to step 1.2.
   1.2. Check the overall lesion response. If it is PD then go to step 1.3. Otherwise go to step 1.4.
   1.3. Check target response. If it's PD then go to step 1.4. Otherwise flag the assessment with ✶ .
   1.4. Calculate the % change from baseline in target lesions.
2. For each subject, go through all valid assessments identified in step 1 and find the assessment with best % change from baseline in target lesions. The "best" means best for the subject, i.e. the largest shrinkage or if a subject only has assessments with tumor growth take the assessment where the growth is minimal.
3. Construct the waterfall graph displaying the best % change from baseline for each subject. Subjects having only ✶ flagged assessment(s) will be displayed separately.

The graph will be constructed using the data from the investigator/local radiologist assessments.

The best overall response (BOR) will be shown above each of the displayed bars in the graph, if the number of subjects displayed in the graph is small enough for the labels to be legible.

The order of the display from left to right will be as follows:

- Bars under the horizontal axis representing tumor shrinkage
- Bars above the horizontal axis representing tumor growth
- "Zero" bars with ✶ symbol.

For each of the 3 categories above, n (%) (where % uses the total number of subjects displayed in the graph) will be displayed. If there are any subjects with zero change they will be as a separate category following subjects with tumor shrinkage.

2.6.1.3 Disease control rate (DCR)

Disease control rate (DCR) is defined as the proportion of subjects with BOR of CR, PR, or SD as per local investigator assessment.

DCR will be evaluated according to RECIST 1.1. DCR based on RECIST1.1 will be calculated based on the FAS and according to the ITT principle. DCR and its 95% exact confidence interval (Clopper 1934) will be presented by treatment group.

2.6.1.4 Time to response (TTR)

Time to response (TTR) is defined as duration of time between the date of randomization and the date of first documented response of either CR or PR, which must be subsequently confirmed (although date of initial response is used, not date of confirmation).

TTR will be evaluated based on local investigator assessment according to RECIST 1.1.

All subjects in the FAS will be included in TTR calculations. Subjects without a confirmed CR or PR will be censored at the study-maximum follow-up time (i.e., LPLV-FPFV) for subjects with a PFS event (i.e., disease progression or death due to any cause), or at the date of the last adequate tumor assessment for subjects without a PFS event. TTR will be listed and summarized by treatment group based on RECIST1.1. The distribution function of TTR will be estimated using the Kaplan-Meier method. The median TTR along with 95% CIs will be presented by treatment arm.

2.6.1.5 Duration of response (DOR)

Duration of response (DOR) is defined as the duration of time between the date of first documented response (CR or PR) and the date of first documented progression or death due to any cause.

DOR will be evaluated based on local investigator assessment according to RECIST 1.1.

Duration of response (DOR) only applies to subjects whose best overall response is complete response (CR) or partial response (PR) based on tumor response data per local review. If a subject has not had an event, DOR is censored at the date of last adequate tumor assessment. Subjects who never achieved a BOR of CR or PR will be excluded from the analysis. The distribution function of DOR will be estimated using the Kaplan-Meier method. The median DOR along with 95% CIs will be presented by treatment arm.

2.6.1.6 Duration of follow-up

Study follow-up will be summarized using the following methods:

- Summary of duration between randomization and cut-off date, and follow-up times for PFS/OS, which are defined as follows:
  - Duration between randomization and data cut-off date = (Cut-off date – Date of randomization + 1) / 30.4375 (months). This item will be summarized overall.
  - Follow-up time = (Date of event or censoring – Date of randomization + 1) / 30.4375 (months) regardless of censoring. Date of censoring is defined as the last adequate tumor assessment date for PFS or last contact date for OS. This item will be summarized by treatment arm.

All summaries will be reported in months. The calculations for PFS will be based on local assessment. Date of censoring is the same as defined for the PFS and OS analysis.

In addition, median time to censoring will be computed by reversing censoring variable and performing Kaplan-Meier analysis based on Schemper 1996.

2.6.2 Safety run-in part (part 1)

Overall response rate (ORR), Disease control rate (DCR), Duration of response (DOR) by investigator's assessment according to RECIST 1.1 will be presented in FAS by canakinumab dose regimen and cohort as described in section 2.6.1 for the randomized part. Waterfall plot will also be generated for each cohort with canakinumab dose regimen as the plot legend.

[redacted]

2.8 Analyses of secondary objective: Safety

Safety set will be used for all safety analyses.

2.8.1 Double-blind randomized part (part 2)

2.8.1.1 Adverse events (AEs)

All safety outputs will use the safety set and will be presented by treatment group. The safety summary tables will include only 'on-treatment' events/assessments. The AEs started before the first dose but worsening during the treatment period are also considered as 'on-treatment' events. All safety events/assessments will be listed and those collected outside the on-treatment window will be flagged.

AEs will be summarized by number and percentage of subjects having at least one AE in each primary system organ class (SOC) and for each preferred term (PT) using MedDRA coding. A subject with multiple occurrences of an AE will be counted only once in the respective AE category. A subject with multiple CTCAE grades (version 5.0) for the same preferred term will be summarized under the maximum CTCAE grade recorded for the event. AEs with missing CTCAE grade will be included in the 'All grades' column of the summary tables.

The following adverse event summaries will be produced by treatment arm:

- Overview of adverse events and deaths (number and % of subjects with any AE, treatment-related AE, SAE, fatal AE, AE leading to any permanent study drug discontinuation and each drug component, AE leading to dose reduction, AE leading to dose interruption, or AE requiring additional therapy),
- AEs by SOC and PT (all AEs and AEs related to at least one study drug component),
- AEs by PT (all AEs and AEs related to at least one study drug component) and by maximum CTC grade

- AEs leading to permanent discontinuation of one of the study drug component, AEs leading to canakinumab permanent discontinuation
- AEs leading to dose reduction (i.e. dose reduced for chemotherapy drugs or dosing interval increase for canakinumab/matching-placebo)
- AEs leading to dose interruption
- AEs requiring additional therapy
- SAEs and SAEs leading to fatal outcome.
- Summary of SAEs and non-SAEs with number of occurrences will be produced (an occurrence is defined as >1 day between start and prior end date of record of same preferred term).

2.8.1.2 Adverse events of special interest / grouping of AEs

An adverse event of special interest (AESI) is a grouping of adverse events that are of scientific and medical concern specific to the compound canakinumab. These groupings are defined using MedDRA terms, SMQs (standardized MedDRA queries), HLGTs (high level group terms), HLT (high level terms) and PTs (preferred terms). Customized SMQs (Novartis MedDRA queries, NMQ) may also be used. A NMQ is a customized group of search terms which defines a medical concept for which there is no official SMQ available or the available SMQ does not completely fit the need. It may include a combination of single terms and/or an existing SMQ, narrow or broad. The groups are defined according to the MedDRA terms defined in the program Case Retrieval Strategy (CRS) document and will be summarized. The latest version of the CRS document available at the time of the analyses will be used. For each specified AESI, number and percentage of subjects with at least one event of the AESI occurring during on treatment period will be summarized.

Summaries of these AESIs will be provided by treatment arm, (specifying grade, SAE, relationship, leading to treatment discontinuation, leading to dose interruption, hospitalization, death etc.).

Table 2-6 provides AESI groupings.

Table 2-6 AESI groupings

AESI grouping
Infections
Opportunistic infections
Neutropenia
Abnormal liver parameters
Thrombocytopenia
Immunogenicity/allergenicity
Autoimmunity reactions
Malignancy
Interactions with vaccines
Interactions with drugs eliminated by CYP450 enzymes

Pulmonary complications: pulmonary hypertension and interstitial lung disease
Injection site reactions

Time to onset of AESI

Time to onset of CTC grade ≧ 2 AESI will be summarized using the Kaplan-Meier method by AESI grouping and treatment arm. Median time to onset and 95% CI will be provided and ascending Kaplan-Meier plots will be generated by treatment arm. This will only be performed for AESI groupings of infections, opportunistic infections, neutropenia, abnormal liver parameters and thrombocytopenia (if there are sufficient number of events).

Time to onset of CTC grade ≧ 2 AESI is defined as the time from the start of treatment to the start date of the first incidence of an event of CTC grade ≧ 2 i.e. time in days is calculated as (start date of first occurrence of the event) – (date of first dose of study treatment) +1.

In the absence of an event during the on-treatment period, the censoring date applied will be the earliest of the following dates:

- end date of on-treatment period (end of study treatment + 130 days).
- death date
- start date of new antineoplastic therapy before experiencing any CTC grade ≧ 2 AESI.
- data cut-off date.
- withdrawal of informed consent date

The same definition of new anticancer therapy as in section 2.5.2.5.2 will be used. The same analysis will be repeated for time to onset of CTC grade ≧ 3 AESI with similar censoring rules applied.

A listing of all grouping levels down to the MedDRA preferred terms used to define each AESI will be generated.

2.8.1.3 Deaths

Separate summaries for on-treatment and all deaths will be produced by treatment arm, system organ class and preferred term.

All deaths will be listed, post treatment deaths will be flagged. A separate listing of deaths prior to starting treatment will be provided for all screened subjects.

2.8.1.4 Laboratory data

When analyzing laboratory related data, all sources (central and local laboratories) will be used and combined. The summaries will include all assessments available for the laboratory parameter collected no later than 130 days after the last study treatment administration date (Section 2.1.3.5). Grading of laboratory values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 5.0.

The following summaries will be produced for hematology and biochemistry laboratory data (by laboratory parameter and treatment):

- Worst post-baseline CTC grade (regardless of the baseline status). Each subject will be counted only for the worst grade observed post-baseline.
- Shift tables using CTC grades to compare baseline to the worst on-treatment value (hypo and hyper worst grade will be summarized separately if applicable)
- For laboratory tests where CTC grades are not defined, shift tables using the low/normal/high/ (low and high) classification to compare baseline to the worst on-treatment value.
- Trends of selected laboratory parameter values (defined in the TFL shells) over time (baseline and selected on-treatment timepoints) will be displayed via boxplots based on time windows and corresponding tables displaying the statistics used for the box plots for the selected time points. For a subject with multiple assessments in a time window, the average value will be used. The time windows are defined in Table 2-7 for selected scheduled assessments

Table 2-7 Time window for laboratory assessments

Assessment	Target day of assessment	Time window definition
Baseline	On or before study day 1[a]	<= Study day 1
During treatment phase		
Week 3	Study Day 22	Study Days 2 to 32
Week 6	Study Day 43	Study Days 33 to 53
Week k Day 1	Study Day = 21*(k/3)+1	Study Day
(with k=9, 12, 15, 18 etc..)		21*(k/3)-9 to 21*(k/3)+11 For last cycle of dosing : from 21*(k/3)-9 to end of treatment visit date +7 "Note: EOT data visit are included if obtained within 7 days of permanent discontinuation of study treatment"
After treatment discontinuation		
Safety follow-up 1*	26 days after treatment discontinuation	[end of treatment visit date+8; end of treatment visit date + 52]
Safety follow-up 3*	78 days after treatment discontinuation	[end of treatment visit +53; end of treatment visit + 104]

Safety follow-up 5*	130 days after treatment discontinuation	[end of treatment visit+105; end of treatment visit + 156]

[a] Study Day 1 = first treatment date

The following listings will be produced for the laboratory data:

- Listings of all laboratory data, with CTC grades and classification relative to the laboratory normal range. Lab data collected during the post-treatment period will be flagged.
- Listing of all CTC grade 3 or 4 laboratory toxicities

Liver function parameters

Liver function parameters of interest are total bilirubin (TBL), ALT, AST and alkaline phosphatase (ALP). The number (%) of subjects with worst post-baseline values as per Novartis Liver Toxicity guidelines will be summarized by treatment arm:

The following summaries will be produced:

- ALT > 3xULN
- ALT > 5xULN
- ALT > 10xULN
- ALT > 20xULN
- AST > 3xULN
- AST > 5xULN
- AST > 10xULN
- AST > 20xULN
- ALT or AST > 3xULN
- ALT or AST > 5xULN
- ALT or AST > 8xULN
- ALT or AST > 10xULN
- ALT or AST > 20xULN
- TBL > 2xULN
- TBL > 3xULN

Combined elevations post-baseline:

For AST and ALT ≤ ULN at baseline

- ALT or AST > 3xULN & TBL > 2xULN
- ALT or AST > 3xULN & TBL > 2xULN & ALP < 2xULN
- ALT or AST > 3xULN & TBL > 2xULN & ALP ≥ 2xULN

For ALT or AST > ULN at baseline (Bsl)

- Elevated ALT or AST (*) & TBL (> 2x Bsl and 2xULN)
- Elevated ALT or AST (*) & TBL (> 2x Bsl and 2xULN) & ALP < 2xULN
- Elevated ALT or AST (*) & TBL (> 2x Bsl and 2xULN) & ALP ≥ 2xULN

* Elevated AST or ALT defined as: >3x ULN if =< ULN at baseline, or (>3x Bsl or 8x ULN) if > ULN at baseline

In addition, a listing of all TBILI, ALT, AST and ALP values for subjects with a post-baseline TBILI > 2xULN, ALT> 3xULN or AST > 3xULN will be provided.

Time to onset of grade 2 or worse and time to onset of grade 3 or worse liver function test abnormalities will be summarized for the following liver function test parameters:

- AST or ALT (whichever occurs first)
- Total bilirubin

Time to onset will be summarized using Kaplan-Meier method. Median time to onset and 95% C.I. will be summarized.

Time to onset of LFT abnormalities is defined as the time from the start of treatment to the start date of the first incidence of grade 2 or worse LFTs post-baseline (or grade 3 or worse), i.e., time in days is calculated as (start date of first occurrence of LFT abnormalities) – (date of first dose of study treatment) +1. A subject will be censored for time to onset if:

- the subject dies without experiencing the LFT abnormality
- the subject receives a new anticancer therapy without experiencing the LFT abnormality or before LFT abnormality has occurred
- the subject discontinues from the study treatment without experiencing the LFT abnormality (up to 130 days after study treatment discontinuation)
- the subject is still ongoing at the analysis cut-off without experiencing the LFT abnormality

The same definition of new anticancer therapy as in section 2.5.2.5.2 will be used. The censoring date will be the earliest date from the following dates: last date of study treatment in the treatment phase + 130 days, analysis cut-off, the day before new anticancer therapy start date, death date and last on-treatment laboratory sampling date (for the particular parameter) during on-treatment period. For the time to onset of grade 2 or worse LFTs, subjects with grade 2 or worse at baseline will be excluded from the analysis. For the time to onset of grade 3 or worse LFTs, subjects with grade 3 or worse at baseline will be excluded from the analysis.

Time to onset of neutropenia and thrombocytopenia

Time to onset of grade 2 or worse and time to onset of grade 3 or worse neutropenia and thrombocytopenia will be summarized respectively. The censoring rule will be the same as used in time to onset of LFT abnormalities analysis.

2.8.1.5 Other safety data

2.8.1.5.1 ECG and cardiac imaging data

ECG abnormality for baseline and any post baseline assessments will be summarized by treatment arm. A listing of all ECG assessments will be produced by treatment arm and notable values will be flagged.

2.8.1.5.2 Vital signs

Vital sign assessments are performed in order to characterize basic body function. The following parameters were collected: height (cm), weight (kg), body temperature (°C), heart rate (beats per minute), systolic and diastolic blood pressure (mmHg).

Data handling

Vital signs collected on treatment will be summarized. Values measured outside of on treatment period will be flagged in the listings.

Data analysis

For analysis of vital signs the clinically notable vital sign criteria are provided in Table 2-8 below.

Table 2-8 Clinically notable changes in vital signs

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
Weight (kg)	increase ≥ 10% from Baseline	decrease ≥ 10% from Baseline
Systolic blood pressure (mmHg)	≥180 with increase from baseline of ≥20	≤90 with decrease from baseline of ≥20
Diastolic blood pressure (mmHg)	≥105 with increase from baseline of ≥15	≤50 with decrease from baseline of ≥15
Pulse rate (bpm)	≥100 with increase from baseline of >25%	≤50 with decrease from baseline of > 25%
Respiratory rate (breaths/minute)	>25	<12
Body temperature (°C)	≥39.1	-

The number and percentage of subjects with notable vital sign values (high/low) will be presented by treatment arm.

A listing of all vital sign assessments will be produced by treatment arm and notable values will be flagged. In the listing, the assessments collected outside of on-treatment period will be flagged.

2.8.1.5.3 Other safety data

Other safety data (e.g. data relating to liver events) will be listed in the safety set.

Data from other tests will be listed, notable values will be flagged, and any other information collected will be listed as appropriate.

All assessments collected later than 130 days after the last treatment date will be flagged in the listings.

Any statistical tests performed to explore the data will be used only to identify any interesting comparisons that may warrant further consideration.

2.8.2 Safety run-in part (part 1)

Summaries of DLTs, AEs, treatment related AEs, SAEs, on-treatment deaths, AESI overview, Laboratory shift table based on key hematologic and biochemistry terms and Liver transaminase abnormality will be summarized by canakinumab dose regimen and by cohort. In addition, serious and non-serious adverse events with number of occurrences will be produced (an occurrence is defined as >1 day between start and prior end date of record of same preferred term). The listings specified for AE, laboratory, death, and other safety data in Section 2.8.1 will be generated by canakinumab dose regimen and by cohort.

2.9 Analyses of secondary objective: Pharmacokinetic endpoints

2.9.1 Double-blind randomized part (part 2)

Pharmacokinetic data analysis will be performed for each drug of the study treatment (canakinumab, pembrolizumab, pemetrexed, carboplatin, cisplatin, paclitaxel, nab-paclitaxel). PK concentration analyses will be repeated for each study drug

PAS-study drug (i.e. PAS-canakinumab, PAS-pembrolizumab, PAS-pemetrexed, PAS-cisplatin, PAS-carboplatin, PAS-paclitaxel, PAS-nab-paclitaxel), will be used in the pharmacokinetic data analysis for each study drug concentrations, separately.

PK concentrations

Descriptive statistics (n, m (number of non-zero concentrations), arithmetic mean, CV% mean, SD, median, geometric mean, CV% geo-mean, minimum and maximum) for study drug concentration will be presented at each scheduled time point for the PAS-study drug.

Individual concentration-time profiles for study drug evaluable concentrations with median will be displayed graphically for PAS-study drug on the semi-log view. In addition, the mean (+/- SD) and geometric mean concentration-time profiles for study drug over time will be displayed graphically for PAS-study drug on the linear and semi-log view. Only time points with $n \geq 4$ observations will be shown on the figure.

All individual plasma study drug concentration data will be listed by treatment arm for the FAS.

PK parameters

The descriptive statistics (n, mean, CV%, standard deviation (SD), median, geometric mean, geometric CV%, minimum and maximum) will be presented by treatment for PK parameters of canakinumab, pembrolizumab, pemetrexed, cisplatin, carboplatin, paclitaxel and nab-paclitaxel except Tmax, where only n, median, minimum and maximum will be presented. PK parameters such as those listed in Table 2-9 will be estimated and reported, when applicable. PK parameters will be derived based on the non-compartmental methods, when derivation of selective PK parameters is feasible, using Phoenix WinNonlin® software version 6.4.

Table 2-9 Noncompartmental pharmacokinetic parameters

AUClast	The AUC from time zero to the last measurable concentration sampling time (tlast) (mass x time x volume-1)
AUCinf	The AUC from time zero to infinity (mass x time x volume-1)
AUCtau	The AUC calculated to the end of a dosing interval (tau) at steady-state (amount x time x volume-1)
Cmin or Ctrough	The minimum observed plasma or serum drug concentration (mass x volume-1)
Cmax	The maximum (peak) observed plasma or serum drug concentration (mass x volume-1)
Tmax	The time to reach maximum (peak) plasma or serum drug concentration (time)
Lambda_z	Smallest (slowest) disposition (hybrid) rate constant (time-1) may also be used for terminal elimination rate constant (time-1)
T1/2	The elimination half-life associated with the terminal slope (λz) of a semi logarithmic concentration-time curve (time). Use qualifier for other half-lives
CL	The total body clearance of drug from the plasma or serum (volume x time-1)
Vz	The apparent volume of distribution during terminal phase (associated with λz) (volume)
Note: Not all PK parameters are applicable to all study drugs	

Handling of PK data below LLOQ or missing

All concentration values below the lower limit of quantitation (LLOQ) are set to zero by the Bioanalyst, and will be displayed in the listings as zero and flagged. LLOQ values will be treated as zero in any calculations of summary statistics, and treated as missing for the calculation of the geometric means and their CV%. The number of non-zero concentrations will also be reported in the summary statistics.

Missing values for any PK data will not be imputed and will be treated as missing.

Population pharmacokinetic analysis

In this analysis, the data from safety run-in and randomization parts will be pooled. If there is adequate amount of data, a mixed-effects model may be applied to the serum canakinumab concentration-time data from this study along with other studies to generate post-hoc estimates of pharmacokinetic parameters using NONMEM to characterize canakinumab exposure and to determine the effects of intrinsic (i.e. demographic factors) and extrinsic covariates (e.g.

concomitant medications, formulation) on canakinumab exposure. If there is sufficient data for analysis, the details of the population pharmacokinetic analyses may be provided in a separate reporting and analysis plan, and the results may be reported in a separate population pharmacokinetic report.

2.9.2 Safety run-in part (part 1)

All analyses except population pharmacokinetic analysis specified in Section 2.9.1 will be performed for canakinumab, pembrolizumab, pemetrexed, cisplatin, carboplatin, paclitaxel concentrations collected in Safety run-in part.

2.10 Analyses of secondary objective: Immunogenicity

In the double-blind randomized part, immunogenicity data will be summarized using the safety set. Immunogenicity will be characterized descriptively by tabulating anti-drug antibodies (ADA) prevalence at baseline and ADA incidence on-treatment. A shift table of subjects with positive or negative anti-canakinumab antibodies and anti-pembrolizumab antibodies overall and by visit will be produced ("positive" corresponds to "worst"). A listing will be provided by subject with supporting information (i.e. ADA sample status at each timepoint (including titer for positive samples) and subject ADA status). In addition, a listing will also be provided for subjects with neutralizing antibodies (NAB) testing results.

Immunogenicity data collected in safety run-in part will be listed by canakinumab dose regimen and by cohort. Immunogenicity data for pembrolizumab collected in safety run-in will also be listed.

2.11 Analyses of secondary analyses: ECOG performance status

The analyses below will be provided for the safety run-in part by canakinumab dose regimen and cohort and for the double-blind randomized part by treatment group.

The performance status will be assessed according to the Eastern Cooperative Oncology Group (ECOG) performance status scale as specified in Table 2-10 below, ranging from 0 (most active) to 5 (least active).

Table 2-10 ECOG performance status

Grade	ECOG status
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature e.g., light house work, office work
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
5	Dead

The time windows as defined in section 2.1.3.10 will be used to group the ECOG PS data over time. If 2 assessments within a time window are equidistant from the target date, the assessment obtained prior to visit will be considered. If the closest assessment to the target date has two ECOG filled out on the same date, then the worst ECOG PS value will be used.

Time windows are applicable for descriptive summary of ECOG data by visit only.

Frequency counts and percentages of subjects in each score category will be provided by treatment arm and time point.

2.12 Analyses of secondary objective: patient-reported outcomes

The following PRO analyses will be performed in the double-blind randomized phase part of the study.

Three patient-reported outcomes (PRO) questionnaires will be assessed: EORTC QLQ-C30, with its QLQ-LC13 lung cancer module, and the EQ-5D-5L. QLQ-C30 and QLQ-LC13 will be considered as the primary scale. Scoring of PRO data and methods for handling of missing items or missing assessments will be handled according to the scoring manual and user guide for each respective subject questionnaire (Fayers 2001, Van Reenen 2015). No imputation procedures will be applied for missing items or missing assessments.

The FAS will be used for analyzing PRO data. For all PRO analysis, nominal p-values will be presented without any statistical inference since there is no adjustment for multiplicity. The baseline is defined as the last PRO assessment on or prior to randomization. Time to definitive 10 point deterioration symptom scores for each of chest pain, cough and dyspnea per QLQ-LC13 questionnaire are the three primary PRO variables of interest. Utilities derived from EQ-5D-5L, together with time to definitive deterioration in global health status/QoL, shortness of breath and pain per QLQ-C30 are the secondary PRO variables of interest. The time to definitive 10-point deterioration is defined as the time from the date of randomization to the date of event, which is defined as at least 10 points absolute worsening from baseline of the corresponding scale score, with no later change below this threshold, i.e. <10 points was observed or if this worsening was observed at the last assessment for the subject, or death due to any cause (whichever occurs earlier). If a subject has not had an event, time to definitive deterioration will be censored at the date of the last adequate assessment. If deterioration occurs at the last adequate assessment, this will also be considered definitive. Subjects receiving any further anti-neoplastic therapy (same definition of new anticancer therapy as in section 2.5.2.5.2) before definitive worsening will be censored at the date of their last assessment before starting this therapy. If a definitive deterioration is observed after two or more missing assessments, subject is censored at the date of their last available questionnaire prior to the deterioration. Subjects with no baseline data will be censored at the randomization date.

Death is considered as a deterioration event when it occurs within a period of time defined by 2 times the period between two assessments as planned in the study protocol. This avoids overestimating the time to definitive worsening in subjects dying after an irregular assessment scheme. Subjects who die after more than twice the planned period between two assessments since the last assessment will be censored at the date of their last adequate assessment.

Censoring reasons for time to definitive deterioration will be summarized.

All assessments will be included in the time to definitive deterioration analysis. The distribution will be presented descriptively using Kaplan-Meier curves. Summary statistics from Kaplan-Meier distributions will be determined, including the median time to definitive 10 point deterioration along with two-sided 95% confidence interval. Log-rank test stratified by randomization stratification factors will be performed. A Cox regression stratified by randomization stratification factors from IRT will be used to estimate the hazard ratio (HR), along with two-sided 95% confidence interval.

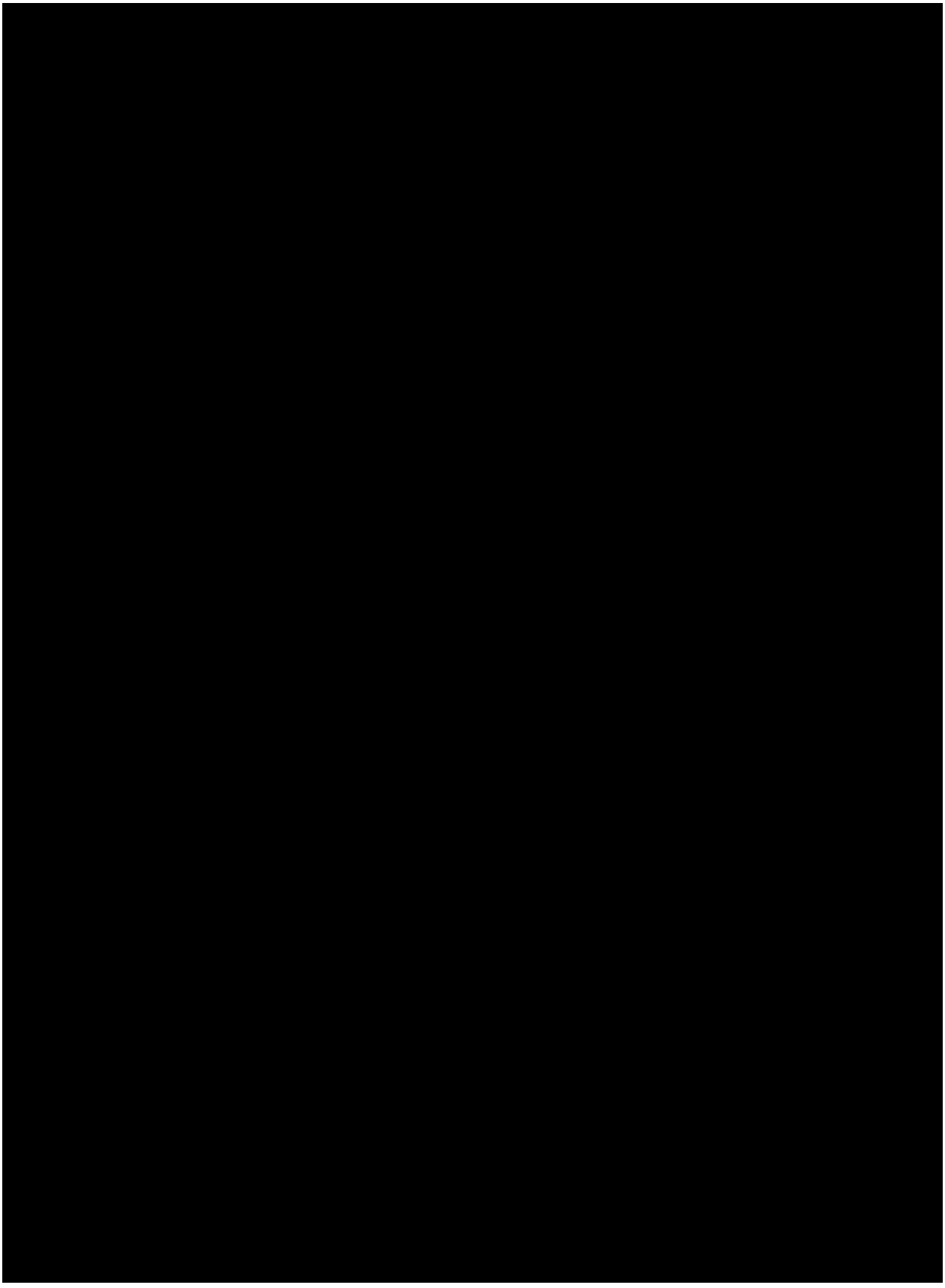
Descriptive statistics will be used to summarize the original scores, as well as change from baseline, of the QLQC30, QLQ-LC13, and EQ-5D-5L at each scheduled assessment time point for each treatment arm using time windows as described in section 2.1.3.10. Additionally, change from baseline in the scale and subscale values at the time of each assessment will be summarized. Subjects with an evaluable baseline score and at least one evaluable post baseline score during the treatment period will be included in the change from baseline analyses.

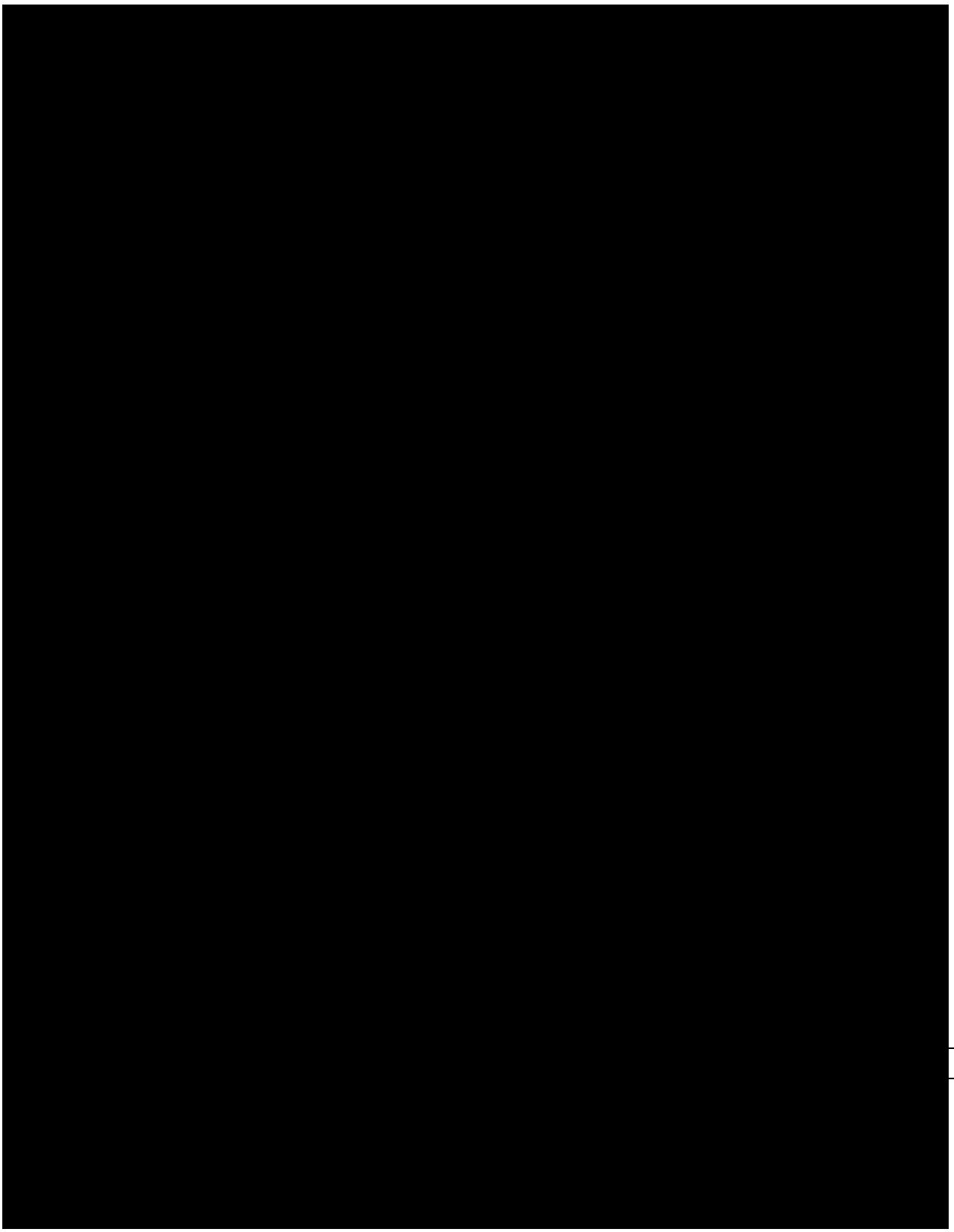
In addition, a repeated measures model for longitudinal data will be used to estimate differences in EORTC QLQ-C30/QLQ-LC13 domains as well as the VAS and utility scores of the EQ-5D-5L between treatment arms. The models will be first based on all assessments. In addition, the models will be repeated to include only assessments collected within 130 days of last study treatment. Any assessments collected after the start of further anti-tumor therapy will not be included in the models.

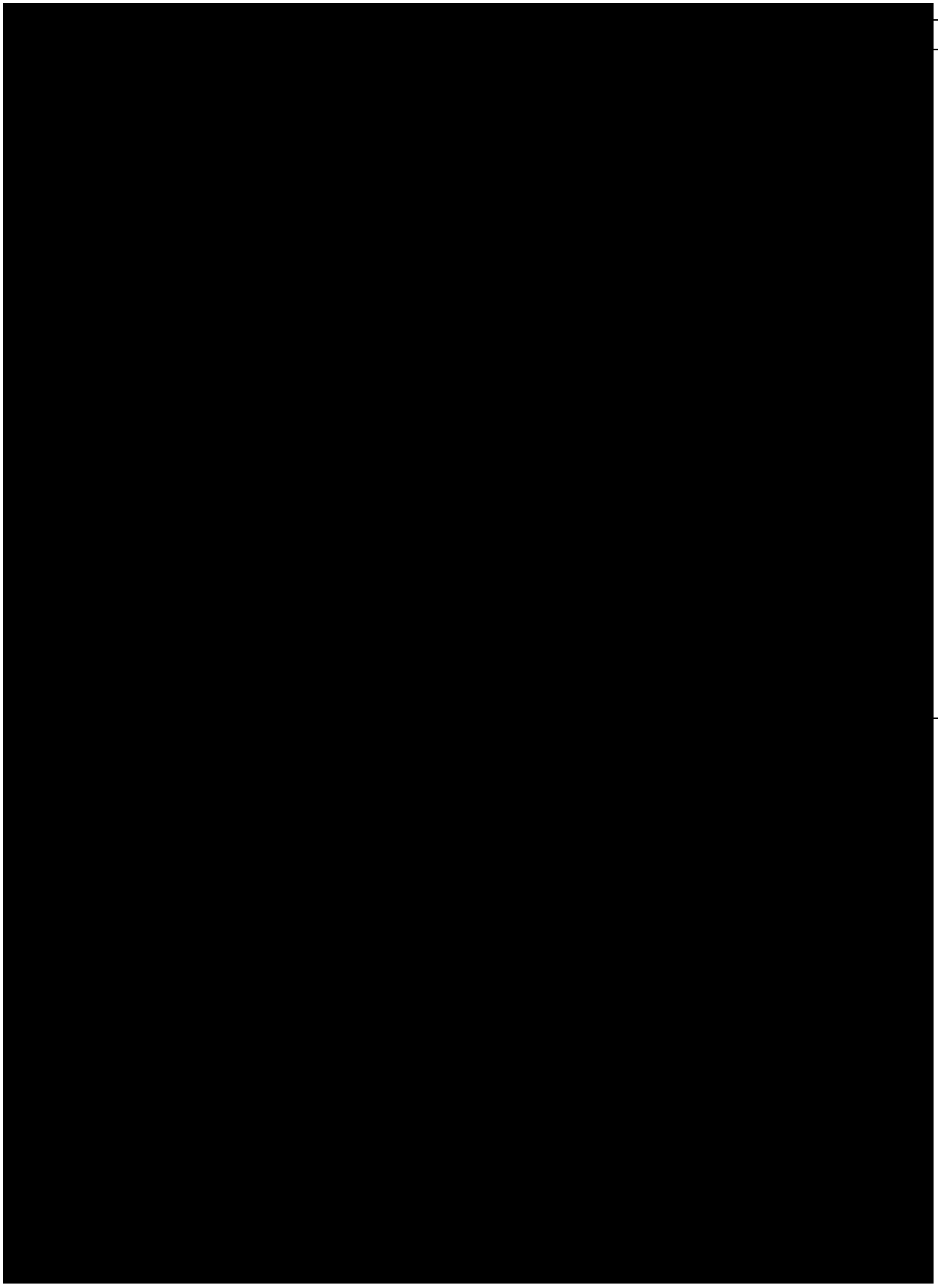
The modeling will mainly be done on the actual score. Note that the modeling of the change in score or the actual score is equivalent since adjustment for baseline score is considered. This repeated measures model will include terms for treatment, the stratification factors, time, baseline value as main effects, and an interaction term for treatment by time. This analysis will be restricted to subjects with an evaluable baseline score and at least one evaluable post-baseline score. The differences in least square means between the treatment arms and corresponding 95% confidence interval at selected time points will be presented.

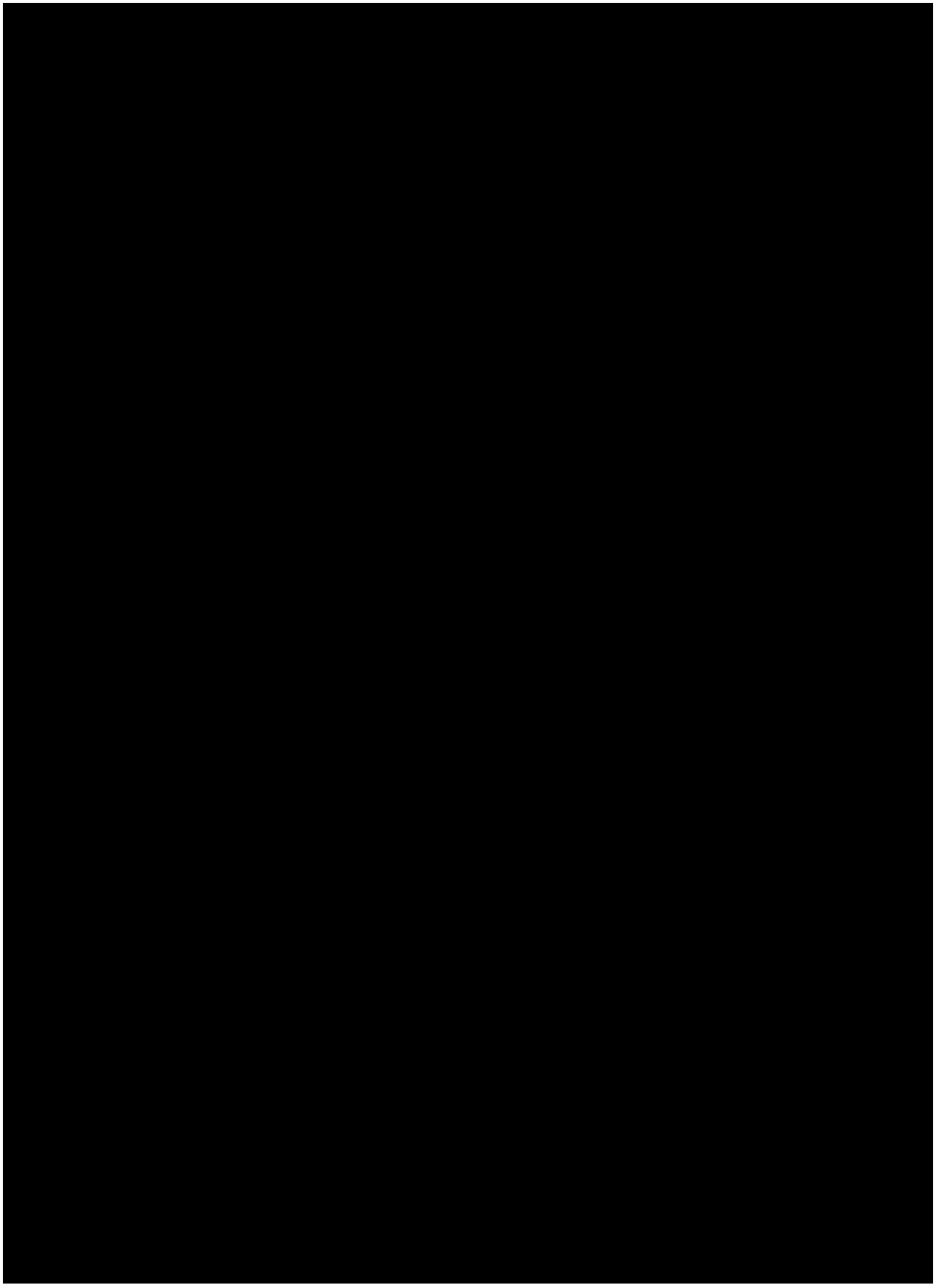
Time will be considered as a continuous variable expressed in weeks, i.e. considering that PRO follow a linear trend. As a first approach, an unstructured correlation matrix will be used to model the correlation within subjects. Other structures of the correlation matrix, including AR(1), may be investigated and simplified using likelihood ratio tested if appropriate.

If PRO is found not to follow a linear trend, the time variable might be considered as a categorical variable instead of continuous in the model.









2.16 Interim analysis

2.16.1 Safety run-in part (part 1)

Not applicable.

2.16.2 Double-blind randomized part (part 2)

There is no interim analysis (IA) for PFS. The primary PFS analysis, is planned after approximately 253 PFS events have been observed. This is expected to occur approximately 18 months after the date of first subject randomized in part 2. The 1st IA for OS will be performed at the same time of primary PFS analysis when approximately 120 over the 304 targeted OS events (39%) deaths are expected to be observed. The intent of this analysis is to assess superior efficacy with either PFS and/or OS results. There is no intent to assess futility at the time of the 1st interim OS analysis.

A maximum of three analyses will be performed for OS. The type I error probability will be controlled by using a Lan-DeMets (O'Brien-Fleming) alpha spending function (Lan and DeMets 1983) at the 1.5% (if PFS was not statistically significant) or 2.5% (if PFS was statistically significant) overall level of significance.

Based on the choice of alpha-spending function described above and if the first interim OS analysis is performed exactly at 120 OS events, the efficacy boundary expressed on the p-value scale (or the Z-statistic scale, HR scale) at the interim is calculated as:

- p <0.0001082 (i.e. Z =3.699 or equivalently HR=0.509) if OS is tested at one-sided 1.5% alpha level
- p<0.0003604 (i.e. Z=3.382 or equivalently HR=0.539) if OS is tested at one-sided 2.5% alpha level

The observed (i.e., nominal) p-value has to be less than the p-value scale efficacy boundary defined above (or equivalently the observed z-statistic has to be > Z-statistic scale boundary or HR < HR scale boundary) to declare statistical significance at 1.5% one-sided alpha level or 2.5% one-sided alpha level.

If the study continues to the 2nd IA for OS, it will be performed when approximately 243 OS events have been documented. If exactly 120 OS events are observed at the 1st IA for OS, the study continued and exactly 243 events are obtained at the 2nd IA for OS, the observed p-value will have to be less than 0.006 (or equivalently Z-statistic has to be >2.485, HR < 0.727) to declare statistical significance at 1.5% one-sided alpha level, or will have to be less than 0.012 (or equivalently Z-statistic has to be >2.255, HR <0.749) to declare statistical significance at

2.5% one-sided alpha level. The 2nd interim OS analysis is anticipated to take place approximately 30 months after the first subject is randomized.

Since the observed number of events at the interim analyses may not be exactly equal to the planned 120 OS events at first interim OS analysis (or 243 OS events at 2nd interim analysis), the efficacy boundaries will need to be re-calculated using the pre-specified α-spending functions and based on the actual number of observed events at the corresponding interim and the total number of targeted events to calculate the exact information fraction. The observed p-value (or Z-test statistic or HR estimate) at each interim analysis will then be compared against the re-calculated efficacy boundaries.

If the study continues to the final OS analysis, the final OS analysis will be performed when approximately 304 OS events have been documented. In practice, the final analysis will be based on the actual number of OS events documented at the cut-off date for the final OS analysis and the alpha already spent at the interim analyses. The boundary for the final analysis will be derived accordingly from the pre-specified α-spending function such that the overall significance level across all analyses is maintained at 0.015 or 0.025 (if PFS results are found statistically significant at 1% level). The final OS analysis is expected to occur approximately 38 months after the first subject is randomized.

The statistical properties of the group sequential design are summarized in Table 2-12 below.

Table 2-12 Simulated probabilities to stop for overall survival efficacy at primary PFS analysis, interim or final OS analyses

Scenario	Look	# deaths	Simulated cumulative probabilities	Simulated incremental probabilities
			Stop for efficacy	Stop for efficacy
If OS is tested at one-sided alpha=0.015				
Under H_0 (HR=1)	Primary PFS (1st interim OS)	120	0.02%	0.02%
	2nd interim OS	243	0.65%	0.63%
	Final OS	304	1.44%	0.80%
Under H_a (HR=0.69)	Primary PFS (1st interim OS)	120	4.5%	4.5%
	2nd interim OS	243	65.0%	60.5%
	Final OS	304	84.6%	19.6%
If OS is tested at one-sided alpha=0.025				
Under H_0 (HR=1)	Primary PFS (1st interim OS)	120	0.04%	0.04%
	2nd interim OS	243	1.25%	1.21%
	Final OS	304	2.40%	1.15%
Under H_a (HR=0.69)	Primary PFS (1st interim OS)	120	8.6%	8.6%
	2nd interim OS	243	72.9%	64.3%

Scenario	Look	# deaths	Simulated cumulative probabilities	Simulated incremental probabilities
			Stop for efficacy	Stop for efficacy
	Final OS	304	89.0%	16.1%

Note: Simulation is performed in East 6.4 with number of simulations=30, 000 and simulation seed = 1234.

3 Sample size calculation

3.1 Safety run-in part (part 1)

No formal statistical power calculations to determine sample size were performed for this part of the study. In the case that the starting dose (canakinumab 200 mg s.c. Q3W) with the fixed dose combination of pembrolizumab plus cohort specific platinum-based doublet chemotherapy is confirmed to be safe and tolerated, the safety run-in part is expected to enroll approximately 9 subjects in each treatment cohort in order to have at least 6 evaluable subjects, so approximately 27 subjects enrolled in total. Otherwise, additional subjects are foreseen to be enrolled to assess additional cohorts at canakinumab dose level-1 Q6W.

3.2 Double-blind, randomized part (part 2)

The sample size of the study is based on the 2 primary endpoints PFS and OS. The hypotheses to be tested and details of the testing strategy are described in Section 2.5.2.2.

3.2.1 Progression-free survival (PFS)

Based on KEYNOTE-189 study (Gandhi et al 2018), the median PFS for pembrolizumab in combination with platinum-based doublet chemotherapy followed by pembrolizumab maintenance treatment (control arm) in subjects with non-squamous NSCLC is approximately 9 months. Based on KEYNOTE-407 (Paz-Ares 2018), the median PFS for pembrolizumab in combination with taxane-based chemotherapy followed by pembrolizumab maintenance treatment (control arm) in subjects with squamous NSCLC is 6.4 months. Based on the assumption that 30% subjects in this current study (CACZ885U2301) will have squamous histology (Rittmeyer et al 2017, Hellmann et al 2018), the median PFS in the control arm for this study is expected to be approximately 8 months based on simulation. Under the assumption that the median PFS in the control arm is 8 months, it is expected that addition of canakinumab will result in a 37.5% reduction in the PFS hazard rate (corresponding to an increase in median PFS from 8 months to 12.8 months under the exponential model assumption). If the true HR=0.625 (under the alternative hypothesis), a total of approximately 253 PFS events are required to have 92% power at a one-sided 1% level of significance if OS is not significant or 96.2% power at 2.5% level of significance if OS is significant to reject the null hypothesis (HR=1) using a log-rank test. If the primary PFS analysis is performed when the targeted 253 PFS events are observed, the observed hazard ratio will have to be <0.746 or <0.782 to declare statistical significance at a one-sided 1% alpha level or 2.5% alpha level, respectively. Assuming a recruitment period of approximately 15 months from the start of the

double-blind, randomized, placebo-controlled part with an accrual rate of 10 subjects/months for first 2 months, 20 subjects/months from 2 to 4 months, 40 subjects/months from 4 to 6 months and 50 subjects/month thereafter, along with an assumed 10% dropout rate/year for PFS, approximately 600 subjects will need to be randomized to the two treatment arms in a 1:1 ratio. Given the above assumptions it is estimated that the events for the primary PFS analysis will occur at approximately 18 months from the date of first subject randomized in the study.

3.2.2 Overall survival

Based on KEYNOTE-189 (Gandhi et al 2018) , the 12-month OS rate for the subjects with non-squamous NSCLC randomized in pembrolizumab in combination with platinum-based doublet chemotherapy followed by pembrolizumab maintenance treatment (control arm) is 69.2%. Assuming an exponential distribution, median OS is derived to be approximately 23 months. Based on KEYNOTE-407 (Paz-Ares 2018), the median OS for pembrolizumab in combination with taxane-based chemotherapy followed by pembrolizumab maintenance treatment (control arm) in subjects with squamous NSCLC at the time of the 2nd IA is 15.9 months (95% CI: 13.2, NE). This median estimate may be immature given the high censoring rate and only ~30% of subjects had events. Therefore, the 9-month OS rate of 73% estimated graphically from the Kaplan-Meier curve was used to derive the median OS for the control arm in squamous NSCLC, which is calculated to be approximately 20 months assuming an exponential distribution. Based on the assumption that 30% subjects in the current study (CACZ885U2301) will have squamous histology, the median OS in the control arm for this study is expected to be approximately 22 months based on simulation. Under the assumption that the median OS in the control arm is 22 months, it is assumed that addition of canakinumab will result in a 31% reduction in the OS hazard rate (corresponding to an increase in median OS from 22 months to 31.9 months under the exponential model assumption). If the true HR=0.69 (under the alternative hypothesis), a total of approximately 304 deaths are required to have 85% power at a one-sided 1.5% level of significance if PFS is not significant or 89.3% power at 2.5% level of significance if PFS is significant to reject the null hypothesis (HR=1) using a log-rank test. If the final OS analysis is performed when the targeted 304 OS events are observed after exactly 120 OS events and 243 OS events have been observed at 1st and 2nd interim analysis respectively, the observed hazard ratio will have to be <0.775 or <0.793 to declare statistical significance at one-sided 1.5% alpha level or 2.5% alpha level respectively. A 5% dropout rate/year for OS is assumed. The interim analyses for OS are planned at the time of primary PFS analysis and when approximately 80% of deaths are observed (expected at approximately 30 months from the date of first subject randomized). Events required for the final OS analysis are expected at approximately 38 months from the date of first subject randomized in the study.

These calculations were performed using EAST 6.4.

3.2.3 Audit size for BIRC assessed PFS

The audit size of the sample-based BIRC assessment will be 40% of all randomized subjects based on the audit size calculation approach proposed by Dodd et al 2011, assuming investigator and BIRC assessments are similar and the estimated log of investigator-based HR is -0.47 (i.e. HR=0.625). The audit size of 40% will ensure that the upper bound of a one-sided 95% CI for BIRC-based log-hazard ratio has 91 % probability of being below 0 (i.e. HR<1) if

the correlation between investigator assessment and BIRC assessment is 0.8 (the estimated correlation approximated based on data from the Novartis [CLDK378A2303] and [CLDK378A2301] studies in NSCLC)

4 Change to protocol specified analyses

Not applicable

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

No imputation of the start date of infusion/injection or end date for infusion will be applied. Complete dates are required as per eCRF.

5.1.2 AE, ConMeds and safety assessment date imputation

Table 5-1 Imputation of start dates (AE, CM) and assessments (LB, EG, VS)

Missing Element	Rule
day, month, and year	• No imputation will be done for completely missing dates

day, month	• If available year = year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY
	  ○ Else set start date = study treatment start date.
	• If available year > year of study treatment start date then 01JanYYYY
	• If available year < year of study treatment start date then 01JulYYYY

day	• If available month and year = month and year of study treatment start date then
	  ○ If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY.
	  ○ Else set start date = study treatment start date.
	• If available month and year > month and year of study treatment start date then 01MONYYYY
	• If available month and year < month year of study treatment start date then 15MONYYYY

Table 5-2 Imputation of end dates (AE, CM)

Missing Element	**Rule** (*=last treatment date plus 130 days not > (death date, cut-off date, withdrawal of consent date))
day, month, and year	• Completely missing end dates (incl. ongoing events) will be imputed by the end date of the on-treatment period*
day, month	• If partial end date contains year only, set end date = earliest of 31DecYYYY or end date of the on-treatment period *
day	• If partial end date contains month and year, set end date = earliest of last day of the month or end date of the on-treatment period*

Any AEs and ConMeds with partial/missing dates will be displayed as such in the data listings.

Any AEs and ConMeds which are continuing as per data cut-off will be shown as 'ongoing' rather than the end date provided.

5.1.2.1 Other imputations

Not applicable.

5.1.3 Handling missing month/day in date of death/last known subject alive from survival eCRF page

Partial date imputation is allowed for event (death)/censoring if coming from 'Survival information' eCRF. Here are the following imputation rules:

- When only the day is missing, the last known subject alive date (from survival page) or date of death will be imputed to the maximum between 1st of the month and year available and any valid date used for last contact derivation + one day
- When the day and months are missing, the last known subject alive date (from survival page) or date of death will be imputed to the maximum between 1st of January of the year and any valid date used for last contact derivation + one day

5.2 AEs coding/grading

Adverse events are coded using the Medical dictionary for regulatory activities (MedDRA) terminology.

AEs will be assessed according to the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0

The CTCAE represents a comprehensive grading system for reporting the acute and late effects of cancer treatments. CTCAE grading is by definition a 5-point scale generally corresponding to mild, moderate, severe, life threatening, and death. This grading system inherently places a value on the importance of an event, although there is not necessarily proportionality among grades (a grade 2 is not necessarily twice as bad as a grade 1).

5.3 Laboratory parameters derivations

Grade categorization of lab values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) version 5.0. The calculation of CTCAE grades will be based on the observed laboratory values only, clinical assessments will not be taken into account. The criteria to assign CTCAE grades are given in Novartis internal criteria for CTCAE grading of laboratory parameters. The latest available version of the document based on the underlying CTCAE version 5.0 at the time of analysis will be used. For laboratory tests where grades are not defined by CTCAE version 5.0, results will be graded by the low/normal/high (or other project-specific ranges, if more suitable) classifications based on laboratory normal ranges.

A severity grade of 0 will be assigned for all non-missing lab values not graded as 1 or higher. For laboratory tests that are graded for both low and high values, summaries will be done separately and labelled by direction, e.g., sodium will be summarized as hyponatremia and hypernatremia.

Imputation Rules

CTC grading for blood differentials is based on absolute values. However, this data may not be reported as absolute counts but rather as percentage of WBC.

If laboratory values are provided as '<X' (i.e. below limit of detection) or '>X', prior to conversion of laboratory values to SI unit, these numeric values are set to X.

The following rules will be applied to derive the WBC differential counts when only percentages are available for a xxx differential

xxx count = (WBC count) * (xxx %value / 100)

Further derivation of laboratory parameters might be required for CTCAE grading. For instance, corrected calcium can be derived using the reported total calcium value and albumin at the same assessment using the following formula:

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

In order to apply the above formula, albumin values in g/L will be converted to g/dL by multiplying by 0.1), calcium values in mmol/L will be converted to mg/dL by dividing by 0.2495. For calculation of laboratory CTC grades 0 and 1, the normal range for derived corrected calcium is set to the same limits (in mg/dL) as for calcium.

CTC grades for the derived absolute WBC differential counts (neutrophils, lymphocytes) and corrected calcium will be assigned as described above for grading

5.4 Statistical models

5.4.1 Primary analysis

The LIFETEST procedure in SAS with the TIME statement including a variable with survival times and a (right) censoring variable, and with STRATA statement including variables of stratification factors and with GROUP option under STRATA statement. As an output of the procedure, the rank statistic S and variance var(S) will be obtained. Under the null hypothesis,

the test statistic $Z=S/\sqrt{[var(S)]}$ is approximately normally distributed. The one-sided p-value will be obtained from normally distributed Z statistic.

One-sided will be obtained using Z statistic.

This Section gives additional details regarding the analyses described in Section 2.5.

Kaplan-Meier estimates

An estimate of the survival function in each treatment arm will be constructed using Kaplan-Meier (product-limit) method as implemented in PROC LIFETEST with METHOD=KM option. The PROC LIFETEST statement will use the option CONFTYPE=LOGLOG.

Median survival for each treatment arm will be obtained along with 95% confidence intervals calculated from PROC LIFETEST output using the method of [Brookmeyer and Crowley 1982]. Kaplan-Meier estimates of the survival function with 95% confidence intervals at specific time points will be summarized. The standard error of the Kaplan-Meier estimate will be calculated using Greenwood's formula [Collett 1994].

Hazard ratio

Hazard ratio will be estimated by fitting the Cox proportional hazards model using SAS procedure PHREG (with TIES=EXACT option in the MODEL statement).

A stratified unadjusted Cox model will be, i.e. the MODEL statement will include the treatment arm variable as the only covariate and the STRATA statement will include stratification variable(s).

Hazard ratio with two-sided 95% confidence interval will be based on Wald test.

Treatment of ties

The STRATA statement in LIFETEST procedure will be used to analyze time to event data with ties. The PHREG procedure in SAS with option TIES=EXACT will be used to fit the Cox proportional hazards model.

Checking proportionality of hazard assumption

Plots (SURVIVAL LOGSURV LOGLOGS) generated by LIFETEST procedure in SAS will be used to provide visual checks of the proportional hazard assumption.

The LOGLOG plot (cumulative hazard) will show parallel curves if hazards are proportional.

[redacted]

[redacted]

[redacted]

5.4.2 Secondary [redacted] analysis

Audit-based BIRC assessment of PFS

NCI method

The auxiliary variable estimator of the NCI audit-based method (Dodd et al. 2011) has the form

$$\tilde{\theta}_C = \hat{\theta}_{CA} + \hat{\lambda}\left(\hat{\theta}_{L\bar{A}} - \hat{\theta}_{LA}\right),$$

where $\hat{\theta}_{CA}$, $\hat{\theta}_{L\bar{A}}$ and $\hat{\theta}_{LA}$ are estimators of the log-hazard ratio based on the central assessment in the audited subset of subjects, the local assessment in the nonaudited subset of subjects, and the local assessment in the audited subset, respectively. $\hat{\lambda}$ is defined as $\hat{\rho}\sqrt{\delta}(1-\delta)\sqrt{(\hat{V}_{CA}/\hat{V}_L)}$, where $\hat{V}_{CA}$ and $\hat{V}_L$ are variance estimators of $\hat{\theta}_{CA}$ and $\hat{\theta}_L$ (the estimator of log-HR based on the local assessment in all subjects) respectively, δ is the proportion of subjects in the audited subset, and $\hat{\rho}$ is an estimator of the correlation between $\hat{\theta}_{LA}$ and $\hat{\theta}_{CA}$. For the latter, a bootstrap approach will be used:

- Within the audited subset of size m, m subjects will be sampled with replacement. Using this sample of m subjects, the log-hazard ratio will be estimated based on the local and central assessments separately;
- This procedure will be repeated 1000 times, giving rise to 1000 pairs (local and central) of estimates of the log-HR;
- The sample correlation coefficient between these pairs of estimates will be used for $\hat{\rho}$.

The log-hazard ratio estimates contributing to the auxiliary variable estimate and corresponding variance estimates will be based on stratified Cox proportional hazards models, with stratification based on the randomization stratification factors. The upper bound of a 95% CI for θ_C will be calculated assuming asymptotic normality of $\tilde{\theta}_C$ and using the variance estimator for $\tilde{\theta}_C$ provided in Dodd et al., 2011, i.e. $\hat{V}_{CA}\{1 - \hat{\rho}^2(1-\delta)\}$.

PhRMA method

The early discrepancy rate (EDR) and late discrepancy rate (LDR) will be calculated using the equations below together with information in Table 5-3.

EDR = $(b + a3)/(a + b)$;

LDR = $(c + a2)/(b + c + a2 + a3)$.

Table 5-3 Local versus central disease progression assessments

		Central
Local	PD	No PD
PD	$a = a1 + a2 + a3$	b
No PD	c	d

$a1$: number of agreements on timing and occurrence of PD
$a2$: number of times local PD declared later than central PD
$a3$: number of times local PD declared earlier than central PD

The timing of local and central response assessment (for subjects with complete agreement of local and central sources) will be considered to agree if they occur within ±7 days of each other, aligned with the protocol-specified window for tumor assessments.

Patient reported outcomes: EORTC QLQ-C30/LC13 and EQ-5D-5L/VAS

The text below gives more detailed instructions and rules needed for programming of the analyses described in Section 2.12.

EORTC QLQ-C30 scale scores will be generated by first obtaining the raw scores adding up the item responses on the questions which make up each domain and then applying the linear transformation to the raw scores in accordance with the respective scoring manual provided by the developers. Scores in each scale will be generated if at least half of the items comprising the scale have been answered. For single item scales with missing responses and scales where less than half of the items have not been answered, scale scores will be set to missing.

The scoring approach for the QLQ-LC13 is identical in principle to that for the symptom scales/single items of the QLQ-C30. The dsypnoea scale of the QLQ-LC13 is the only multi item scale (all others are single item scales) and should only be used if all items comprising the scale have been answered.

For the calculation of EQ-5D-5L index value, the EQ-5D crosswalk value set for the UK using the time trade-off method will be used. If any one of the five dimensions of health state is missing, the index value will be set to missing.

A repeated measures model for longitudinal data will be used to compare the two treatment arms in terms of the domain scores over time. This longitudinal model will include terms for treatment, the randomization stratification factors, time of visit (duration in days from the time of baseline measurement to the time of a particular post baseline measurement), baseline value as fixed effects, and an interaction term for treatment by time. This analysis will be restricted to subjects with an evaluable baseline score and at least one evaluable post-baseline score. Time will be considered as a continuous variable in this analysis. As a first approach, an unstructured correlation matrix will be used to model the correlation within subjects. Other structures of the correlation matrix, including AR(1), will be investigated and simplified using likelihood ratio tested if appropriate. In particular situations, the non-convergence of the model may be caused by few subjects with assessments at later timepoints. The possibility of removing few assessments later than a certain time point will be investigated if appropriate.

For the model considered, the SAS code will therefore be:

```
PROC MIXED data=dataset method=reml;
     CLASS subject trt timeC strat_factors;
     MODEL score = trt strat_factors time score_B trt*time /
noint s ddfm=kr;
     REPEATED timeC / subject=subject type=un;
RUN;
/* score refers to the observed scores after Baseline
score_B refers to the baseline score
trt is the treatment variable
timeC refers to the time window as class variable
time refers to the time window
method=reml specifies that the restricted Maximum Likelihood Estimation method
used
type=un specifies that an unstructured covariance matrix is used
```

[redacted]

6 Reference

Amit O, Mannino F, Stone AM, Bushnell W, Denne J, Helterbrand J and Burger HU. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. European Journal of Cancer 2011; 47: 1772-1778

Bretz F, Maurer W, Brannath W, et al (2009) A graphical approach to sequentially rejective multiple test procedures. Stat Med; 586-604.

Bretz F, Posch M, Glimm E, et al (2011) Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biom J; 894-913.

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

Clopper, C.; Pearson, E.S. (1934). "The use of confidence or fiducial limits illustrated in the case of the binomial". Biometrika 26: 404–413

Collet D (1994). Modelling survival data in medical research. London, Chapman & Hall

Dodd E L, Korn L E, Freidlin B, Gray R, Bhattacharya S. An audit strategy for progression-free survival. Biometrics, Vol 67, Issue 3; 1092-99, 2011

Eisenhauer E, et al (2009). New response evaluation criteria in solid tumors: revised RECIST guideline (version 1.1). European Journal of Cancer, Vol.45: 228-47

European Medicines Agency (2020). Guidance on the management of clinical trials during the COVID-19 (coronavirus) pandemic

Fayers PM, Aaronson NK, Bjordal K, et al, on behalf of the EORTC Quality of Life Group (2001) The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001

Food and Drug Administration (2020). FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic: Guidance for Industry, Investigators, and Institutional Review Boards (2020)

Fleming TR, Harrington DP, O'Brien PC (1984). Designs for group sequential tests. Controlled Clinical Trials 5: 348-361.

Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al (2018) Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J Med; 378:2078-2092

Glimm, E., Maurer, W., and Bretz, F. (2010) Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine 29, 219-228

Hellmann MD, Ciuleanu TE, Pluzanski A, et al (2018) Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. N Engl J Med; 2093-2104.

Lan, K. K. G. & Demets, D. L. (1983). Discrete sequential boundaries for clinical trials. Biometrika 70, 659-63

Paz-Ares LZ, Luft A, Trafreshi A et al (2018) KEYNOTE 407 Phase 3 study of carboplatin-paclitaxel/nab-paclitaxel (Chemo) with or without pembrolizumab (Pembro) for patients (Pts) with metastatic squamous (Sq) non-small cell lung cancer (NSCLC). JCO 2018; ASCO Abstract 105#.

Rittmeyer A, Barlesi F, Waterkamp D, et al (2017) Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet; 255-265.

Rogatko A, Schoeneck D, Jonas W, et al (2007) Translation of innovative designs into phase I trials. J Clin Oncol; 4982-6.

[redacted]

Schemper M1, Smith TL (1996) A note on quantifying follow-up in studies of failure time, Control Clin Trials. 1996 Aug;17(4):343-6

Van Reenen M et al (2017) EQ-5D-5L User Guide – Basic information on how to use the EQ-5D- 5L instrument.http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf (Accessed on 10-July-2018).

Zhang JJ, Chen H, He K, Tang S, Justice R, Keegan P, Pazdur R and Sridhara R. Evaluation of blinded independent central review of tumor progression in oncology clinical trials: a meta-analysis. Therapeutic innovation & regulatory science: 2012, 47(2): 167-174.