NOVARTIS

Clinical Development

ACZ885/Canakinumab

CACZ885V2301 / NCT03626545

A randomized, double-blind, placebo-controlled, phase III study evaluating the efficacy and safety of canakinumab in combination with docetaxel versus placebo in combination with docetaxel in subjects with non-small cell lung cancer (NSCLC) previously treated with PD(L)1 inhibitors and platinum-based chemotherapy

Statistical Analysis Plan (SAP)

Author:	Trial statistician [redacted]

Document type:	SAP Documentation
Document status:	Amendment 1.0
Release date:	September 11, 2020
Number of pages:	70

Document History – Changes compared to previous final version of SAP

Date/Version	Section	Changes
18-Dec-2018/ Final		N/A - First version
27-Jul-2020/ Amendment 1	Section 1	Updated to latest protocol amendment version 2.0 prior to OS interim analysis.
	Section 1.1	Updated study design diagram with including the option to transition to the Open Label Extension.
	Section 2.1	Clarified that 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 2.1.1	Specified dose cohort for safety run-in: Dose Level 1 (DL1) cohort.
	[redacted]	[redacted]
	Section 2.2.3.2	Added subgroup of geographic region (East Asia vs. North America + Western Europe vs. Rest of the world).
	Section 2.2.3.3	Removed the requirement of minimum number of subjects in Japan for the conduct of Japanese subgroup analyses; clarified an un-stratified Cox model will be used given the Japan subgroup is expected to have fewer patients.
	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 2.4.1.1.3	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 2.5.2.1	Updated the estimand language and aligned 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; clarified that intercurrent events due to COVID-19 will be handled by treatment policy strategy.

Date/Version	Section	Changes
	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; additional supplementary analyses were added to assess the impact of COVID-19 on OS. Hypothetical strategy was used to handle intercurrent events due to COVID-19. Supplementary analyses based on key inclusion and exclusion criteria were removed, as it is not recommended under estimand framework.
	Section 2.5.2.5.3	China subgroup was added for primary OS analyses to support China submission.
	Section 2.6.1.1	The secondary estimand based on the secondary endpoint PFS was specified.
	Section 2.6.1.1.1	Additional analysis on "New anticancer therapy resulting in censoring of PFS" " was added.
	Section 2.6.1.2	Unknown (UNK) response was updated to Not evaluable (NE) based on latest RECIST macro. This update was applied throughout the document
	Section 2.7.1.1.1	AESI term 'DILI (Hepatic transaminase and bilirubin elevations)' was updated to 'Abnormal liver parameters' based on latest eCRS. These updates were also made in other applicable sections.
	Section 2.7.1.3	Clarified that in the analysis by visit, for 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 2.7.1.4.2	Clarified the clinical notable criteria for weight in Table 2.8.
	Section 2.8.1	PK parameter analyses for docetaxel were added for randomized part (Detailed contents was moved from SRI part as it referred to randomized part).
	Section 2.10	Clarified that if 2 assessments within a time window are equidistant from the target date, the assessment obtained prior to visit will be considered.
	[redacted]	[redacted]
	[redacted]	[redacted]

Date/Version	Section	Changes
	Section 6	Two new references were added.

Table of contents

List of tables

List of figures

List of abbreviations

ADA	Anti-drug antibodies
AE	Adverse event
AESI	Adverse events of special interest
ATC	Anatomical Therapeutic Classification
BLRM	Bayesian Logistic Regression Model
BMI	Body Mass Index
CSR	Clinical Study report
CTC	Common Toxicity Criteria
CTCAE	Common Terminology Criteria for Adverse Events
DAR	Dose Administration Record
DDS	Dose-Determining Set
DI	Dose Intensity
DL	Dose Level
DLT	Dose Limiting Toxicity
DLRT	Dose Level Review Team
DMC	Data Monitoring Committee
DRL	Drug Reference Listing
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
FAS	Full Analysis Set
DCR	Disease control rate
DOR	Duration of response
eCRF	Electronic Case Report Form
IA	Interim Analysis
i.v	intravenous
MedDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
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
PPS	Per-Protocol Set
PRO	Patient-reported Outcomes
PT	Preferred Term
qd	Qua'que di'e / once a day
QoL	Quality of Life
RDI	Relative Dose Intensity
RP3R	Recommended Phase 3 dose Regimen
SAP	Statistical Analysis Plan

SC	Steering committee
SOC	System Organ Class
TA	Tumor assessment
TFLs	Tables, Figures, Listings
TTR	Time to response
WHO	World Health Organization
WHO-DD	World Health Organization-Drug Dictionary

1 Introduction

This statistical analysis plan (SAP) describes the planned analyses for the primary Clinical Study Report (CSR) of study CACZ885V2301, a randomized, double-blind, placebo-controlled, Phase III study evaluating the efficacy and safety of canakinumab in combination with docetaxel versus placebo in combination with docetaxel in patients with non-small cell lung cancer (NSCLC) previously treated with PD(L)1 inhibitors and platinum-based chemotherapy. As specified in Section 12.7 of the study protocol (original version dated 17-July-2018), the primary analysis of overall survival (OS) will occur when approximately 137 deaths are observed, unless OS is declared statistically significant at the interim OS analysis, in which case the interim OS analysis will be called the primary analysis of OS (approximately 96 deaths are expected at the interim OS analysis). At this time, the primary CSR will be produced. In the primary CSR, the safety run-in data will also be reported.

The content of this SAP is based on protocol CACZ885V2301 amendment 02 (dated 10-Jun-2020). All decisions regarding 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 multicenter Phase III study evaluates the efficacy and safety of canakinumab in combination with docetaxel versus placebo in combination with docetaxel, as second or third line treatment, in adult subjects with advanced NSCLC who have one prior platinum-based chemotherapy and one prior PD-(L)1 inhibitor therapy (either given together with the PD-(L)1 inhibitor or sequentially) and have progressed in at least one prior therapy.

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

Safety run-in part

The objective of the safety run-in part is to confirm the Recommended Phase 3 dose Regimen (RP3R) of the combination of canakinumab and docetaxel.

A minimum of 6 subjects will be treated with full doses of docetaxel and canakinumab dose level 1 (DL1): docetaxel 75mg/m^2 i.v. + canakinumab 200 mg s.c., on Day 1 of each 21-day cycle. Subjects will be assessed for at least 2 complete cycles of treatment (21 days per cycle; a total of 42 days) for safety evaluation (DLT-Dose Limiting Toxicities) to define RP3R. Once this dose is confirmed, the randomized part of the study will begin.

The determination of RP3R will be guided by a Bayesian analysis of dose limiting toxicities (DLT), which occurred in the first 42 days during which subjects will receive the combination of canakinumab and docetaxel. The dose-toxicity relationship of canakinumab in combination with docetaxel 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.

[redacted] If judged necessary, additional subjects might be enrolled in the Dose Level 1 (DL1) cohort, 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.

[redacted]

Double-blind, randomized, placebo controlled part

Figure 1-2 provides an overview on the randomization part.

Approximately 226 subjects will be randomized in a 1:1 ratio to either docetaxel with canakinumab or docetaxel with placebo after the subject has met all inclusion/exclusion criteria:

- Arm A: canakinumab s.c at RP3R + docetaxel 75mg/m2 i.v. Q3W
- Arm B: placebo s.c. at RP3R + docetaxel 75mg/m2 i.v. Q3W

The randomization will be stratified based on number of prior lines of therapy in the advanced setting (1 prior line of therapy versus 2 prior lines of therapy) and histology (squamous vs. non-squamous).

[redacted]

Overall survival (OS) is the primary endpoint in the randomization part.

One interim analysis is planned after approximately 96 of the 137 targeted OS events (i.e., at approximately 70% information fraction) have been observed.. The primary intent of this interim analysis is to stop early for superior efficacy. There is no intent to assess futility at this interim analysis. The interim analysis will only be carried out after all subjects have been randomized and (if not withdrawn early) should have at least one post baseline assessment.

The final analysis for OS will occur when approximately 137 deaths are expected.

This study will have a data monitoring committee (DMC), which will function independently of all other individuals associated with the conduct of this clinical trial, including the site investigators participating in the study. The DMC will review at defined intervals the safety data as well as the efficacy data from the interim OS analyses of the double-blind, randomized, placebo-controlled part of the study. The DMC meeting for safety review will be held approximately every six months. Additional DMC reviews may be performed if considered appropriate by DMC. DMC will recommend to the sponsor whether to continue, modify or terminate a trial. The interim analyses will be performed by an independent statistician external

to Novartis and the results will be provided to the DMC by the independent statistician. The final OS analysis will be performed by the sponsor's clinical team.

A separate DMC SAP specifies the analyses to be performed for the DMC safety/efficacy reviews.

1.2 Study objectives and endpoints

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• **Safety run-in part**: To confirm the Recommended Phase 3 Regimen (RP3R) of the combination of canakinumab and docetaxel • **Randomized part**: To compare the overall survival (OS) in the docetaxel plus canakinumab arm versus docetaxel plus placebo arm	• Incidence rate of dose limiting toxicities (DLTs) in the first 42 days associated with the administration of canakinumab in combination with docetaxel. • OS

Objective(s)	Endpoint(s)
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**
Safety run-in part:	
• To assess the preliminary clinical anti-tumor activity of canakinumab and docetaxel combination	• ORR, DOR and DCR by investigator's assessment according to RECIST 1.1
• To characterize the safety and tolerability of the combination of canakinumab and docetaxel	• Type, frequency and severity of adverse events, changes in laboratory values, vital signs, ECGs
• To characterize the pharmacokinetics (PK) of canakinumab and docetaxel when given in combination	• Concentration of canakinumab/docetaxel and PK parameters
Randomized part:	
• To evaluate the 2 treatment arms with regards to progression-free survival (PFS), overall response rate (ORR), disease control rate (DCR), Time to response (TTR) and duration of response (DOR) based on local investigator assessment per RECIST1.1	• PFS, ORR, DCR, TTR and DOR based on local investigator assessment per RECIST 1.1
• To characterize the safety profile of the combination of docetaxel and canakinumab	• AEs (CTCAE[redacted]), ECGs, vital signs and laboratory abnormalities
• To assess the effect of docetaxel plus canakinumab vs docetaxel plus placebo arms on PROs (EORTC QLQ-C30 , lung specific module QLQ-LC13 and EQ-5D-5L) including lung cancer symptoms, health-related quality of life and health status	• Time to definitive 10 point deterioration symptom scores of chest pain, cough and dyspnea per QLQ-LC13 questionnaire are primary PRO variables of interest. Time to definitive deterioration in global health status/QoL, shortness of breath and pain per QLQ-C30 are secondary PRO variables of interest. Change from baseline in EORTC-QLQ C30 and LC13, EQ-5D-5L
• To characterize the pharmacokinetics of canakinumab when given in combination	• Concentration of canakinumab/docetaxel and PK parameters
• To characterize the immunogenicity (anti-drug antibodies, ADA) of canakinumab	• Antidrug antibodies (ADA) prevalence at baseline and ADA incidence on-treatment
• To assess the effect of docetaxel plus placebo vs docetaxel plus canakinumab arms on ECOG performance status.	• Time to definitive deterioration of the ECOG performance status of the score from baseline.
[redacted]	

2 Statistical methods

2.1 Data analysis general information

The interim OS analyses will be performed by an independent statistician external to Novartis. If the DMC determined that study met the OS endpoint at the interim analysis, these results will be the basis of primary CSR. Otherwise, the final OS analysis will be the basis of 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 - 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).

2.1.1.1 Safety Run-in Part

The analysis cut-off date for **the primary analysis** of study data for safety run-in part will be established after all enrolled subjects who have completed two cycles of treatment with canakinumab in combination with docetaxel (between 3 and 6 weeks of treatment, dependent on selected dosing regimens) or have discontinued the study. The primary analysis will comprise the DLT data to determine the RP3R using the BLRM. No CSR will be written based on the primary analysis of safety run-in part alone and will be presented in a CSR only at the final analysis of safety run-in part.

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 the randomized part, when approximately 137 OS events are expected to have occurred or when OS is declared statistically significant at the interim OS analysis (approximately 96 deaths are expected at the interim OS analysis). At this time, data for all subjects from safety run-in part collected until the data cut-off date will be included and will be analyzed for Dose Level 1 (DL1) cohort..

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.

2.1.1.2 Randomization Part

There is one interim efficacy analysis and a final analysis planned for the primary efficacy endpoint of OS. A unique **cut-off date** will be established after the approximately targeted number of events for the planned interim 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. In case, there are subjects ongoing at the time of the primary analysis of the randomized part, a separate TFL shells will detail the planned outputs for any planned close-out CSR.

2.1.1.3 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.

2.1.2 General definitions

2.1.2.1 Study drug and study treatment

In the safety run-in part, Study drug refers to canakinumab and docetaxel. In the randomized part, study drug refers to canakinumab, canakinumab matching placebo and docetaxel.

In the safety run-in part, study treatment refers to canakinumab in combination with docetaxel. In the randomized part, study treatment refers to canakinumab in combination with docetaxel or canakinumab matching placebo in combination with docetaxel.

Canakinumab matching placebo will be referred to as "placebo" in the remainder of this document.

2.1.2.2 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 (canakinumab/placebo or docetaxel) is administered and recorded on the Dosage Administration Record (DAR) (e)CRF. The date of first administration of study drug will also be referred as ***start of study drug***.

Note : Dates from "DAR –PK sampling" eCRF will not be used for this derivation.

2.1.2.3 Date of last administration of study drug

The date of last administration of study drug is defined as the last date when a non-zero dose of study drug (canakinumab/placebo or docetaxel) is administered and recorded on DAR eCRF. The date of last administration of study drug will also be referred as end of study drug.

Note 1: Dates from "DAR –PK sampling" eCRF will not be used for this derivation.

Note 2: Last date of study drug exposure may not be the same as the last date of study drug

2.1.2.4 Date of first administration of study treatment

The date of first administration of study treatment is derived as the first date when a non-zero dose of any component of study treatment (canakinumab/placebo or docetaxel) was administered as per the Dosage Administration (e)CRF. The date of first administration of study treatment will also be referred as ***start of study treatment***.

For example: if the first dose of canakinumab/placebo is taken on 05JAN2019, and first dose of docetaxel is taken on 03JAN2019, then the date of first administration of study treatment is 03JAN2019.

2.1.2.5 Date of last administration of study treatment

The date of last administration of study treatment is derived as the last date when a non-zero dose of any component of study treatment (canakinumab/placebo or docetaxel) was administered as per Dose Administration (e)CRF.

For example: if the last dose of canakinumab/placebo is taken on 15APR2019, and the last dose of docetaxel is taken on 17APR2019, then the date of last administration of study treatment is on 17APR2019.

2.1.2.6 Last date of exposure to study drug/treatment

The study treatment schedule is organized in cycles of 21 days. The last date of exposure to study treatment is derived to be the latest date of the last date of exposure to canakinumab/placebo and docetaxel. The last date of exposure to study drug (canakinumab/placebo or docetaxel) will be derived as follows:

- The last date of exposure to study drug is calculated as (last date of administration of study drug) + (length of time interval - 1) i.e. [last date of study drug administration+ (21-1)].
    - If the dose frequency for canakinumab shifts from Q3W to Q6W or from Q6W to Q9W, the last date of exposure to canakinumab will be derived using the same definition specified above by replacing 20 days with 41 days (if the subject is on Q6 dose schedule) and 62 days (if the subject is on Q9 dose schedule).
- If the subject died or was lost to follow-up within last date of administration of study drug + 21 days, the last date of exposure to study drug is the date of death or the date of last contact, respectively.

- If the dose frequency for canakinumab shifts from Q3W to Q6W or from Q6W to Q9W, the last date of exposure to canakinumab needs to take into account whether the subject dies or was lost to follow-up within last date of administration of study drug + 42 days (if the subject is on Q6 dose schedule) or within last date of administration of study drug + 63 days (if the subject is on Q9 dose schedule).

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. 'Date of last administration of study drug' and 'Date of last contact' are defined in sections Section 2.1.2.3 and Section 2.1.2.12 respectively.

2.1.2.7 Study day

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

The study day is defined as:

- 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 Randomization part, the reference start date for safety assessments (e.g. adverse event onset, laboratory abnormality occurrence, vital sign measurement, dose interruption, PK etc.) is the start of study treatment; and the reference start date for all other, non-safety assessments (i.e., survival, 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.

2.1.2.8 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.

2.1.2.9 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 **Randomization part**, the last non-missing assessment, including unscheduled assessments on or before the date of randomization is taken as "baseline" assessment for efficacy. In the context of baseline definition, the efficacy evaluations also include PRO and ECOG performance status. For PRO, if the randomization date is not Cycle 1 Day 1, 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.

For cases where time of assessment and time of treatment start is captured (e.g. pre-dose ECG, laboratory assessments), the last available assessment before the treatment start date/time is used for baseline.

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.

2.1.2.10 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***: from date of first administration of study treatment to 130 days (approximately five terminal half-lives of canakinumab) after date of last administration of study treatment (including start and stop date)
3. ***post-treatment period***: starting at day 131 after date of last administration of study treatment.

Safety summaries (tables, figures) 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, a separate summary for death including on treatment and post treatment deaths will be provided. In particular, summary tables for adverse events (AEs) will summarize only on-treatment events, with a start date during the on-treatment period (***treatment-emergent*** AEs).

However, all safety data (including those from the post-treatment period) will be listed and those collected during the pre-treatment and post-treatment period will be flagged.

2.1.2.11 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

Time Window	Planned Visit Timing	Time Window Definition

Treatment phase		
Baseline	On or before Study Day 1[a]	≤ Study Day 1
Week 3	Study Day 22	Study Days 2 to 32
Week 6	Study Day 43	Study Days 33 to 53
.....	.....	.....
Week k (with k = 9, 12…)	Study Day 21*(k/3)+1	Study Day 21*(k/3)-9 to 21*(k/3)+11 If last dose date is in the window, upper bound of this time window will be EOT visit date +7 "Note: EOT will be included if EOT is performed within 7 days of permanent discontinuation of study treatment"
Safety follow-up (ECOG PS only)	Post treatment study day 130	Assessment taken at the safety follow-up [EOT visit date+8; EOT date +130]
Post treatment efficacy follow-up		
Post treatment follow-up (Efficacy follow up)	Every 6 or 12 weeks depending the study stage	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 (PRO only)	Within 7 days post disease progression	[disease progression date+1; disease progression date +21]
28 days post disease progression (PRO only)	28 days post disease progression	[disease progression date+22; disease progression date +42]
[a] Study Day 1 = randomization date and PRO assessment on Cycle 1 Day1 will be baseline for PRO if randomization date is not Cycle 1 Day1.		

2.1.2.12 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

Source data	Conditions
Last date subject was known to be alive from Survival Follow-up page	Subject status is reported to be alive, lost to follow-up or unknown.
Start/End dates from further antineoplastic therapy	Non-missing medication/procedure term.
Start/End* dates from drug administration record	Non-missing dose. Doses of 0 are allowed.
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
Imaging assessment date	Imaging marked as done

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 from 'Survival information' eCRF (Section 5.1.2.1).

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

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 study treatment. Subjects will be analyzed according to the dose regimen they have been assigned.

All efficacy data will be analyzed using the FAS.

All safety data will be analyzed using the Safety 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 of dosing.

A subject meets the minimum exposure criterion if the subject receives at least one dose of canakinumab and one dose of docetaxel within first 42 days, or if the subject didn't receive the planned number of doses due to dose limiting toxicity. Subjects who do not experience a DLT during the first 42 days are considered to have sufficient safety evaluations if they have been observed for ≥ 42 days following the first dose.

Subjects will be analyzed according to the study treatment received as defined for the safety 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. The following analysis sets will be derived.

The canakinumab pharmacokinetic analysis set (**PAS-canakinumab**) includes all subjects who provide at least one evaluable canakinumab PK concentration. For a concentration to be evaluable, subjects are required to:

- 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
- 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 (624 hours) (47 days for Q6W) after the last 200 mg canakinumab dose administration

The docetaxel pharmacokinetic analysis set (**PAS-docetaxel**) includes all subjects who provide at least one evaluable docetaxel PK concentration. For a concentration to be evaluable, subjects are required to:

- Receive at least one dose of docetaxel prior to sampling except C1D1 predose sample.
- Have the pre-dose samples collected within 2 hour before the infusion begins
- Receive 75 mg of docetaxel prior to post-dose PK sampling

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-docetaxel	No written informed consent	See definition of PAS-docetaxel

2.2.2 Randomization Part

Full Analysis Set

The Full Analysis Set (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 (Line of therapy: 1 line of prior therapy versus 2 lines of prior therapy; Histology: squamous versus non-squamous) to which they have been assigned during the randomization procedure.

Safety Set

The Safety Set includes all subjects who received at least one dose of study treatment. Subjects will be analyzed according to the study treatment they received, either canakinumab+docetaxel or placebo+docetaxel. The treatment received is defined as the randomized treatment if the subject took at least one dose of that treatment or the first treatment received if the randomized treatment was never received.

Pharmacokinetic analysis set (PAS)

Please see Safety run-in part for the canakinumab pharmacokinetic analysis set (PAS-canakinumab) and the docetaxel pharmacokinetic analysis set (PAS-docetaxel).

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.

Withdrawal of Informed Consent

Any data collected in the clinical database after a subject withdraws informed consent from all further participation in the trial, will not be included in the analysis. The date on which a subject withdraws full consent is recorded in the eCRF.

Death events may be used in the analysis if they are captured from public records (registries), local law and subject informed consent permitting.

Additional data for which there is a separate informed consent, e.g. PK, [redacted] etc., collected in the clinical database without having obtained that consent will not be included in the analysis. These data will be excluded by the presence of the appropriate protocol deviation criterion.

2.2.3 Subgroup of interest

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

2.2.3.1 Efficacy

The primary endpoint of OS will be summarized by the following subgroups to examine the homogeneity of treatment effect provided OS is statistically significant:

- Line of therapy: 1 line of prior therapy versus 2 lines of prior therapy based on randomization data from IRT
- Line of therapy: 1 line of prior therapy versus 2 lines of prior therapy based on eCRF data.
    - 1 line of prior therapy: platinum-based chemotherapy and PD-(L)1 inhibitor therapy were received sequentially without PD in between; or were given together with PD; or adjuvant or neoadjuvant platinum-doublet chemotherapy (after surgery and/or radiation therapy) and a PD-(L)1 inhibitor and developed recurrent or metastatic disease while on or within 12 months of completing therapy; or with recurrent disease > 12 months after adjuvant or neoadjuvant platinum based chemotherapy, subsequently platinum-based chemotherapy and PD-(L)1 inhibitor therapy were received sequentially without PD in between, or were given together with PD.
    - 2 line of prior therapy: platinum-based chemotherapy and PD-(L)1 inhibitor therapy were received sequentially with PD for each therapy; or with recurrent disease > 12 months after adjuvant or neoadjuvant platinum based chemotherapy, subsequently platinum-based chemotherapy and PD-(L)1 inhibitor therapy were received sequentially with PD for each therapy.
- Histology (squamous vs. non-squamous) based on randomization data from IRT
- Histology (squamous vs. non-squamous) based on eCRF data
- 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"))
- Age (<65 vs. ≥65 years)
- ECOG performance status (0 vs. ≥ 1)
- Smoking history (Former/Current vs. Never)
- hs-CRP at baseline (<2mg/L vs. ≥2mg/L, <10mg/L vs. ≥10mg/L, <50 mg/L vs. ≥50 mg/L)

Geographic region (East Asia vs. North America + Western Europe vs. Rest of the world)No formal statistical test of hypotheses will be performed for the subgroups. [redacted] Hazard ratio (HR) for the treatment effect and 95% confidence intervals will be summarized based on Cox

regression model stratified by randomization stratification factors. The objective of the efficacy subgroup analysis is to demonstrate homogeneity of treatment effect in the above subgroups.

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"))
- Age (<65 vs., ≥65 years)
- ECOG PS status (0 vs. ≥ 1)
- Smoking history (Former/Current vs. Never)

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 subgroups specified above:

- AEs, regardless of study treatment, by preferred term and maximum CTC grade
- AEs with suspected relationship to study treatment 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 analysis. These will be specified in the TFL shells.

2.3 Subject disposition, demographics and other baseline characteristics

2.3.1 Randomized part

The Full Analysis Set (FAS) will be used for all baseline and demographic summaries and listings unless otherwise specified. Summaries will be reported by treatment arm and for all subjects. In addition, listings will be reported by treatment arm. No inferential statistics will be provided.

2.3.1.1 Enrollment status

The following summaries will be provided in each cohort separately for the FAS overall, and for both treatment groups:

1. Number (%) of subjects who were randomized

2. 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

All demographic and baseline disease characteristics data will be summarized and listed by treatment arm using the FAS and safety set. Categorical data (e.g. gender, age groups: <65 years vs. vs. ≥65 years, race, WHO performance status, smoking history) will be summarized by frequency counts and percentages; the number and percentage of subjects with missing data will be provided. Continuous data (e.g. age, weight, height, body mass index (BMI)) will be summarized by descriptive statistics (e.g. N, mean, median, and standard deviation, percentiles, minimum and maximum).

BMI (kg/m^2) 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 randomization stratum based on data obtained from the IRT system will be summarized overall and by treatment arm for the FAS. 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. This analysis will include the following: primary site of cancer, details of tumor histology/cytology and stage.

2.3.1.5 Medical history

Medical history and ongoing conditions, including cancer-related conditions and symptoms entered on (e) CRF will be summarized and listed by treatment arm. The summaries will be presented by primary system organ class (SOC), preferred term (PT) and treatment arm. 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, including source of subject referral and other informed consents than main study informed consent, will be listed.

2.3.1.7 Subject disposition

Enrollment by country and center will be summarized for all screened subjects and also by treatment arm using the FAS. The number (%) of randomized subjects included in the FAS will be presented overall and by treatment arm. The number (%) of screened and not-randomized

subjects 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 discontinuation will be presented overall and by treatment arm.

The following summaries will be provided (with % based on the total number of FAS subjects):

- Number (%) of subjects who were randomized (based on data from IRT system)
- Number (%) of subjects who were randomized but not treated (based on 'DAR' eCRF page not completed for any study treatment component)
- Primary reason for not being treated (based on "Treatment disposition" eCRF page)
- Number (%) of subjects who were treated (based on 'DAR' eCRF pages of each study treatment component completed with non-zero dose administered)
- Number (%) of subjects who are still on-treatment (based on the 'Treatment disposition' page not completed);
- Number (%) of subjects who discontinued the study treatment phase (based on the 'Treatment disposition' page)
- Primary reason for study treatment phase discontinuation (based on the 'Treatment disposition' page)
- Number (%) of subjects who have discontinued from the post-treatment follow-up (based on the 'post-treatment follow-up disposition' page);
- Reasons for discontinuation from the post-treatment follow-up (based on 'post-treatment follow-up' page)

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-check 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. 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-2 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

Summaries will be reported by dose cohort and for all subjects; and listings will be reported by dose cohort. 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 and dose cohort
- Basic demographic and background data
- Diagnosis and extent of cancer
- Medical history will be listed
- Subject disposition and screening disposition
- Protocol deviations
- Analysis sets

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

2.4.1 Randomization part

2.4.1.1 Study treatment / compliance

Duration of study treatment exposure, cumulative dose, dose intensity (DI) and relative dose intensity (RDI) will be summarized by treatment. The number of subjects with dose change, interruptions, discontinuation and the corresponding reasons, will be summarized and listed. Details of the derivations and summaries are provided in the following sections.

The safety set will be used for all summaries and listings of study treatment.

For the missing date of last administration, use the data handling rule in Section 5.1.2.2

2.4.1.1.1 Duration of study treatment exposure

Duration of exposure to study drug (for canakinumab/placebo and docetaxel) is defined according to dosing regimen for each study drug as outlined in Section 2.1.2.1.

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

Duration of exposure to study treatment is considered by taking into account the duration of exposure to each study drug:

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

The duration includes the periods of temporary interruption. 'Date of first administration of study drug/treatment' and 'last date of exposure to study drug/treatment' are defined in Sections 2.1.2.2 to 2.1.2.6.

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

Note: If the last record in DAR CRF is a zero dose, this record will not be used in the analyses. For subjects who have ongoing treatment, see Section 5.1.2.2.

2.4.1.1.2 Cumulative dose

Cumulative dose for any component of study treatment is defined as the total dose of the medication given during the study treatment exposure.

Cumulative dose will be summarized using descriptive statistics by treatment arm for each component of study treatment. For subjects who do not receive any drug the cumulative dose will be set to zero.

The cumulative dose is defined according to the type of dosing schedule and is calculated from the DAR eCRF.

The cumulative dose for canakinumab/placebo and docetaxel is defined based on the days when the subject is assumed to have taken a non-zero dose during dosing periods.

2.4.1.1.3 Dose intensity and relative dose intensity

Actual dose intensity (DI) for subjects with non-zero duration of exposure is defined as follows:

For canakinumab (/placebo)

- DI (mg/cycle) = Actual Cumulative dose (mg) / (last date of administration of study drug - first date of administration of study drug + length of dose frequency)*21 (days/cycle)

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 (planned dose intensity) is the planned dose as per protocol in mg (i.e. 200 mg Q3W)
- RDI (%) = DI (mg/cycle) / PDI (mg/cycle)

For docetaxel

- DI (mg/m^2 /cycle) = Actual Cumulative dose (mg) / (last date of administration of study drug - first date of administration of study drug + 21)*21 (days/cycle)

- PDI is the planned dose as per protocol in mg (i.e. 75 mg/m^2 Q3W)
- RDI (%) = DI (mg/m^2 /cycle) / PDI (mg/m2 /cycle)

2.4.1.1.4 Dose changes, interruptions or permanent discontinuations

The number of subjects who have dose changes, dose interruptions and dose permanent discontinuations and the corresponding reasons, will be summarized separately for each of the study treatment components.

The duration of the interruption will be summarized for each observed interruption by time intervals in weeks : <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.

For the purpose of summarizing interruptions and reasons, in case multiple entries for interruption that are entered on consecutive cycles with different reasons will be counted as separate interruptions. However, if the reason is the same in this mentioned multiple entries on consecutive cycles, then it will be counted as one interruption.

For docetaxel, dose change is defined as a reduction of 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/placebo, dose change is defined as any dosing interval increase from Q3W to Q6W, or from Q6W to Q9W.

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 (e.g. chemotherapy, etc.), setting (e.g. adjuvant, etc.) 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 therapy. For example, a combination therapy of chemotherapy and immunotherapy will be counted under chemotherapy and immunotherapy. In addition, the number of lines of prior therapy and the type of regimen within each line of therapy (i.e. platinum-based chemotherapy, PD-(L) 1 inhibitor therapy or platinum-based chemotherapy+PD-(L) 1 inhibitor therapy) will be summarized.

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:

1. Medications starting on or after the start of study treatment but no later than 130 days after last date of administration of study treatment and
2. 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

The analyses for study treatment / compliance in Section 2.4.1.1will be repeated for safety run-in part but by dose cohort. All prior, concomitant and post therapies in safety run-in part will be listed by dose cohort.

2.5 Analysis of the primary objective

2.5.1 Safety run-in Part

The primary objective is to determine the RP3R of the combination of canakinumab and docetaxel for the randomized part.

2.5.1.1 Primary endpoint

The primary endpoint is the incidence of dose limiting toxicities in the first 42 days of dosing associated with administration of canakinumab in combination with docetaxel.

2.5.1.2 Statistical model, hypothesis, and method of analysis

Identification of recommended regimen

Determination of the RP3R of canakinumab in combination with docetaxel will be based upon the estimation of the probability of DLT up to 42 days following first dose for subjects in the dose-determining set. A lower recommended regimen may be identified based on other safety and PK data from the current study. For the purposes of confirming the recommended dosing regimen for canakinumab in combination with docetaxel to be used in the randomized part of the study, each cohort will consist of a minimum of 6 newly enrolled subjects who will be treated at the specified dose level.

The recommended dose regimen is confirmed when the following conditions are met:

1. At least 6 subjects at this dose from DDS
2. This dose satisfies the EWOC criteria
3. The selected regimen is recommended either per the model or by review of all clinical data by the members of the Dose Level Review Team (DLRT).

If the DLRT cannot confirm the DL1 (docetaxel 75mg/m2 i.v. Q3W+ canakinumab 200 mg s.c.Q3W) based on the available data, the team will make a recommendation to either expand the cohort for additional subjects in Dose Level 1 or enroll a minimum of 6 subjects in DL-1 (docetaxel 75mg/m2 i.v. Q3W+ canakinumab 200 mg s.c.Q6W) .

Bayesian adaptive approach

The determination of RP3R will be guided by a Bayesian analysis of DLT data for the first 42 days during which subjects receive the combination of canakinumab and docetaxel.

The dose-toxicity relationship of canakinumab in combination with docetaxel will be modeled by a 5-parameter BLRM for each dose regimen that comprises single agent toxicity parts and interaction part. Single agent toxicity is modelled 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 two single agent toxicities, and interaction is accounted for by adjusting these odds with additional model parameters (odds multipliers). BLRM details are specified in the appendix of the protocol.

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 principle (Rogatko et al 2007). [redacted]

Starting dose

The starting dosing regimen is 200 mg sc Q3W canakinumab and 75 mg/m^2 i.v. Q3W docetaxel. For this starting dose level of canakinumab (i.e. 200 mg Q3W), the prior risk of excessive toxicity is 15.7%, which satisfies the EWOC criterion.

After each dose cohort of subjects, the posterior distribution of the model parameters will be updated via simulation with emerging DLT data and will be used to derive the posterior distribution of the probability of a DLT occurring at a given dose level of canakinumab in combination with docetaxel.

Summaries of the posterior distribution of model parameters and posterior distribution of DLT rates based on the DLT data from all subjects in the DDS will be produced. DLTs will be listed and their incidence summarized by primary system organ class, preferred term and worst toxicity grade based on the CTCAE [redacted]. The DDS will be used for these summaries.

2.5.1.3 Handling of missing values/censoring/discontinuations

Subjects who are ineligible for the DDS will be excluded from the primary analysis of safety run-in part, 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 Randomization part

The primary objective is to compare the OS in the combination of canakinumab + docetaxel versus placebo + docetaxel arms.

2.5.2.1 Primary endpoint

The primary efficacy variable of the study is OS. If a subject is not known to have died, then OS will be censored at the last contact date (Section 2.1.2.12) (on or before the cut-off date). The primary analysis will be based on FAS and will include all data observed up-to the cut-off date. Censoring conventions are provided below in Section 2.5.2.3.

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 and docetaxel compared to placebo and docetaxel arm, for the target population irrespective of any post-treatment anti-neoplastic therapies received. The primary estimand is described by the following five attributes:

A. The **target population** is defined by all subjects randomized in the study (FAS)-- adult subjects with locally advanced (stage IIIB) or metastatic (stage IV) non-small cell lung cancer, who have been previously treated and progressed on platinum-based chemotherapy and PD-(L)1-inhibitor (either in combination or sequentially).

B. The **primary variable** is OS, defined as the time from the date of randomization to the date of death due to any cause.
C. The treatment of interest is the randomized treatment (canakinumab+docetaxel arm or the matching placebo+docetaxel arm) with or without any new anti-neoplastic therapy post randomization as needed.
D. 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
    - Any unforeseen intercurrent events due to COVID-19 pandemic will be handled by treatment policy strategy.
E. The **summary measure** is the hazard ratio (HR) for OS between two treatment arms. It will be estimated using Cox proportional hazard model stratified by randomization stratification factors. The primary comparison will be performed using log-rank test stratified by randomization stratification factors.

2.5.2.2 Statistical hypothesis, model, and method of analysis

Assuming proportional hazards model for OS, the null hypothesis will be tested at one-sided 2.5% level of significance:

H01 (null hypotheses): $\Theta 1 \geq 0$ vs. Ha1 (alternative hypotheses): $\Theta 1 < 0$

Where $\Theta 1$ is the log hazard ratio of OS in the docetaxel plus canakinumab (investigational) arm vs. docetaxel plus placebo (control) arm.

The primary analysis to test this hypothesis and compare the two treatment groups will consist of a stratified log-rank test at an overall one-sided 2.5% level of significance in favor of the docetaxel plus canakinumab arm. The stratification will be based on following randomization stratification factors (line of therapy: 1 prior line of therapy vs. 2 prior lines of therapy; and histology: squamous vs. non-squamous). The primary efficacy variable, OS, will be analyzed at the interim analysis and final analysis of a group sequential design, using a Lan-DeMets (O'Brien-Fleming) α-spending function (Section 2.15).

Analyses will be based on the FAS population according to the randomized treatment group and strata assigned at randomization. The OS distribution will be estimated using the Kaplan-Meier method, and Kaplan-Meier curves, median and associated 95% confidence intervals will be presented for each treatment group. The OS Kaplan-Meier estimate along with 95% confidence intervals will be presented at different time points (e.g. 6 months, 1 year) for each of the two treatment arms. The hazard ratio for OS will be calculated, along with its 95% confidence interval, from a stratified Cox model using the same stratification factors as for the log-rank test.

2.5.2.3 Handling of missing values/censoring/discontinuations

If a subject is not known to have died, then OS will be censored at the last contact date (on or before the cut-off date). 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 the

protocol specified interval between the survival follow-up assessments plus 2 weeks, i.e., 14 weeks for this study. Otherwise subjects will be censored as 'Alive'.

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 proportion 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 analysis, nominal p-values will be presented without any statistical inference. The following summaries will be provided: Kaplan-Meier estimates, estimate of the median OS along with 95% confidence interval, and hazard ratio obtained using the stratified Cox proportional hazards model.

2.5.2.5.1 Sensitivity estimand analysis related to primary endpoint:

First Sensitivity analysis estimand: the target population, the primary variable, treatment attribute, handling of 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 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 OS sensitivity analysis'.

Second sensitivity analysis estimand: the target population, the primary variable, treatment attribute, handling of 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. 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 'OS sensitivity analyses with stratification factors from eCRF'.

2.5.2.5.2 Supplementary estimand analysis related to primary endpoint:

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 '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 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 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). Details to identify medications in Section 2.5.2.5.3.
- **Death due to COVID-19**: OS will be censored on the date of death due to COVID-19 reason (hypothetical strategy)

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

2.5.2.5.3 Medications treating suspected or confirmed COVID-19

The identification of the drug used to treat COVID-19 depends on whether COVID-19 is confirmed or suspected.

For **confirmed COVID-19**, the drug treating COVID-19 is selected based on

1. Drug indication that contains key word "COVID" or "SARS-COV" ;
2. Drug class (CM ATC level 4) is aminoquinolines, protease inhibitors, glucocorticoids, macrolides, antiviral or antiretroviral; or drug name (CM standard medicine name) contains key word "hydroxychloroquine" or "chloroquine".

For **suspected COVID-19**, the drug treating COVID-19 is selected based on drug and AE condition

1. AE with preferred term of "SARS-CoV-2 test negative" or "suspected COVID-19";
2. Drug class (CM ATC level 4) is aminoquinolines, protease inhibitors, glucocorticoids, macrolides, antiviral or antiretroviral; or drug name (CM standard medicine name) contains key word "hydroxychloroquine" or "chloroquine"; or drug class (CM ATC level 4) is glucocorticoids and drug indication contains key word "PNEUM" or equals to "DYSPNEA";
3. Select the first drug with start date within a time window of +/- 7 days selected AE start date.

2.5.2.5.4 Subgroup analyses for the primary endpoint

If the primary efficacy analysis is statistically significant, the primary endpoint of OS will be summarized for the subgroups, which are specified in Section 2.2.3.1 and China subgroup, defined as subjects randomized in sites from China 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 of OS

Number of subjects with an OS event and 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, details in Section 2.5.2.3.

2.6 Secondary objective: Efficacy Analysis

2.6.1 Randomization part

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 Progression survival free (PFS)

PFS based on local investigator assessment as per RECIST 1.1 will be one of the secondary efficacy endpoints. The analysis will be based on the FAS and will include all data observed up-to the cut-off date.

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

- The **target population** is defined by all subjects randomized in the study (FAS))-- adult subjects with locally advanced (stage IIIB) or metastatic (stage IV) non-small cell lung cancer, who have been previously treated and progressed on platinum-based chemotherapy and PD-(L)1-inhibitor (either in combination or sequentially).
- The **secondary 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 censored if no PFS event is observed before the analysis cut-off date. The censoring date will be the date of the last adequate tumor assessment prior to cut-off. PFS will be analyzed in the FAS population according to the randomized treatment group and strata assigned at randomization. The PFS distribution will be estimated using the Kaplan-Meier method, and the Kaplan-Meier curves, medians and 95% confidence intervals of the medians will be presented for each treatment group. The PFS Kaplan-Meier estimate along with 95% confidence intervals will be presented at different time points (e.g. 2 months, 4 months and 6 months) for each of the two treatment arms. The hazard ratio for PFS will be calculated, along with its 95% confidence interval, using a stratified Cox model.

2.6.1.1.1 PFS censoring

The analysis of PFS will be based on the local radiological assessments up 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 2-4 based on options A(1), B(1), C1(1), C2(1), D(1), E(1), and F(1). In particular, PFS will be censored at the last adequate tumor assessment if a subject didn't have an event or if the event occurred after two or more missing tumor assessments. PFS will not be censored if a new antineoplastic therapy is started; instead, an Intent-To-Treat 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 2-4 will be used. Discontinuation of study treatment (for any reason) will not be considered as a reason for censoring. A new anticancer therapy is defined in Section 2.6.1.1.2.

Table 2-4 Options for event dates used in PFS, duration of response

Situation		Options for end-date (progression or censoring)[1] (1) = default unless specified differently in the protocol or RAP	Outcome
A	No baseline assessment	(1) Date of randomization/start of treatment[3]	Censored
B	Progression at or before next scheduled assessment	(1) Date of progression (2) Date of next scheduled assessment[2]	Progressed Progressed
C1	Progression or death after **exactly one** missing assessment	(1) Date of progression (or death) (2) Date of next scheduled assessment[2]	Progressed Progressed
C2	Progression or death after **two or more** missing assessments	(1) Date of last adequate assessment[2] (2) Date of next scheduled assessment[2] (3) Date of progression (or death)	Censored Progressed Progressed
D	No progression	(1) Date of last adequate assessment	Censored
E	Treatment discontinuation due to 'Disease progression' without documented progression, i.e. clinical progression based on investigator claim	(1) Ignore clinical progression and follow situations above (2) Date of discontinuation (visit date at which clinical progression was determined)	As per above situations Progressed
F	New anticancer therapy given	(1) Ignore the new anticancer therapy and follow situations above (ITT approach) (2) Date of last adequate assessment prior to new anticancer therapy (3) Date of secondary anti-cancer therapy (4) Date of secondary anti-cancer therapy	As per above situations Censored Censored Event
G	Deaths due to reason other than deterioration of 'Study indication'	(1) Date of last adequate assessment	Censored (only TTP and duration of response)

[1].=Definitions can be found in the Protocol Appendix.Section 16.1..3.25[2].=After the last adequate tumor assessment. "Date of next scheduled assessment" is defined in Protocol Appendix 3 Section 16.1..3.25.
[3].=The rare exception to this is if the subject dies no later than the time of the second scheduled assessment as defined in the protocol in which case this is a PFS event at the date of death

2.6.1.1.2 Supplementary analysis for PFS

Supplementary analysis secondary estimand : the target population, the variable, intercurrent events and the summary measure of this endpoint are the same as for the secondary 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 table, this approach is referred as 'new anticancer therapy leading to censoring PFS supplementary analysis'.

In the supplementary analysis second estimand, 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 docetaxel monotherapy 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.

2.6.1.1.3 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 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 analysis of PFS, an event occurring after two or more missing assessments or non-adequate tumor assessments is censored at the last adequate tumor assessment.

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. 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.

Schedule for tumor assessment and time windows

Assessment schedule		Scheduled date – 1 week	Scheduled date (weeks from randomization)	Scheduled date +1 week	Threshold (weeks)*
Every 6 weeks for the first 12 months	Baseline	0	0^	1	n/a
	C3D1	5	6	7	9
	C5D1	11	12	13	15
	C7D1	17	18	19	21
	C9D1	23	24	25	27
	C11D1	29	30	31	33
	C13D1	35	36	37	**39**
	C15D1	41	42	43	**45**
	C17D1	47	48	49	**54**
Every 12	C21D1	59	60	61	66
	C25D1	71	72	73	78

weeks after 12 months	C29D1	83	84	85	90
	C33D1	95	96	97	102

* The mid-point between current and next visit (except for baseline) and 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).

Two additional thresholds, 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.
- 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 C15D1, in which case D2 is defined in Rule 2 below).

Since there is a change of schedule for tumor assessments after 12 months, D1 and D2 are defined differently depending on when LATA occurs.

Rule 1: if LATA happens within 39 weeks from randomization (the matched scheduled tumor assessment is C13D1 or before). For example, D1=6+2=8 weeks and D2=2*6+2=14 weeks.

Rule 2: if LATA happens after 39 weeks but within 45 weeks from randomization (the matched scheduled tumor assessment is C15D1). For example, D1=6+2=8 weeks and D2=6+12+2=20 weeks.

Rule 3: if LATA happens after 45 weeks from randomization (the matched scheduled tumor assessment is C17D1 or later). For example, D1=12+2=14 weeks and D2=2*12+2=26 weeks.

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 is smaller or equal to D2 days:

Analysis cut-off date

Date of consent withdrawal

Visit 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.

Then the PFS censoring reason will be respectively:

- 'Ongoing'
- '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'.

2.6.1.1.4 No baseline tumor assessments

As described in Table 16-5 in the Appendix 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. All deaths will be counted in the overall survival analysis regardless of presence or absence of the baseline tumor assessment.

2.6.1.2 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 review. 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. In addition, ORR in the subset of subject with measurable disease at baseline will be presented. Waterfall plot will also be generated (Section 2.6.1.3).

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. any additional secondary anti-neoplastic therapy with the exception of palliative radiotherapy) will be considered in the assessment of BOR. A new anticancer therapy is defined in Section 2.6.1.1.2.

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).
- Non-CR/Non-PD = at least one non-CR/non-PD assessment (or better) > 5 weeks after randomization (and not qualifying for CR).
- 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 unknown status. The following reasons will be used:

a. No valid post-baseline assessment
b. All post-baseline assessments have overall lesion response UNK
c. New anti-neoplastic therapy started before first post-baseline assessment
d. SD or non-CR/non-PD too early
e. PD too late

Note 1: A SD or Non-CR/Non-PD is considered as "SD too early"' if the SD or Non-CR/Non-PD 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 or Non-CR/Non-PD 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".

2.6.1.3 Construction of waterfall graphs

Waterfall graphs will be used to depict the anti-tumor activity. 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 unknown 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	UNK	Any	No, exclude assessment
2	Any	UNK	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 UNK 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:

1. Bars under the horizontal axis representing tumor shrinkage
2. Bars above the horizontal axis representing tumor growth
3. "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.

2.6.1.4 Disease control rate (DCR)

Disease control rate (DCR) is defined as the proportion of subjects with BOR of CR, PR, or SD or Non-CR/Non-PD.

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.

2.6.1.5 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 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.

2.6.1.6 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.

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.

2.6.1.7 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 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 last contact date for OS. This item will be summarized by treatment arm.

All summaries will be reported in months. Date of censoring is the same as defined for the OS analysis.

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

2.6.2 Safety run-in Part

Overall response rate (ORR), and Disease control rate (DCR) by investigator's assessment according to RECIST 1.1 will be presented by dose cohort. Duration of response (DOR) will be summarized by dose cohort. Waterfall plot will also be generated in one plot with dose cohort as the plot legend.

2.7 Secondary objective: Safety analyses

Safety set will be used for all safety analyses.

2.7.1 Randomization part

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

In AE summaries, the primary system organ class will be presented alphabetically and the preferred terms will be sorted within primary SOC in descending frequency. The sort order for the preferred term will be based on their frequency in the investigational treatment.

The following adverse event summaries will be produced by treatment arm; overview of adverse events and deaths, AEs by SOC and PT, AEs by PT summarized by relationship, seriousness, leading to study drug discontinuation, leading to dose interruption, leading to dose change, requiring additional therapy and leading to fatal outcome (SAE). In addition, a summary of 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).

2.7.1.1.1 Adverse events of special interest / grouping of AEs

Data analysis of AESIs

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.).

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

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 $\geq$ 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 defined in Section 2.6.1.1.2 before experiencing any CTC grade ≥ 2 AESI.
- data cut-off date.
- withdrawal of informed consent date

The same analysis will be repeated for time to onset of CTC grade ≥ 3 AESI with similar censoring rules applied.

2.7.1.2 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.

2.7.1.3 Laboratory data

On analyzing laboratory, data from all sources (central and local laboratories) will be combined. The summaries will include all assessments available for the lab parameter collected no later than 130 days after the last study treatment administration date (Section 2.1.2.10). Grading of laboratory values will be assigned programmatically as per NCI Common Terminology Criteria for Adverse Events (CTCAE) [redacted].

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 lab parameter values (details will be specified in the TFL shell) 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 by the selected time points. For a subject with multiple assessments in a time window, the average value will be used. If there are multiple assessments as baseline candidate, the central lab assessment will be defined as “baseline”. If there are multiple assessments

from the central lab or the local lab, the worst assessment will be defined as "baseline." The time windows are defined in Table 2-7 for selected scheduled assessments:

Table 2-7 Time windows for lab assessments

Time Window	Planned Visit Timing	Time Window Definition
On treatment		
Baseline	On or before Study Day 1[a]	≤ Study Day 1
Week 3	Study Day 22	Study Days 2 to 32
Week 6	Study Day 43	Study Days 33 to 53
.....	.....	.....
Week k (with k = 9,12 …)	Study Day 21*(k/3)+1	Study Day 21*(k/3)-9 to 21*(k/3)+11
Safety follow-up	Post treatment study day 130	All on-treatment assessment after EOT(Last dose administration date)
[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 ALT and AST ≤ 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 defined in Section 2.6.1.1.2 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 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.

2.7.1.4 Other safety data

2.7.1.4.1 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.

2.7.1.4.2 Vital sign

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

Vital sign (unit)	Clinically notable criteria	
	above normal value	**below normal value**
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%
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.

2.7.1.4.3 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.

2.7.2 Safety run-in part

Summaries of DLTs, AEs, SAEs, treatment related AEs, on-treatment deaths, AESI overview, Laboratory shift table based on key hematologic and biochemistry terms and Liver transaminase abnormality will be summarized by dose 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, LAB, death, and other safety data in Section 2.7.1 will be generated by dose cohort.

2.8 Secondary objective: Pharmacokinetic endpoints

2.8.1 PK analysis in Randomization part

Pharmacokinetic data analysis will be performed for canakinumab and docetaxel concentrations. PK concentration analyses will be performed for each study drug.

PAS-canakinumab and PAS-docetaxel will be used in the pharmacokinetic data analysis for canakinumab and docetaxel concentrations, respectively.

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 for the FAS.

PK parameters for docetaxel

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 docetaxel, 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 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 canakinumab or docetaxel.	

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.

2.8.2 PK analysis in Safety run-in part

All analyses except population pharmacokinetic analysis specified in Section 2.8.1 will be performed for canakinumab and docetaxel concentrations collected in Safety run-in part. In addition, PK parameters for canakinumab will be summarized.

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, 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 using Phoenix WinNonlin® software version 6.4.

2.9 Secondary objective: Immunogenicity

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 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 dose cohort

2.10 Secondary objective: ECOG performance status

The ECOG PS scale (Table 2-10) will be used to assess physical health of patients, ranging from 0 (most active) to 5 (least active):

Table 2-10 ECOG Performance Scale

Score	Description
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

Time window has been defined in Table 2-1 . 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. For time to deterioration analysis described hereafter all post-baseline assessments will be considered.

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

Time to definitive deterioration of the ECOG PS is the number of days between the date of randomization and the date of the assessment at which definitive deterioration is seen. The ECOG PS deterioration is considered definitive if there is an increase in the performance status by at least one category relative to the baseline or death due to any cause and if no improvement in ECOG PS is observed subsequent to the deterioration.

Baseline is the last available assessment on or before date of randomization. If a patient has 2 ECOG PS values at the same date, the worst ECOG PS value will be taken as ‘baseline’.

Example: If the ECOG PS is 1 at baseline and then 1, 2, 1, 2, 3 at D22, D43, D64, D85, and D106 respectively, then the time to definitive worsening is D85.

Example: if the ECOG PS is 1 at baseline and then 1, 1, 2 at D22, D43, and D64 respectively, with no assessment of the ECOG PS after D64 then the time to definitive worsening is 64 days.

If a definitive deterioration is observed after any missing assessments, this event will be backdated to the first of the missing assessments before the deterioration. The first missing assessment date is calculated as the last available assessment before the definitive deterioration plus X, where X corresponds to the planned scheduled time point for ECOG PS (21 days).

For example, if a patient has an assessment at week Cycle 2 Day 1, misses the following two assessments on weeks Cycle 3 Day 1 and Cycle 4 Day1 and a definitive deterioration is observed on week Cycle 5 Day 1 , then the event will be backdated to Cycle 2 Day 1+21 days.

In addition, death is considered as a worsening of performance status if it occurs close to the last available assessment, where “close” is defined as twice the planned (i.e. protocol scheduled) period between two assessments. This avoids overestimating the time to definitive worsening

in patients dying after an irregular assessment scheme. Patients who die after more than twice the planned period between two assessments are censored at the date of their last available assessment of the performance status.

For example, if the last assessment is at Cycle 2 Day 1 and the patient dies at Cycle 3 Day 15, the definitive deterioration date will be Cycle 3 Day 15. On the other hand, if the last assessment is at Cycle 2 Day 1 and the patient dies at cycle 5 Day 15, which is after more than twice the planned period between two assessments since the last assessment (Cycle 2 Day 1), then the definitive deterioration date will be week 3.

Patients receiving any further anti-neoplastic therapy before definitive worsening will be censored at the date of their last assessment before the start date of the therapy. Patients that have not worsened as of the cutoff date will be censored at the date of their last assessment before the cutoff.

Patients without baseline ECOG PS or without any post-baseline ECOG PS will be censored at the date of randomization with censoring reason being 'No baseline score' or 'No post-baseline score', respectively. However, patients without post-baseline ECOG PS who die within 48 days after date of randomization will be counted as having a definitive deterioration of the ECOG PS at the date of death.

This threshold corresponds to twice the protocol defined ECOG assessment interval plus twice the time window around each assessment: (2 x 21 days)+(2 x 3 days), i.e. 48 days.

Kaplan-Meier estimates will be constructed for each treatment arm in each cohort. The median, 25th and 75th percentiles for time to definitive deterioration for each treatment group will be obtained along with 95% confidence intervals.

2.11 Secondary objective: Patient-reported outcomes

The analyses below are for Randomization part.

Three patient-reported outcomes (PRO) questionnaires will be assessed: EORTC QLQ-C30, with it's 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. In the absence of a better definition to define clinically relevant changes in this population (lung cancer, adjuvant setting) a 10 point deterioration will be assumed to be clinically meaningful for the QLQ-LC13 and QLQ-C30. 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, time to definitive deterioration in global health status/QoL, shortness of breath and pain per QLQ-C30 are 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 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 day 1.

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.

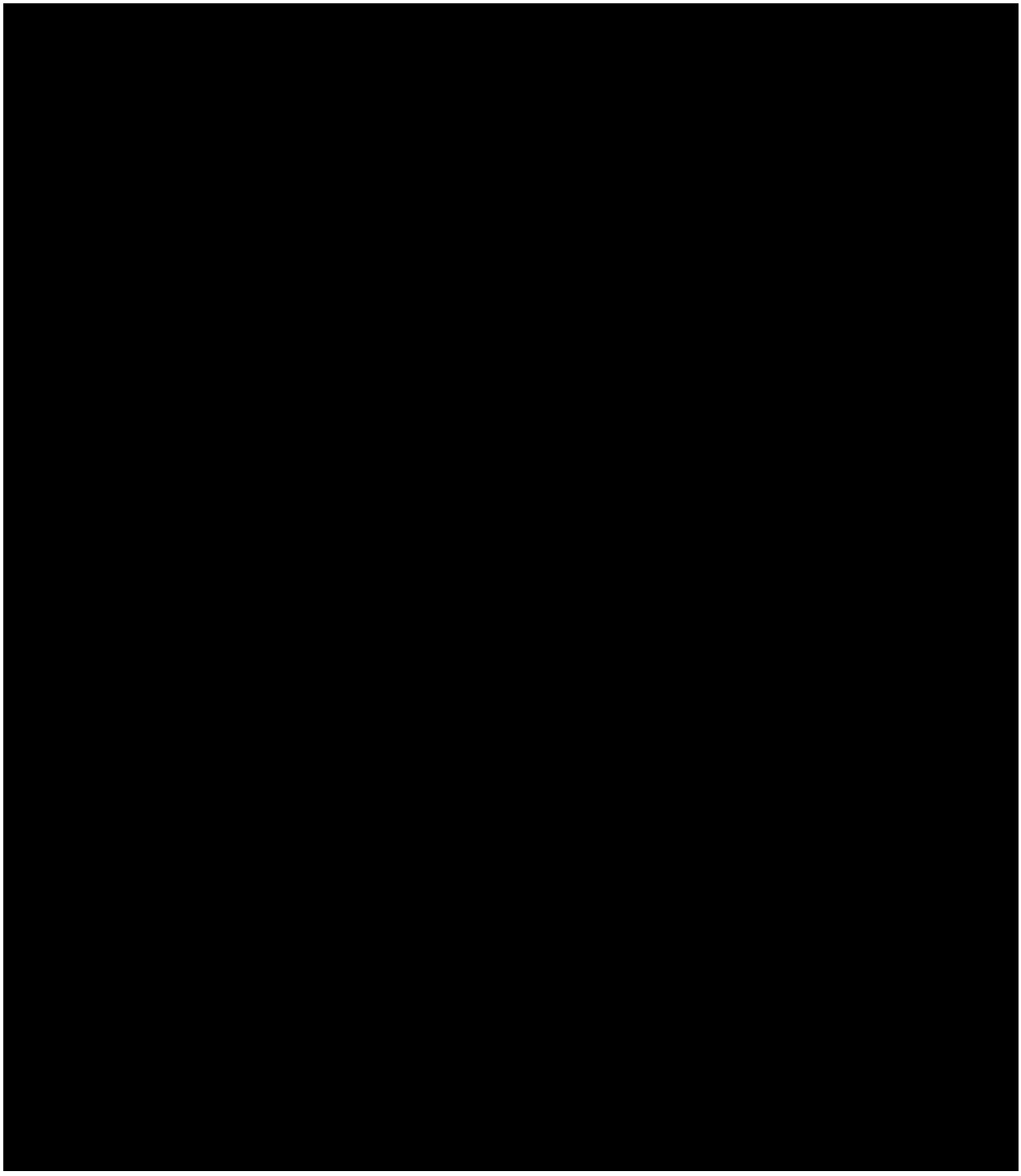
Descriptive statistics will be used to summarize the original scores, as well as change from baseline, of the QLQC30, QLQ-LC13, and EQ-5D-5L at each scheduled assessment time point for each treatment arm using time windows as described in Section 2.1.2.11. Additionally, change from baseline in the scale 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.

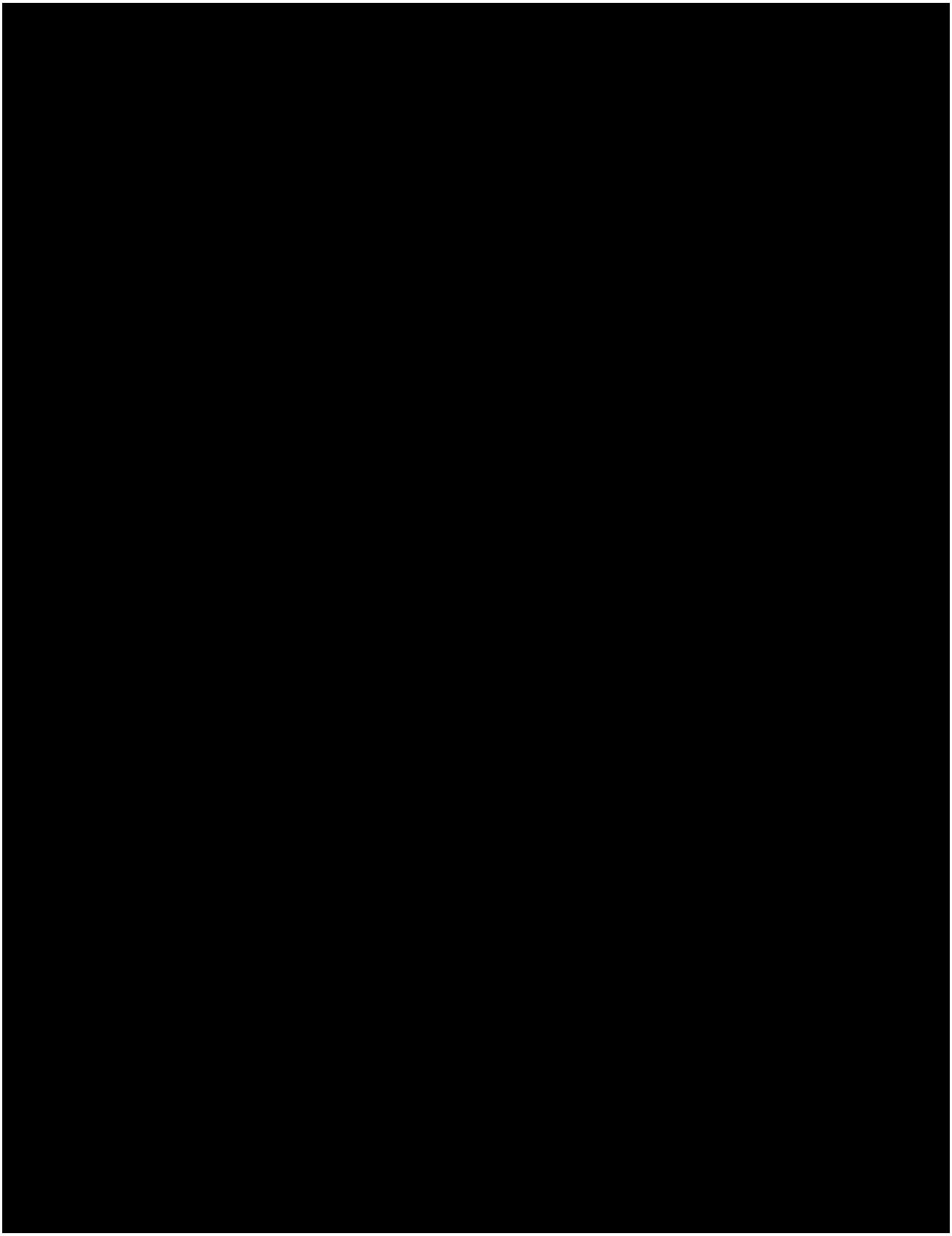
In addition, a repeated measures model for longitudinal data will be used to estimate differences in EORTC QLQ-C30/QLQ-LC13 domains as well as the VAS and utility scores of the EQ-5D-5L between treatment arms. The models will be first based on all assessments. In additional, the models will be repeated 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.

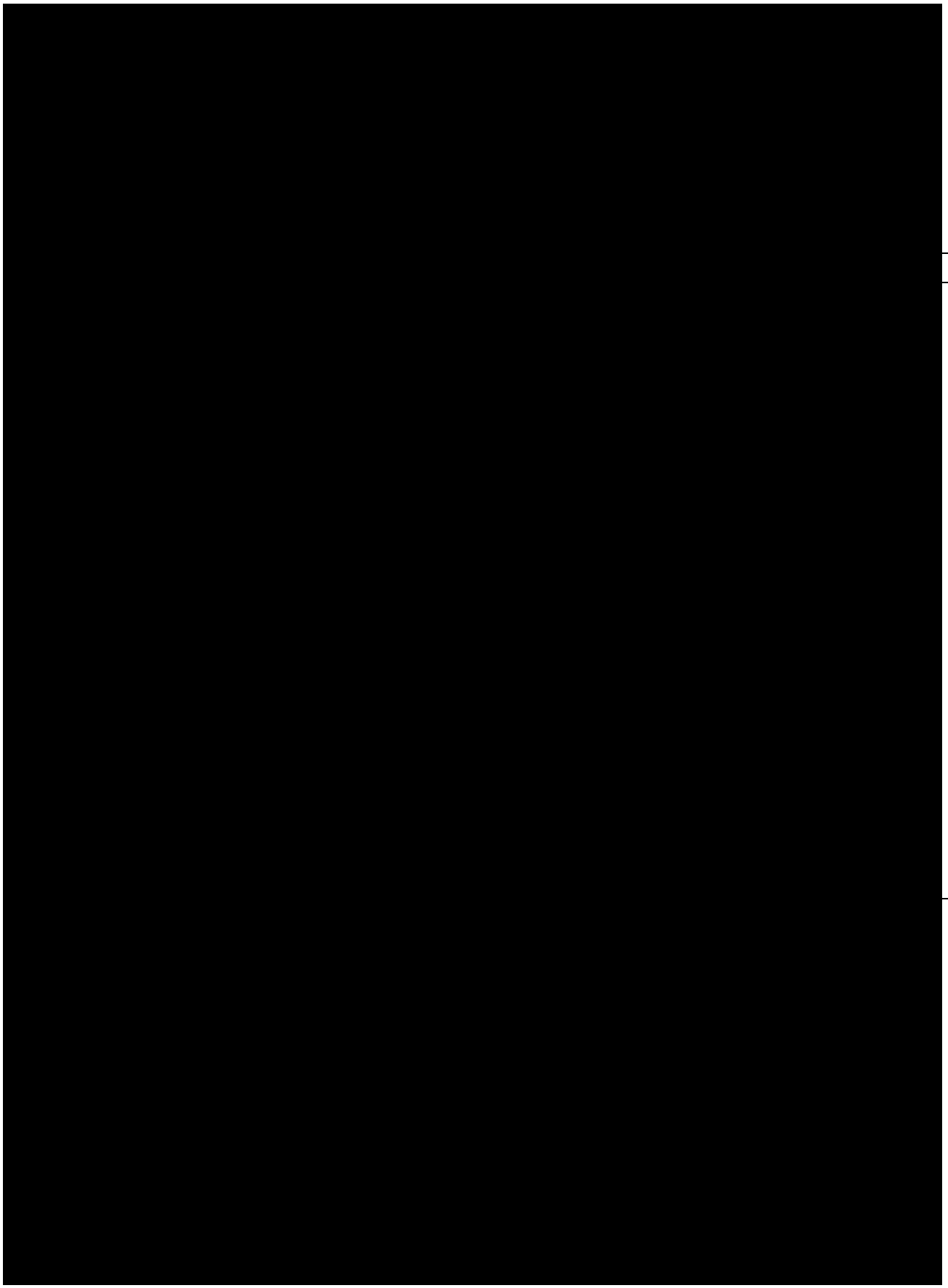
The modeling will mainly be done on the actual score. Note that the modeling of the change in score or the actual score is equivalent since adjustment for baseline score is considered. This repeated measures model will include terms for treatment, the stratification factors, time, baseline value as main 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.

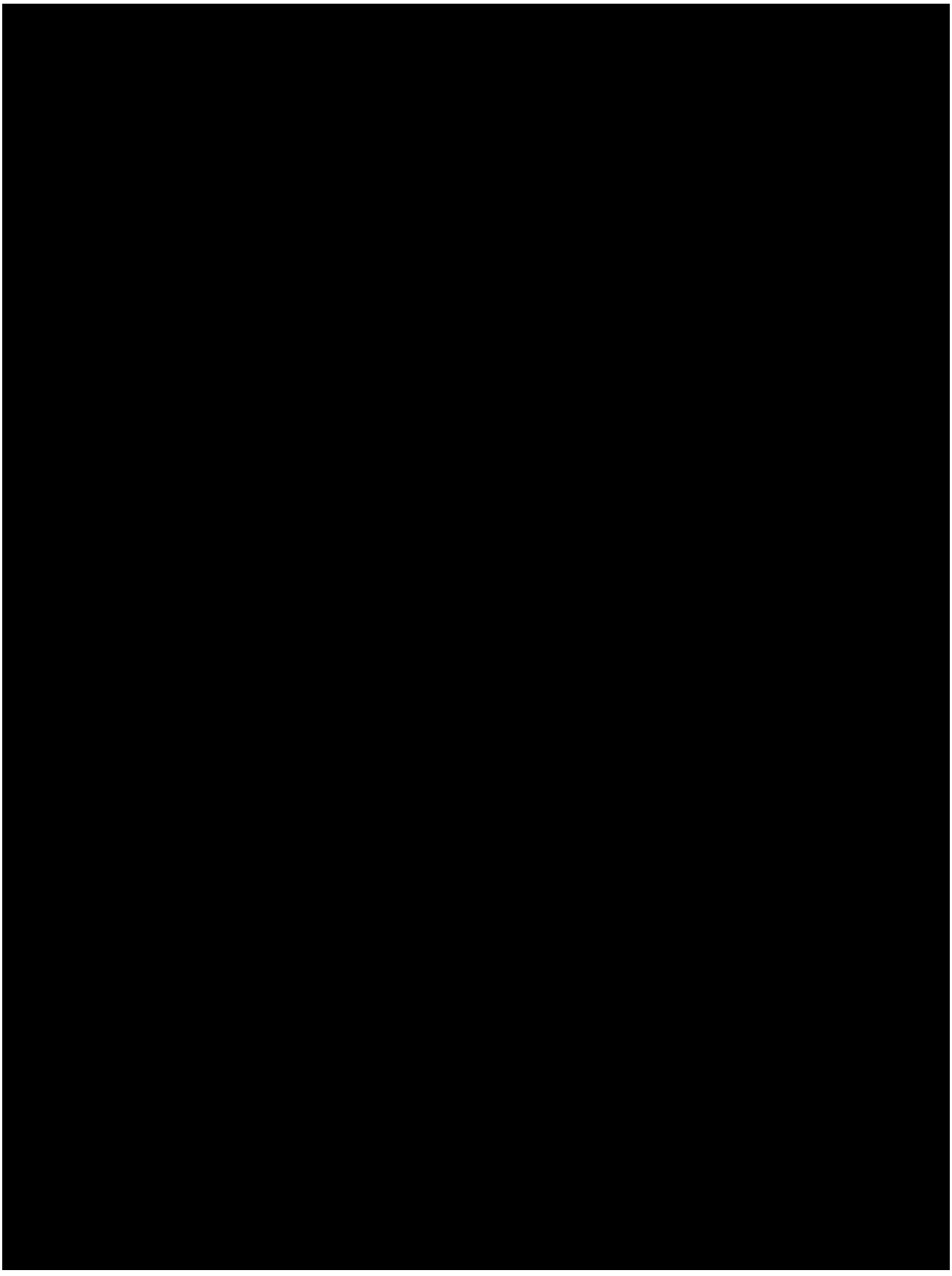
Time will be considered as a continuous variable expressed in weeks. i.e. considering that PRO follow a linear trend. As a first approach, an unstructured correlation matrix will be used to model the correlation within subjects. Other structures of the correlation matrix, including AR(1), will be investigated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









2.15 Interim analysis

Safety run-in part

Not Applicable

Randomized part

One interim analysis is planned after approximately 96 of the 137 targeted OS events (i.e., at approximately 70% information fraction) have been observed.. The primary intent of this interim analysis is to claim superior efficacy. There is no intent to assess futility at this interim analysis. The interim analysis will only be carried out after all subjects have been randomized and (if not withdrawn early) should have at least one post baseline assessment.

An α-spending function according to a two-look (Lan-DeMets) group sequential design with (O'Brien-Fleming) type stopping boundary (as implemented in East 6.4) will be used to construct the efficacy stopping boundaries (Lan et al 1983). Based on the choice of α-spending function described above and if the interim analysis is performed exactly after 96 OS events, the efficacy boundary in terms of p-value scale (or equivalently Z-statistic scale) at the interim is calculated as p = 0.0074 (or Z = 2.436 or HR=0.608). The observed (i.e. nominal) p-value has to be smaller than 0.0074 (or equivalently the observed Z-statistic has to be > Z-statistic scale boundary = 2.436 or equivalently the observed HR has to be smaller than 0.608) to conclude superior efficacy at the interim analysis.

Since the observed number of events at the interim analysis may not be exactly equal to the planned 96 OS events, the efficacy boundary will need to be recalculated using the pre-specified α-spending function and based on the actual number of observed events at interim and the total number of targeted events to calculate the exact information fraction. The observed p-value (or Z-test statistic) at the interim analysis will then be compared against the re-calculated efficacy boundary.

If the study continues to the final OS analysis, the final OS analysis will be performed when approximately 137 OS events have been observed. In practice, the final analysis will be based on the actual number of OS events observed at the cut-off date for the final OS analysis and α 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 0.025.

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

Table 2-12 Simulated probabilities to stop for efficacy at the interim or final OS analysis

Scenario	Look	# OS events	Simulated cumulative probabilities (%)	Simulated incremental probabilities (%)
Under H_0 (HR=1)	Interim	96	0.8	0.8
	Final	137	2.4	1.6
Under H_a (HR=0.57)	Interim	96	61.2	61.2
	Final	137	89.7	28.5
Under HR=0.67	Interim	96	32.6	32.6
	Final	137	65.0	32.4

Note: Simulation is performed in East 6.4 with number of simulations = 10,000 and randomization seed =1234.

Only the investigational drug (canakinumab) will be blinded in this study.

OS interim analysis will be performed by an independent statistician and reviewed by a data monitoring committee (DMC). Unblinded results from the Interim Analysis (IA) for OS will not be communicated to the Sponsor's clinical team or to any party involved in the study conduct (apart from the independent statistician and DMC members) until the DMC has determined that either: (i) interim OS analyses has crossed the pre-specified boundary for efficacy or (ii) the study needs to be terminated due to any cause including safety reasons. Investigators and subjects will remain blinded to study treatment until OS is statistically significant at the interim or final analyses. The final OS analysis will be performed by the sponsor’s clinical team.

Randomization data are kept strictly confidential until the time of unblinding and will not be accessible by anyone else involved in the study except for the independent biostatistician and programmer who will perform DMC analysis, the PK bioanalyst, and modeler and modeling programmer. The study bioanalyst will receive a copy of the randomization schedule to facilitate analysis of the samples. The independent biostatistician and programmer and bioanalyst will keep treatment allocation information confidential until clinical database lock.

3 Sample size calculation

3.1.1 Primary endpoint(s)

Safety run-in part

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 sc Q3W with the fixed dose combination of docetaxel 75 mg/m^2 i.v. Q3W) is confirmed to be safe and tolerated, the safety run-in part is expected to enroll approximately 9 subjects in a cohort in order to have at least 6 evaluable subjects (i.e. who met the minimum exposure criterion and had sufficient safety evaluations during the first 6 weeks of canakinumab in combination with docetaxel

dosing). Otherwise, up to 18 additional subjects are foreseen to be enrolled to assess additional cohorts.

Randomized part

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

Based on available data (Herbst et al 2016, Rittmeyer et al 2017), the median OS in the docetaxel plus placebo arm is expected to be around 8 months. It is expected that treatment with docetaxel plus canakinumab will result in a 43% reduction in the hazard rate for OS, i.e., an expected hazard ratio of 0.57 (which corresponds to an increase in median OS to 14 months under the exponential model assumption).

Then in order to ensure 90% power to test the null hypothesis: OS hazard ratio = 1, versus the specific alternative hypothesis: OS hazard ratio = 0.57, it is calculated that a total of 137 deaths need to be observed. This calculation assumes analysis by a one-sided log-rank test at the overall 2.5% level of significance, subjects randomized to the two treatment arms in a 1:1 ratio, and a 2-look group sequential design with a Lan-DeMets (O'Brien-Fleming) alpha spending function using an information fraction of 70%. If the final analysis is performed when the targeted 137 OS events are observed after exactly 96 OS events have been observed at IA, the observed hazard ratio will have to be < 0.711 to declare statistical significance. Assuming that enrolment will continue for 18 months with an accrual rate of approximately 5 subjects/month in the first 3 months, and approximately 10 subjects/month for the next 3 months, and approximately 15 subjects/months till the completion of enrollment, along with an assumed 5% dropout rate/year for OS, a total of 226 subjects will need to be randomized to observe the targeted 137 death events at about 8 months after the randomization date of the last subject, i.e., 26 months after the randomization date of the first subject. These calculations were made using the software package East 6.4.

4 Change to protocol specified analyses

Not applicable

5 Appendix

5.1 Imputation rules

5.1.1 Study treatment

The following rule should be used for the imputation of the dose end date for a given study treatment component:

Use the study treatment start date

For subjects with missing study treatment start dates, no imputation will be made. If start date is missing then end-date should not be imputed.

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 o If stop date contains a full date and stop date is earlier than study treatment start date then set start date = 01JanYYYY o 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 o If stop date contains a full date and stop date is earlier than study treatment start date then set start date= 01MONYYYY. o 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

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.

For partial date imputation in PR, randomization date instead treatment start date will be used in the above imputation algorithm.

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

For rare cases when either day is missing or both month and day are missing for the date of death/last known subject alive from survival eCRF page, the follow imputation rules will be implemented:

- If only day is missing, then impute max [(1 mmm-yyyy), any valid date from database used for deriving last contact date +1].
- If both day and month are missing, then impute max [(1 Jan-yyyy, any valid date from database used for deriving last contact date +1].

5.1.2.2 Other imputations

Incomplete date of initial diagnosis of cancer and date of most recent recurrence

5.2 Missing day is defaulted to the 1st of the month and the missing month and day is defaulted to 01-Jan.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) [redacted]

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) [redacted]. 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 [redacted] at the time of analysis will be used. For laboratory tests where grades are not defined by CTCAE [redacted], 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 an 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 (mmol/L) = Calcium (mmol/L) + 0.020 [40 - albumin (g/L)]]

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

Analysis of time to events Data

Hypothesis and test statistic

The null hypothesis stating that OS survival distributions of the two treatment arms are equivalent will be tested against one-sided alternative.

H_{01} (null hypotheses): $\Theta_1 \geq 0$ vs. H_{a1} (alternative hypotheses): $\Theta_1 < 0$

Where Θ_1 is the log hazard ratio of OS in the canakinumab (investigational) arm vs. placebo (control) arm. Log-rank test will be used to test the difference between the treatment arms. 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.

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.

LOGLOGS plots log (cumulative hazard)

The LOGLOG plot will show parallel curves if hazards are proportional.

5.4.2 Other secondary analysis

5.4.2.1 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 2.11.

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 dyspnea 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 */
```

6 Reference

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

2. Clopper C.J. and Pearson E. S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, Biometrika Vol. 26, No. 4 (Dec., 1934), pp. 404-413
3. Collett D (1994). Modelling survival data in medical research. London, Chapman & Hall.
4. European Medicines Agency (2020). Guidance on the management of clinical trials during the COVID-19 (coronavirus) pandemic
5. Fayers P.M. et al. on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd edition). Published by: European Organization for Research and Treatment of Cancer, Brussels 2001.
6. 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)
7. Herbst RS, Baas P, Kim DW, et al (2016) Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet p. 1540-50. .
8. Lan KKG, DeMets DL (1983). Discrete sequential boundaries for clinical trials. Biometrika, 70, 649-53.
9. 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 p. 255-265.
10. Rogatko, Xu, Tighiouart (2007) EWOC 2.0 Interactive Software for Dose Escalation in Cancer Phase I Clinical Trials.
11. 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
12. Van Reenen M. et al. 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 Oct 2017.