# STATISTICAL ANALYSIS PLAN

## Randomized, open-label, Phase 3 study of SAR408701 versus docetaxel in previously treated, metastatic, nonsquamous, non-small-cell lung cancer patients with CEACAM5-positive tumors

**SAR408701-EFC15858**

NCT04154956

---

**STATISTICIAN:** [redacted]

**DATE OF ISSUE: 28-JUL-2023**

---

Total number of pages: 117

## TABLE OF CONTENTS

# LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

| | |
|---|---|
| ADI: | actual dose intensity |
| AEs: | adverse events |
| AESI: | adverse events of special interest |
| ALT: | alanine aminotransferase |
| AST: | aspartate aminotransferase |
| ATC: | atonomical therapeutic chemical |
| AUC: | area under the plasma concentration-versus-time curve |
| BCVA: | best corrected visual acuity |
| BOR: | best overall response |
| Ceoi: | concentration at end of infusion |
| CI: | confidence interval |
| CL: | total body clearance from plasma |
| $C_{max}$: | maximum concentration |
| CO: | crossover |
| CP: | conditional power |
| CR: | complete response |
| CSR: | clinical study report |
| CTCAE: | common terminology criteria for adverse events |
| DD: | differential discordance |
| DOR: | duration of response |
| ECG: | electrocardiogram |
| ECOG: | Eastern Cooperative Oncology Group |
| eCRF: | electronic case report form |
| EDR: | early discrepancy rate |
| EOI: | end of infusion |
| FA: | final analysis |
| FDG-PET: | fluorodeoxyglucose-positron emission tomography |
| GHS: | global health status |
| HLGT: | high-level group term |
| HLT: | high-level term |
| HR: | hazard ratio |
| HRQOL: | health related quality of life |
| IA: | interim analysis |
| ICI: | immune checkpoint inhibitor |
| IHC: | immunochemistry |
| IMP: | investigational medicinal product |
| INR: | international normalized ratio |
| IPCW: | inverse probability of censoring weighting |
| IRC: | Independent Review Committee |
| IRT: | interactive response system |
| ITT: | intent-to-treat |

| | |
|---|---|
| LDH: | lactate dehydrogenase |
| LDR: | late discrepancy rate |
| LLOQ: | lower limit of quantification |
| LLT: | lower-level term |
| MedDRA: | Medical Dictionary for Regulatory Activities |
| MRT: | mean residence time |
| NCI: | National Cancer Institute |
| NE: | non evaluable |
| NSCLC: | non small cell lung cancer |
| ODR: | outcome discrepancy rate |
| ORR: | objective response rate |
| PCSA: | potentially clinically significant abnormality |
| PD: | progressive disease |
| PDI: | planned dose intensity |
| PF: | physical functioning |
| PP: | predictive power |
| PR: | partial response |
| PRO: | patient reported outcome |
| PS: | performance status |
| PT: | preferred term |
| Q2W: | every 2 weeks |
| Q3W: | every 3 weeks |
| RDI: | relative dose intensity |
| RF: | role functioning |
| RMST: | restricted mean survival time |
| RoW: | rest of the world |
| RPSFTM: | Rank Preserving Structural Failure Time Model |
| RS: | raw score |
| SAEs: | serious advserse events |
| SD: | stable disease |
| SDF: | survival distribution function |
| SoA: | schedule of activities |
| SOC: | system organ class |
| $t_{1/2z}$: | terminal half-life associated with the terminal slope |
| Tinf: | duration of infusion |
| TKI: | tyrosine kinase inhibitor |
| TNM: | tumor node metastasis |
| TTD: | time to deterioration |
| ULN: | upper limit of normal |
| $V_{ss}$: | volume of distribution in the terminal phase |
| WHO-DD: | World Health Organization-Drug Dictionary |

# 1 OVERVIEW AND INVESTIGATIONAL PLAN

## 1.1 STUDY DESIGN AND RANDOMIZATION

This is a prospective multicenter, multinational, randomized, open label, parallel group, 2-arm study comparing tusamitamab ravtansine (SAR408701) 100 mg/m² every two weeks (Q2W) versus Docetaxel at 75 mg/m² every three weeks (Q3W) in participants with metastatic Non-squamous - Non Small Cell Lung Cancer (NSCLC) expressing CEACAM5 and previously treated with platinum-based chemotherapy and an immune checkpoint inhibitor (ICI).

After confirmation of eligibility criteria, participants will be randomly assigned to one of the two following arms in a 1:1 ratio using an Interactive Response Technology (IRT):

- Tusamitamab ravtansine Q2W at the dose of 100 mg/m$^2$ (experimental arm)
- Docetaxel Q3W at the dose of 75 mg/m$^2$ (control arm).

Screening visit should be performed in prescreened participant whose tumor tissue analyses at central laboratory meets the CEACAM5 expression of ≥2+ in intensity involving ≥ 50% of the tumor cell population.

The duration of the study for a participant prescreened and CEACAM5 positive will include a period for screening of up to 28 days. The cycle duration is 14 days for tusamitamab ravtansine and 21 days for docetaxel therapy.

Participants will continue to receive their assigned treatment until disease progression, unacceptable toxicity, or upon participant's request to stop treatment, or Investigator's decision, whichever occurs first.

For participants who have progressive disease, follow up visits will be performed every 12 weeks until Overall Survival (OS) study cut-off date or death (whichever comes first), unless the participant is allocated to crossover (CO) tusamitamab ravtansine treatment. In addition, if the participant discontinues study intervention for reasons other than progression, the participant should undergo a tumor assessment and a follow-up visit every 8 weeks until radiological disease progression, death, final study cut-off (OS cut-off date), or withdrawal of participant's consent, whichever comes first.

Randomization will be stratified according to:

- Eastern Cooperative Oncology Group (ECOG) performance status (0 versus 1),
- Previous ICI treatment administration (sequential versus combination with chemotherapy)
- Geographical region (Asia versus Western Europe + Australia + North America versus Rest of the World [RoW]).

Randomization will be performed via a stratified blocked randomization as described above. Randomization blocks will not be pre-allocated to strata; instead, a central (study-level) blocked randomization schedule will be utilized with randomization blocks allocated dynamically (on-demand) to each of the stratification levels as needed (based on the enrollment in each stratum). 50 blocks will be assigned to a stratum when a first participant is to be randomized to that stratum or when a participant is to be randomized to the stratum but there are no remaining randomization entries from the blocks already assigned to that stratum.

Approximately 450 participants (225 per treatment arm) will be randomized from approximately 240 sites.

**Crossover treatment phase (if implemented)**

Crossover to tusamitamab ravtansine treatment may be implemented by the Sponsor for participants randomized to docetaxel upon progression, in the case that:

- the final Progression Free Survival (PFS) outcome is statistically significant, and
- statistical power for final analysis of the primary endpoint of OS is expected not to be significantly altered

The decision for each participant to receive crossover treatment with tusamitamab ravtansine will be based on the Investigator's discretion after documented progressive disease (PD) per RECIST 1.1 as assessed by the Investigator and confirmed by the Sponsor after review of the eCRF (Electronic Case Report Form) data.

Intolerance to chemotherapy, withdrawal of consent, or clinical progression without documented radiological progression per RECIST 1.1 is not sufficient for eligibility for crossover to tusamitamab ravtansine. Participants allocated to crossover tusamitamab ravtansine treatment will come to the clinic every 2 weeks for treatment and follow-up procedures as detailed in the Schedule of Activities (SoA) of the protocol (Section 1.3.2).

Treatment with tusamitamab ravtansine during the crossover treatment phase will continue until disease progression, unacceptable Adverse events (AEs),participant's decision to discontinue the study, or Investigator discretion, whichever occurs first. An End Of Treatment (EOT) visit will be performed 30 (±5) days after the last crossover dose for safety follow-up.

## 1.2 OBJECTIVES

### 1.2.1 Primary objectives

**Table 1 - Primary objectives and endpoints**

| Objectives | Endpoints |
|---|---|
| Study is designed with two primary endpoints that will be analyzed on randomized participants at the time of the cutoff date for each given analysis (progression free survival and overall survival).<br>Study success is defined either on PFS or OS | |
| • The primary objective is to determine whether tusamitamab ravtansine improves the progression free survival (PFS) when compared to docetaxel in participants with metastatic non-squamous NSCLC expressing CEACAM5 ≥2+ in intensity in at least 50% of the tumor cell population and previously treated with standard-of-care platinum-based chemotherapy and an immune checkpoint inhibitor (ICI) | • The primary endpoint is PFS as assessed by independent blinded review committee (IRC) and based on RECIST 1.1 assessments. PFS will be defined as the time from randomization to the date of the first documented disease progression or death due to any cause whichever comes first |
| • The primary objective is to determine whether tusamitamab ravtansine improves the overall survival (OS) when compared with docetaxel in participants with metastatic non-squamous NSCLC expressing CEACAM5 ≥2+ in intensity in at least 50% of the tumor cell population and previously treated with standard-of-care platinum-based chemotherapy and an immune checkpoint inhibitor | • The primary endpoint is OS and defined as the time from randomization to the date of death due to any cause |

### 1.2.2 Secondary objectives

**Table 2 - Secondary objectives and endpoints**

| Objectives | Endpoints |
|---|---|
| • To compare the objective response rate (ORR) to tusamitamab ravtansine with docetaxel | • Objective response rate will be defined as the proportion of participants who have a complete response (CR) or partial response (PR), as best overall response derived from Overall Response (OR) determined by the IRC per RECIST 1.1 |
| • To compare the health related quality of life (HRQOL) to tusamitamab ravtansine with docetaxel | HRQOL will be analyzed through three different endpoints:<br>• Time to deterioration (TTD) in disease related symptoms (composite endpoint of cough, chest pain and dyspnea) as determined by EORTC QLQ-LC13<br>• Time to deterioration (TTD) in physical function as determined by EORTC QLQ C-30<br>• Time to deterioration in role function as determined by EORTC QLQ C30<br>The TTD is defined as the time from baseline (Cycle 1, Day 1) until the first ≥10-point change from baseline up to the End-of-Treatment assessment |
| • To evaluate the safety of tusamitamab ravtansine compared to docetaxel | • Incidence of TEAEs and SAEs and laboratory abnormalities according to NCI CTCAE V5 |
| • To assess the duration of response (DOR) of tusamitamab ravtansine with docetaxel | • Duration of response (DOR) is defined as the time from first documented evidence of CR or PR until progressive disease (PD) determined by the IRC per RECIST 1.1 or death from any cause, whichever occurs first |

### 1.2.3 Tertiary/exploratory objectives

**Table 3 - Tertiary/exploratory objectives and endpoints**

| Objectives | Endpoints |
|---|---|
| • To characterize the pharmacokinetic (PK) of tusamitamab ravtansine | • Plasma concentrations of tusamitamab ravtansine during the treatment period |
| • To assess the potential immunogenicity of tusamitamab ravtansine | • Incidence of anti-therapeutic antibodies (ATAs) against tusamitamab ravtansine |
| • To assess the relationship between the tumor mutation profiles detected in the circulating free DNA (cfDNA) at baseline with efficacy outcome | • Plasma analysis for tumor cfDNA at baseline |
| • To explore modulations of circulating CEA as a potential pharmacodynamics biomarker of response to tusamitamab ravtansine treatment | • Circulating CEA at baseline and during the treatment period |
| • To explore PK and pharmacodynamics (PD) (PK/PD) relationship | • Exploration of relationship between relevant safety and efficacy endpoints with PK parameters during treatment period |
| • To identify disease and participant characteristics that could be associated to CEACAM5 expression | • To evaluate disease and participant characteristics in each CEACAM5 expression group (CEACAM5 positive ≥50%; CEACAM5 positive 1% to <50%; and CEACAM5 negative) and CEACAM5 expression patterns |
| **Tertiary endpoints associated with the crossover treated population:** | |
| • To assess safety of tusamitamab ravtansine in terms of TEAEs, SAEs, and treatment-related AEs in participants previously treated with docetaxel | • Incidence of TEAEs, SAEs, treatment-related AEs, and laboratory abnormalities according to NCI CTCAE v5.0 in the crossover treatment phase |
| • To assess tusamitamab ravtansine efficacy in terms of ORR in participants receiving crossover tusamitamab ravtansine treatment after documented disease progression on docetaxel | • Objective response rate in the crossover treatment phase, defined as the proportion of participants who have a complete response (CR) or partial response (PR), as best overall response reported by the investigator per RECIST 1.1 |

## 1.3 DETERMINATION OF SAMPLE SIZE

The sample size calculation incorporates multiplicity adjustment for the analyses of the multiple primary efficacy endpoints: progression free survival and overall survival.

Multiplicity of primary endpoints to calculate the sample size is taken into account using Bonferroni adjustment with different weights assigned to endpoints: PFS is tested at the 1-sided 0.01 level and OS is tested at the 1-sided 0.015 level for a strong control of the overall type-I error at the 2.5% level.

The graphical method of Maurer and Bretz will be applied to provide a strong type-I error control for multiple hypotheses Section 2.4.5.3.

**Overall survival**:

For OS, the following assumptions were used:

- OS has an exponential distribution in both treatment arms.
- The Docetaxel treatment group will have a median OS of 9 months.
- The tusamitamab ravtansine arm will have 28% risk reduction in hazard rate in comparison to Docetaxel treatment group. The targeted hazard ratio (HR) is 0.72, which corresponds to an improvement in the true median OS time from 9 months to 12.5 months.
- The first interim analysis (IA) for futility and efficacy assessment on OS is planned when approximately 221 PFS events or approximately 210 deaths (approximately 58% of OS events) are observed, whichever comes first.
- A second interim analysis of OS for assessment of efficacy is planned when 80% of the OS events (ie, approximately 290 deaths) are observed. Control of the type I error will be ensured using an O'Brien and Fleming spending function (Section 3).
- A randomization ratio of 1:1.

With a total of 363 deaths, a log-rank test at an overall 1-sided 1.5% significance level will have 81% power to detect a risk reduction of 28% for tusamitamab ravtansine compared to Docetaxel in OS.

**Progression free Survival:**

For PFS, the following assumptions were used:

- Progression-free survival has an exponential distribution in both treatment arms.
- The docetaxel treatment group will have a median PFS of 4 months.
- The tusamitamab ravtansine arm will have 38.5% risk reduction in hazard rate in comparison to Docetaxel treatment group. The targeted HR is 0.615, which corresponds to an improvement in the true median PFS time from 4 months to 6.5 months.
- A randomization ratio of 1:1.

With a total of 221 PFS events, a log-rank test at a 1-sided 1% significance level will have 90% power to detect a risk reduction of 38.5% for tusamitamab ravtansine compared to Docetaxel in PFS.

In case of reallocation of alpha following the multiplicity rules defined in Section 2.4.5.3,

- with 363 OS events at final OS analysis the study has ~ 86 % power for detecting a HR of 0.72 at one-sided alpha 0.0249 (see Table 14).

- with 221 PFS events at final PFS analysis the study has ~ 95% power for detecting a HR of 0.615 at one −sided alpha 0.02485 (see Table 13).

Approximately 450 randomized participants (225 in each arm) would be adequate to achieve the targeted number of events for both PFS and OS.

Approximately 3534 participants will be prescreened for CEACAM5 status, of whom 530 participants will be screened to achieve 450 randomized participants with estimated failure rates of 85% for prescreening and 15% for screening.

Calculations consider the observed current enrollment at 3 participants per month during the first 7 months and assume 10 participants per month afterwards based on observed enrollment in the following year. Based on these enrollment assumptions, cut-off dates (CODs) for final analyses of PFS and OS would be approximately 40 and 58 months after FPI, respectively. These cutoffs assume a dropout rate (censoring reason other than cutoff reached) of 17% at 4 months for the final analysis of PFS, and a probability of being dropout of 3% at 58 months for the final analysis of OS.

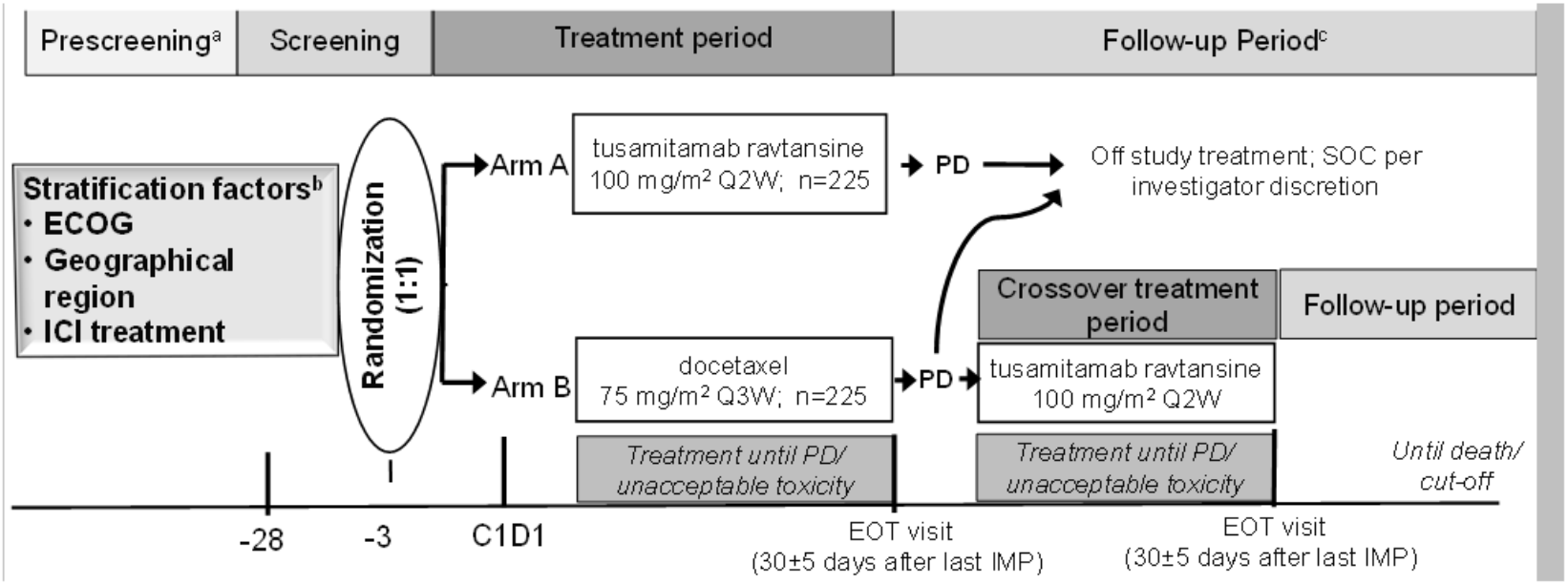
Based on these assumptions, cut-off dates for 58% and 80% IAs of OS are planned approximately 40 months and 48 months after FPI, respectively.

Calculations were made using East 6.5.2 software.

## 1.4 STUDY PLAN

The duration of the study for a participant will include a period for screening of up to 28 days. Once successfully screened, participants may receive study intervention until disease progression, unacceptable AE, participant's or Investigator's decision to stop the treatment. Tumor assessment will be made at every 8 weeks interval (± 5 days window) and scheduled assessment time point will not be modified in case of a cycle delay. After discontinuing study intervention, participants will return to the study site approximately 30 days (± 5 days) after the last investigational medicinal product (IMP) administration or before the participant receives another anti-cancer therapy, whichever is earlier, for end-of- treatment assessments. For participants who have progressive disease, follow up visits will be performed every 12 weeks until OS study cut-off date or death (whichever comes first), unless the participant is allocated to CO tusamitamab ravtansine treatment. In addition, if the participant discontinues study intervention for reasons other than progression, the participant should undergo a tumor assessment and a follow-up visit every 8 weeks until radiological disease progression, death, final study cut-off (OS cut-off date), or withdrawal of participant's consent, whichever comes first.

**Figure 1 - Graphical study design**




a. Patients' tumor tissue (Archival or if not available as new fresh biopsy) will be analyzed for CEACAM5 expression, and only participants with tumors showing CEACAM5 expression level of 2+ intensity in 50% of tumor cells will be screened. Patients who have additional tumor tissue samples available during the prescreening phase may be re-assessed for CEACAM5 expression level; the most recent tissue sample assessed for CEACAM5 expression will be used to determine eligibility of participants for screening.
b. Stratification factors: ECOG (Eastern Cooperative Oncology Group) 0 vs 1; Geographical Region (Asia vs Western Europe+Australia+NA vs ROW); prior ICI (immune checkpoint inhibitors) treatment: Sequential versus combination with chemotherapy
c. A participant discontinuing main study treatment due to a reason other than PD will be followed every 8 weeks until documented disease progression. After disease progression, the participant may be treated with SOC or, if treated under docetaxel arm and if the crossover phase is implemented, tusamitamab ravtansine, per investigator discretion.
Abbreviations: C1D1=treatment Cycle 1 Day 1; ECOG=Eastern Cooperative Oncology Group; EOT=End of Treatment; ICI=immune checkpoint inhibitor; IMP=investigational medicinal product; NA=North America; PD=progressive disease; Q2W=every 2 weeks; Q3W=every 3 weeks; ROW=rest of world.

The complete schedules of activities (SoA) for main study phase and crossover treatment phase are presented in Section 1.3 of the protocol.

## 1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

The protocol amendment history table below gives the timing, rationale, and key details of major changes to the protocol statistical section. Changes performed in the amendments number 1 and 2 were prior to randomization of any participant but one participant was prescreened at the time of the amendment 2. Changes of the amendment 3, 4, 5 and 6 were performed after the first participant randomized and before the first planned 58% interim analysis of OS/final PFS analysis (2023).

**Table 4 - Protocol amendment statistical changes**

| Amendment Number | Date Approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 1 | 27-Sept-2019 | To reflect the study population. This criterion considered as best reflect the study population in terms of prior PDL1 expression level. | Stratification factor related to prior ICI treatment updated as “Sequential versus combination with chemotherapy” instead of “Yes versus No” |
| 1 | 27-Sept-2019 | Feedback from regulatory authorities with consideration of achieving more mature OS interim analysis at the time of final PFS analysis. | Updated to reflect new design of final PFS analyses at the time of 50% of OS events instead of 38% of OS events. Update of boundary properties of OS. |
| 1 | 27-Sept-2019 | To avoid introducing informative censoring and bias in the estimated treatment effect in case of imbalance of informative censoring between arms and to not favor the arm with more informative censoring | Censoring rules updated: For participants alive without PD at the COD, PFS will be censored at the date of the last valid assessment before the COD without taking into account the response of the tumor assessment after the COD if any. PFS will not be censored at the COD anymore |
| 2 | 17-Dec-2019 | For clarity | Precise that the BOR will be derived from OR as determined by IRC per RECIST1.1. |
| 3 | 20-Jul-2020 | For clarity | Details added for HRQOL endpoints and statistical analysis. |
| 3 | 20-Jul-2020 | For clarity | Wording added related to analysis on impact of the regional or national emergency situation (eg, impact of COVID-19) |
| 4 | 21-Jul-2021 | To reflect current enrollment status. | Planned cut-off dates for final PFS and OS analyses were modified to account for observed enrollment up to the amendment. |

| Amendment Number | Date Approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 4 | 21-Jul-2021 | To provide preliminary efficacy data | Addition of an interim PFS analysis at the time of 60% of estimated final PFS events observed. Number of PFS events at final analysis and corresponding power were also updated.<br>A table is added to detail PFS hypothesis testing. Table on boundary properties for planned analyses of OS is updated with current cut-off dates and estimations |
| 5 | 18-Oct-2022 | Anticipated that PFS data will be immature at the planned cut-off date based on regulatory authority input | PFS IA at 60% information fraction was removed |
| 5 | 18-Oct-2022 | To reduce the total sample size, while maintaining an adequate statistical power | The sample size, the number of OS events at the final analysis, the power and the efficacy boundaries were updated |
| 5 | 18-Oct-2022 | To change the estimated number of events at the 50% information fraction IA of OS and to add a second IA of OS, based on a new strategy for interim analyses of OS and adjustments to the total number of events for the final OS analysis | The number of OS events at the 50% information fraction IA of OS was updated and a new IA of OS at 80% information fraction was added. |
| 5 | 18-Oct-2022 | To adjust timelines based on the reduced total sample size, the addition of the 80% OS interim analysis, and the current enrollment curve. | The estimated cut-off dates and timelines for the final analysis of PFS and interim and final analyses of OS were updated. |
| 5 | 18-Oct-2022 | To maintain the same OS futility criterion based on observed HR as initially planned | The beta-spending function planned to be used for the assessment of futility at the time of the first IA of OS was removed and replaced by a rule based on observed OS HR |
| 5 | 18-Oct-2022 | To reflect FDA's feedback on censoring rules for primary analysis of PFS | Change in censoring rules for the primary PFS analysis to censor the events occurring after the initiation of further systemic anti-cancer therapy. This new censoring rule has also been applied for time to deterioration for PROs |

| Amendment Number | Date Approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 5 | 18-Oct-2022 | To reflect addition of new provisional tusamitamab ravtansine crossover treatment phase for participants who progress during or after docetaxel treatment in the main study phase | Tertiary objectives and endpoints were added to specify safety and efficacy analysis for participants in the tusamitamab ravtansine crossover treatment phase.<br>Subsections added to define minimal entrance criteria for allocation to crossover treatment phase. Specified that dose reductions/delays and retreatment criteria for crossover participants would be according to the same rules for participants in the main study phase.<br>Added text in other sections, wherever needed, describing provisional crossover treatment phase, required assessments and analyses of safety and efficacy for crossover treatment participants; wording adjusted as necessary throughout to reflect distinction between main study phase and newly added crossover phase |
| 5 | 18-Oct-2022 | Clarification of how PRO data will be handled if data are missing or not evaluable | Modified description of analysis population and censoring rules for data analyses of time to deterioration in PRO endpoints |
| 5 | 18-Oct-2022 | To adjust to match observed OS dropout rates | Estimated rate of OS dropout for sample size calculation was increased to 3% |
| 6 | 13-Mar-2023 | To reflect increased dropout rate for PFS | The PFS probability of dropout rate by 4 months was updated from 10% to 17%<br>The information fraction at the first IA of OS was updated from 50% to 58%.<br>Respective cut-off dates for the first IA, second IA, and final analysis of OS were updated from 33 months, 44 months, and 53 months to 40 months, 48 months, and 58 months.<br>The estimated power for analysis of OS was updated from 82% to 81%; and the expected number of OS events at first IA was increased from 183 to 210 deaths<br>Definition of the cut-off date for final analysis of PFS was updated from “approximately 183 deaths” to “approximately 221 PFS events or approximately 210 deaths, whichever comes first.” |
| 6 | 13-Mar-2023 | To reflect updated cut-off date for final analysis of PFS / first IA for OS due to increased dropout rate for PFS | Probabilities of crossing the futility boundary on overall survival under different assumptions for the true hazard ratio were updated<br>Boundary properties for planned analyses of PFS and OS were updated |

| Amendment Number | Date Approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 6 | 13-Mar-2023 | Due to uncertainty regarding the information fraction and overall α level available for testing those key secondary endpoints | For ORR and HRQoL, time to deterioration endpoints, the Lan-DeMets O'Brien-Fleming α-spending function is replaced by a user-defined α-spending function. |
| 6 | 13-Mar-2023 | To reflect new statistical assumption | Updated previous estimate of enrollment rate from 12 to 10 participants per month, and removed previously provided alternative estimate based on 10 participants per month as redundant |
| 6 | 13-Mar-2023 | Updated per observed enrollment | Total estimated enrollment duration was updated from 43 months to 50 months. |
| 6 | 13-Mar-2023 | To clarify considerations to be used in the futility decision | Added statement that predictive power for the final OS analysis given the IA data could be used as supportive information in the futility decision |

## 1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

The statistical analysis plan history table below gives the timing, rationale, and key details for major changes to the statistical analysis features in the statistical analysis plan.

The initial SAP was approved on December 20, 2019.

The main changes for SAP version number 2 are the updating of the SAP with the addition of the exploratory/tertiary endpoints and analyses that were not described in the first version of the SAP, the addition of dense PK analysis for China in first 12 participants treated with tusamitamab ravtansine, which was implemented with Amended Protocol 02 and the addition of COVID-19 related analyses.

The first participant was randomized on February 11th, 2020. In March 2023, approximately 320 participants are randomized.

The main changes for SAP version number 3 are

- to reduce the total sample size and number of OS events at the final OS analysis, while maintaining an adequate statistical power for the final analysis of OS: to ensure integrity of the study, the total sample size has been reduced from 554 to 450 participants based on non-squamous-NSCLC treatment landscape and Sponsor considerations. Final PFS analysis will be conducted with a lower number of events than initially planned (221 instead of 271), while maintaining adequate statistical power for this final analysis.
- to redefine the cut-off date for the first IA of OS as the time when approximately 221 PFS events or approximately 210 deaths (approximately 58% information fraction) are observed, whichever comes first in order to maintain an adequate power for PFS analysis while not considerably delaying the futility analysis on OS planned at the same time.

- To replace the beta-spending function for OS futility analysis by a criterion based on observed HR using a criterion based on observed HR (HR>0.9) to maintain the probability of stopping under $H_0$ at 58% IA for OS.
- To add a second interim analysis for OS, when enrollment is expected to be close to completion, at 80% information fraction due to the longer time interval between the cut-off date for 58% IA of OS and final OS analysis.
- To modify the censoring rules for the primary PFS analysis, as per FDA feedback.
- to adjust timelines based on the reduced total sample size and the current enrollment curve.
- to include provisional implementation, under conditions detailed in the protocol amendment #5, of a crossover tusamitamab ravtansine treatment phase for participants with documented progressive disease during or after the docetaxel treatment. Indeed in the anticipated case that analysis of PFS shows a statistically significant advantage of tusamitamab ravtansine over docetaxel, participants randomized to docetaxel may be expected to benefit from the access to treatment with tusamitamab ravtansine after objective disease progression.

In July 2023 approximately 370 participants are randomized. The main change for SAP version number 4 is to add criteria for the futility assessment at the time of final PFS analysis / 58% IA OS, based on predictive power for the final OS analysis and on the observed HR for primary PFS analysis.

**Table 5 - Statistical analysis plan changes**

| SAP version number | Date approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 2 | 07-Jul-2021 | Completion of SAP | Adding or completion of description of endpoints and analyses for tertiary/exploratory endpoints |
| 2 | 07-Jul-2021 | Completion of SAP | Adding of specific PK parameters and analyses for the first 12 participants in China treated with tusamitamab ravtansine after implementation of the Protocol Amendment 2 |
| 2 | 07-Jul-2021 | Subgroups analyses were added to further evaluate the efficacy in specific subgroups | The following efficacy subgroups have been added:<br>- PD1 / PDL1 expression<br>- EGFR mutation<br>- Circulating CEA at baseline<br>- Tumor burden at baseline |
| 2 | 07-Jul-2021 | For clarity | SAR408701 was replaced by the INN Tusamitamab ravtansine |
| 2 | 07-Jul-2021 | For clarity | Method for estimating the standard error of the probability of being event-free was clarified |

| SAP version number | Date approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 2 | 07-Jul-2021 | For the time to deterioration endpoint, the time to censoring was artificially prolonged for participants whose baseline scores do not allow further deterioration | For participants whose baseline scores do not allow further deterioration, the censoring time was changed from the date of last assessment up to end of treatment to the date of first administration intake |
| 2 | 07-Jul-2021 | A separate quality of life SAP will be written to detail further sensitivity analyses and supportive analyses (using different definition of the clinically meaningful change, subgroup analyses) | Deletion of supportive and additional analyses for quality of life variables |
| 2 | 07-Jul-2021 | To adapt the analyses to the schedule of assessments | The troponin analyses were modified to consider the schedule of assessments and the limitations of the analyses |
| 2 | 07-Jul-2021 | To precise that further anti-cancer therapies considered as censor in the PFS sensitivity analysis #1 and also used for the determination of the BOR are further anti-cancer treatments.<br>To describe the further anti-cancer surgeries and radiotherapies | Further anti-cancer therapies defined for BOR analysis were modified.<br>Summaries of further anti-cancer surgeries and radiotherapies were added. |
| 2 | 07-Jul-2021 | Supportive analyses of OS were reviewed and modified to use statistical methods more easily interpretable | Supportive analyses for OS were replaced by inverse probability of censoring weighting and Rank Preserving Structural Failure Time Model methods |
| 2 | 07-Jul-2021 | Data corrections/precisions or handling conventions were added or clarified | Handling of adverse events with missing or partial date of onset was added; Correction of the BSA formula; Primary location is counted in the number of organs involved as per eCRF; Dose ranges were too modified for the dose level -2 of tusamitamab ravtansine and Docetaxel; Definition of the screened population was modified by deleting “concomitantly fulfill the following requirements: CEACAM5 assay assessment of their tumor biopsy showing CEACAM5 positivity (≥50% of membrane cells at intensity ≥2+). The wording “Post-treatment” for the anti-cancer therapy was replaced by “Further” for homogenization through the document. |
| 3 | 24-Mar-2023 | To reduce the sample size. | The total sample size, the number of OS events at the final OS analysis, and the corresponding statistical power and efficacy boundaries were revised. |
| 3 | 24-Mar-2023 | Due to the longer time interval between the cut-off date for 58% IA of OS and final OS analysis | To add a second IA of OS at 80% of Information fraction |

| SAP version number | Date approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 3 | 24-Mar-2023 | Timelines were adjusted based on the reduced total sample size, the addition of the 80% OS interim analysis, and the current enrollment curve | The estimated cut-off dates and timelines for the final analysis of PFS and interim and final analyses of OS were updated |
| 3 | 24-Mar-2023 | FDA feedback | Change in censoring rules for the primary PFS analysis to censor the events after the initiation of further systemic anti-cancer therapy.<br>This has also been applied for time to deterioration for PROs.Sensitivity and supportive PFS analyses were updated following this new censoring rule. |
| 3 | 24-Mar-2023 | Modification of the futility criterion for first interim OS futility at the time of final PFS analysis to maintain the probability of stopping under H0 at 58% IA for OS | The beta-spending function planned to be used for the assessment of futility at the time of the first IA of OS was removed and replaced by a rule based on observed OS HR (HR>0.9) |
| 3 | 24-Mar-2023 | Due to uncertainty regarding the information fraction and overall α level available for testing those key secondary endpoints | For overall response rate and health-related quality of life, time to deterioration endpoints, the Lan-DeMets O'Brien-Fleming α-spending function is replaced by a user-defined α-spending function |
| 3 | 24-Mar-2023 | For clarity on cross-tabulation of IRC and Investigator assessments of PFS outcome | Documented PD and Death are not distinguished anymore |
| 3 | 24-Mar-2023 | Method to calculate the Odds ratio and its α-adjusted two-sided CIs for ORR analyses was added | Stratified Cochran Mantel Haenszel test will be used to calculate the odds ratio and its 95% and α-adjusted two-sided CIs for primary and sensitivity analyses |
| 3 | 24-Mar-2023 | To reflect addition of new provisional tusamitamab ravtansine crossover treatment phase for participants who progress during or after docetaxel treatment in the main study phase | Tertiary objectives and endpoints were added to specify safety and efficacy analysis for participants in the tusamitamab ravtansine crossover treatment phase.<br>Text clarifications throughout the SAP to adjust definition of baseline, CO treatment emergent period and describe planned analyses of safety and efficacy for crossover treated participants |
| 3 | 24-Mar-2023 | Adjustment to match observed dropout rates | Estimated rate of dropout for sample size calculation was increased to 3% |
| 3 | 24-Mar-2023 | To adjust to the population described in the dedicated SAP HRQOL at the time of the interim analyses | The analysis population for HRQOL endpoints at the time of the OS interim analyses is ITT population |
| 3 | 24-Mar-2023 | Efficacy subgroup analysis was added to further evaluate the efficacy in specific subgroup | The following subgroup analysis has been added: Prior taxanes |

| SAP version number | Date approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 3 | 24-Mar-2023 | Efficacy subgroup analyses were removed because not deemed relevant based on correlation with other covariates and/or further clinical evaluation | The following subgroup analyses have been removed<br>• Tumor burden at baseline as per eCRF<br>• Time from initial diagnosis to randomization |
| 3 | 24-Mar-2023 | Some efficacy subgroup analyses were moved to exploratory biomarker analyses and one subgroup for biomarker analyses was added | The following subgroup analyses have been moved:<br>• PD1/PDL1 expression<br>• Circulating CEA at baseline<br>CEACAM5 expression IHC subgroup was added for exploratory biomarker analyses |
| 3 | 24-Mar-2023 | For clarity | The reference modality of each subgroup was specified |
| 3 | 24-Mar-2023 | For clarity | Only systemic treatments are taken into account as potential censoring by further anticancer therapy for efficacy endpoints. Analyses for further radiotherapies and surgeries were defined |
| 3 | 24-Mar-2023 | For a more pragmatic approach | Actual treatment arm was modified in a specific case: for participants receiving more than one study intervention during the study, they will be analyzed according to the randomized arm if the participant has received at least one administration of the as-randomized intervention |
| 3 | 24-Mar-2023 | For clarity on immunogenicity analyse | Immunogenicity endpoints section was updated with new terminology and some notions were added as participants with treatment-emergent ATA or participants with inconclusive ATA. Wording of the immunogenicity section was updated and treatment-boosted ATA are included in the analyses of ATA kinetics. Descriptive statistics at baseline were added for immunogenicity population. Safety parameters to be considered for the impact on immunogenicity were clarified and/or modified |
| 3 | 24-Mar-2023 | For clarity | The wording of the baseline definition was slightly modified to take into account the time of the assessment (if any) in the evaluation of the baseline. |
| 3 | 24-Mar-2023 | To clarify supportive analysis that may be performed in case of nonproportional hazard. | Restricted mean survival time (RMST) in case of non-proportional hazard has been added |
| 3 | 24-Mar-2023 | For clarity | The post-treatment period was clarified to include AEs that worsen or become serious during the post-treatment period |

| SAP version number | Date approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 3 | 24-Mar-2023 | To describe more accurately the QOL quality of completion and baseline status | Outputs on compliance rate and descriptive statistics at baseline were added for the three QOL efficacy scores |
| 3 | 24-Mar-2023 | Safety analyses added to better characterize the safety profile | Specifications on analyses of neutropenic complications were added<br>Analyses on infection localization were added |
| 3 | 24-Mar-2023 | Focus on analyses per SMQ/CMQ | Safety analyses by category of AESI were removed |
| 3 | 24-Mar-2023 | Simplification of the section related to analyses of pharmacodynamic/ogenomics variables and harmonization with other studies from the project | The SAP "A pooled study for investigating novel approaches to identify prognostic factors of CEACAM5 expression" focused primarily on CARMEN LC03 study, so the analyses were added directly in the SAP. Some analyses were clarified, parameters with the same analyses were gathered, and other analyses were removed for simplification |
| 3 | 24-Mar-2023 | To better characterize the safety profile according to premedication | Analysis on premedications was added |
| 3 | 24-Mar-2023 | Focus on key analyses on pre-screening, screening and post-treatment adverse events | Analyses on pre-screening, screening and post-treatment adverse events were clarified |
| 3 | 24-Mar-2023 | To better describe the analyses on ocular examinations | Analses on ocular examinations were clarified and detailed |
| 3 | 24-Mar-2023 | Data corrections/precisions or handling conventions were added or clarified | Added imputation rules for partial dates for start of adverse event and recovery date of adverse event for time to event analyses; Added formula for QTc Fridericia; Added formula for corrected calcium; Deletion of conjugated bilirubin, Deletion of heart rate as vital signs as not assessed in the protocol, clarification of the baseline value for the stratification factors; added analysis of fresh/archived biopsy; clarification on the date of last contact; added data convention for QOL assessment, Precision that analyses on PCSA are based on the worst value during the on-treatment period, clarification of the exposure section, standard unit of hematocrit was updated.Some sections were reworded and codelist were completed for some variables |
| 4 | This version | To add futility rules on PFS and OS endpoints for the final PFS analysis / 58% IA OS | Adding two futility criteria based on the predictive power of final OS analysis and on the observed HR in the primary analysis of PFS in the case where PFS is not statistically significant |
| 4 | This version | To correct an erroneous deletion on previous version (version 3.0) regarding exploratory biomarker subgroups | Putting back the following thresholds for biomarker subgroups of interest:<br>• PD-L1 (<1% vs. ≥1%,)<br>• circulating CEA (<100 vs. ≥100 ng/mL) |

| SAP version number | Date approved | Rationale | Description of statistical changes |
|---|---|---|---|
| 4 | This version | To assess impact of baseline PRO questionnaires collected on the same day as the first IMPadministration but time after the IMP administration time | Adding a sensitivity analysis including only as baseline value the HRQOL assessments performed the day and the time before the first IMP administration |
| 4 | This version | To provide a more detailed description of the immunogenicity population at baseline | Demographics, medical or surgical history, and overall summary of prior anti-cancer treatments analyses were added |
| 4 | This version | To modify the analysis period of ATA samples | ATA samples are analyzed at any time after first IMP administration instead of the treatment-emergent period |
| 4 | This version | To better characterize the dose modifications | The analyses as per Kaplan Meier method of the time to first modification and time to first dose reduction were added |
| 4 | This version | For a more conservative approach | Visual acuity test section was clarified (taking the worst case scenario) |
| 4 | This version | For a more pragmatic approach | Analyses by cycles are removed for electrocardiogram and vital signs parameters as the schedule of administration is different between tusamitamab ravtansine and docetaxel |
| 4 | This version | Other corrections/clarifications /Typography | Clarification of the categories of ICI monotherpay and ICI+ICI for the last regimen; correction of terminology for by incidence rate per participant-years (instead of event rate); modification that “SAR408701” is replaced by “tusa rav” in the tables/listings/figures; for further systemic anti-cancer therapy, the number of participants at risk as well as the probabilities of being free of will be provided every 3 months; clarification of the definition of the adjusted completion rate in HRQOL analyses; α-adjusted two-sided confidence intervals of the odds ratio for response rate and of the hazard ratio for time to deterioration of the three HRQOL endpoints might be provided if feasible; modification of the output displaying the number of participants experiencing SOC coronary cardiac TEAEs and myocardial TEAEs according to troponin values; clarification of the Nab incidence; clarification of the timing and reason of censoring for PFS and HRQOL analyses; modification of the descriptive graph on troponin values |

# 2 STATISTICAL AND ANALYTICAL PROCEDURES

## 2.1 ANALYSIS ENDPOINTS

### 2.1.1 Demographic and baseline characteristics

The baseline value is defined as the last available value or measurement before the first dose of study treatment. Unscheduled visit measurements or repeated tests of laboratory data, vital signs, and electrocardiogram (ECG) data will be used for the computation of the baseline.

For participants initially randomized to docetaxel who cross over to tusamitamab ravtansine (if any), the crossover baseline value is defined as the last available value or measurement before or on the crossover date, which is defined as the first tusamitamab ravtansine administration in the crossover part. Unscheduled visit measurements or repeated tests of laboratory data, vital signs, and ECG will be used for the computation of the crossover baseline.

This definition applies to all variables unless otherwise specified.

For participants randomized and not treated, the baseline value is defined as the last available value obtained before or on the date of the randomization.

The value used to define stratification factor is the last available value before randomization.

All baseline safety and efficacy parameters (apart from those listed below) are presented along with the on-treatment summary statistics in the safety and efficacy sections (and Section 2.4.6 and Section 2.4.5).

***Demographic characteristics***

Demographic variables are gender (Male, Female), race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Multiple, Not reported, Unknown), ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not reported, Unknown), age in years (quantitative and qualitative variable: <65, [65 - 75[ and ≥75 years), weight (kg) and BSA ($m^2$).

***Medical or surgical history***

Medical (including detailed history of cardiovascular risk factors) or surgical history includes relevant history of previous or associated pathologies other than the tumor.

This information will be coded using the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

***Disease characteristics at diagnosis***

The following disease characteristics at initial diagnosis will be described:

- Time from initial diagnosis of lung cancer to randomization date (in years)
- Location (as collected in eCRF)
- Histopathology type (as collected in eCRF)
- Stage of the disease derived according to "the 8th lung cancer TNM classification changes and clinical implication" from TNM collected in the eCRF

***Disease characteristics at study entry***

The following disease characteristics at study entry will be described:

- Extent of the disease as per eCRF
- Measurability of disease (at least one target lesion reported) as per eCRF
- Number of organ(s) involved as per eCRF (including primary site of the tumor)
- Type of organ(s) involved as per eCRF (including primary site of the tumor)
- Smoking history (in pack-years for current and former smokers - 1 pack-year is 1 pack of cigarettes smoked per day for 1 year).

***Prior anticancer therapies***

- Number of prior anti-cancer therapies including neoadjuvant/adjuvant/radiosensitizer (ie, concurrent chemoradiotherapy with or without maintenance treatment) and advanced regimen,
- Number of participants with intent: neoadjuvant and adjuvant and chemoradiotherapy and advanced, neoadjuvant and/or adjuvant and advanced, chemoradiotherapy and advanced, neoadjuvant or adjuvant or chemoradiotherapy only, advanced only,
- Number of prior anti-cancer therapies in the advanced setting,

  A regimen in the advanced setting consists of a single agent, combination or a sequential therapeutic strategy with several drugs, given until a PD is documented. Participant who had adjuvant /neoadjuvant treatment and relapsed as metastatic disease during or within 6 months of treatment will be considered as first line treatment in the advanced setting,
- Type of prior treatment including neoadjuvant/adjuvant/chemoradiotherapy and advanced regimen
  - Biologics/Small molecules:
    - Tyrosine kinase inhibitor: EGFR inhibitor, Anti-angiogenic (Bevacizumab…), BRAF Kinase inhibitors, ALK inhibitors, RAS/RAF/MEK/ERK signaling pathway inhibitors
    - ICI: anti PD1/PDL1, CTL4A inhibitors

    - Others
  - Chemotherapy: Platinum, Pemetrexed, Taxanes, Vinca alcaloids…, Others
  - Antibody Drug Conjugate
  - Others
- Prior Immune checkpoint inhibitor (anti PD1/PDL1) as per eCRF: sequential vs in combination with chemotherapy,
- Summary of last prior anti-cancer therapy:
  - Time from completion of last regimen of treatment to randomization (months)
  - Main treatments of last regimen
    - ICI monotherapy (either anti-PD1/PD-L1 or anti-CTLA-4 compounds)
    - ICI and ICI (anti-PD1/PD-L1 and anti CTLA-4 compounds)
    - ICI in combination with chemotherapy
      - Chemotherapy and ICI and ICI
    - ICI and other biologics and small molecules
      - Other biologics and small molecules and ICI and ICI
    - No ICI: Chemotherapy, Biologic, Chemotherapy and biologics
  - Reason for discontinuation of the last line
  - Best response to the last line
- Prior surgery: number (n, %) of participants with any prior surgery related to lung cancer, type of procedure (Preferred Term) and time from the last surgery to the randomization date (months),
- Prior radiotherapy: number (n, %) of participants with any prior radiotherapy related to lung cancer, location, intent, analgesic intent if palliative and time from last radiotherapy to the randomization date (months). Radiotherapies performed the day of the first administration are considered as on-treatment.

***Other baseline characteristics***

Randomization strata (as defined in Section 1.1) as per IRT and as per eCRF will be described.

Biomarker data will be collected, if available, at prescreening including PD-L1 status by immunohistochemistry (IHC), type of sample (fresh or archived tumor biopsy), tumor mutation/gene alteration status (eg, EGFR, ALK, ROS, MET, RET, BRAF) and circulating CEA.
CEACAM 5 IHC is assessed at prescreening for eligibility of participant for screening to determine if the participant is positive according to the protocol definition: membrane staining at ≥ 2+ intensity of ≥ 50% of tumor cells. In case of re-assessment of CEACAM5 IHC, the most recent tumor tissue samples assessed for CEACAM5 expression will be used for baseline characteristics. All these parameters including type of sample (fresh or archived tumor biopsy) will be described.

Any technical details related to computation, dates, and imputation for missing dates are described in Section 2.5.4.

### 2.1.2 Prior or concomitant medications (other than anticancer therapies)

All medications taken within 28 days before randomization and up to 30 days after the last study treatment administration are to be reported in the case report form pages.

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) using the version currently in effect at Sanofi at the time of database lock.

- Prior medications are those the participant used prior (from 28 days) to first study treatment administration intake. Prior medications can be discontinued before first administration or can be ongoing during treatment phase.
- Concomitant medications are any treatments received by the participant concomitantly to the IMP, from first study treatment administration to the end of treatment + 30 days or to crossover date –1 day (if earlier) for participants in the docetaxel treatment group who crossover. Concomitant medications for crossover treatment phase are any treatments received by the participants concomitantly to the IMP, from first tusamitmab ravtansine administration to the end of treatment in the crossover part +30 days.

A given medication can be classified both as a prior medication and as a concomitant medication. Concomitant medications do not include medications started during the post treatment period (as defined in the observation period in Section 2.1.4).

***Premedications***

Participants will routinely receive premedications prior to tusamitamab ravtansine (main study phase and crossover treatment phase) and docetaxel infusions.

Any technical details related to computation, dates, imputation for missing dates are described in Section 2.5.4.

### 2.1.3 Efficacy endpoints

#### *2.1.3.1 Primary efficacy endpoints*

OS and PFS are two multiple primary endpoints.

##### *2.1.3.1.1 Overall survival*

Overall survival is defined as the time from date of randomization to date of death due to any cause. In the absence of observation of death, survival time will be censored to last date the participant is known to be alive (last contact date, see Section 2.5.1).

The cutoff date for the final analysis of OS will be the date when approximately 363 deaths are observed.

#### *2.1.3.1.2 Progression free survival*

An independent review committee (IRC) blinded to the randomization arm and to the participants' characteristics will review the tumor assessments to evaluate overall response at each tumor assessment and determine progression status per RECIST 1.1 definitions (Appendix E). The full details regarding the determination of the progressive disease are provided in protocol and the IRC charter.

Progression-free survival is defined as the time from the date of randomization to the date of the first documentation of objective PD as assessed by IRC according to RECIST 1.1 definitions (Appendix F) or death due to any cause before the PFS analysis cutoff date, whichever occurs first.

The following censoring rules will be used:

- Participants without documented progression or death before the PFS analysis cut-off, will be censored at the date of the last valid disease assessment before the PFS cut-off date with no evidence of a disease progression performed prior to the initiation of any further systemic anticancer therapies (if any).
- A participant without PFS event (death or documented progression) and without any evaluable post-baseline tumor assessments will be censored at the day of randomization (Day 1).
- A participant with an event documented immediately after two or more non-evaluable tumor assessments will be censored at the date of the last evaluable tumor assessment documenting no progression prior to the initiation of a further systemic anticancer therapy.

Clinical/Non-radiological progression (as collected in the eCRF) will not be considered as documented progression for primary PFS analysis.

The cutoff date for the final analysis of PFS will be the date when approximately 221 PFS events or approximately 210 deaths are observed, whichever comes first.

Additional details regarding the definition of PFS and handling of events and censoring are given in Section 2.4.5.1.2 and Appendix F.

### *2.1.3.2 Secondary efficacy endpoint(s)*

#### *2.1.3.2.1 Overall response rate*

In the main study phase, the Best Overall Response will be derived according to RECIST v1.1 definitions (Appendix E), based on Overall Responses determined by the IRC (Appendix D). BOR will be the best tumor response observed from the date of the randomization until documented disease progression, death, analysis cutoff date, initiation of further systemic anticancer therapy, whichever occurs first. For a supportive analysis, BOR according to the Investigator's assessment will also be determined.

For participants from the docetaxel treatment group who cross over, the first systemic further anti-cancer therapy is the tusamitamab ravtansine treatment.

The ORR on each randomized treatment arm will be estimated by dividing the number of participants with objective response (CR or PR as BOR derived from RECIST 1.1 based on overall response determined by IRC) by the number of participants from the analysis population of the respective treatment arm.

For a supportive analysis in the main study phase, ORR according to the Investigator's assessment will also be determined.

In the crossover treatment period, for participants who cross over to tusamitamab ravtansine after progression on docetaxel treatment, tumor assessment will be performed per site local practice starting from first dose of crossover tusamitamab ravtansine treatment. There will be no central reading of tumor imaging, and efficacy assessments will be performed by the Investigator per RECIST 1.1. The BOR will be reported by the investigator in the eCRF as the date of progression.

Post crossover efficacy data (eg, ORR, BOR) will be analyzed separately and will be considered as exploratory.

#### *2.1.3.2.2 Health-Related Quality of life*

Health-related quality of life will be assessed using:

- The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Cancer specific module with 30 items (EORTC QLQ-C30).
- The EORTC QLQ-LC13 (LC13) - lung cancer-specific module with 13 items.

**EORTC QLQ-C30.**

The EORTC QLQ-C30 (C30) is a validated, self-administered, 30-item, cancer specific health related quality of life (HRQOL) questionnaire (1). The recall period for the C30 is the past week. The C30 assesses global health status/health related quality of life (GHS/QoL), functioning, symptoms and financial difficulties due to disease or treatment related to cancer. For the GHS/QoL scale and the five functional scales (physical, role, emotional, cognitive, and social) higher scores indicate better GHS/QoL or function (higher scores better). The C30 also assesses symptoms commonly reported by cancer participants and financial difficulties due to disease or treatment. For the three C30 symptom scales (fatigue, nausea & vomiting, and pain), six symptom items (dyspnea, insomnia, appetite loss, constipation, diarrhea and perceived financial difficulties item), lower scores indicate fewer symptoms/difficulties (lower scores better). Items on the C30 were scored using the C30 scoring algorithms which standardize and transform the raw scores to a 0-100 range (1).

***EORTC LC 13***

The EORTC QLQ-LC13 (LC13) is the lung cancer module of the C30. The recall period for the LC13 is the past week. The LC13 assesses lung cancer- associated symptoms (cough, dyspnea, pain, hemoptysis) and side-effects from conventional chemotherapy and radiotherapy (sore mouth, hair loss, dysphagia and neuropathy) that can impact the HRQOL of lung cancer participants (1). The EORTC QLQ-LC13 contains 13 items. Items on the LC13 were scored using the LC13 scoring algorithms which standardize the raw scores to a 0-100 range (1). Lower scores indicate lower symptomology/symptom burden (lower scores better).

Raw scores (RS) for the EORTC are calculated using EORTC QLQ-C30 and EORTC QLQ-LC13 scoring formulas (detailed in Appendix G):

- RS is the mean score of the items that contribute to the scale.
- For functional scales: Score = [1− ((RS− 1)/range)] x 100.
- For symptom scales/single items scales and the Global health status (GHS) /QoL scale: Score = [(RS − 1)/ range] x 100.

The range is the difference between the maximum and the minimum response to individual items.

HRQL is analyzed as TTD through three different secondary endpoints:

- TTD in disease related symptoms (composite endpoint of cough (item 1), chest pain (item 10) and dyspnea (items 3, 4 and 5)) as determined by EORTC QLQ-LC13
- TTD in physical functioning (PF) as determined by EORTC QLQ-C30 (items 1 to 5)
- TTD in role functioning (RF) as determined by EORTC QLQ-C30 (items 6 and 7)

The baseline value is defined as the last questionnaire assessment before or on the first IMP day.

For the disease related symptoms (cough, dyspnea, chest pain) from EORTC QLQ LC-13, a baseline is defined as missing if all of the symptoms scores (cough score and chest pain score and dyspnea score) are missing at baseline.

The TTD is defined as the time from baseline (Cycle 1 Day 1) until the first ≥10-point change from baseline (considered as clinically meaningful change) up to the end of treatment assessment before the analysis cutoff date.

For the TTD in disease related symptoms (cough, dyspnea, chest pain) from EORTC QLQ LC-13, a deterioration will be defined as an increase from baseline score of at least 10 points in any one of these three symptoms. In case of one or several missing scores, no imputation of missing scores will be done, and the assessment will be done according to the scales available. For the TTD in PF and RF from EORTC QLQ- C30, a deterioration will be defined as a decrease from baseline score of at least 10 points.

Participants with a non-missing baseline assessment will be censored at the last assessment up to end of treatment before the start of further systemic anticancer therapy or before the analysis cutoff date (whichever is earlier), provided their symptoms/functional scale had not deteriorated up to that point before the analysis cutoff date. Participant without baseline or post-baseline electronic patient-reported outcomes (ePROs) questionnaire or whose baseline scores do not allow further deterioration will be censored at the first administration intake (or at randomization date if the participant is randomized and not treated).

Missing assessment

Missing data for EORTC QLQ-C30 and EORTC QLQ-LC13 will follow missing data/handling procedures recommended in the EORTC QLQ-C30 scoring manual for the C30 and cancer specific modules; no additional imputations will be conducted for missing data. A questionnaire is evaluable as soon as one scale is complete. A scale is defined as complete when at least half of the items from the scale has been answered.

In addition to the composite endpoint for disease-related symptoms, the TTD on each scale (cough, chest pain and dyspnea) will be analyzed separately to provide supportive information.

Electronic PROs will not be assessed in the crossover arm.

#### *2.1.3.2.3 Duration of response*

The duration of response (DOR) is only summarized on the subgroup of participants who have achieved objective response.

DOR is defined as the time from the date of first initial occurrence of a CR or PR to the date of first documentation of objective PD according to RECIST 1.1 or death due to any cause before the initiation of any further systemic anticancer therapy, whichever occurs first.

In the absence of disease progression or death before the analysis cut-off date, DOR will be censored at the date of the last evaluable tumor assessment not showing PD before the initiation of a further systemic anticancer therapy or before the analysis cutoff date, whichever is earlier.

For a supportive analysis, DOR according to the Investigator assessment will also be determined.

### 2.1.4 Safety endpoints

The safety analysis will be based on the reported adverse events (AEs) and other safety information, such as clinical laboratory data, vital signs, weight, electrocardiogram (ECG) and Eastern Cooperative Oncology Group (ECOG) performance status (PS).

***Observation period***

The observation period will be divided into 4 segments:

- The pre-screening period is defined as the time from when the participants give prescreening informed consent to the day before the screening informed consent.

- The screening period is defined as the time from when the participants give screening informed consent to the first administration of the IMP.
- The treatment period is defined as the time from the first dose of IMP up to 30 days after the last dose of IMP. For participants randomized to docetaxel who enter the crossover phase to receive tusamitamab ravtansine treatment, the treatment period is defined as the time from the first dose of IMP up to the last dose of IMP before the crossover +30 days, or up to the first dose of tusamitamab ravtansine –1 day, whichever is earlier.
- The crossover treatment period is the period from first dose of tusamitamab ravtansine administration after crossover up to 30 days after the last dose of crossover study treatment.
- The post-treatment period is defined as the time starting 31 days after the last dose of IMP (including the crossover treatment phase) to study closure or death, whichever comes first.

#### *2.1.4.1 Adverse events variables*

For participants who have fresh biopsy, only AE related to the fresh biopsy itself is reported with reporting time frame interval up to 30 days from the date of biopsy, unless the occurrence of later event is assessed as related to the intervention by the investigator. Otherwise, all AEs (including serious adverse events [SAEs] and AEs of special interest [AESI]) will be collected from the signing of screening informed consent until at least 30 days after the last study intervention.

During follow-up period, SAEs and AESI regardless of relationship to IMP and IMP-related AEs, which are ongoing at the end of study treatment, and new IMP related AE/SAE/AESI, will be followed until resolution or stabilization.

***Adverse event observation period***

- Pre-screening AEs are defined as AE occurring during the pre-screening period.
- Screening AEs are defined as any AE occurring during the screening period.
- Treatment-emergent AEs are defined as AEs that develop, worsen (according to the Investigator's opinion), or become serious during the treatment period.
- Crossover treatment-emergent AEs are defined as AEs that developed, worsen or become serious during the crossover treatment phase
- Post-treatment AEs are defined as AEs that are reported during the post-treatment period.

To be noted, AEs that worsen or become serious during the post-treatment period are also counted as post-treatment AEs.

All adverse events (including SAE and AESI) will be graded according to National cancer institute common terminology for adverse events (NCI-CTCAE) v5.0 and coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the version of MedDRA currently in effect at Sanofi at the time of database lock.

***Adverse events of special interest***

AESI will be defined using eCRF AESI specific tick box on the AE page.

***Adverse events as per SMQ/CMQ:***

- Corneal events (reported as AEs) (CMQ: "Corneal events compound level" containing the preferred terms included in both SOC "Eye disorders" and SMQ "Corneal disorders" [Narrow]),
- Ocular/visual adverse events (excluding corneal disorders) (CMQ "Eye disorders exclude corneal disorders" containing PTs included in SOC "Eyes disorders" and excluding PTs in SMQ "Corneal disorders" [Narrow]),

In addition, analysis of ocular/visual symptoms will be performed based on the symptoms page of the CRF linked to the ocular AE, designed to collect reported symptom linked to the event).

- Cardiac conduction defects (using SMQ "Conduction defects" [Narrow]),
- Peripheral neuropathy events (using SMQ "Peripheral neuropathy" [Narrow and Broad]),
- Colitis events (excluding infective) (using CMQ "Colitis [excluding infective]" containing PTs included in HLT "Colitis [excl infective]"),
- Hypersensitivity events (using SMQ "Hypersensitivity" [Narrow] and defined as events occurring on the day or the day after infusion),
- Hepatic disorders adverse events (using SMQ "Hepatic Disorders" [Narrow and Broad]),
- Hematological adverse events (using SMQ "Haematopoietic cytopenias" [Narrow and Broad]),
- AE related to Covid-19 illness (using SMQ "COVID-19" [Narrow]).
- Neutropenia and neutropenic complications: Neutropenia (from laboratory abnormalities) will be displayed along with neutropenic complications (febrile neutropenia and neutropenic infections).

Neutropenic complications will be analyzed using the following data source:

- Febrile neutropenia selected using CMQ "Febrile neutropenia"
- Neutropenic infections: defined as a NCI-CTCAE Grade ≥2 infections from SOC "Infections and Infestations" (selected using CMQ "GLB_SOC infections and infestations") concomitant with NCI-CTC grade 3-4 neutropenia from laboratory results. Infection and Grade 3-4 neutropenia will be considered as concomitant if one of the following conditions is met:
  - Grade3-4 neutrophils count value measured the day of the start of the AE infection,
  - the last Grade3-4 neutrophils count value measured before the start date of the AE infection is within 7 days before the start of the AE infection,
  - the first Grade 3-4 neutrophils count value measured after the start date of the AE infection is within 2 days after the start of the AE infection.

#### *2.1.4.2 Deaths*

The deaths will be analyzed in the pre-screening, screening, treatment, the crossover treatment period (if applicable) and post-treatment periods.

#### *2.1.4.3 Laboratory safety variables*

Clinical laboratory data that will be analyzed consists of blood analyses (including hematology and biochemistry). Clinical laboratory values will be converted into standard international units that will be used in all listings and tables.

Blood samples for clinical laboratories will be taken as defined in the study flowchart and as clinically indicated. The laboratory parameters will be classified as follows:

- Hematology
  - **Red blood cells and platelets and coagulation:** hemoglobin, hematocrit, platelet count and international normalized ratio (INR)
  - **White blood cells:** white blood cell count, neutrophils, lymphocytes, monocytes, basophils, eosinophils
- Clinical chemistry
  - **Metabolism:** glucose, total protein, albumin
  - **Electrolytes:** sodium, potassium, chloride, corrected calcium, phosphate, magnesium, bicarbonate
  - **Renal function:** creatinine, blood urea nitrogen or urea,
  - **Liver function:** alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase, lactate dehydrogenase (LDH), total bilirubin

The troponin was also assessed as a biomarker for cardiac safety.

Technical formulas are described in Section 2.5.1.

#### *2.1.4.4 Vital signs variables*

Vital signs include: systolic and diastolic blood pressure, weight.

#### *2.1.4.5 Electrocardiogram variables*

ECG parameters include heart rate, PR interval, QRS interval, QT interval. QTcF interval will be derived according to the Fridericia's formula provided in Section 2.5.1.

#### *2.1.4.6 Other safety endpoints*

Other safety endpoints will include ECOG PS.

### 2.1.5 Pharmacokinetic variables of tusamitamab ravtansine

#### *2.1.5.1 Whole tusamitamab ravtansine arm*

In addition to the description of concentration observed at the end of intravenous infusion ($C_{eoi}$) and observed predose concentration ($C_{trough}$), plasma concentrations of tusamitamab ravtansine collected as described in section 8.5 of the protocol will be used for population PK analysis by non-linear mixed effects modeling. Data from previous conducted studies might be added for model development. This analysis will involve an estimation of inter- participant PK variability, the determination of the population PK parameters estimates and the quantitative evaluation of potential effect of participant characteristics on the main PK parameters. Empirical Bayesian estimation of individual exposure parameters such as maximum concentration ($C_{max}$), and area under the curve (AUC) will also be performed. Those individual exposure parameters will then be investigated as predictive factors for clinical outcomes including safety and efficacy endpoints, if possible.

The population PK analyses will be described in the Population PK analysis plan provided by PKDM Modeling and Simulation group. The results of the population PK analysis will be presented separately from the main clinical study report (CSR).

#### *2.1.5.2 Participants treated in China*

Intensive PK sampling as detailed in Protocol Section 10.8 will be done in the first 12 participants in China treated with tusamitamab ravtansine after implementation of the protocol amendment 2. The following pharmacokinetic parameters will be calculated using non-compartmental method from tusamitamab ravtansine plasma concentrations at Cycle 1.

The parameters will include, but may not be limited to the following:

**Table 6 - List of pharmacokinetic parameters and definitions**

| PK Parameter | Definition/Calculation |
|---|---|
| $C_{max}$ | Maximum observed concentration |
| $t_{max}$ | Time to reach $C_{max}$ |
| $C_{eoi}$ | Concentration observed at the end of intravenous infusion |
| $AUC_{0\text{-}14d}$ | Area under the curve (AUC) of plasma concentrations over time calculated using trapezoidal method from time zero to Day 14 after dosing. |
| AUC | AUC extrapolated to infinity according to the following equation:<br>$AUC = AUC_{last}+(C_{last}/\lambda_z)$<br>$AUC_{last}$ being the AUC from time zero to $t_{last}$ (time of last concentration observed above the LLOQ, ie, $C_{last}$)<br>If $AUC_{ext}$ >30%, the AUC value will not be taken into account in descriptive statistics and derived parameters will not be calculated. |

| PK Parameter | Definition/Calculation |
|---|---|
| CL | Total body clearance of the drug from plasma calculated using the following equation from AUC, on Cycle 1.<br>$CL = Dose/AUC$ |
| $V_{ss}$ | Volume of distribution in the terminal phase calculated according to the following equation:<br>$Vss = CL \times MRT$<br>with MRT = AUMC/AUC(-Tinf /2 if infusion), MRT being the Mean Residence Time of a molecule in the body; AUMC is the AUC of the moments and Tinf is the duration of infusion |
| $t_{1/2z}$ | Terminal half-life associated with the terminal slope determined according to the following equation:<br>$t_{1/2z} = 0.693/\lambda_z$<br>Where $\lambda_z$ is the slope of the regression line of the terminal phase of the plasma concentration versus time curve, in semi-logarithmic scale. Half life is calculated by taking the regression of at least 3 points on the terminal slope |

Abbreviations: LLOQ = Lower Limit of Quantification

No PK sample is planned to be collected during the crossover treatment phase.

### 2.1.6 Analyses endpoints for PK/PD

A separate PK/PD SAP will present the details of these PK/PD endpoints.

### 2.1.7 Immunogenicity endpoints

Anti-Therapeutic Antibodies (ATA) are drug-reactive antibodies, including pre-existing host antibodies that are cross-reactive with the administered biotherapeutic (baseline ATA). It comprises neutralizing and non-neutralizing ATA. ATA against tusamitamab ravtansine will be collected according to the pharmacokinetics/pharmacodynamics (PK/PD) flowcharts in Section 1.3 of the protocol for the tusamitamab ravtansine arm only and will be analyzed using the Immunogenicity population (see Section 2.3.5).

***Participant's ATA status***

- Participants with **pre-existing ATA**s correspond to participants with ATAs present in samples drawn before first administration of intervention. Participants with missing ATA sample at baseline will be considered as without pre-existing ATA.
- Participants with **treatment-emergent ATA**s correspond to participants with at least one treatment-induced/boosted ATA.
    - Participants with **treatment-induced ATA**s correspond to participants with ATAs that developed at any time after first IMP administration and without pre-existing ATA (including participants without pre-treatment samples).
    - Participants with **treatment-boosted ATA**s correspond to participants with pre-existing ATAs that are boosted at any time after first IMP administration to a significant higher titer than the baseline. If a 2-fold serial dilution schema is used during titration, so at least a 4-fold increase will be considered as significant.

  - Participants with **unclassified ATA** correspond to participants with pre-existing ATA that cannot be classified as treatment-boosted ATA because of missing titer(s) (ie, a positive ATA sample at any time after first IMP administration in a participant with pre-existing ATA but with missing titer at this sample or at baseline).

- Participants **without treatment-emergent ATA** correspond to participants without treatment-induced/boosted ATA and without any inconclusive sample nor unclassified ATA at any time after first IMP administration.
- Participants **with inconclusive ATA** are defined as participants which cannot irrefutably be classified as with or without treatment-emergent ATA.

***Neutralizing ATA / Non-neutralizing ATA***

- **Neutralizing ATA (NAb):** ATA that inhibits or reduces the pharmacological activity of the biotherapeutic, regardless of its in vivo clinical relevance (ie, whether or not test method results relate to clinical impact in the participant).
- **Non-neutralizing ATA (non-NAb):** ATA that binds to the biologic drug molecule but does not inhibit its pharmacological activity in an in vitro test, regardless of its in vivo clinical relevance (ie, whether or not test method results relate to clinical impact in the participant).

***Overview of the ATA response***

One main category can be reported for the epidemiology of an ATA immune response: ATA incidence:

- **ATA incidence** defines the proportion of participants found to either have seroconverted (treatment induced ATAs) or boosted their pre-existing ATA response (treatment-boosted ATA) at any time after first IMP administration.
- **NAb:** When applicable, report pre-existing NAb, and incidence of treatment induced NAbs as described above. If all ATA are neutralizing in all participants, a separate analysis is obviously redundant.

***Kinetics of ATA response***

Kinetics of ATA response will be derived for participants with treatment-induced/boosted ATA.

- **Time to onset of ATA response** is defined as the time period between the first IMP administration and the first treatment-induced/boosted ATA.
- **Duration of ATA response** is defined as the time between the first treatment-induced/ boosted ATA and the last treatment-induced/boosted ATA, irrespective of negative samples or positive samples not reaching the boosted threshold in-between. ATA duration will be summarized only for participants with persistent ATA response.
- **Persistent ATA response** is defined by treatment-induced/boosted ATA with a duration of ATA response of at least 16 weeks.

- **Transient ATA response** is defined by treatment-induced/boosted ATA with a duration of ATA response of less than 16 weeks and the last sample is not treatment-induced/boosted.
- **Indeterminate ATA response** is defined by treatment-induced/boosted ATA that are neither persistent nor transient.

For Nab, the same analysis will be applied when applicable.

No ATA sample is planned to be collected during the crossover treatment phase.

### 2.1.8 Pharmacodynamic/genomics endpoints

Blood samples (cfDNA and germline DNA) collected at pre infusion of Cycle 1 Day 1 are planned to be transferred to a centralized laboratory for circulating free DNA (cfDNA) / DNA extraction.

DNA samples will be analyzed for determination of tumor mutation profile on plasma cfDNA. Subtractive mutation/gene alteration analysis will be performed with germline DNA data to identify tumor-specific somatic genetic aberrations.

The extraction will be performed on a subset of samples which correspond to all samples of responders when available and samples from matched non-responders based on baseline characteristics. A targeted approach will be considered and different types of genetic alternations will be investigated such as single nucleotide variants, indel mutations, abnormal copy number variants and fusion genes. Depending on the genetic alteration, results will be provided as percentage (%) or as score and as presence/absence. In addition, data on locally assessed genetic alternations will be collected in CRF in all participants at prescreening: AKT1, ALK, BRAF, CDKN1B, CDKN2A, EGFR, ESR1, FGF4, HER2, KRAS, MDM2, MED12, MET, NRAS, PIK3CA, PTEN, RB1, RET, ROS1 or TP53 (list not exhaustive).

For pharmacodynamic purpose, the circulating CEA levels will be assessed using local testing at baseline, during the treatment and follow-up period at the time of laboratory assessment as close as possible to tumor assessment (and no more than 1-2 weeks) until disease progression. The circulating CEA will be considered as a continuous measure as well as considering different binary thresholds: eg, <5, <50 or <100 ng/mL.

### 2.1.9 Health-related quality-of-life endpoints

As presented in “secondary endpoints” Section 2.1.3.2.2, HRQOL is analyzed as Time to Deterioration (for disease related symptoms, physical functioning, and role functioning).

A separate QOL SAP will present additional, supportive analyses on the EORTC-QLQ-C30 and EORTC-QLQ-LC13 scales.

### 2.1.10 Further therapy after discontinuation of investigational medicinal product administration during the study

Further anti-cancer therapies after discontinuation of IMP include further systemic anti-cancer therapy, radiotherapy and surgery.

***Time to first further systemic anti-cancer therapy***

The first further systemic anti-cancer therapy is defined as the first systemic anti-cancer therapy starting after the first dose of study treatment.

The time to first further systemic anti-cancer therapy is defined as the time from date of randomization to start of the first further systemic anti-cancer therapy. In the absence of first further systemic anti-cancer therapy, time will be censored to last contact date (Section 2.5.1) or the cutoff date, whichever occurs first.

Only further systemic anti-cancer therapy will be considered as further anti-cancer therapy for PFS, BOR analyses and OS supportive analyses. For participants from the docetaxel treatment group who cross over, the first systemic further systemic anticancer therapy is the tusamitamab ravtansine treatment.

Further anticancer radiotherapies and surgeries will be described separately.

## 2.2 DISPOSITION OF PARTICIPANTS

This section describes participant disposition for both participant study status and the participant analysis populations.

Prescreened population consists of all participants who signed the prescreening informed consent for CEACAM5 assay assessment of their biopsy.

Screened participants are defined as all participants who signed the main study informed consent for study participation.

Randomized participants consist of all participants with a signed main study informed consent form who have had a treatment kit number allocated and recorded in the IRT database, regardless of whether the treatment kit was used or not.

For participant study status, the total number of participants in each of the following categories will be presented in the clinical study report using a flowchart diagram or summary table:

- Prescreened participants
- Prescreen failure participants and reasons for prescreen failure (if any)
- Screened participants
- Screen failure participants and reasons for screen failure
- Randomized participants
- Randomized but not treated participants

- Randomized and treated participants
- Participants who discontinued study treatment
- Participants in the docetaxel treatment arm who discontinued treatment due to PD and who did and did not cross over
- Participants still on treatment
- Status at last study contact

To be noted, for a rescreened participant who is thereafter screen failed a second time, the data of the first screening visit will be kept in the analysis. For a rescreened participant who is thereafter randomized, the data of the second screening visit will be kept in the analysis.

Number of prescreened participants, number of prescreen failed participants and reasons for prescreen failure, number of screened participants, number of screen failure participants and reasons for screen failure will be summarized in a separate table. In addition, number and percentage of randomized participants by country and sites will be presented.

A summary of the reasons for study treatment and study discontinuation by treatment group will be provided. Same summaries will be provided in the crossover treated population.

Reasons for permanent study treatment and study discontinuation “adverse event” and “other reasons” will be split as related versus not related to Covid-19, if applicable.

For all categories of participants (except for the prescreened, screened and non-randomized categories) percentages will be calculated using the number of randomized participants as the denominator.

In addition, the number and percentage of randomized participants will be presented for each of the 12 stratification levels and by treatment group.

A participant is considered lost to follow-up at the end of the study if he/she is not assessed at the last protocol planned visit and if the time from the last contact (Section 2.5.1) to the last protocol planned visit is greater than 12 weeks (+2 weeks) (greater than 98 days).

All critical or major deviations potentially impacting efficacy analyses, randomization, and drug-dispensing irregularities and other major or critical deviations will be summarized in tables giving numbers and percentages of deviations by treatment group, as well as displayed separately as related versus not related to Covid-19 if applicable.

Additionally, the following analysis populations will be summarized in a table by number of participants on the randomized population:

- Efficacy population: intent-to-treat (ITT) population
- Safety population
- Pharmacokinetics population
- Crossover treated population
- ATA population.
- Population without trial impact (disruption) due to COVID-19

Definitions of study populations are provided in Section 2.3.

Reasons for exclusion from the population without trial impact (disruption) due to Covid-19 will be summarized.

### 2.2.1 Randomization and drug dispensing irregularities

Randomization and drug-dispensing irregularities occur whenever:

1. A randomization is not in accordance with the protocol-defined randomization method, such as a) a participant is randomized based on an incorrect stratum, b) a participant is randomized twice.

   OR

2. A participant is dispensed an IMP kit not allocated by the protocol-defined randomization, such as a) a participant at any time in the study is dispensed a different treatment kit than as randomized (when not containing the correct-as-randomized IMP), or
   b) a nonrandomized participant is treated with IMP reserved for randomized participants.

Randomization and drug-dispensing irregularities to be prospectively identified include but are not limited to:

| Randomization and drug allocation irregularities |
|---|
| Kit dispensation without IRT transaction |
| Erroneous kit dispensation |
| Randomization by error |
| Participant randomized twice |
| Stratification error |

All randomization and drug-dispensing irregularities will be documented in the clinical study report. If the number of irregularities is large enough to make a tabular summary useful, the irregularities will be categorized and summarized among randomized participants (number and percentages). Nonrandomized, treated participants will be described separately.

## 2.3 ANALYSIS POPULATIONS

Participants treated without being randomized will not be considered randomized and will not be included in any efficacy population.

The safety experience of participants treated and not randomized will be reported separately, and these participants will not be in the safety population.

The randomized population is defined in Section 2.2.

For any participant randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be assessed separately.

### 2.3.1 Efficacy populations

#### *2.3.1.1 Intent-to-treat population*

The ITT population is the randomized population: all participants who have given their informed consent and who have had a treatment kit number allocated and recorded in the IRT database. Participants will be analyzed according to the treatment group allocated by randomization.

This population is the primary population for all efficacy parameters.

### 2.3.2 Safety population

The safety population is defined as all participants randomly assigned to study intervention and who have actually received at least one dose of study intervention. This population is the primary population for the analysis of safety parameters, for analysis of exposure and safety data before crossover for the docetaxel participants who cross over. Participants will be analyzed according to the treatment arm they actually received.

This population is the primary population for the analysis of all safety parameters

In addition:

- Nonrandomized but treated participants will not be part of the safety population; however, their safety data will be presented separately.
- Randomized participants for whom it is unclear whether they took the IMP will be included in the safety population as randomized.
- For participants receiving more than one study intervention during the main study phase, they will be analyzed according to the randomized arm if they received at least one administration of the as-randomized intervention.

### 2.3.3 Crossover treated population

The crossover treated population is defined as all participants randomized to docetaxel treatment group and who have received at least one dose of tusamitamab ravtansine in the crossover treatment phase. Exploratory analyses of safety and efficacy on the crossover treatment phase will be performed in this population.

### 2.3.4 Pharmacokinetic population

The PK population will include all participants from the safety population who have actually received at least one dose or a part of a dose of tusamitamab ravtansine with at least one evaluable concentration post-baseline and adequate documentation of sampling and dosing date and time.

Note: No PK sample is planned to be collected during the crossover treatment phase.

### 2.3.5 Immunogenicity population

The Immunogenicity population will include all participants from the safety population who have actually received at least one dose or a part of a dose of tusamitamab ravtansine with at least one evaluable ATA result post baseline.

Note: No ATA sample is planned to be collected during the crossover treatment phase.

### 2.3.6 Population without trial impact (disruption) due to COVID-19

The population without trial impact (disruption) due to COVID-19 will include all participants from the randomized population fulfilling the following conditions:

- without any critical or major deviation related to Covid-19,
- and who did not permanently discontinue treatment due to Covid-19
- and who did not permanently discontinue study due to Covid-19

## 2.4 STATISTICAL METHODS

Continuous data will be summarized using the number of available data, mean, standard deviation, median, minimum and maximum for each treatment group; Q1 and Q3 can be added as necessary. Categorical and ordinal data will be summarized using the number and percentage of participants in each treatment group. One-sided tests will be used in this study for the primary endpoints and key secondary endpoints. Confidence intervals will be estimated as two-sided and will be used for descriptive purposes only.

For the main study phase, data will be presented by treatment arm (as randomized or as treated): tusa rav arm and docetaxel treatment group until the crossover date for exposure and safety. For the participants docetaxel who cross over (the docetaxel - tusa rav crossover arm), the data after crossover will be presented separately. The objectives of the crossover arm are exploratory and no testing will be performed.

### 2.4.1 Demographics and baseline characteristics

Parameters described in Section 2.1.1 will be summarized on the randomized population by treatment group (as allocated by IRT) and overall using descriptive statistics.The demographics and key baseline characteristics will also be presented for the crossover treated population.

Analyses for the safety population will be included in the appendices if the size of the safety population is different (>10%) from the size of the randomized population for any treatment group.

Medical or surgical history will be summarized by SOC and PT (SOC will be sorted according to the internationally agreed [Appendix C] order and PT by decreasing frequency).

### 2.4.2 Prior or concomitant medications

The prior and concomitant medications will be presented for the randomized population. The concomitant medication will also be presented for the crossover treated population.

Medications will be summarized by treatment group according to the WHO-DD, considering the first digit of the anatomical therapeutic chemical (ATC) class (anatomic category) and the first 3 digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized, and participants will be counted once in each ATC category (anatomic or therapeutic) linked to the medication. Therefore, participants may be counted several times for the same medication.

The table for prior medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the overall incidence across treatment groups. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

The tables for concomitant medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the incidence in the tusamitamab ravtansine arm. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

***Premedication***

The number (%) of participants with premedications received before each cycle will be provided among the participants who received an infusion. For the participants with hypersensitivity events (using SMQ “Hypersensitivity” (Narrow) and defined as events occurring on the day or the day after infusion), a listing by cycle will detail the premedication status and the TEAE details.

### 2.4.3 Prior anticancer therapies

In addition, the following specific medications will be summarized. A table for prior anticancer therapies will be provided using the ATC code (chemical class) and the standardized medication name. This table will be sorted by decreasing frequency of ATC followed by medication names based on the overall incidence. In case of equal frequency regarding ATCs, the alphabetical order will be used.

### 2.4.4 Extent of investigational medicinal product exposure and compliance

The extent of IMP exposure will be assessed and summarized by actual treatment within the safety population (ie, taking into account all cycles administered before the crossover date) and on the crossover treated population (ie, taking into account all cycles administered after the crossover date) (Section 2.3.2). In addition, summaries will be provided by trial impact (disruption) due to Covid-19.

#### *2.4.4.1 Extent of investigational medicinal product exposure*

The dose information will be assessed by the following variables:

- Number of cycles started, defined by maximum of cycles started on tusamitamab ravtansine or docetaxel respectively.
- Duration of overall exposure (in weeks) is defined as following
  - (last day of the last cycle - first dosing day of the first cycle + 1) / 7.

  The first dosing day of the first cycle is defined as the date of the first dose of study treatment at Cycle 1. The last day of the last cycle is defined as follows:
  - Date of last dose of tusamitamab ravtansine + 14 -1 days
  - Date of last dose of Docetaxel + 21 -1 days

  In addition, for crossover participants, duration of exposure will be calculated as follows:
  - Before crossover: [(min (crossover date-1, last day of last cycle of docetaxel) – first dosing day of the first cycle) +1 /7],
  - After crossover: [(last day of tusamitamab ravtansine of (docetaxel - tusa rav CO) – first administration of tusamitamab ravtansine + 1)/7].

  Where the last day of last cycle of Doc + tusa rav CO is defined as follows:
  - Date of last dose of tusamitamab ravtansine during the crossover treatment phase + 14 -1 days
- Cumulative exposure to treatment (in participant - years) derived by summing the duration of exposure by treatment groups.

The total number of cycles started, number of cycles started by participants as a quantitative variable and by category, duration of overall exposure and cumulative exposure to treatment will be summarized by descriptive statistics.

#### *2.4.4.2 IMP exposure*

The dose information will be assessed by the following:

- Actual dose (mg/m$^2$): for a given cycle, the actual dose in mg/m$^2$ corresponds to the actual dose in mg administered during the infusion. In case of dose interruption during a cycle, actual dose per cycle will be sum of the actual dose administered before and after the dose interruption.
- Cumulative dose (mg/m$^2$): the cumulative dose is the sum of all actual doses of IMP, expressed in mg/m$^2$, given from first to last administration
- Actual dose intensity (ADI, in mg/m$^2$/week): defined as the cumulative dose divided by the duration of IMP exposure (in weeks)
- Planned dose intensity (PDI, in mg/m$^2$/week): corresponds to the planned dose at Cycle 1 Day 1 in the main study phase or at Cycle Day 1 CO in the crossover treatment phase multiplied by the theoretical total number of doses started and divided by the theoretical cycle duration expressed in weeks (ie, 2 weeks per cycle started for tusamitamab ravtansine and 3 weeks per cycle started for Docetaxel)

- Relative dose intensity (RDI, in %):

$$100 \times \frac{\text{ADI (mg/m}^2\text{/week)}}{\text{PDI (mg/m}^2\text{/week)}}$$

The ADI and RDI (quantitative and qualitative variable: <80%, [80% – 100%] and >100%) will be summarized by descriptive statistics.

***Dose or cycle modifications***

The following variables will be derived to describe dose modifications and dose interruptions:

- Cycle delay: A cycle is deemed as delayed if
  - For tusamitamab ravtansine randomized participants in the main study phase and for docetaxel randomized participants who cross over in the crossover treatment phase: the start date of the current cycle – 14 – start date of the previous cycle is >3 days.
  - For docetaxel randomized participants: the start date of the current cycle – 21 – start date of the previous cycle is >4 days.

Cycle delay is not defined for the first cycle.

- Dose reduction: The first administration will not be counted as a dose reduction. For the second and subsequent IMP administrations, dose reduction will be determined using the dose level intervals provided in Table 7 and Table 8, by comparing the current dose level to the previous dose level. If the current dose level is not within the same or higher dose level interval as the previous dose administration, then the current dose level is considered reduced.

**Table 7 - Tusamitamab ravtansine dose reduction criteria ( in the main study or in the crossover treatment phase)**

| Actual dose level ($mg/m^2$) | Dose level interval ($mg/m^2$) |
|---|---|
| Initial dose (100) | >90 |
| Dose level -1 (80) | >72.5 and ≤90 |
| Dose level -2 (65) | >57.5 and ≤72.5 |
| Dose level (low dose) | >0 and ≤57.5 |

**Table 8 - Docetaxel dose reduction criteria**

| Actual dose level ($mg/m^2$) | Dose level interval ($mg/m^2$) |
|---|---|
| Initial dose (75) | >67.5 |
| Dose level -1 (60) | >55 and ≤67.5 |
| Dose level -2 (50) | >40 and ≤55 |
| Dose level (low dose) | >0 and ≤40 |

- Dose interruption: A dose will be considered to be interrupted if the IMP administration is stopped during an infusion before it is completed regardless of whether it is further restarted or not.

Dose modifications and dose interruptions will be analyzed by participant, and by cycle as follows:

- Participant (for the following variables, the number of participants treated will be used for % calculation, unless otherwise noted)
  - Number (%) of participants with at least 1 dose modification:
    - Number (%) of participants with at least 1 cycle delayed:
      - Number (%) of participants with a cycle delayed ≤ 7 days, between 8 and 14 days (included) and ≥15 days
    - Number (%) of participants with at least 1 dose reduction
      - Number (%) of participants with exactly 1 dose reduction
      - Number (%) of participants with at least two dose reductions
  - Number (%) of participants with a least 1 dose interruption
- Cycle (for the following variables, the number of cycles started will be used for % calculation, unless noted otherwise)
  - Total number of cycles started (used for % calculation)
  - Number (%) of cycles with at least 1 dose modification:
    - Number (%) of cycles with at least one dose delayed
      - Number (%) of cycles delayed ≤ 7 days and >7 days
    - Number (%) of cycles with at least 1 dose reduction
  - Number (%) of cycles with at least 1 dose interruption.
    - Number (%) of cycles with at least 1 dose interrupted and not re-started

In addition, time to first dose modification and time to first dose reduction will be described using Kaplan-Meier method by treatment group. Time to first dose modification is defined as the time from the date of first IMP administration to the date of first dose modification. Time to first dose reduction is defined as the time interval from first IMP administration to the date of first dose reduction. In the absence of dose modification or dose reduction respectively before the analysis cut-off date, it will be censored at the last date of the last IMP administration, or analysis cut-off date, whichever occurs first.

### 2.4.5 Analyses of efficacy endpoints

All efficacy analyses will be performed on the ITT population unless stated otherwise. All analyses using the stratification factors will be performed using the stratification factor as per IRT unless stated otherwise.

All primary and secondary efficacy endpoints based on radiological assessments of tumor burden (ie, PFS, BOR, ORR and DOR) will be derived using the IRC tumor assessment. Analyses based on local radiologist's/Investigator's assessment will be considered as supportive analyses.

#### *2.4.5.1 Analysis of primary efficacy endpoints*

This study is designed to test the following statistical set of hypotheses on PFS and on OS respectively as multiple primary endpoints:

H0: the survival distribution function (SDF) of the tusamitamab ravtansine treatment group is lower or the same as the survival distribution function for the docetaxel treatment group
SDF (tusamitamab ravtansine) ≤ SDF (docetaxel)

versus

H1: the survival distribution function of tusamitamab ravtansine treatment group is superior to the survival distribution function of docetaxel treatment group
SDF(tusamitamab ravtansine) > SDF (docetaxel)

where SDF denotes the survival distribution function of the parameter PFS or OS respectively.

##### *2.4.5.1.1 Overall survival*

###### 2.4.5.1.1.1 Primary analysis

Primary efficacy analysis will consist of OS comparison between the tusamitamab ravtansine arm and docetaxel treatment group through a log-rank test procedure stratified by the stratification factors as entered in the IRT. See Section 2.4.5.3 for details about the type-I error level to be used for statistical testing.

In addition, the following estimates will be provided for OS:

- The hazard ratio (HR) and its corresponding 95% and α-adjusted two-sided CIs will be estimated using the Cox proportional hazards model stratified by the same stratification factors as those used for the log-rank test. Ties will be handled using the exact method. Underlying assumptions of the Cox proportional hazard model will be assessed by graphical methods (ie, log-log graphical methods.)
- OS data will be summarized using the Kaplan-Meier method by treatment group:
  - Kaplan-Meier estimates of the 25th, 50th, and 75th percentiles and their associated 95% CIs will be provided. The 95% CIs will be constructed using a log-log transformation of the survival function and the method of Brookmeyer and Crowley.
  - Number of participants at risks as well as the probabilities of surviving for example at least 3, 6, 9, 12 ,15, 18, 21, 24 months with 95% CIs (if possible depending on the time of the analysis) will be estimated for each treatment group using the Kaplan-Meier method and a log-log approach based on a normal approximation following the Greenwood's formula.

- The number of censored participants, the reasons for their censoring (ie, alive at the cut-off date, alive at the last contact (Section 2.5.1) before the cut-off date, and lost to follow-up) and the time between the date of last contact and the cut-off date will be summarized by treatment group.
- Kaplan-Meier curves will be plotted. These plots will include the number of participants at risk at key time points by treatment group

- Follow-up duration (months) will be defined as the time interval from the date of randomization to the date of last contact (Section 2.5.1). Participants who have died will be censored on their date of death. Kaplan-Meier estimates of the 25$^{th}$, 50$^{th}$ and 75$^{th}$ percentiles will be provided.

#### 2.4.5.1.1.2 Sensitivity analyses

The same statistical methods used in the primary analysis will be applied using stratification rules as defined below.

*Sensitivity analysis #1 (OS analysis using stratification factors derived from eCRF data)*

If more than 10% participants have a discordance between the strata as entered in the IRT system and as derived from eCRF data, OS endpoint will be analyzed stratified by the stratification factors as derived from eCRF data.

*Sensitivity analysis #2 (sensitivity to stratification)*

In order to assess the robustness of the primary analysis to stratification factors, analysis of OS by using unstratified log-rank test procedure and unstratified Cox's regression model will be performed.

***Non-proportional hazard***

As mentioned above, underlying assumptions of the Cox Proportional hazards model will be evaluated to assess the relevance of estimating the treatment effect from hazard ratio. In case the proportional hazards assumption may not be valid, Restricted Mean Survival Time (RMST) method (2) may be conducted for OS. The RMST methodology is valid under any distribution of the time to event in the treatment groups and provides an estimate of the expected OS between randomization and a common timepoint denoted by $\tau$. The timepoint $\tau$ should be limited to the largest event time:

- $\tau_{max}$ = minimum of (largest observed OS event time for tusamitamab ravtansine arm, largest observed OS event time for docetaxel treatment group).

The RMST estimate up to $\tau_{max}$ and associated 95% CI for each treatment arm will be provided. The treatment effect will be estimated based on the difference between the two treatment arms in RMST up to $\tau_{max}$. Similar analysis will be performed using clinical meaningful truncation points (eg, up to 18 months if possible based on the largest event time within each treatment arm). Additionally, the RMST estimate within each arm will be plotted against time $\tau$, with $\tau$ varying

from 0 to the largest observed OS event time for the respective arm. Treatment effect based on the difference in RMST between the two treatment arms and associated 95% CI will also be provided against time τ, varying from 0 to $\tau_{max}$.

#### 2.4.5.1.1.3 Supportive analyses

Supportive analyses adjusting OS for switch to subsequent anticancer therapies (including the switch to tusamitamab ravtansine for docetaxel randomized participants who cross over) could be performed at interim and/or final analyses (eg, using inverse probability of censoring weighting (IPCW) and Rank Preserving Structural Failure Time Model (RPSFTM) methods) (3).

#### 2.4.5.1.1.4 Subgroups analyses

***Evaluation of Consistency***

The consistency of the results from the primary analysis will be evaluated across pre-defined subgroups. The definition of each subgroup is defined in Table 9. Depending upon the study results, additional subgroups may be examined, and subgroups with small sample sizes may be pooled to create a larger meaningful subgroup. Some subgroup analyses may not be conducted if not relevant based on number of participants. For each subgroup, Kaplan-Meier estimates of the median and its associated 95% CI will be provided for each treatment arm along with the HR and its 95% CI estimated using the unstratified Cox proportional hazards model. A forest plot summarizing the results for each subgroup will be provided.

**Table 9 - Subgroups analyses: covariates investigated**

| Prognostic factor | Description |
|---|---|
| ECOG Performance status as per IRT | 0 vs 1 |
| Geographical region as per IRT | Asia vs Rest of the World vs Western Europe + North America+ Australia |
| ICI treatment administration as per IRT | Sequential vs combination with chemotherapy |
| Age | <65 years vs ≥65 years |
| Sex | Female vs Male |
| Race | White vs Other |
| EGFR alteration | Yes vs No |
| Brain metastases at baseline as per eCRF | No vs Yes |
| Smoking status | Never versus Smokers (former and current) |
| Number of organs involved as per eCRF (Including the primary site of the tumor) | <3 versus ≥3 |
| Prior taxanes | No vs Yes |

***Evaluation of interactions***

For each pre-defined factor defined in Table 9, OS will be analyzed using an unstratified Cox proportional hazards model with terms for the factor, treatment and their interaction.

***Evaluation of confounding factors***

Since the results from the primary analysis could be impacted by confounding factors, any potential issues will be examined and, if confirmed, exploratory analysis of the primary endpoint will be done accordingly. A multivariate Cox proportional hazards model will be used to identify prognostic factors among the demographic and baseline characteristics factors described in the Table 9 using a stepwise selection procedure with a 15% significance level for entering effects and a 10% significance level for removing effects. For significant prognostic factors identified in the multivariate model, the balance between treatment groups will be assessed. If major confounding factor is identified through screening for treatment group imbalances in a prognostic factor at baseline), adjustments to the model will be performed. The treatment effect for OS will be re-estimated after adjusting for the prognostic factors in the multivariate Cox proportional hazards model. An exploratory analysis of OS will be done after adjusting for the prognostic factors in the multivariate Cox proportional hazards model. Differences between the adjusted and unadjusted models will be discussed in the clinical study report.

#### *2.4.5.1.2 Progression Free Survival*

##### 2.4.5.1.2.1 Primary analysis

Primary efficacy analysis of PFS will consist of PFS comparison between the tusamitamab ravtansine arm and docetaxel treatment group through a log-rank test procedure stratified by the stratification factors as entered in the IRT. See Section 2.4.5.3 for details about the type-I error level to be used for statistical testing.

In addition, the following estimates will be provided:

- The hazard ratio (HR) and its corresponding 95% and α-adjusted two-sided CIs will be estimated using the Cox proportional hazards model stratified by the same stratification factors as those used for the log-rank test described above. Ties will be handled using the exact method. Underlying assumptions of the Cox Proportional hazards model will be assessed by graphical methods (ie, log-log graphical methods).
- PFS data will be summarized using the Kaplan-Meier method by treatment group:
  - Kaplan-Meier estimates of the 25th, 50th and 75th percentiles and their associated 95% confidence interval (CIs) will be provided. The 95% CIs will be constructed using a log-log transformation of the survival function and the method of Brookmeyer and Crowley.
  - Number of participants at risk as well as the probabilities of being event-free for example at least 2, 4, 6, 8, 10 and 12 months with 95% CIs will be estimated for each treatment group using the Kaplan-Meier method and a log-log approach based on a normal approximation following the Greenwood's formula.
  - Kaplan-Meier curves will be plotted. These plots will include the number of participants at risk at key time points by treatment group.

- For participants with event, the type of event (documented disease progression or death without documented disease progression) will be summarized by treatment group using counts and percentages.
- For participants who died without documented disease progression, the time from the last evaluable disease assessment/randomization (if no evaluable post-baseline tumor assessment) to the death will be summarized by treatment group using descriptive statistics.
- The number (%) of censored participants, the reason and timing of their censoring (ie, censored at randomization (if no evaluable post-baseline tumor assessment), censored at last evaluable tumor assessment before the initiation of further systemic anticancer therapy, or censored at last evaluable tumor assessment before the cut-off date), and the time from the last evaluable disease assessment/randomization (if no postbaseline tumor assessment) to the cut-off date will be summarized by treatment group.
- Follow-up time duration (months) will be estimated by treatment group using the reverse Kaplan-Meier method, where censored data are treated as events and events are treated as censored data. Kaplan-Meier estimates of the 25$^{th}$, 50$^{th}$ and 75$^{th}$ percentiles will be provided.

#### 2.4.5.1.2.2 Sensitivity analyses

The same statistical methods used in the primary analysis will be applied using different censoring and event rules or stratification rules as defined below.

***Different censoring and events rules***

The sensitivity analyses will include the following censoring rules:

- Ignoring further systemic anti-cancer therapy and considering events (documented progression or death) occurring immediately after two or more missing or non-evaluable tumor assessments as events.
- Considering events (documented progression or death) occurring immediately after two or more missing or non-evaluable tumor assessments as events and back-dated to the next schedule assessment.

Additional details are provided in Appendix F.

*Sensitivity analysis #1 (PFS considering events occurring immediately after two or more missing or non-evaluable tumor assessments as events and ignoring further systemic anti-cancer therapy)*

PFS endpoint will be analyzed based on IRC assessment, including events (documented progression or death) occurring immediately after two or more missing or non-evaluable tumor assessment as event. Events occurring after the start of any further systemic anticancer therapy will be included. The date of progression (or death) will be used for date of outcome.

*Sensitivity analysis #2 (PFS considering events occurring immediately after two or more missing or non-evaluable tumor assessments as events and back-dating at the next schedule assessment)*

If more than 10% of participants have two or more consecutive missing or non-evaluable assessments prior to PFS event, PFS endpoint will be analyzed based on IRC assessment, including events (documented progression or death) occurring immediately after two or more non-evaluable or missing tumor assessments as events. The date of the next schedule assessment will be used for date of outcome.

***Stratification factors***

*Sensitivity analysis #3 (PFS analysis using stratification factors derived from eCRF data)*

If more than 10% participants have a discordance between the strata as entered in the IRT system and as derived from eCRF data, PFS endpoint will be analyzed based on IRC assessment stratified by the stratification factors as derived from eCRF data.

*Sensitivity analysis #4 (sensitivity to stratification)*

In order to assess the robustness of the primary analysis to stratification factors, analysis of PFS by using unstratified log-rank test procedure and unstratified Cox's regression model will be performed.

***Non-proportional hazard***

Same analyses as for OS apply for PFS with the primary analysis censoring rule.

#### 2.4.5.1.2.3 Supportive analyses

PFS analysis will be supported by the PFS according to Investigator/local radiologist's assessment, with the same censoring rules as for the PFS primary analysis. A sensitivity analysis will be performed considering the non-radiological progression as event, with the date of non-radiological progression as date of outcome. In addition sensitivity analyses could be performed using different censoring and event rules or stratification rules as defined in Section 2.4.5.1.2.2, if relevant.

***Concordance of PFS outcome***

A comparison of PFS outcome (ie, "Event", "Censored") between the IRC and the investigator assessments will be summarized for each treatment group.

**Table 10 - Cross-tabulation of IRC and Investigator assessments of PFS outcome**

| Investigator/local radiologist | IRC | |
|---|---|---|
| | Event | Censored |
| Event | $n_{11}$ | $n_{12}$ |
| Censored | $n_{21}$ | $n_{22}$ |

The PFS Outcome Discrepancy Rate (PFS ODR) will be calculated for each treatment arm as follow.

$$\text{PFS ODR} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

***Concordance of tumor assessment evaluation***

The differential discordance between the IRC and the Investigator assessment of documented progression will be assessed using the Pharmaceutical Research and Manufacturers of America (PhRMA) method (4). The early discrepancy rate (EDR) and late discrepancy rate (LDR) differences between the two treatment arms will be calculated as follows:

**Table 11 - Cross-tabulation of IRC and Investigator assessments of documented PD**

| Investigator/local radiologist | IRC | |
|---|---|---|
| | **Documented PD** | **No documented PD** |
| Documented PD | a = $a_1$ + $a_2$ + $a_3$ | b |
| No documented PD | c | d |

Only documented PD component of PFS is considered; if death occurs without prior PD, a participant is counted under "No documented PD"
$a_1$: number of agreement on timing and occurrence of documented PD
$a_2$: number of times Investigator declares documented PD later than IRC
$a_3$: number of times Investigator declares documented PD earlier than IRC

The timing of Investigator/local radiologist and IRC (for participants with agreement on documented PD) will be considered to agree if they occur within ±5 days of each other, aligned with the protocol-specified window for tumor assessments.

The EDR is defined as:

$$EDR = \frac{b + a_3}{a + b}$$

The EDR quantifies the frequency with which the Investigator declares progression early relative to IRC as a proportion of the total number of Investigator assessed PDs, within each arm.

The LDR is defined as:

$$LDR = \frac{c + a_2}{b + c + a_2 + a_3}$$

The LDR quantifies the frequency with which the Investigator declares progression later than IRC as a proportion of the total number of discrepancies, within each arm.

If the distribution of discrepancies is similar between the arms then this suggests the absence of evaluation bias favoring a particular arm.

The EDR and LDR will be calculated for each treatment arm and the differential discordance (DD) for each measure will be summarized as the rate on the tusamitamab ravtansine arm minus the rate on the docetaxel arm. A negative differential discordance for the EDR and/or positive differential discordance for the LDR are suggestive of a bias in the Investigator favoring the experimental arm.

##### 2.4.5.1.2.4 Additional analyses linked to COVID-19

The potential impact of COVID-19 on missed TAs can also be addressed by comparing the different PFS analyses. In the primary PFS analysis, two or more consecutive missing or non-evaluable TAs lead to censoring. PFS sensitivity #1 will assess the impact of considering events (documented progression or death) after two or more missing or non-evaluable consecutive TAs as an event.

##### 2.4.5.1.2.5 Subgroups analyses

Same analyses as for OS (evaluation of consistency/interactions/of confounding) apply for PFS.

### *2.4.5.2 Analyses of secondary efficacy endpoints*

Analysis of response-based endpoints (ie, ORR, DOR) will be performed primarily on the ITT population. Quality of life analyses will be conducted on the ITT population.

#### *2.4.5.2.1 Overall response rate*

##### 2.4.5.2.1.1 Primary analysis

Of note, the BOR for each participant will also be summarized by treatment arm according to IRC and Investigator assessments. BOR according to IRC will be derived from Overall response determined by IRC.

ORR according to IRC and Investigator assessments (supportive analysis) will be summarized by treatment arm with descriptive statistics at the time of the final analysis on PFS/58% IA of OS, at the time of the 80% IA of OS (including participants from the ITT population with randomization date ≥8 weeks prior to analysis cut-off date or with early PD prior to the cut-off date) and at the time of the final OS analysis. In addition, 95% confidence intervals will be computed using the Clopper-Pearson method.

For the comparison between the tusamitamab ravtansine arm and the docetaxel treatment group, the odds ratio and its 95% and α-adjusted (if feasible) two-sided CIs and the p-value will be provided from a Cochran-Mantel Haenszel test stratified according to the stratification factors as entered in the IRT. See Section 2.4.5.3 for details about the type-I error level to be used for statistical testing.

#### 2.4.5.2.1.2 Sensitivity analyses

If more than 10% participants have a discordance between the strata as entered in the IRT system and as derived from eCRF data, a sensitivity analysis with stratification factors derived from eCRF data will be performed based on a Cochran-Mantel Haenszel test.

Additional analyses of ORR will be provided according to subgroups of interest (Table 9).

#### 2.4.5.2.1.3 Crossover analyses

Post crossover efficacy endpoints (eg, ORR, BOR) will be analyzed separately in the population of crossover treated participants and will be considered as exploratory.

BOR for each participant will also be summarized according to Investigator assessments. ORR computed from BOR as reported by the investigator will be summarized with descriptive statistics. In addition, 95% confidence intervals will be computed using the Clopper-Pearson method.

### *2.4.5.2.2 Health-Related Quality of life*

The analysis will be performed at the time of the final analysis on PFS/58% IA of OS, at the time of the 80% IA of OS and at the time of the OS final analysis.

#### 2.4.5.2.2.1 Completion rate in disease symptoms, PF and RF scores

Instrument completion rate by week will be reported for each score (each disease symptoms, PF and RF scores):

- Unadjusted completion rate by week will be calculated as the number of participants meeting at least the minimum requirements for scoring of the instrument divided by the number of participants in the ITT population
- Adjusted completion rate by week will be calculated as the number of participants meeting at least the minimum requirements for scoring of the instrument among those who were expected to complete the questionnaire. To be noted that a participant is expected to complete the PRO assessment if the participant is alive and still on-treatment.

The criteria that will be used to define the PRO assessment study weeks are displayed in Table 12. In order to include the EOT visits in the compliance rate outputs, an analysis visit variable will be derived for each participant and assessment, so that the EOT visit will be remapped to a study week timepoint.

The time windows will be exhaustive so that data recorded at any timepoint have the potential to be summarized, including the EOT assessment. Inclusion within the visit window will be based on the actual date and not the intended date of the visit. The window for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half-way between two consecutive visits.

If there is more than one value per participant within an assessment window then the closest to the planned study day value should be summarized, or the earlier in the event the values are equidistant from the planned study day. The value at a given timepoint will be missing if no assessment was reported within the specified assessment window around the planned study day.

**Table 12 - Analysis window during study (including treatment period and EOT)**

| Analysis visit | Start day | Target day | End day |
|---|---|---|---|
| Baseline | Up to day 1 | 1 | 1 |
| Week 7 | 22 | 43 | 63 |
| Week 13 | 64 | 85 | 105 |
| Week W | (W-1) *7 – 20 | (W-1) * 7 + 1 | (W-1) *7 + 21 |

#### 2.4.5.2.2.2 Descriptive statistics at baseline in disease symptoms, PF and RF scores

The five QOL Scores at baseline will be summarized using descriptive statistics.

#### 2.4.5.2.2.3 Time-to-deterioration in disease symptoms, PF and RF scores

Primary efficacy analysis will consist of TTD comparison between the tusamitamab ravtansine arm and docetaxel treatment group through a log-rank test procedure stratified by the stratification factors as entered in the IRT. See Section 2.4.5.3 for details about the type-I error level to be used for statistical testing.

In addition, the following estimates will be provided for TTD:

- The hazard ratio and its corresponding 95% and α-adjusted (if feasible) two-sided CIs will be estimated using the Cox proportional hazards model stratified by the same stratification factors as those used for the log-rank test. Ties will be handled using the exact method. Underlying assumptions of the Cox proportional hazard model will be assessed by graphical methods (ie, log-log graphical methods).
- TTD data will be summarized using the Kaplan-Meier method by treatment group:
  - Kaplan-Meier estimates of the 25th, 50th, and 75th percentiles and their associated 95% CIs will be provided. The 95% CIs will be constructed using a log-log transformation of the survival function and the method of Brookmeyer and Crowley
  - Number of participants at risks as well as the probabilities of being event-free at for example least 3, 6, 9, 12 ,15, 18 months with 95% CIs (if possible depending on the time of the analysis) will be estimated for each treatment group using the Kaplan-Meier method.
  - The number of censored participants, the reasons and timing for their censoring (censored at first IMP administration or randomization (if participant randomized and not treated), censored at last non-missing assessment before the initiation of a further systemic anti-cancer therapy, censored at last non-missing assessment before the cut-off date, censored at last non-missing assessment up to end of treatment), and the time from the last non-missing assessment to the cut-off date for participants still on-

treatment will be summarized by treatment group. Kaplan-Meier curves will be plotted. These plots will include the number of participants at risk at key time points by treatment group.

#### 2.4.5.2.2.4 Sensitivity analyses

The same statistical methods used in the primary analysis will be applied using different timing rules for events or stratification rules as defined below.

The sensitivity analyses will include:

- Events occurring after one or more missing assessments are back-dated to the next schedule assessment.
- Sensitivity to stratification.
- Sensitivity to the definition of baseline questionnaire

*Sensitivity analysis #1 (TTD considering events occurring after one or more missing assessment back-dated at the next scheduled assessment)*

In case of one or more consecutive missing PRO assessments, TTD endpoints will be analyzed considering events occurring after one or more non-evaluable PRO assessment as event backdated at the next schedule assessment. The date of the next schedule assessment will be used for date of outcome (Section 2.5.3).

*Sensitivity analysis #2 (TTD analyses using stratification factors derived from eCRF data)*

If more than 10% participants have a discordance between the strata as entered in the IRT system and as derived from eCRF data, TTD endpoint will be analyzed stratified by the stratification factors as derived from eCRF data.

*Sensitivity analysis #3 (sensitivity to stratification)*

In order to assess the robustness of the primary analysis to stratification factors, analysis of TTD by using unstratified log-rank test procedure and unstratified Cox's regression model will be performed.

*Sensitivity analysis #4 (sensitivity to baseline value)*

In order to assess the robustness of the primary analysis to the definition of the baseline value, analysis of TTD by defining baseline value as the last questionnaire with day and time before the first IMP administration will be performed.

#### 2.4.5.2.2.5 Supportive analysis

Further sensitivity analyses and supportive analyses (using different definition of the clinically meaningful change, subgroup analyses) will be detailed in specific QOL SAP.

#### *2.4.5.2.3 Duration of response*

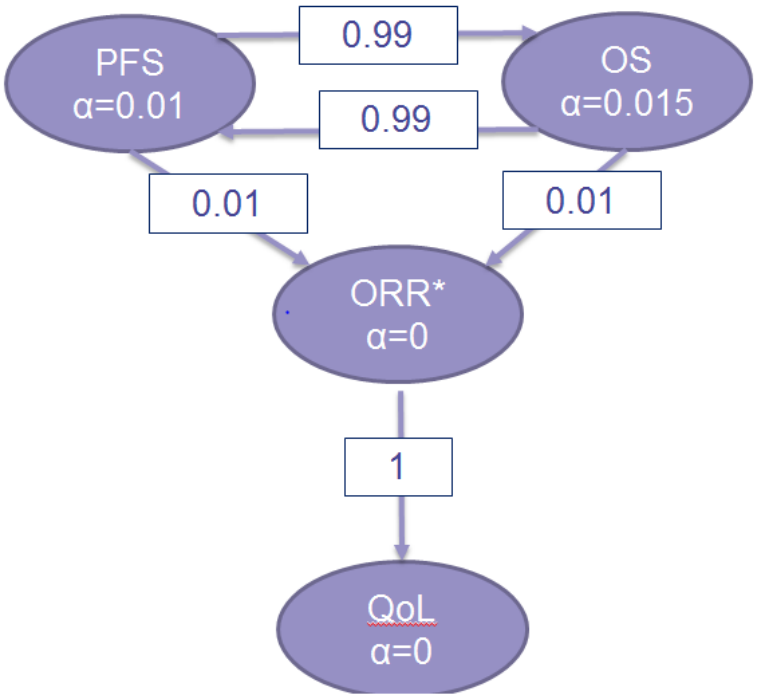
The DOR will only be summarized on the subgroup of participants who have achieved objective response. DOR by treatment will be summarized using Kaplan-Meier methods and displayed graphically, if appropriate. The median DOR and associated 95% CI will be provided.

As a supportive analysis, DOR according to the Investigator's assessment will also be determined.

### *2.4.5.3 Multiplicity issues*

The trial uses the graphical method of Maurer and Bretz (5) to provide a strong type-I error control for multiple hypotheses. Figure 2 shows the initial one-sided α-allocation for each hypothesis (in the ellipse) on the endpoints of PFS, OS, ORR and QoL. The weights for reallocation from each hypothesis to the others are represented in the boxes on the lines connecting hypotheses.

**Figure 2 - Type I error reallocation strategy**




*If both PFS and OS are significant, based on the graphical testing procedure, the entire α=0.025 will be reallocated to the testing of ORR

**PFS**

PFS hypothesis will be tested at α=0.01. When OS test is significant, the PFS hypothesis may be tested at α=0.02485 (0.99*0.015+0.01) (re-allocated α). With 221 PFS events at final analysis the study has ~95% power for detecting a HR of 0.615 at one-sided alpha 0.02485. Table 13 demonstrates the boundary properties for PFS hypothesis testing. The HR of PFS between the experimental group and control group is assumed to be 0.615.

**Table 13 - Boundary properties for planned analyses of PFS**

| Analysis | Boundary | α=0.01 | α=0.02485 |
|---|---|---|---|
| IF: 100% | Critical value | -2.326 | -1.963 |
| N: 351 | | | |
| Events: 221 | Threshold for the observed HR to be significant | 0.731 | 0.768 |
| Month: 40 | | | |
| | Prob of significant if true HR=0.615 | 0.901 | 0.951 |

## OS

The OS hypothesis will be tested at α=0.015. When PFS test is significant, the OS hypothesis may be tested at α=0.0249 (0.99*0.01+0.015) (re-allocated α). Table 14 demonstrates the bounds and boundary properties for OS hypothesis testing. The HR of OS between the experimental group and control group is assumed to be 0.72. Table 14 summarizes the alpha spending for interim and final analyses based on the planned number of events. The actual alpha spending will be based on the actual number of events included in the analyses and determined by the Lan-DeMets O'Brien-Fleming spending function at the time of interim and final analyses.

**Table 14 - Boundary properties for planned analyses of OS**

| Analysis | Boundary | α=0.015 | α=0.0249 |
|---|---|---|---|
| First **IA** | | | |
| IF: 58% | Critical value | -2.992 | -2.728 |
| N: 351 | Significance level (one-sided) | 0.0014 | 0.0032 |
| Events: 210 | Threshold for the observed HR to be significant | 0.662 | 0.686 |
| Month: 40 | Prob significant if true HR=1 | 0.0014 | 0.0032 |
| | Prob significant if true HR=0.72 | 0.2702 | 0.3641 |
| Second IA | | | |
| IF: 80% | Critical value | -2.508 | -2.287 |
| N:436 | Significance level (one-sided) | 0.0061 | 0.0111 |
| Events: 290 | Threshold for the observed HR to be significant | 0.745 | 0.764 |
| Months: 48 | Cumulative prob of significant if true HR=1 | 0.0065 | 0.0120 |
| | Cumulative prob of significant if true HR=0.72 | 0.6178 | 0.6997 |
| **Final** | | | |
| IF: 100% | Critical value | -2.227 | -2.031 |
| N: 450 | Significance level (one-sided) | 0.0130 | 0.0211 |
| Events: 363 | Threshold for the observed HR to be significant | 0.792 | 0.808 |
| Month: 58 | Cumulative prob of significant if true HR=1 | 0.0145 | 0.0236 |
| | Cumulative prob of significant if true HR=0.72 | 0.8143 | 0.8573 |

IF: Information fraction; IA: Interim Analysis

**ORR**

Following the graphical testing procedure, the ORR hypothesis will be tested at the following α level shown in Table 15 depending on the results from testing PFS and OS**.** ORR will be tested if either PFS or OS is significant. Then, a user-defined spending function will be applied to the α level using a fixed α level of 0.00001 for ORR hypothesis testing at each IA, and a fixed α level for final analysis, computed as (α level for testing ORR [as defined in Table 15] - 2 × 0.00001).

**Table 15 - ORR hypothesis testing level**

| α level for testing ORR hypothesis | Condition |
|---|---|
| 0.0001 | if only PFS is significant, and OS is not significant |
| 0.00015 | if only OS is significant at either IA or final analysis, and PFS is not significant |
| 0.025 | if PFS is significant and OS is significant at either IA or final analysis |

Abbreviations: IA=interim analysis; OS=overall survival; PFS=progression-free survival.

**TTD endpoints for QoL**

If ORR is significant, then QoL key secondary endpoints will be tested at the same significance level as ORR following a hierarchical order: TTD in disease-related symptoms, TTD in physical functioning, TTD in role functioning.The procedure, including the user-defined α-spending function, is similar to the hypothesis testing of ORR.

### 2.4.6 Analyses of safety data

The summary of safety results will be presented by treatment group by using actual treatment.

***General common rules***

All safety analyses will be performed on the safety population and separately for the crossover treated population as defined in Section 2.3.2, unless otherwise specified. Summaries for the crossover treated population will only be performed on the crossover treatment period displaying one treatment arm (docetaxel - tusa rav CO) unless specified otherwise.They will only be provided where specifically indicated. Depending on the number of participants who cross over, some analyses on the crossover treated population may not be performed.

The following common rules will be applied:

- Safety data in participants who do not belong to the safety population (eg, exposed but not randomized) will be listed separately.
- The analysis of the safety variables will be essentially descriptive and no systematic testing is planned. Relative risks versus docetaxel and their 95% confidence intervals may be provided, if relevant.
- Descriptive statistics for vital signs and ECG parameters will be displayed by week as the duration of a cycle is different between the two treatment arms.Summaries will be presented by theoretical week from C1D1 when the assessments correspond to the same

theoretical weeks for the two schedules (eg, Week 7 will correspond to Cycle 4 Day 1 for tusamitamab ravtansine and Cycle 3 Day 1 for Docetaxel).

- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review and defined by the Sponsor for clinical laboratory tests, vital signs, and ECG (PCSA list currently in effect at Sanofi at the time of the database lock).
- The PCSA criteria will determine which participants had a PCSA at baseline (crossover baseline respectively) and which participants had at least 1 PCSA during the treatment period (crossover treatment period respectively), taking into account all evaluations performed during the treatment period (crossover treatment period respectively), including non-scheduled or repeated evaluations. The number of all such participants will be the numerator for the on-treatment PCSA percentage.
- The treatment-emergent (crossover treatment-emergent respectively) PCSA denominator by group for a given parameter will be based on the number of participants assessed for that given parameter in the treatment-emergent adverse event period (crossover treatment-emergent respectively) by treatment group on the safety population (cross-over treated population respectively).

#### *2.4.6.1 Analyses of adverse events*

***Generalities***

The primary focus of adverse event reporting will be on treatment-emergent adverse events. Pre-screening, screening, crossover treatment-emergent events and post-treatment adverse events will be described separately.

If an adverse event date of onset (occurrence, worsening, or becoming serious) is incomplete, an algorithm will be used to classify the adverse event as pre-screening, screening, treatment-emergent, crossover treatment-emergent or post-treatment. The algorithm for imputing date of onset will be conservative and will classify an adverse event as treatment emergent and/or crossover treatment emergent unless there is definitive information to determine it is pre-screening, screening or post-treatment. Details on classification of adverse events with missing or partial onset dates are provided in Section 2.5.4.

The grade will be taken into account in the summary. For participants with multiple occurrences of the same event, the maximum (worst) grade by period of observation is used. Summaries will be provided for all grades combined and for Grade ≥3 (including Grade 5). Missing grades, if any, will be included in the “all grades” category (see Section 2.5.4).

The AE tables will be sorted as indicated in Table 16.

**Table 16 - Sorting of AE tables**

| AE presentation | Sorting rules |
|---|---|
| SOC, HLGT, HLT and PT | By the internationally agreed SOC (Appendix C) order and by alphabetic order of HLGTs, HLTs and PTs. |
| SOC and PT | By the internationally agreed SOC order and decreasing frequency of PTs[a, b] |
| [SMQ/CMQ] and PT | By decreasing frequency of [SMQs/CMQs] and PTs[a] |

*a* Sorting will be based on the experimental study intervention group.
*b* The table of all TEAEs presented by SOC and PT will define the presentation order for all other tables (eg, treatment-emergent SAE) presented by SOC and PT, unless otherwise specified.

A summary of tables provided for analyses of adverse events is provided in Appendix B.

The all TEAE summary by Primary SOC and PT (and other safety summaries (eg, SAEs, deaths), if deemed needed after TEAE evaluation) will be performed by trial impact (disruption) due to Covid-19.

***Analysis of all treatment-emergent adverse events***

The following treatment-emergent adverse event and crossover TEAE summaries will be generated for the safety population and the crossover treated population, respectively.

- Overview of treatment-emergent adverse events, summarizing the number (%) of participants with any
  - Treatment-emergent AEs
  - Grade ≥3 TEAEs
  - Grade 5 TEAEs (any TEAE with a fatal outcome during the treatment period)
  - Serious TEAEs
  - Treatment-emergent AEs leading to treatment discontinuation
  - TEAEs related to IMP
  - TEAEs related to IMP Grade ≥3
  - Serious TEAEs related to IMP
  - Treatment-emergent AESI.

The following frequency distributions of AEs (incidence tables) will also be provided for the safety population and the crossover treated population, respectively,for all grades combined and Grade ≥3. Overall, unless otherwise specified, the analyses of TEAEs and crossover TEAEs will be similar and are described below using the generic collective term of “TEAE”.

- All treatment-emergent adverse events by primary SOC, HLGT, HLT, and PT, showing number (%) of participants with at least 1 TEAE.
- All treatment-emergent adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 TEAE.

- All treatment-emergent adverse events presented by PT, sorted by decreasing incidence of PT, showing the number (%) of participants with at least 1 TEAE.
- Most frequent (≥5% of participants in any group) treatment emergent adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 TEAE (only the main study phase).
- All related treatment-emergent adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 related TEAE.
- Most frequent (≥5% of participants in any group) TEAEs related to IMP by primary SOC and PT, showing the number (%) of participants with at least 1 related TEAE (only the main study phase).

***Analysis of all treatment emergent serious adverse event(s)***

- All treatment-emergent serious adverse events by primary SOC, HLGT, HLT, and PT, showing the number (%) of participants with at least 1 serious TEAE.
- All treatment-emergent serious adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 serious TEAE.
- Most frequent (≥2% of participants in any group) serious TEAEs by primary SOC and PT (only the main study phase)
- All serious TEAEs related to treatment, by primary SOC and PT, showing the number (%) of participants with at least 1 serious TEAE.
- Most frequent (≥2% of participants in any group) serious TEAEs related to IMP by primary SOC and PT, showing the number (%) of participants with at least 1 serious TEAE (only the main study phase).

***Analysis of all treatment-emergent adverse event(s) leading to treatment discontinuation***

- All treatment-emergent adverse events leading to treatment discontinuation, by primary SOC, HLGT, HLT, and PT, showing the number (%) of participants with at least 1 TEAE leading to treatment discontinuation.
- All TEAEs leading to treatment discontinuation, by primary SOC and PT, showing the number (%) of participants with at least 1 TEAE leading to treatment discontinuation.

***Analysis of all treatment-emergent adverse event(s) leading to dose modification or dose interruption***

- All treatment-emergent adverse events leading to dose modification, by primary SOC and PT, showing the number (%) of participants with at least 1 TEAE leading to dose modification.
- All treatment-emergent adverse events leading to dose reduction, by primary SOC and PT, showing the number (%) of participants with at least 1 TEAE leading to dose reduction.

- All treatment-emergent adverse events leading to cycle delay, by primary SOC and PT, showing the number (%) of participants, showing the number (%) of participants with at least 1 TEAE leading to cycle delay.
- All treatment-emergent adverse events leading to dose interruption, by primary SOC and PT, showing the number (%) of participants with at least 1 TEAE leading to dose interruption.

***Exposure-adjusted analyses of TEAEs (only the main study phase)***

- Incidence rate per participant-years (number of participants with an event in question divided by total participant-years) will be provided by SOC and PT for most frequent (≥5% of participants in any group) TEAEs, most frequent (≥5% of participants in any group) Grade ≥3 TEAEs, serious TEAEs, TEAEs leading to dose modification, TEAE leading to treatment discontinuation and fatal AEs in context other than progressive disease (Section 2.4.6.2) . For participants with an event, the number of participant-years is truncated at the time of the first event; for participants without an event, the number of participant-years corresponds to the length of the treatment-emergent period.

***Subgroup analyses (only the main study phase)***

- To assess the homogeneity of the intervention effect across subgroups, the number (%) of participants experiencing at least one TEAE will be provided by age category: (<65 years, ≥65 years), ECOG PS (0 vs 1) as per IRT and by race.

***Analysis of adverse events of special interest***

Number (%) of participants experiencing at least one treatment-emergent AESI mentioned in Section 2.1.4.1 will be provided overall.

A listing of participants will also be provided including the treatment period, the preferred term, the cycle of occurrence (start day), the day of seriousness, the outcome and end day, the worst grade, the action taken, the relatedness to the treatment and the corrective treatment.

***Adverse events as per SMQ/CMQ***

For each AEs grouping terms defined in Section 2.1.4.1, the summary, by PT and by CMQ or SMQ, showing number (%) of participants will be displayed.

Analysis of adverse events of ocular toxicities

An overview of corneal TEAE will be provided with the following AE categories: any corneal TEAE, Grade ≥3 corneal TEAE, treatment-emergent corneal SAE, corneal TEAE related to IMP, Grade ≥3 corneal TEAE related to IMP, corneal TEAE leading to treatment discontinuation, and corneal TEAE leading to dose modification (cycle delay or dose reduction). In addition, a summary table of corneal events will be displayed by grade. A summary of treatment-emergent corneal events will be provided.

- Cycle of first onset of corneal event regardless of the grade

- Cycle of first onset of corneal event with the worst grade
- Number (%) of participants by worst grade
- Relationship to the study intervention: in case of multiple events with different relationships, if any event is related, then the relationship will be considered as related
- Action taken with the study intervention: in case of multiple events with different actions, the most severe action taken will be tabulated and selected according to the following order of severity: drug withdrawn, dose reduced and delayed, dose reduced, dose delayed, drug interrupted, dose not changed
- Outcome: in case of multiple events with different outcomes, the most severe outcome will be tabulated and selected according to the following order of severity: fatal, not recovered or resolved, recovering or resolving, recovered or resolved with sequelae, recovered or resolved, unknown

In addition, analyses on occurrence and recurrence of corneal events will be provided.
An occurrence of corneal event is defined as one or a group of concomitant corneal events.
A recurrence is defined as any new occurrence of corneal event starting after a previous resolved occurrence.

- The number of occurrences by participant
- The time to first onset of corneal event will be described using Kaplan-Meier curves. Time to first onset is defined as the time from the date of first IMP administration to the date of the first occurrence of the event. In the absence of an event before the analysis cut-off date, it will be censored at the end date of the treatment-emergent period, analysis cut-off date or the date of death, whichever occurs first.
- The time to recovery will be summarized using descriptive statistics in participants who had had at least one recovered or resolved occurrence of corneal event (with or without sequelae), considering the longest duration among all occurrences by participant.
- The time to recurrence will be summarized using descriptive statistics in participants who had had at least one recurrence, considering the shortest time among all recurrences by participant.

Beside the AE categorized as ocular/visual adverse events, all ocular symptoms (coded term) recoded in eCRF will be reported as descriptive analysis. Same analysis will be done separately on ocular symptoms associated to treatment-emergent corneal events (CMQ).

Neutropenia and neutropenic complications (only the main study please)

Neutropenia (from laboratory abnormalities) will be displayed along with neutropenic complications (febrile neutropenia and neutropenic infections, see Section 2.1.4.1).

If relevant, duration of Grade 3/4 neutropenia episode, cumulative duration of Grade 3/4 neutropenia by participant and time to first Grade 3/4 neutropenia will be analyzed using laboratory data.

The start date of a Grade 3/4 laboratory neutropenia episode is defined as the date of first Grade 3/4 assessment for that episode. The end date of a Grade 3/4 neutropenia episode is defined as the first date of neutropenia assessment afterwards of Grade 1/2 or with no abnormality for that episode assuming there will be at least 3 days between the first Grade ≤2 neutropenia and the next Grade ≥3 assessment (if any). If the start date of a new episode is within 3 days of the previous episode, then the two episodes will be considered as one episode. The worst grade of an episode is the worst grade of all assessments included in that episode.

Duration of a Grade 3/4 neutropenia episode (in days) is defined as end date of an episode – start date of an episode +1. If a participant does not have an end date before the cutoff date, then the duration of the episode will be censored at the last neutrophil assessment of Grade 3/4, or the cutoff date, whichever comes first.

Time to first Grade 3/4 neutropenia (in days) is defined as: date of the first on-treatment Grade 3/4 neutropenia assessment – date of first study IMP +1. If a participant does not have Grade 3/4 neutropenia, time to first Grade 3/4 neutropenia will be censored at the last on-treatment assessment of neutropenia of Grade 1/2 or with no abnormality until the cutoff date.
If a participant does not have any on-treatment assessment of neutropenia, then the participant will be censored at the day of the first IMP.

***Infection location***

For the purpose of the analysis of infection location, a specific output summarizing the SOC "Infection and infestations" by HLT and PT will be provided.

***Pre-screening, screening and post-treatment adverse events***

The following frequency distributions of AEs (incidence tables) will be provided for the safety population, for all grades combined and Grade ≥3:

- Overview of adverse events, summarizing the number (%) of participants with any prescreening/screening/post-treatment AEs, any screening/post-treatment serious AEs, any prescreening/screening/post-treatment AEs leading to death, any screening AEs leading to treatment discontinuation.
- All prescreening adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 prescreening adverse event.
- All screening adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 screening adverse event.
- All post-treatment adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 posttreatment adverse event.
- All post-treatment serious adverse events by primary SOC and PT, showing the number (%) of participants with at least 1 posttreatment serious adverse event.

#### *2.4.6.2 Deaths*

An overview of Grade 5 AEs (excluding pre-treatment ones and displaying treatment period and crossover treatment period) will be provided by treatment groups (docetaxel, tusa rav, docetaxel – tusa rav CO) summarizing respectively the number (%) of participants with any:

- Grade 5 AE (TEAE, crossover TEAE and post-treatment)
- Fatal TEAE (regardless date of death/period)
  - Grade 5 TEAE (TEAE with a fatal outcome during the treatment period)/Grade 5 CO TEAE (CO TEAE with a fatal outcome during crossover treatment period),
  - Any grade TEAE with a fatal outcome during the crossover or the post-treatment period/Any Grade CO TEAE with a fatal outcome during the post-treatment period.
- Post-treatment Grade 5 AE (excluding a TEAE that worsened to Grade 5 during the crossover or the post-treatment period).

The following summaries of deaths will be generated for the safety population and the crossover treated population:

- Number (%) of participants who died by study period (treatment period or crossover treatment period, or post-treatment period) and reasons for death (disease progression, AE, other) by treatment received.
- All TEAEs leading to death (regardless date of death/period) by primary SOC and PT
- All TEAEs related to IMP and leading to death (regardless date of death/period) by primary SOC and PT.
- Number (%) of participants with TEAE(s) leading to death regardless date of death/period and regardless of relationship and related to study treatment by primary SOC, HLGT, HLT, and PT.
- Summary of AEs leading to death (fatal outcome), by primary SOC and PT:
  - In context of disease progression (death within 30 days from last study treatment administration and the cause of death is disease progression).
  - In context other than disease progression (death within 30 days from last study treatment administration and for whom cause of death is not disease progression, or the death occurred more than 30 days from last study treatment administration and the cause of death is AE).

In addition, a listing of deaths in non-randomized participants or randomized but not treated participants will be provided on the prescreened participants.

#### *2.4.6.3 Analyses of laboratory variables*

Laboratory (hematology and biochemistry) values will be analyzed after conversion into standard international units.

International units will be used in all listings and tables. Hematological and biochemistry results will be graded according to NCI-CTCAE version 5.0, when applicable. For participants with multiple occurrences of the same laboratory variable during the treatment period (the crossover treatment period, respectively), the maximum grade (worst) per participant will be used for the safety population (the crossover treated population, respectively).

For hematological parameters and for some selected biochemistry parameters, Sanofi generic ranges (lower limit of normal, upper limit of normal) are defined (see list of parameters in Appendix H) and will be used for grading. For other biochemistry parameters (eg, for hepatic enzymes ALT, AST, alkaline phosphatase, total bilirubin), local laboratory normal ranges will be used.

The number and proportion of participants with abnormal laboratory tests at baseline (crossover baseline, respectively) will be presented by grade and all grades together (except for liver function). A similar table showing abnormalities during the treatment period (the crossover treatment period, respectively) will be provided. For liver function, baseline status (crossover baseline status, respectively) will be provided according to multiples of Upper limit of normal (ULN).The denominator used for percentage calculation is the number of participants with at least one evaluation of the laboratory test during the considered observation period.

The frequency of participants in each grade of laboratory tests during treatment (the crossover period, respectively) will be summarized. For participants with multiple occurrences of the same laboratory variable during the treatment (the crossover period, respectively), the maximum grade per participant will be used. For a given laboratory variable, a participant contributes 1 to the numerator for each cycle in which an episode occurred. An episode occurs during a cycle if the date of sampling is after the first day of the cycle, but prior or equal to the first day of the next cycle.

When appropriate, the summary table will present the frequency of participants with any grade of abnormal laboratory tests and with Grade 3-4 abnormal laboratory tests.

In addition, for hematology and biochemistry toxicities except for liver function, shift tables showing the number of participants in each grade at baseline (crossover baseline, respectively) by worst grade during the on-treatment period (the crossover period, respectively) will be provided.

For laboratory tests for which NCI-CTCAE V5.0 scale is not applicable, potentially clinically significant abnormality (PCSA) values will be derived. Analyses according to PCSA will be performed based on the worst value during the on-treatment period (the crossover treatment period, respectively), using all measurements (either scheduled, unscheduled or repeated). The incidence of PCSA will be summarized by treatment group at baseline (crossover baseline, respectively) as well as at any time during the treatment period (the crossover treatment period, respectively) irrespective of the baseline (crossover baseline, respectively) level in the safety population (the crossover treated population, respectively).

For laboratory tests for which NCI-CTCAE V5.0 scale and PCSA are not applicable, frequency of evaluable participants outside normal ranges will be provided.

***Cardiac function parameters***

The summary statistics of specific troponin as reported troponin I and troponin T (relative change from baseline) will be provided for each study assessment (baseline, on-treatment value and EOT value) by treatment arm. In addition, a graph describing relative change from baseline from baseline and associated ± SD will also be provided.

The number and proportion of participants with abnormal value at baseline will be presented by treatment group.

In addition, a table showing at least the number of participants with relative change from baseline >20% during the on-treatment period will be provided according to baseline status.

Number of participants experiencing SOC coronary cardiac TEAEs and myocardial TEAEs will be displayed by treatment group and by cardiac medical history according to the troponin I (respectively troponin T) values.

Similar analyses on the crossover treatment period will be performed in the crossover treated population.

***Drug-induced liver injury***

A listing of possible Hy's law cases identified (eg, participants with any elevated AST or ALT of >3 ULN and elevated Total bilirubin >2 ULN, occurring on the same day or 1-2 days apart will be considered), will be provided, displaying ALT, AST, Total bilirubin and ALP values.

#### *2.4.6.4 Analyses of vital sign variables*

Vital signs parameters are described in Section 2.1.4.4.

For blood pressure, and weight parameters, PCSA will be derived.

The incidence of PCSAs will be summarized based on the worst value during the on-treatment period (the crossover treatment period, respectively) by treatment group in the safety population (the crossover treated population, respectively) irrespective of the baseline (the crossover baseline, respectively) and/or according to the following baseline (the crossover baseline, respectively) categories:

- Normal/missing.
- Abnormal according to PCSA criterion or criteria.

The summary statistics of blood pressure (change from baseline) will be provided for baseline and worst on-treatment value. In addition, for blood pressure a graph describing mean changes from baseline and associated ± SEM will also be done throughout the on-treatment period.

#### *2.4.6.5 Analyses of electrocardiogram variables*

ECG parameters are described in Section 2.1.4.5.

For HR, PR interval, QRS interval, QT interval and QTcF interval (Section 2.5.1), PCSA will be derived.

The summary statistics of all QRS, QT and QTcF variables (values and change from baseline) will be provided for each study assessment (baseline, each postbaseline time point, and worst on-treatment value). Mean changes from baseline with the corresponding standard error will be plotted over time.

QTcF prolongation will be graded according to NCI-CTCAE version 5.0. The frequency of participants in each grade of QTcF prolongation during the on-treatment period (the crossover treatment period, respectively) will be summarized. For participants with multiple occurrences of QTcF prolongation during the treatment (the crossover treatment period, respectively), the maximum grade per participant will be used.

The incidence of PCSAs will be summarized based on the worst value during the on-treatment period (the crossover treatment period, respectively) irrespective of the baseline (the crossover baseline, respectively) level and/or according to the following baseline (the crossover baseline, respectively) status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

The number (%) of participants with QTcF abnormality worsening during the on-treatment period (the crossover treatment period, respectively) (worst value per participant) according to baseline (the crossover baseline, respectively) value will be displayed.

#### *2.4.6.6 Analyses of other safety endpoints*

##### *2.4.6.6.1 Analyses of ocular examinations*

Schirmer test

Participants reported Schirmer's test will be classified into three classes at baseline: normal ≥10 mm, moderate >5 and <10 mm, severe ≤5 mm (6).The worst classification between the laterality will be considered. A frequency table of the baseline status with and without anaesthetics will be provided by treatment group for the participants with corneal events (CMQ) during the treatment-emergent period and for the participants without event.

Visual acuity test

Analyses are done on the best corrected visual acuity assessed using Snellen Chart method.

Frequency tables on best corrected visual acuity (BCVA) measured during on-treatment period will be performed separately on all participants and on participants who experienced at least one treatment-emergent corneal event:

- on the worst change from baseline: No worsening versus baseline (no line decrease), Worsening versus baseline (1 to 3 lines decrease versus baseline, or >3 lines decrease versus baseline)
- on the worst absolute value: No change versus baseline, Worsening versus baseline (BCVA equals to 20/40 or better, BCVA worse than 20/40 up to 20/200, BCVA 20/200 or worse).
- on the CTCAE worst vision decrease: BCVA equals to 20/40 or better or 1 to 3 lines decrease versus baseline, BCVA worse than 20/40 up to 20/200 or >3 lines decrease versus baseline, BCVA 20/200 or worse.
- For participants who had worsening versus baseline (at least one line decreased) on BCVA, the worst outcome on the last BCVA value will be displayed: Recovered to baseline, Not recovered to baseline, Lost to follow-up or dead with ongoing corneal events

A shift table of the category of the last BCVA value (Normal (20/20 or better), worse than 20/20 up to 20/40, worse than 20/40 up to 20/200 or 20/200 or worse) versus the category of the worst BCVA value (20/40 or better, worse than 20/40 up to 20/200, 20/200 or worse) will be performed for participants who experienced at least one treatment-emergent corneal event.

For the summary table on participants who experienced at least one treatment-emergent corneal event, the worst value and the worst change from baseline are measured during any of the treatment-emergent corneal events experienced by the participant (between start date and end date of the corneal events). For the summary table on all participants, all on-treatment BCVA values are considered.

The worst classification is considered for the worst value, for the change from baseline and the last value even if not measured on the same eye. If the worst absolute value is the same for both eyes, then the eye with the worst change from baseline is considered, and if identical, then the eye with the worst last value is considered.

Participants whose baseline visual acuity had been reported in naked eye instead of BCVA (ie, whenever baseline value was worse than the values reported during the during the treatment period/ treatment-emergent corneal events) will be excluded from this analysis.

Slit lamp examination

Descriptive statistics of slit lamp examination will be provided separately at baseline and at the time of first abnormal slit lamp after occurrence of a treatment-emergent corneal event and at the time of the worst BCVA value during a treatment-emergent corneal event for participants experiencing treatment-emergent corneal events (CMQ). The outcome (normal, abnormal), and for abnormal findings, the type of lesions and the distribution will be described by laterality (unilateral, bilateral, all).

#### *2.4.6.6.2 ECOG PS*

A shift table of baseline ECOG PS versus last and worst ECOG PS on treatment will be provided.

### 2.4.7 Analyses of pharmacokinetic and pharmacodynamic variables

#### *2.4.7.1 Analysis of pharmacokinetic*

Individual observed predose concentrations ($C_{trough}$) and concentrations observed at end of infusion (Ceoi) of tusamitamab ravtansine will be tabulated and summarized with standard descriptive statistics by visit. For the descriptive statistics, $C_{trough}$ following any dose modification (delay or reduction as defined in Section 2.4.4.2) will be excluded. Ctrough and Ceoi will be kept for the descriptive statistics if sampling occurs respectively during Cycle 1 at predose (within 48 hours before Start of Infusion), end-of-infusion (EOI; ±10 min), and at D3 (ie, whenever between 24 hours and 72 hours after administration), at predose (within 24 hours before infusion) from Cycle 2 to Cycle 7, then every 6 cycles (ie, Cycles 13, 19, 25, etc). At Cycle 3, in addition to the predose sample, an EOI+1 hour (±10 minutes) sample is collected. $C_{eoi}$ will be excluded following dose reduction. This analysis will be under the responsibility of B&P.

Exposure parameters, as estimated by population PK analysis (see Section 2.1.5.1) will be tabulated with same descriptive statistics.

All concentration values below the lower limit of quantitation (LLOQ) will be treated as zero in all summary statistics. Geometric mean will not be computed in case at least one concentration is below LLOQ. For the first 12 participants treated with tusamitamab ravtansine in China, individual concentration and PK parameters computed by non-compartmental analysis will be tabulated with same standard descriptive statistics under the responsibility of PKDM.

#### *2.4.7.2 Analyses of pharmacokinetic/pharmacodynamic*

A separate PK/PD SAP will present the details of these analyses.

### 2.4.8 Analyses of immunogenicity variables

For tusamitamab ravtansine randomized and treated participants, demographic characteristics (such as gender, race, ethnicity, age in years, weight [kg] and BSA [$m^2$]), medical or surgical history (summarized by SOC and PT) and also overall summary of prior anti-cancer treatments will be described in the immunogenicity population.

In addition, the followings will be provided:

- An individual data listing with ATA samples status (positive, negative or inconclusive), the titer if applicable, the neutralizing ability status and the titer if available, duration of exposure, cycle, time point and day/time of sampling along with tusamitamab ravtansine $C_{trough}$ value will be provided.
- A summary table with the number (%) of participants with and without pre-existing ATAs, participants with unclassified ATA, participants with inconclusive ATA and participants with treatment-emergent ATA (treatment-induced ATA or boosted-treatment ATA) will be reported, along with descriptive statistics of titer. Incidence will also be provided.

In addition, for participants with treatment-emergent ATA, time to onset, duration of ATA response, and the characterization of the immune response (transient, persistent, indeterminate) will be provided.

Bivariate plots of ATA onset versus ATA duration and transient versus persistent ATA frequency will be illustrated.

If applicable (by the presence of Nabs), further summaries will be provided for participants exhibiting Nabs.

#### *2.4.8.1 Impact of immunogenicity on PK, safety and efficacy endpoints*

Descriptive and graphical methods will be displayed to analyze the impact on immunogenicity on PK, safety and efficacy endpoints. The impact of immunogenicity on these endpoints will be modeled only if a sufficient number of participants are ATA positive.

##### *2.4.8.1.1 Impact of immunogenicity on PK*

The impact of immunogenicity on PK will be assessed graphically by plotting individual $C_{trough}$ profiles of participants with treatment-emergent ATA along with mean (± SD) $C_{trough}$ profile of participants without treatment-emergent ATA over cycles. In addition, for participants with treatment-emergent ATA, each ATA positive, negative and inconclusive sample status will be flagged. The neutralizing ability will be also flagged differently. A boxplot for participants without treatment-emergent ATA will also be displayed at each timepoint along with individual concentrations for participants with treatment-emergent ATA . A flag for ATA positive, negative and inconclusive sample status and its neutralizing ability will also be used.

Descriptive statistics of $C_{trough}$ for participants without treatment-emergent ATA will be provided by cycle. Participants with treatment-emergent ATA will be detailed in a listing as described in Section 2.4.8.1.3.

$C_{trough}$ following any dose modification (delay or reduction as defined in Section 2.4.4.2) will be excluded from this analysis.

##### *2.4.8.1.2 Impact of immunogenicity on safety endpoints*

Safety parameters that will be included in the analyses are the following:

- Hypersensitivity events (using SMQ “Hypersensitivity” [Narrow] and defined as events occurring on the day or the day after infusion),
- Anaphylactic reactions (using SMQ “Anaphylaxis” [Narrow])
- Infusion associated reactions (using CMQ “Infusion related reaction_single PT”).

In addition, in case any impact of PK profile is shown, then the following safety endpoints will also be analyzed:

- All TEAEs, if any relevant event
- Treatment-emergent AESI (depending on the number of participants experiencing a TE AESI)

Impact of immunogenicity on the frequency and intensity of each defined safety parameters will be displayed by a histogram for participants without treatment-emergent ATA, participants with treatment-emergent ATA , and subcategories of participants with treatment-emergent ATA may also be used, depending on the number of participants in each category: participants with and without Neutralizing ATA (Nab and non Nabs, respectively). Each bar will be stratified by frequency of participants experiencing each toxicity grade or no toxicity. Increasing intensity of the gray color represents increasing toxicity grade. In addition, for participants with treatment-emergent ATA, the frequency of participants with events occurring after the ATA sample became positive will also be displayed.

For participants with treatment-emergent ATA and selected adverse events, $C_{trough}$ will be plotted over time, along with symbols for non-neutralizing ATA samples and neutralizing-ATA and indicating the grade of the selected adverse event over time.

#### *2.4.8.1.3 Impact of immunogenicity on efficacy endpoints*

Following efficacy endpoints will be of interest:

- Overall Survival
- Progression-free Survival
- Best Overall Response, based on Overall Responses determined by the IRC

Kaplan-Meier curves of PFS and OS will be plotted for participants with treatment-emergent ATA and participants without treatment-emergent ATA.

A bar chart will be displayed showing separately progression free survival and overall survival in participants with treatment-emergent ATA, circles will represent at the time of the corresponding weeks ATA negative, positive or inconclusive status sample and marked as NAb for neutralizing antibody. Censored participants will be represented by triangles and Best Overall response will be provided in bracket.

Impact of immunogenicity on Best Overall Response will be displayed by a histogram for participants without treatment-emergent ATA, participants with treatment-emergent ATA, and subcategories of participants with treatment-emergent ATA may also be used, depending on the number of participants in each category: participants with and without Neutralizing ATA (Nab and non Nabs, respectively). Each bar will be stratified by frequency of participants with PD, Non-CR/non-PD and SD, PR or CR as BOR.

A listing of participants with treatment-emergent ATA will be provided over time with the ATA status of each sample (positive, negative, inconclusive), as well as the titer and the ability of neutralizing, $C_{trough}$, the best overall response and grades of selected TEAEs.

### 2.4.9 Analyses of pharmacodynamic/genomics variables

Several exploratory objectives related to pharmacodynamic/genomic variables will be considered in this study.

#### *2.4.9.1 Characterization of CEACAM5 expression using disease and participant characteristics*

The CEACAM5 expression will be categorized into three groups: negative expressors (0% of tumoral cells expressing CEACAM5 with intensity ≥2+), moderate expressors (1 to 49% of tumoral cells expressing CEACAM5 with intensity ≥2+) and high expressors (≥50% of tumoral cells expressing CEACAM5 with intensity ≥2+).

In the pre-screened population, the CEACAM5 expression subgroups will be presented using descriptive statistics overall and according to some disease and participant characteristics (eg, genomic alterations, PD-L1 status, circulating CEA).

The circulating CEA collected during the prescreening will be correlated with the CEACAM5 IHC expression (as % of tumoral cells expressing CEACAM5 with intensity ≥2+ or H-score). A correlation coefficient will be calculated to evaluate the association strength. The time from tumor biopsies and pre-screening CEA assessment date will be also described in the correlation between pre-screening circulating CEA levels and IHC CEACAM5 expression.

In case of re-assessment of CEACAM5 IHC, a listing with the number of days between the two assessments, location and CEACAM5 expression for each sample of the participant will be displayed as well as a scatter plot describing the difference of expression between the samples of the participant.

#### *2.4.9.2 To assess the relationship between baseline biomarkers with efficacy endpoints*

In addition to the subgroups defined in Table 9 for association with efficacy endpoints, some additional exploratory biomarker subgroups will be tested such as CEACAM5 expression IHC ( 50%-79% vs ≥80%), for PD-L1 (<1% vs. 1-49% vs. ≥50%, <1% vs. ≥1%,) and circulating CEA (<5 vs. ≥5; < 50 vs. ≥50; <80 vs. ≥80, <100 vs. ≥100 ng/mL). Additional genomics alterations (yes/altered vs. no/wild-type) will be also tested such as: ALK, BRAF, KRAS, MET ROS1 and TP53.

The analysis will be performed on time-to-event (OS and PFS) and binary endpoints (ORR), similarly to what has been described in Section 2.4.5.1.1.1 and Section 2.4.5.2.1.1, respectively.

#### *2.4.9.3 To explore modulation of circulating CEA as a potential pharmacodynamic biomarker of response to tusamitamab ravtansine treatment and association with efficacy endpoints*

The circulating CEA, as assessed by local laboratory, at different timepoints and its relative change from baseline will be presented using descriptive statistics for each treatment arm. Graphical visualization such as lineplots can be considered.

Circulating CEA collected during the crossover treatment phase will also be described.

### 2.4.10 Analyses of quality of life

A separate quality of life SAP (QOL SAP) will present the details of these additional analyses.

Further additional analyses (analysis on deterioration on disease related symptom, physical functioning or role functioning domains linked to the reported AEs [eg, impact of ocular safety on physical functioning and role functioning], analysis of GHS and pain score versus change in analgesic use, etc) will be detailed in this specific QOL SAP.

### 2.4.11 Further therapy after discontinuation of investigational medicinal product administration during the study

A summary table will be provided for further systemic anti-cancer therapies based on WHO-DD coding.

**Time to first further systemic anti-cancer therapy**

The following estimates will be provided for the time to first further systemic anti-cancer therapy after disease progression:

- Time to first further systemic anti-cancer therapy data will be analyzed using the Kaplan-Meier method by treatment group in the ITT population:
  - Kaplan-Meier estimates of the 25th, 50th and 75th percentiles and their associated 95% CIs will be provided. The 95% CIs will be constructed using a log-log transformation of the survival function and the method of Brookmeyer and Crowley.
  - Number of participants at risk as well as the probabilities of being free of further systemic anti-cancer therapy for example at least 3, 6, 9,12, 15, 18, 21 and 24 months with 95% CIs will be estimated for each treatment group using the Kaplan-Meier method.
- The number of censored participants, the reasons for their censoring (ie, alive without further systemic anti-cancer therapy, death, and lost to follow-up) will be summarized by treatment group.
- Kaplan-Meier curves will be plotted.

Further anticancer surgeries and radiotherapies (eg, therapies administered during the follow-up period) will be summarized separately.

## 2.5 DATA HANDLING CONVENTIONS

### 2.5.1 General conventions

The following formulas will be used for computation of parameters.

**Time unit**

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

**Duration**

Unless otherwise specified, difference between two dates (Date A ≤ Date B) will be calculated as follows:

Duration (days) = Date B – Date A + 1

**Demographic formula**

Body surface area value will be derived using the variation of DuBois and DuBois formula:

$$BSA = 0.007184 \times Weight(kg)^{0.425} \times Height(cm)^{0.725}$$

**QTcF formula**

QT will be derived using the Fridericia formula: $QTcF=QT/RR^{0.33}$

where RR interval (in seconds) = 60 / HR and HR = Heart rate in beats per minute (from ECG form)

**Corrected calcium formula**

Corrected calcium (mmol/L) will be derived using the following formula:

measured total calcium (mmol/L) + 0.8 x 0.25 x (4.0 - [serum albumin (g/L) x 0.1]),

where 4.0 represents the average albumin level

**Date of last contact**

The last contact date is derived for all participants based on the latest date among the following:

- Date of visits
- Assessment dates (eg, laboratory, vital signs, ECOG performance status, ECG, tumor assessment, PK assessment, EOT completion, death form etc).
- Medication and procedures dates including study medication, concomitant medications, surgical and medical procedures, further anti-cancer therapies administered after treatment discontinuation.

- Adverse event start, seriousness and end dates.
- Study treatment start and end dates.
- Randomization date.
- Date of Last Available Information" collected on the "Subject status" page.

The last contact date is defined as the latest complete date from the above list or the cut-off date (for participants with subject status eCRF form or date of death available after the cut-off date), whichever comes first.

### 2.5.2 Data handling conventions for primary efficacy variables

The following formulas will be used for computation of PFS endpoint.

**Date of tumor assessment**

It is acknowledged that an assessment may include several methods of evaluation performed over a period of several days within a window of time around an expected assessment date. For each tumor assessment, a date will be derived according to the overall response of that assessment.

When the overall response is different from PD, the date of tumor assessment is defined as the date of the last evaluation included in the series of evaluations performed within that time point (ie, target, non-target and new lesion evaluations).

When the overall response is PD, the date of tumor assessment is the date when progression was first demonstrated as specified below:

- For progression based on new lesion(s) or a non-target lesion, the date of progression is the date of the first observation that a new lesion was detected or the date of the first non-target lesion with PD.
- If multiple assessments based on the sum of diameters (SOD) measurements are performed at different times, the date of progression is the date of the last observation or radiological or clinical assessment of target lesions that show a pre-defined (ie, at least, 20%) increase in the SOD from nadir.
- In the instance where the progression identified on one modality is confirmed by a subsequent exam, the date of progression is the date the lesion was first suspected and not the date of the confirmatory exam.

Note: if there are multiple progression events within the same time point (eg, new lesion seen along with non-target progression), the earliest date of the two (or more) instances, according to the rules listed above, will be the date of progression.

**Evaluable tumor assessment**

An evaluable tumor assessment is defined as a non-missing tumor assessment with an overall response different from non-evaluable (NE).

**Date of documented progression**

The date of documented progression is defined as the first date of tumor assessment at which the overall response was recorded as progressive disease.

**Date of death**

The date of death is defined as the date of death recorded in the eCRF.

**Date of next scheduled assessment**

The date of next scheduled tumor assessment is the date of the last evaluable tumor assessment plus the protocol specified time interval for assessments equal to 56 days or the randomization date plus 56 days in case of no evaluable tumor assessment. This date may be used if back-dating is considered when the event occurred beyond the acceptable time window for the next tumor assessment (eg, 1 or 2 non-evaluable tumor assessment).

**Date of last evaluable tumor assessment**

The date of the last evaluable tumor assessment is defined as the last date of tumor assessment at which the overall response was recorded as CR, PR, SD or Non-CR/Non-PD before a censoring reason occurred.

**Non-evaluable cases and missing visits:**

In general, unless there is clear evidence of progression (independent of a missing assessment or component of an assessment), the overall assessment will be NE when at least one assessment is not evaluated or missing.

Documented progression or death occurring after exactly one tumor assessment with an overall response equal to NE will be considered as an event in the primary analysis and sensitivity analyses. Documented progression or death occurring after two or more tumor assessments with an overall response equal to NE will be considered as censored in the primary analysis. Modifications to the censoring and event rules that incorporate the number of consecutive visits with an NE response or missing visits will be considered in the sensitivity analyses as described in Appendix F.

### 2.5.3 Data handling conventions for secondary efficacy variables

**Missing QOL assessment**

A QOL assessment is considered as missing if the time between the last QOL assessment or the the baseline (in case of no previous QOL assessment) is greater than 6*7 days + 21 days (half between two theoretical assessments) = 63 days.

**Date of next scheduled QOL assessment**

The date of next scheduled QOL assessment is the date of the last evaluable QOL assessment plus the protocol specified time interval for assessments equal to 42 days or the baseline plus 42 days in case of no evaluable QOL assessment. This date may be used if back-dating is considered when the event occurred beyond the acceptable time window for the next QOL assessment.

### 2.5.4 Missing data

The analyses and summaries of continuous and categorical variables will be based on observed data only. Percentages will be calculated using as denominator the number of participants with non-missing observation in the considered population. When relevant, the number of participants with missing data is presented.

***Handling of disease characteristics missing/partial dates***

- If the day is missing, it will be imputed by 1.
- If the month is missing, it will be imputed by 1 (only for medical history variables).
- If the year is missing, no imputation will be performed.

Incomplete date of cancer diagnosis:

- If the day of the cancer diagnosis is missing, the date will be imputed to the first day of the month.
- If day and month of the cancer diagnosis are missing, no imputation will be done.

***Handling of medication missing/partial dates***

No imputation of medication (other than anti-cancer therapies) start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered a prior and concomitant medication.

For prior anti-cancer therapies, following rules will be applied:

- Incomplete start date of prior anti-cancer therapy: if the day of the start date of the prior regimen is missing, the date will be imputed to the first day of the month; if the month is missing, the date will be imputed to the first month of the year.
- Incomplete end date of prior anti-cancer therapy: if the day of the end date of last prior regimen is missing, the day will be imputed to the end day of the month or to randomization date, whichever is first, if the month is missing, the date will be imputed to the last month of the year or to randomization date, whichever is first.

***Imputation of incomplete date for post anti-cancer treatment start date***

For post anti-cancer treatments, if the medication start date is missing, it will be imputed as follows:

- If the medication start day and month are missing and the medication start year is the same as treatment end year, the medication start date will be set equal to treatment end date (last IMP administration date) + 1.
- If the medication start day and month are missing and the medication start year is after the treatment end year, the medication start day and month will each be set to 01.
- If the medication start day is missing and medication start year and month is the same as the treatment end year and month, the medication start date will be set to the treatment end date + 1.
- If the medication start day is missing and medication start month is before the treatment end month and the medication start year is the same as treatment end year, the medication start day will be set to 01.
- If the medication start day is missing and the medication start month is after the treatment end month and the medication start year is the same as treatment end year, the medication start day will be set to 01.
- If the medication start day is missing and the medication start month is not missing and the medication start year is after the treatment end year, the medication start day will be set to 01.
- If the medication start day, start month and start year is missing, the medication start date will be set equal to the treatment end date + 1.

***Handling of other missing dates***

Incomplete date of progression for the last prior regimen:

- If the day of the progression for the last prior regimen is missing, the day will be imputed to the end day of the month or to randomization date, whichever is first.
- If day and month of the progression for the last prior regimen are missing, no imputation will be done.

Incomplete date of prior surgery:

- If the day of the last prior surgery is missing, the date will be imputed to the end day of the month or to randomization date, whichever is first.
- If day and month of the last prior surgery are missing, no imputation will be done.

Incomplete date of prior radiotherapy:

- If the end date of the last prior radiotherapy is missing, the date will be imputed to the end day of the month or to randomization date, whichever is first.

- If start date or day and month of the last prior radiotherapy are missing, no imputation will be done.

***.Handling of missing/partial death dates***

- If the day of the death date is missing, it will be imputed as the first day of the month, except if the date of the participant's last contact is in the same month as the death date. In this case, the death date is imputed as the date of last contact + 1 day.
- If the day and month of the death date is missing, the date of death will be imputed to the first of January of the year, except if the date of the participant's last contact is in the same year as the death date. In this case, the death date will be imputed as the date of last contact + 1 day.

If the death date is missing, no imputation will be done and the participant will be censored at the last contact date.

***Handling of adverse events when date and time of first investigational medicinal product administration is missing***

When the date and time of the first IMP administration is missing, all adverse events that occurred on or after the day of randomization should be considered as treatment-emergent adverse events.The exposure duration should be kept as missing.

***Handling of adverse events with missing or partial date of onset***

Missing or partial adverse event onset dates (occurrence or becoming serious) will be imputed so that if the partial adverse event onset date information or visit number does not indicate that the adverse event started prior to treatment or after the treatment-emergent adverse event period, the adverse event will be classified as treatment-emergent. In case of AEs worsening during the study, the emergence will also be based on the cycle of worsening. Imputations may be performed for time to event analyses and participants-year analyses. These data imputations will not be used in listings.

<u>Imputation of the start date of an adverse event</u>

When the day is missing, the AE start date is imputed by:

- If AE cycle ≤ C1, AE start day = first IMP day when month and year of the AE and IMP are the same, otherwise screening day when month and year of the screening and the AE are the same, otherwise day = 01
- If AE cycle > C1, AE start day = day of the cycle when the corresponding cycle and the AE have the same month and year, otherwise day = 01

In any case start date of AE should not be greater than end date of AE. If that is the case, then AE start date = AE end date.

When the day and the month are missing, the AE start date is imputed as following.

- If AE cycle ≤ C1, AE start day and month = first IMP day and month when the year of AE and IMP is the same, otherwise day and month of screening if the year of the screening and the AE is the same, otherwise AE day = 01 and AE month = 01
- If AE cycle > C1, AE start day and month = day and month of the cycle when the year of the AE and the cycle is the same, otherwise day of the AE = 01 and month of the AE = 01

In any case start date of AE should not be greater than end date of AE. If that is the case, then AE start date = AE end date.

To be noted that the start date imputation of the first occurrence cycle of the AE is applicable for all further cycles where this AE occurred.

***Missing grade***

If the grade is missing for one of the treatment emergent occurrences of an AE, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences, no imputation will be done and missing grades will be summarized in the "all grades" category.

***Handling of missing assessment of relationship of adverse events to investigational medicinal product***

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the adverse event considered as such in the frequency tables of related adverse events.

***Handling of potentially clinically significant abnormalities***

A participant with a missing baseline will be grouped in the category "normal/missing at baseline."

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is >0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

### 2.5.5 Windows for time points

***Laboratory data***

An episode occurred during a cycle if the date of sampling is after (>) the first day of the cycle, but prior or equal (≤) to the first day of the next cycle.

### 2.5.6 Unscheduled visits

Unscheduled visit measurements of laboratory data, vital signs, and ECG will be used for computation of baseline and worst values and/or grades.

### 2.5.7 Pooling of centers for statistical analyses

Data from all sites will be pooled together for analyses.

### 2.5.8 Statistical technical issues

Not applicable.

# 3 INTERIM ANALYSIS

A first IA for efficacy and futility assessment on OS will be performed when approximately 221 PFS events or approximately 210 deaths (approximately 58% of OS events) are observed, whichever comes first. At that time, final PFS analysis will also be performed. The cut-off date for final analysis of PFS and 58% interim analysis of OS will be approximately 40 months after FPI.

A second IA for efficacy assessment of OS will be performed when approximately 80% of targeted number of deaths will be available (ie, approximately 290 deaths). The cut-off date for the second IA of OS will be approximately 48 months after FPI.

Futility assessment for the first IA

The stopping boundary for the nonbinding futility assessment will be based on the observed OS HR using a Cox proportional-hazard model stratified by the stratification factors as entered in the IRT. If the observed OS HR is >0.90 the study may be stopped (nonbinding futility assessment). Table 17 summarizes the probabilities of crossing the futility boundary under different assumptions for the true HR.

**Table 17 - Probabilities of crossing the futility boundary on overall survival under different assumptions for the true hazard ratio**

| $H_0$ (HR=1) | HR=0.95 | HR=0.85 | HR=0.80 | $H_1$ (HR=0.72) |
|---|---|---|---|---|
| 77.7% | 65.2% | 33.9% | 19.7% | 5.3% |

Abbreviations: $H_0$=null hypothesis (nonsuperiority of tusamitamab ravtansine treatment to docetaxel treatment); $H_1$: superiority of tusamitamab ravtansine treatment to docetaxel treatment; HR=hazard ratio.

In addition, the predictive power (PP) for the final analysis of OS given the interim OS data, as well as the observed HR of primary PFS analysis may be used as supportive information to guide the decision. Indeed, if primary PFS analysis is not statistically significant and first interim analysis of OS does not cross efficacy boundary, the study may also be terminated if the predictive power for final OS analysis is <15% or the observed HR PFS is >0.9. The methodology to calculate the predictive power for the final OS analysis is detailed in Appendix I.

**Table 18** summarizes the probabilities of having primary PFS analysis not statistically significant and first interim analysis of OS at the 58% IA not crossing the efficacy boundary with OS HR ≤0.9, but crossing the thresholds of PP OS <15% (HR OS >0.868) or PFS HR >0.9 under the null and alternative assumptions for the true hazard ratio.

**Table 18 - Probabilities of having primary PFS analysis not statistically significant and first interim analysis of OS at the 58% IA not crossing the efficacy boundary with OS HR ≤0.9, but crossing (or not) the thresholds of PP OS <15% (HR >0.868) or PFS HR >0.9 under the null and alternative assumptions for the true hazard ratios**

| | True OS HR=1<br>True PFS HR=1 | True OS HR=0.72<br>True PFS HR=0.615 |
|---|---|---|
| Probability of having primary PFS analysis not statistically significant and first interim analysis of OS at the 58% IA not crossing the efficacy boundary with OS HR ≤0.9, and: | | |
| Crossing any threshold PP OS <15% or PFS HR >0.9 | 14.9% | 1.0% |
| Crossing the threshold PFS HR >0.9 and PP OS ≥15% | 8.4% | 0.1% |
| Crossing the threshold PP OS <15% and PFS HR ≤0.9 | 2.1% | 0.9% |
| Crossing both thresholds PP OS <15% and PFS HR >0.9 | 4.4% | 0.0% |
| Not crossing the thresholds PP OS ≥15% and PFS HR ≤0.9 | 6.5% | 7.2% |

Boundaries for overwhelming efficacy

The stopping boundary for efficacy on OS endpoint will be derived based on the O'Brien and Fleming α-spending function and depends on the actual number of events observed at the time of the interim analysis.

If PFS test is not statistically significant, the OS hypothesis will be tested at α=0.015. If exactly 210 deaths (58% information fraction) are observed at the time of the first OS IA, the efficacy boundary for OS will be crossed if the 1-sided p-value is ≤0.0014 (corresponding to a hazard ratio ≤0.662). If exactly 290 deaths (80% information fraction) are observed at the time of the second OS IA, the efficacy boundary for OS will be crossed if the 1-sided p-value is ≤0.0061 (corresponding to a hazard ratio ≤0.745).The 1-sided nominal significance level to declare superiority of tusamitamab ravtansine at the final analysis (approximately 363 deaths) is 0.0130 (the corresponding threshold for the observed HR to reach this significance level is 0.792).

If PFS test is statistically significant, the OS hypothesis will be tested at α=0.0249. If exactly 210 deaths (58% information fraction) are observed at the time of the first OS IA, the efficacy boundary will be crossed if the 1-sided p-value is ≤0.0032 (corresponding to a hazard ratio ≤0.686). If exactly 290 deaths (80% information fraction) are observed at the time of the second OS IA, the efficacy boundary for OS will be crossed if the 1-sided p-value is ≤0.0111 (corresponding to a hazard ratio ≤0.764). The 1-sided nominal significance level to declare superiority of tusamitamab ravtansine at the final analysis (approximately 363 deaths) is 0.0211 (the corresponding threshold for the observed HR to reach this significance level is 0.808).

If exactly 210 deaths are observed (58% information fraction), and exactly 290 deaths (80% information fraction) are observed respectively at the time of the first and second IAs of OS, boundary properties for the planned IAs of overwhelming efficacy for OS are detailed in Table 14.

PFS analysis performed at the same time as the 58% IA of OS will be the final one. However the significance level of PFS analysis may be increased to 0.02485 if OS is significant at 1.5% level at the time of the final OS analysis.

Regardless of the PFS data, the study will continue until any subsequent OS analysis, unless either interim analyses of OS showed overwhelming efficacy or futility. Note that if futility is shown at the time of the 58% OS IA, the Sponsor may still decide to continue the study (nonbinding properties).

At the time of the 58% interim analysis of OS, if PFS test is statistically significant, the potential impact of the crossover phase on the power for the final OS analysis will be assessed, considering the conditional power for final analysis of OS based on the observed OS data until that time point, to assess whether this tusamitamab ravtansine crossover phase can be implemented.

ORR can be tested at the time of interim OS analyses if PFS or OS is significant based on user-defined alpha-spending function as defined in Section 2.4.5.3.

If ORR is significant at the time of the OS IA then the QoL key secondary endpoints will be tested at the same significance level as ORR following an hierarchical order: TTD in disease-related symptoms, TTD in physical functioning, TTD in role functioning. The procedure, including the user-defined alpha -spending function, is similar to the hypothesis testing of ORR.

## 3.1 DATA MONITORING COMMITTEE

The first DMC meeting will be set up to review early safety results (eg, after approximately 50 participants enrolled into the study or after 6 months after first participant randomized whichever comes earlier), and then periodically. Ad hoc DMC meetings may also be held if a significant safety issue or an issue deemed important for discussion arises on this or other tusamitamab ravtansine studies. After each meeting, the DMC will make recommendations to the Sponsor's representatives regarding the continued safety of treating ongoing and future study participants, as well as the course of action regarding the conduct of the study.

During the course of the study an external statistician (independent from the sponsor) will perform the unblinded final analysis of PFS /58% interim analysis of OS, the 80% interim analysis of OS, as well as unblinded safety and efficacy analyses for the purpose of the DMC data review. Access to these data and analyses will be restricted to the DMC members and, only for interim analyses, to limited personnel from the Sponsor involved in the submission if the analysis is positive and DMC recommends unblinding.

For the final analysis of PFS / 58% IA of OS and for the 80% IA of OS, after review of the results by arm, the DMC will share their recommendations on continuation of the study or not, and on unblinding or not, according to the criteria pre-specified by the Sponsor. Criteria pre-specified by the Sponsor for recommendation by the DMC are fully defined in the DMC charter appendices.

# 4 DATABASE LOCK

Calculations assume the observed current enrollment at 3 participants per month during the first 7 months and assume 10 participants per month afterwards based on observed enrollment in the following year. Based on these enrollment assumptions, the CODs for 58% IA OS (and final analysis of PFS), 80% IA OS, and final analysis of OS would be approximately 40 months, 48 months and 58 months, respectively, after FPI.

# 5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS version 9.4 or higher. East 6.4 or higher was used for the computation of the number of events and boundary properties for PFS and OS at interim and final analyses.

# 6 REFERENCES

1. Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. EORTC QLQ-C30 Scoring Manual (3rd edition). Brussels: EORTC. 2001.

2. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013;13(1):152.

3. Watkins C, Huang X, Latimer N, Tang Y, Wright EJ. Adjusting overall survival for treatment switches: commonly used methods and practical application. Pharm Stat. 2013;12(6):348-57.

4. Amit O, Mannino F, Stone AM, Bushnell W, Denne J, Helterbrand J, et al. Blinded independent central review of progression in cancer clinical trials: results from a metaanalysis. Eur J Cancer. 2011;47(12):1772-8.

5. Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Stat Biopharm Res. 2013;5(4):311-20.

6. Akpek EK, Amescua G, Farid M, Garcia-Ferrer FJ, Lin A, Rhee MK, et al. Dry Eye Syndrome Preferred Practice Pattern®. Ophthalmology. 2019;126(1):P286-334.

7. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). 2009;45(2):228-47.

8. Rufibach K, Jordan P, Abt M. Sequentially updating the likelihood of success of a Phase 3 pivotal time-to-event trial based on interim analyses or external information. J Biopharm Stat. 2016;26(2):191-201.

# 7 LIST OF APPENDICES

# Appendix A Summary of statistical analyses

## EFFICACY ANALYSIS

| Endpoint | Analysis population | Primary analysis | Sensitivity analysis | Supportive analysis | Subgroup analysis | Other analyses |
|---|---|---|---|---|---|---|
| **Primary endpoints** | | | | | | |
| OS | ITT | Stratified Log-rank test<br>Kaplan-Meier<br>Cox proportional hazards model | #1 OS analysis using stratification factors derived from eCRF data<br>#2 OS unstratified | IPCW and RPSFTM models | Yes | Multivariate Cox proportional hazards model |
| PFS according to IRC | ITT | Stratified Log-rank test<br>Kaplan-Meier<br>Cox proportional hazards model | #1 PFS considering events occurring after two or more non-evaluable tumor assessments as events and ignoring further systemic anti-cancer therapies<br>#2 PFS considering events occurring after two or more non-evaluable tumor assessment as events and back-dating at the next schedule assessments<br>#3 PFS analysis using stratification factors derived from eCRF data<br>#4PFS unstratified | - PFS according to Investigator assessment (same censoring rules as for the PFS primary analysis, PFS considering the non-radiological progression as event, and same rules as the sensitivity analyses as per IRC if relevant)<br>- Concordance of PFS outcome (PFS ODR)<br>- Concordance of tumor assessment evaluation (EDR and LDR) | Yes | Multivariate Cox proportional hazards model |

| Endpoint | Analysis population | Primary analysis | Sensitivity analysis | Supportive analysis | Subgroup analysis | Other analyses |
|---|---|---|---|---|---|---|
| **Secondary endpoints** | | | | | | |
| ORR according to IRC OR | ITT | Clopper-Pearson 95% CI<br>Stratified Cochran-Mantel Haenszel test | Cochran-Mantel Haenszel test with stratification factors derived from eCRF data | ORR according to Investigator | Yes | - Best Overall Response |
| TTD in disease related symptoms, RF, PF | ITT population | Stratified Log-rank test<br>Kaplan-Meier<br>Cox proportional hazards model | #1 TTD considering events occurring after one or more non-evaluable assessment as event back-dated at the next schedule assessment<br>#2 TTD analysis using stratification factors derived from eCRF data<br>#3 TTD unstratified<br>#4 TTD analysis using different definition for baseline value | | No | |
| DOR according to IRC OR | ITT with BOR=CR or PR | Kaplan-Meier | No | DOR according to Investigator | No | |
| **Tertiary/exploratory endpoints** | | | | | | |
| Pharmacokinetics of tusamitamab ravtansine | Pharmacokinetic-evaluable population | Descriptive statistics | No | No | No | PK parameters according to safety/efficacy endpoints |

| Endpoint | Analysis population | Primary analysis | Sensitivity analysis | Supportive analysis | Subgroup analysis | Other analyses |
|---|---|---|---|---|---|---|
| **Tertiary endpoints asociated with the crossover** | | | | | | |
| ORR | CO treated population | Clopper-Pearson 95% CI | | | | |

BOR: Best overall response, CI: Confidence interval, CR: Complete response, EDR: Early discrepancy rate, ITT: intent-to-treat, LDR: Late discrepancy rate, ODR: Outcome discrepancy rate, OR: Overall response, ORR: Overall response rate, OS: overall survival, PFS: Progression free survival, PR: Partial response, IRC: Independent Review Committee, DOR: Duration of Response

Stratified analyses are performed with the stratification factors as entered in the IRT ie, ECOG PS, Geographical region, ICI administration treatment

## *SAFETY ANALYSES*

| *Endpoint* | *Analysis population* | *Primary analysis* | *Supportive analysis* | *Subgroup analysis* |
|---|---|---|---|---|
| Adverse events | Safety | Descriptive statistics by treatment group | | Age , Race and ECOG PS |
| Deaths | Safety | Descriptive statistics by treatment group | | No |
| Laboratory | Safety | Descriptive statistics by treatment group | | No |
| Vital signs | Safety | Descriptive statistics by treatment group | | No |
| ECG | Safety | Descriptive statistics by treatment group | | No |
| ECOG PS | Safety | Descriptive statistics by treatment group | | No |
| Ocular exam | Safety | Descriptive statistics by treatment group | | No |

## Appendix B Summary of TEAES analyses

| MedDRA coding variables | Sorting | Layout | Events |
|---|---|---|---|
| SOC, HLGT, HLT, and PT | • Primary SOC: internationally agreed order<br>• HLGT, HLT, PT: alphabetical order | Treatment groups: n (%) of participants with any event (all grades combined) and n (%) of participants with event of Grade ≥3 | • All TEAEs<br>• Serious TEAEs<br>• TEAEs leading to treatment discontinuation |
| SOC and PT | • Primary SOC: internationally agreed order<br>• PT: decreasing order of frequency within each SOC defined by the all TEAEs table | Treatment groups: n (%) of participants with any event (all grades combined) and n (%) of participants with event of Grade ≥3 | • All TEAEs<br>• Most frequent (≥5% of participants in any group) TEAEs (only the main study phase).<br>• TEAEs related to IMP<br>• Most frequent (≥5% of participants in any group) TEAEs related to IMP (only the main study phase).<br>• Serious TEAEs<br>• Most frequent (≥2% of participants in any group) serious TEAEs (only the main study phase).<br>• Serious TEAEs related to IMP<br>• Most frequent (≥2% of participants in any group) serious TEAEs related to IMP (only the main study phase).<br>• TEAEs leading to treatment discontinuation<br>• TEAEs leading to dose interruption<br>• TEAEs leading to dose delay<br>• TEAEs leading to dose reduction<br>• Pre-screening AEs<br>• Screening AEs<br>• Serious screening AEs<br>• Post-treatment AEs<br>• Serious post-treatment AEs |

| MedDRA coding variables | Sorting | Layout | Events |
|---|---|---|---|
| CMQ/SMQ and PT | • SMQ<br>• PT: decreasing order of frequency within each SMQ | Treatment groups: n (%) of participants with any event (all grades combined) and n (%) of participants with event of Grade ≥3 | |
| HLT and PT | • HLT, PT: alphabetical order | **Treatment groups:** n (%) of participants with any event (all grades combined) and n (%) of participants with event of Grade ≥3 | • All TEAEs from the SOC "Infections and Infestations" |
| PT | • Decreasing order of frequency | **Treatment groups:** n (%) of participants with any event (all grades combined) and n (%) of participants with event of Grade ≥3 | • All TEAEs |

AE=adverse event; HLGT=high-level group term; HLT=high-level term; MedDRA=Medical Dictionary for Regulatory Activities; n (%)=number and percentage of participants; PT=preferred term; SOC=system organ class; TEAE=treatment-emergent adverse event.

In addition, an overview of TEAEs, pre-screening AEs, screening AEs, post-treatment AEs, specific ocular toxicities analyses and exposure-adjusted analyses of TEAEs will be provided.

Similar analyses will be performed for CO TEAEs (except if specified and except for exposure-adjusted analyses of TEAEs).

## Appendix C Internationally agreed SOC order

The internationally agreed order (Guideline on summary of product characteristics, December 1999, European commission) for SOC:

1. Infections and infestations
2. Neoplasms benign and malignant (including cysts and polyps)
3. Blood and the lymphatic system disorders
4. Immune system disorders
5. Endocrine disorders
6. Metabolism and nutrition disorders
7. Psychiatric disorders
8. Nervous system disorders
9. Eye disorders
10. Ear and labyrinth disorders
11. Cardiac disorders
12. Vascular disorders
13. Respiratory, thoracic and mediastinal disorders
14. Gastrointestinal disorders
15. Hepato-biliary disorders
16. Skin and subcutaneous tissue disorders
17. Musculoskeletal, connective tissue and bone disorders
18. Renal and urinary disorders
19. Pregnancy, puerperium and perinatal conditions
20. Reproductive system and breast disorders
21. Congenital and familial/genetic disorders
22. General disorders and administration site conditions
23. Investigations
24. Injury and poisoning
25. Surgical and medical procedures
26. Social circumstances
27. Product Issues

The other terms are sorted by dictionary code order.

## Appendix D Determination of Best Overall Response

A participant's best tumor response will depend on the achievement of documented tumor measurement assessments. For each participant the BOR will be derived based on the following algorithm:

A) SD will be determined from the overall tumor response at each timepoint. If SD is the overall response and it occurs less than 42 days from the randomization date, then overall response will not be SD.

B) Convention for PD after several NE cycles: in the case where the assessment at study entry is followed by several visits (or cycles) with no assessment (or only NE assessments or SD <42 days from the randomization date) for more than twice the periodicity of tumor assessments (ie, for more than 117 days (2*8weeks +5 days of time window ), and the next assessment is PD, the best overall response will be NE. If the PD occurs within this period of time, the best overall response will be PD.

Using the outcome from a) and b) above, the BOR will be determined based on the following order of Overall Responses: CR, PR, SD, non-CR/non PD (if only non measurable lesions at baseline), PD and NE.

In the BOR determination, tumor assessments performed the day after a participant started further systemic anti-cancer anti-tumor therapy will not be considered.

# Appendix E Response Evaluation Criteria in Solid Tumors (Recist 1.1)

Details provided in bibliographic reference (7).

**Measurability of tumor at baseline**

At baseline, tumor lesions/lymph nodes will be categorized measurable or non-measurable as follows.

Measurable lesions must be accurately measured in at least 1 dimension (longest diameter in the plane of the measurement to be recorded) with a minimum size of:

- 10 mm by CT scan (CT scan slice thickness no greater than 5 mm).
- 10 mm caliper measurement by clinical exam (lesions which cannot be accurately measured with calipers should be recorded as non-measurable).
- 20 mm by chest X-ray.

Malignant lymph nodes: To be considered pathologically enlarged and measurable, a lymph node must be ≥15 mm in short axis when assessed by CT scan (CT scan slice thickness recommended to be no greater than 5 mm). At baseline and in follow-up, only the short axis will be measured and followed.

Non-measurable lesions are all other lesions, including small lesions (longest diameter <10 mm or pathological lymph nodes with ≥10 to <15 mm short axis), as well as non-measurable lesions. Lesions considered non-measurable include: leptomeningeal disease, ascites, pleural or pericardial effusion, inflammatory breast disease, lymphangitic involvement of skin or lung, abdominal masses/abdominal organomegaly identified by physical exam that is not measurable by reproducible imaging techniques.

Special considerations regarding lesion measurability:

- Bone lesions:

1. Bone scan, positron emission tomography scan or plain films are not considered adequate imaging techniques to measure bone lesions. However, these techniques can be used to confirm the presence or disappearance of bone lesions.
2. Lytic bone lesions or mixed lytic-blastic lesions, with identifiable soft tissue components, that can be evaluated by cross sectional imaging techniques such as CT or MRI can be considered as measurable lesions if the soft tissue component meets the definition of measurability described above.
3. Blastic bone lesions are non-measurable.

- Cystic lesions:

4. Lesions that meet the criteria for radiographically defined simple cysts should not be considered as malignant lesions (neither measurable nor non-measurable) since they are, by definition, simple cysts.

5. “Cystic lesions” thought to represent cystic metastases can be considered as measurable lesions, if they meet the definition of measurability described above. However, if non-cystic lesions are present in the same participant, these are preferred for selection as target lesions.

- Lesions with prior local treatment:

6. Tumor lesions situated in a previously irradiated area, or in an area subjected to other loco-regional therapy, are usually not considered measurable unless there has been demonstrated progression in the lesion.

**Method of assessment**

All measurements should be recorded in metric notation, using calipers if clinically assessed. All baseline evaluations should be performed as close as possible to the treatment start and never more than 4 weeks before the beginning of the treatment.

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Imaging based evaluation should always be performed rather than clinical examination unless the lesion(s) being followed cannot be imaged but are assessable by clinical examination.

- **Clinical lesions:** Clinical lesions will only be considered measurable when they are superficial and ≥10 mm diameter as assessed using calipers. For the case of skin lesions, documentation by color photography including a ruler to estimate the size of the lesion is suggested. As noted above, when lesions can be evaluated by both clinical exam and imaging, imaging evaluation should be undertaken since it is more objective and may be reviewed at the end of the study.
- **Chest X-ray:** Chest CT is preferred over chest X-ray, particularly when progression is an important endpoint, since CT is more sensitive than X-ray, particularly in identifying new lesions. However, lesions on chest X-ray may be considered measurable if they are clearly defined and surrounded by aerated lung.
- **CT, MRI:** CT is the best currently available and reproducible method to measure lesions selected for response assessment. Measurability of lesions on CT scan is based on the assumption that CT slice thickness is 5 mm or less. When CT scans have slice thickness greater than 5 mm, the minimum size for a measurable lesion should be twice the slice thickness.
- **Ultrasound:** Ultrasound is not useful in assessment of lesion size and should not be used as a method of measurement. If new lesions are identified by ultrasound in the course of the study, confirmation by CT or MRI is advised.
- **Endoscopy, laparoscopy:** The utilization of these techniques for objective tumor evaluation is not advised.
- **Tumor markers:** Tumor markers alone cannot be used to assess objective tumor response.
- **Cytology, histology:** These techniques can be used to differentiate between PR and CR in rare cases if required by protocol.

**Baseline documentation of "target" and "non-target" lesions**

When more than 1 measurable lesion is present at baseline all lesions up to a maximum of 5 lesions total (and a maximum of 2 lesions per organ) representative of all involved organs should be identified as target lesions and will be recorded and measured at baseline.

Target lesions should be selected on the basis of their size (lesions with the longest diameter), be representative of all involved organs, and should lend themselves to reproducible repeated measurements.

Lymph nodes merit special mention since they are normal anatomical structures which may be visible by imaging even if not involved by tumor. Pathological nodes which are defined as measurable and may be identified as target lesions must meet the criterion of a short axis of ≥15 mm by CT scan. Only the short axis of these nodes will contribute to the baseline sum. All other pathological nodes (those with short axis ≥10 mm but <15 mm) should not be considered non-target lesions. Nodes that have a short axis <10 mm are considered non-pathological and should not be recorded or followed.

A sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all target lesions will be calculated and reported as the baseline sum diameters. The baseline sum diameters will be used as reference to further characterize any objective tumor regression in the measurable dimension of the disease.

All other lesions (or sites of disease) including pathological lymph nodes should be identified as non-target lesions and should also be recorded at baseline. Measurements are not required and these lesions should be followed as "present", "absent", or "unequivocal progression". In addition, it is possible to record multiple non-target lesions involving the same organ as a single item on the case (eg, "multiple enlarged pelvic lymph nodes" or "multiple liver metastases").

**Response criteria**

Response criteria are described in Table 19.

**Table 19 - Response criteria**

| Response criteria | Evaluation of target lesions |
|---|---|
| CR | Disappearance of all target lesions. Any pathological lymph nodes (whether target or non-target) must have reduction in short axis to <10 mm. |
| PR | At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters. |
| PD | At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. (Note: the appearance of 1 or more new lesions is also considered progression). |
| SD | Neither sufficient shrinkage from the baseline study to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters while on study. |

Abbreviations: CR=complete response; PD=progressive disease; PR=partial response; SD=stable disease.

**Special notes on the assessment of target lesions**

Lymph nodes identified as target lesions should always have the actual short axis measurement recorded and should be measured in the same anatomical plane as the baseline examination, even if the nodes regress to below 10 mm on study. This means that when lymph nodes are included as target lesions, the "sum" of lesions may not be zero even if CR criteria are met, since a normal lymph node is defined as having a short axis of <10 mm. For PR, SD and PD, the actual short axis measurement of the nodes is to be included in the sum of target lesions.

Target lesions that become "too small to measure": All lesions (nodal and non-nodal) recorded at baseline should have their actual measurements recorded at each subsequent evaluation, even when very small (eg, 2 mm). However, sometimes lesions or lymph nodes which are recorded as target lesions at baseline become so faint on CT scan that the radiologist may not feel comfortable assigning an exact measure and may report them as being "too small to measure". When this occurs it is important that a value be recorded on the CRF. If it is the opinion of the radiologist that the lesion has likely disappeared, the measurement should be recorded as 0 mm. If the lesion is believed to be present and is faintly seen but too small to measure, a default value of 5 mm should be assigned.

When non-nodal lesions "fragment", the longest diameters of the fragmented portions should be added together to calculate the target lesion sum. Similarly, as lesions coalesce, a plane between them may be maintained that would aid in obtaining maximal diameter measurements of each individual lesion. If the lesions have truly coalesced such that they are no longer separable, the vector of the longest diameter in this instance should be the maximal longest diameter for the "coalesced lesion".

**Evaluation of non-target lesions**

While some non-target lesions may actually be measurable, they need not be measured and instead should be assessed only qualitatively at the time points specified in the protocol.

- CR: Disappearance of all non-target lesions and normalization of tumor marker level. All lymph nodes must be non-pathological in size (<10 mm short axis).
- Non-CR/Non-PD: Persistence of 1 or more non-target lesion(s) and/or maintenance of tumor marker level above the normal limits.
- Progressive Disease: Unequivocal progression of existing non-target lesions. (Note: the appearance of 1 or more new lesions is also considered progression).

The concept of progression of non-target disease requires additional explanation as follows:

When the participant also has measurable disease; in this setting, to achieve "unequivocal progression" on the basis of the non-target disease, there must be an overall level of substantial worsening in non-target disease such that, even in presence of SD or PR in target disease, the overall tumor burden has increased sufficiently to merit discontinuation of therapy. A modest "increase" in the size of 1 or more non-target lesions is usually not sufficient to qualify for unequivocal progression status.

When the participant has only non-measurable disease; to achieve "unequivocal progression" on the basis of the non-target disease, there must be an overall level of substantial worsening such that the overall tumor burden has increased sufficiently to merit discontinuation of therapy. A modest "increase" in the size of 1 or more non-target lesions is usually not sufficient to qualify for unequivocal progression status. Because worsening in non-target disease cannot be easily quantified (by definition: if all lesions are truly non-measurable) a useful test that can be applied when assessing participants for unequivocal progression is to consider if the increase in overall disease burden based on the change in non-measurable disease is comparable in magnitude to the increase that would be required to declare PD for measurable disease: ie, an increase in tumor burden representing an additional 73% increase in "volume" (which is equivalent to a 20% increase diameter in a measurable lesion). Examples include an increase in a pleural effusion from "trace" to "large", an increase in lymphangitic disease from localized to widespread, or may be described in protocols as "sufficient to require a change in therapy". If "unequivocal progression" is seen, the participant should be considered to have had overall PD at that point.

**New lesions**

The appearance of new malignant lesions denotes disease progression. The finding of a new lesion should be unequivocal: ie, not attributable to differences in scanning technique, change in imaging modality or findings thought to represent something other than tumor (for example, some "new" bone lesions may be simply healing or flare of preexisting lesions). This is particularly important when the participant's baseline lesions show PR or CR. For example, necrosis of a liver lesion may be reported on a CT scan report as a "new" cystic lesion, which it is not.

A lesion identified on a follow-up study in an anatomical location that was not scanned at baseline is considered a new lesion and will indicate disease progression. An example of this is the participant who has visceral disease at baseline and while on study has a CT or MRI brain ordered which reveals metastases. The participant's brain metastases are considered to be constitute PD even if he/she did not have brain imaging at baseline.

If a new lesion is equivocal, for example because of its small size, continued therapy and follow-up evaluation will clarify if it represents new disease. If repeat scans confirm that there is a new lesion, then progression should be declared using the date of the initial scan.

While fluorodeoxyglucose-positron emission tomography (FDG-PET) response assessments need additional study, it is sometimes reasonable to incorporate the use of FDG-PET scanning to complement CT scanning in assessment of progression (particularly possible "new" disease). New lesions on the basis of FDG-PET imaging can be identified according to the following algorithm:

A) Negative FDG-PET at baseline, with a positive FDG-PET at follow-up is a sign of PD based on a new lesion.

B) No FDG-PET at baseline and a positive FDG-PET at follow-up:

   If the positive FDG-PET at follow-up corresponds to a new site of disease confirmed by CT, this is PD.

If the positive FDG-PET at follow-up is not confirmed as a new site of disease on CT, additional follow-up CT scans are needed to determine if there is truly progression occurring at that site (if so, the date of PD will be the date of the initial abnormal FDG-PET scan). If the positive FDG-PET at follow-up corresponds to a preexisting site of disease on CT that is not progressing on the basis of the anatomic images, this is not PD.

**Evaluation of best overall response**

Time point response: At each protocol specified time point, a response assessment should occur. Table 20 provides a summary of the overall response status calculation at each time point for participants who have measurable disease at baseline.

**Table 20 - Response in participants with target disease**

| Target lesions | Non-target lesions | New lesions | Overall response |
|---|---|---|---|
| CR | CR | No | CR |
| CR | Non-CR/non-PD | No | PR |
| CR | Not evaluated | No | PR |
| PR | Non-PD or not all evaluated | No | PR |
| SD | Non-PD or not all evaluated | No | SD |
| Not all evaluated | Non-PD | No | Not evaluable |
| PD | Any | Yes or No | PD |
| Any | PD | Yes or No | PD |
| Any | Any | Yes | PD |

Abbreviations: CR=complete response; PD=progressive disease; PR=partial response; SD=stable disease.

When participants have non-measurable (therefore non-target) disease only, Table 21 is to be used.

**Table 21 - Response in participants with non-target disease only**

| Non-target lesions | New lesions | Overall response |
|---|---|---|
| CR | No | CR |
| Non-CR/non-PD | No | Non-CR/non-PD |
| Not all evaluated | No | Not evaluable |
| Unequivocal PD | Yes or No | PD |
| Any | Yes | PD |

Abbreviations: CR=complete response; PD=progressive disease; PR=partial response; SD=stable disease.

Missing assessments and not evaluable designation: When no imaging/measurement is done at all at a particular time point, the participant is not evaluable at that time point.

If only a subset of lesion measurements are made at an assessment, usually the case is also considered NE at that time point, unless a convincing argument can be made that the contribution of the individual missing lesion(s) would not change the assigned time point response. This would be most likely to happen in the case of PD.

**Special notes on response assessment**

When nodal disease is included in the sum of target lesions and the nodes decrease to "normal" size (<10 mm), they may still have a measurement reported on scans. This measurement should be recorded even though the nodes are normal in order not to overstate progression should it be based on increase in size of the nodes. As noted earlier, this means that participants with CR may not have a total sum of "zero" on the CRF.

In trials where confirmation of response is required, repeated "NE" time point assessments may complicate best response determination. The analysis plan for the trial must address how missing data/assessments will be addressed in determination of response and progression. For example, in most trials it is reasonable to consider a participant with time point responses of PR-NE-PR as a confirmed response.

Participants with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be reported as "symptomatic deterioration". Every effort should be made to document objective progression even after discontinuation of treatment. Symptomatic deterioration is not a descriptor of an objective response: it is a reason for stopping study therapy.

The objective response status of such participants is to be determined by evaluation of target and non-target disease. For equivocal findings of progression (eg, very small and uncertain new lesions; cystic changes or necrosis in existing lesions), treatment may continue until the next scheduled assessment. If at the next scheduled assessment, progression is confirmed, the date of progression should be the earlier date when progression was suspected.

**Duration of response**

The duration of overall response is measured from the time measurement criteria are first met for CR/PR (whichever is first recorded) until the first date that recurrent or PD is objectively documented (taking as reference for PD the smallest measurements recorded on study).

The duration of overall CR is measured from the time measurement criteria are first met for CR until the first date that recurrent disease is objectively documented.

**Duration of stable disease**

Stable disease is measured from the start of the treatment until the criteria for progression are met, taking as reference the smallest sum on study (if the baseline sum is the smallest, this is the reference for calculation of PD).

Reproduced from: Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228-47. (7)

# Appendix F Description of censoring and event rules for primary and sensitivity analyses of PFS

**Table 22 - PFS Primary analysis**

| Situation | Date of outcome | Outcome | Category |
|---|---|---|---|
| No baseline tumor assessments[a] | Date of randomization | Censored | No baseline tumor assessments |
| No evaluable post-baseline tumor assessments[a] | Date of randomization | Censored | No evaluable post-baseline tumor assessments |
| Alive and no documented progression | Date of the last evaluable tumor assessment documenting no progression | Censored | Alive without documented progression |
| Documented progression (or Death) at or between scheduled visits | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after exactly one missing or non-evaluable tumor assessment[b] | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after two or more missing or non-evaluable tumor assessments[c] | Date of the last evaluable tumor assessment documenting no progression | Censored | Event occurred after two or more non-evaluable tumor assessments |
| Clinical/Non-radiological progression | Ignored | Ignored | |
| Initiation of further systemic anti-cancer therapy | Date of the last evaluable tumor assessment before start date of further systemic anti-cancer therapy (if any), date of randomization otherwise | Censored | Initiation of further systemic anti-cancer therapy |

*a* Except if the participant dies within 117 days (2*8 weeks plus 5-day window after) after the date of randomization in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

*b* An event occurring at least 66 days (excluded) after the last evaluable tumor assessment, documenting no progression (or 61 days after randomization date in case of no evaluable post-baseline TA) and at most 122 days (included) after the last evaluable tumor assessment, documenting no progression (or 117 days after randomization date in case of no evaluable post-baseline TA). 66 days corresponds to once the time between two disease assessments per protocol (every 1*8 weeks), plus the 5-day window before and after. 122 days corresponds to twice the time between two disease assessments per protocol (every 2*8 weeks), plus the 5-day window before and after

*c* An event occurring at least 122 days (excluded) after the last evaluable tumor assessment, documenting no progression or 117 days after randomization date in case of no evaluable post-baseline TA.

Note: for participants falling in several possible outcomes, the event outcome takes the priority to the censored outcome (eg, death without baseline tumor assessment)

**Table 23 - PFS sensitivity analysis #1 (PFS considering events occurring after two or more non-evaluable tumor assessment as event and ignoring further systemic therapy)**

| Situation | Date of outcome | Outcome | Category |
|---|---|---|---|
| No baseline tumor assessments[a] | Date of randomization | Censored | No baseline tumor assessments |
| No evaluable post-baseline tumor assessments[a] | Date of randomization | Censored | No evaluable post-baseline tumor assessments |
| Alive and no documented progression | Date of the last evaluable tumor assessment documenting no progression | Censored | Alive without documented progression |
| Documented progression (or Death) at or between scheduled visits | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after exactly one missing or non-evaluable tumor assessment[b] | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after two or more missing or non-evaluable tumor assessments[c] | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Clinical/ Non-radiological progression | Ignored | Ignored | |
| Initiation of further systemic anti-cancer therapy | Ignored | Ignored | |

*a* Except if the participant dies in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

*b* An event occurring at least 66 days (excluded) after the last evaluable tumor assessment, documenting no progression (or 61 days after randomization date in case of no evaluable post-baseline TA) and at most 122 days (included) after the last evaluable tumor assessment, documenting no progression (or 117 days after randomization date in case of no evaluable post-baseline TA). 66 days corresponds to once the time between two disease assessments per protocol (every 1*8 weeks), plus the 5-day window before and after. 122 days corresponds to twice the time between two disease assessments per protocol (every 2*8 weeks), plus the 5-day window before and after

*c* An event occurring at least 122 days (excluded) after the last evaluable tumor assessment, documenting no progression or 117 days after randomization date in case of no evaluable post-baseline TA

Note: for participants falling in several possible outcomes, the event outcome takes the priority to the censored outcome (eg, death without baseline tumor assessment)

**Table 24 - PFS sensitivity analysis #2 (PFS considering events occurring after two or more non-evaluable tumor assessment as event and back-dating at the next scheduled assessment)**

| Situation | Date of outcome | Outcome | Category |
|---|---|---|---|
| No baseline tumor assessments[a] | Date of randomization | Censored | No baseline tumor assessments |
| No evaluable post-baseline tumor assessments[a] | Date of randomization | Censored | No evaluable post-baseline tumor assessments |
| Alive and no documented progression | Date of the last evaluable tumor assessment documenting no progression | Censored | Alive without documented progression |
| Documented progression (or Death) at or between scheduled visits | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after exactly one missing or non-evaluable tumor assessment[b] | Date of the next schedule assessment | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after two or more missing or non-evaluable tumor assessments[c] | Date of the next schedule assessment | Event | Documented progression (or Death without documented progression) |
| Clinical/ Non-radiological progression | Ignored | Ignored | |
| Initiation of further systemic anti-cancer therapy | Date of the last evaluable tumor assessment before start date of further systemic anti-cancer therapy (if any), date of randomization otherwise | Censored | Initiation of further systemic anti-cancer therapy |

*a* Except if the participant dies in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

*b* An event occurring at least 66 days (excluded) after the last evaluable tumor assessment, documenting no progression (or 61 days after randomization date in case of no evaluable post-baseline TA) and at most 122 days (included) after the last evaluable tumor assessment, documenting no progression (or 117 days after randomization date in case of no evaluable post-baseline TA). 66 days corresponds to once the time between two disease assessments per protocol (every 1*8 weeks), plus the 5-day window before and after. 122 days corresponds to twice the time between two disease assessments per protocol (every 2*8 weeks), plus the 5-day window before and after

*c* An event occurring at least 122 days (excluded) after the last evaluable tumor assessment, documenting no progression or 117 days after randomization date in case of no evaluable post-baseline TA

Note: for participants falling in several possible outcomes, the event outcome takes the priority to the censored outcome (eg, death without baseline tumor assessment)

**Table 25 - PFS sensitivity analysis for Investigator assessment (taking clinical/non-radiological progression as event)**

| Situation | Date of outcome | Outcome | Category |
|---|---|---|---|
| No baseline tumor assessments[a] | Date of randomization | Censored | No baseline tumor assessments |
| No evaluable post-baseline tumor assessments[a] | Date of randomization | Censored | No evaluable post-baseline tumor assessments |
| Alive and no documented progression | Date of the last evaluable tumor assessment documenting no progression | Censored | Alive without documented progression |
| Documented progression (or Death) at or between scheduled visits | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after exactly one missing or non-evaluable tumor assessment[b] | Date of progression (or Date of death) | Event | Documented progression (or Death without documented progression) |
| Documented progression (or Death) occurring immediately after two or more missing or non-evaluable tumor assessment[c] | Date of the last evaluable tumor assessment documenting no progression | Censored | Event occurred after two or more non-evaluable tumor assessment |
| Clinical/ Non-radiological progression | Date of non-radiological progression | Event | Non-documented progression |
| Initiation of further systemic anti-cancer therapy | Date of the last evaluable tumor assessment before start date of further anti-cancer therapy (if any), date of randomization otherwise | Censored | Initiation of further anti-cancer therapy |

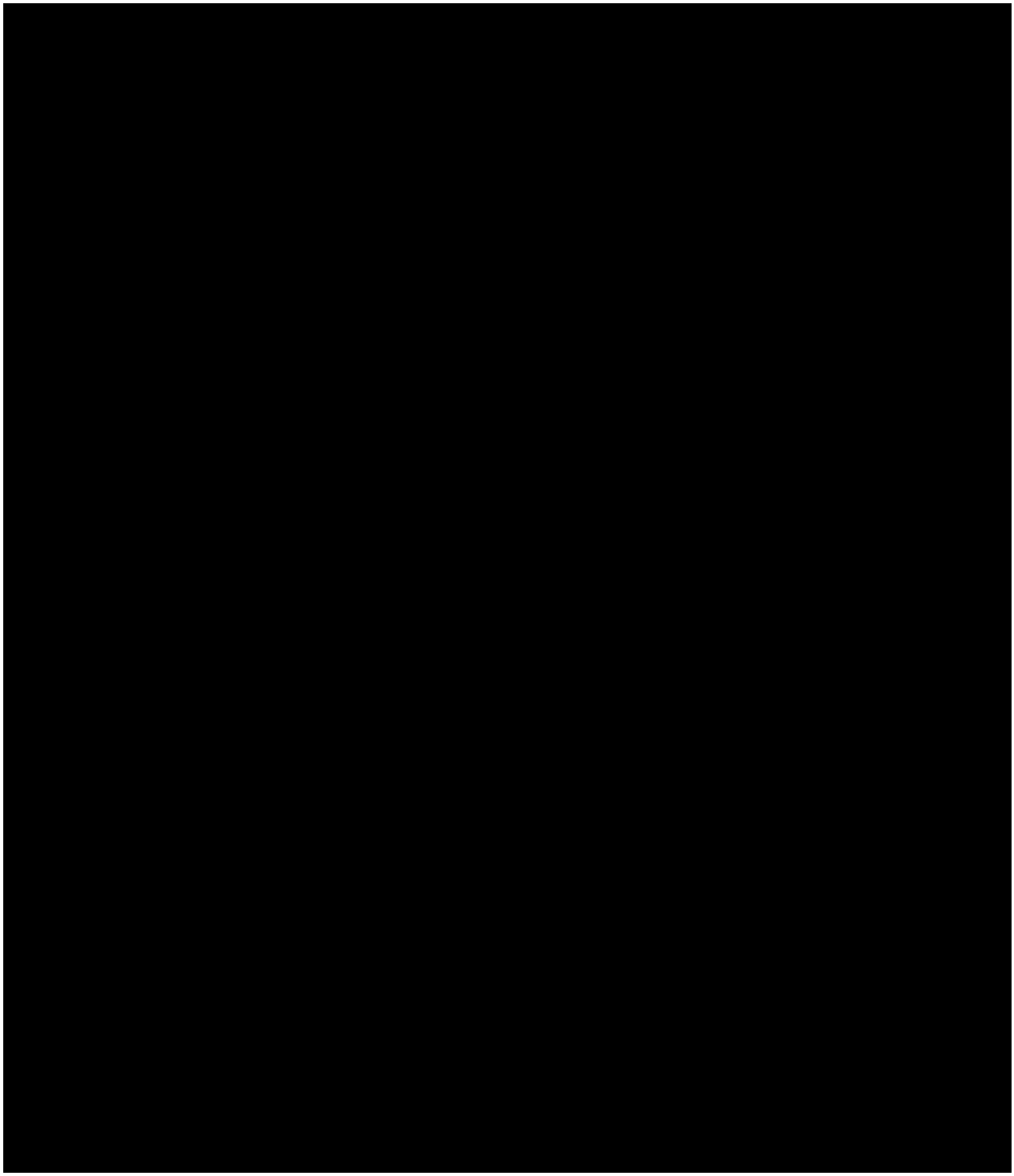
*a* Except if the participant dies within 117 days (2*8 weeks plus 5-day window after) after the date of randomization in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

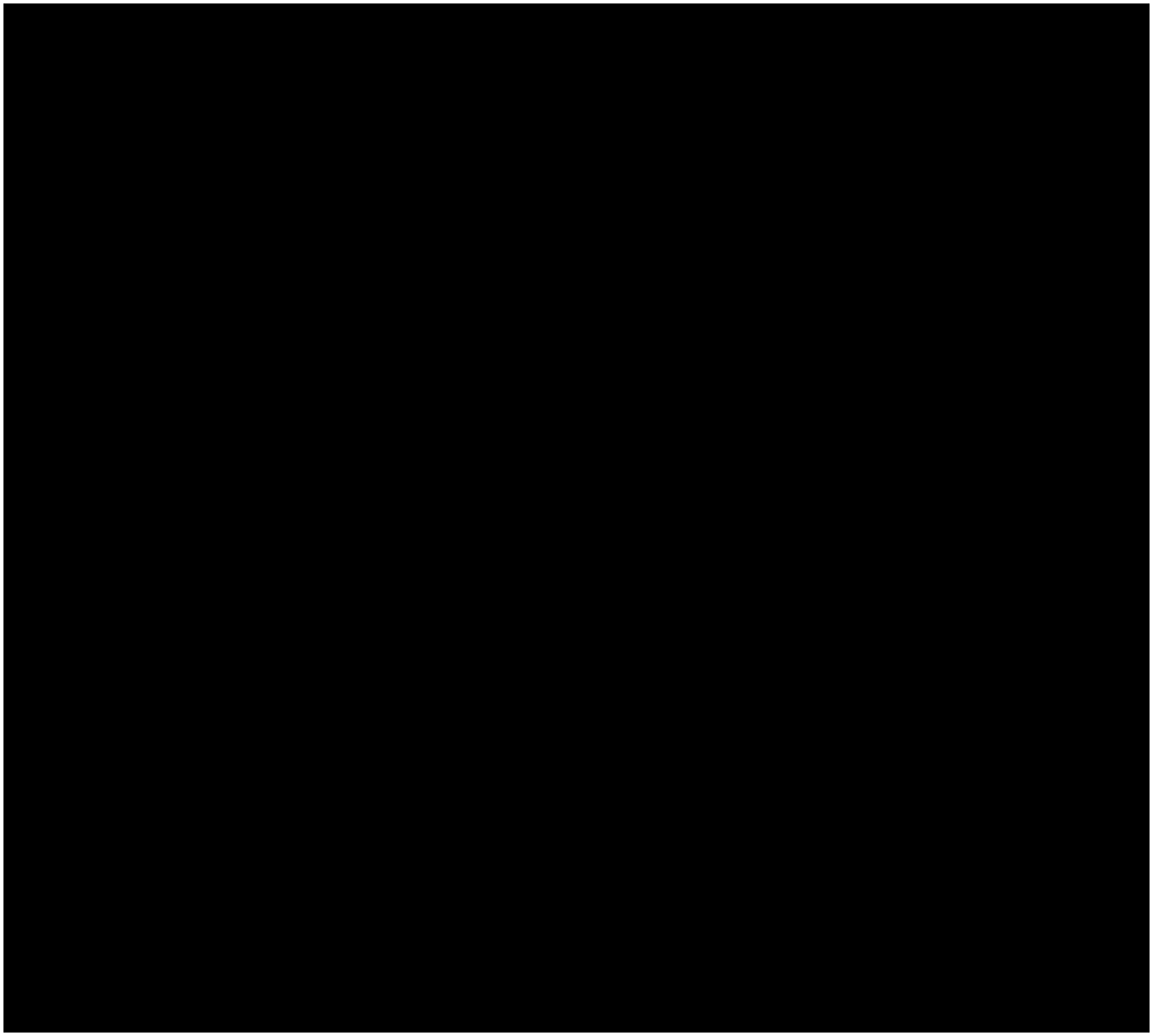
*b* An event occurring at least 66 days (excluded) after the last evaluable tumor assessment, documenting no progression (or 61 days after randomization date in case of no evaluable post-baseline TA) and at most 122 days (included) after the last evaluable tumor assessment, documenting no progression (or 117 days after randomization date in case of no evaluable post-baseline TA). 66 days corresponds to once the time between two disease assessments per protocol (every 1*8 weeks), plus the 5-day window before and after. 122 days corresponds to twice the time between two disease assessments per protocol (every 2*8 weeks), plus the 5-day window before and after

*c* An event occurring at least 122 days (excluded) after the last evaluable tumor assessment, documenting no progression or 117 days after randomization date in case of no evaluable post-baseline TA

Note: for participants falling in several possible outcomes, the event outcome takes the priority to the censored outcome (eg, death without baseline tumor assessment)

## Appendix G EORTC QLQ-C30 and LC 13 items, scales and scores





## Appendix H Generic ranges for hematological and biochemistry parameters

The current list of generic ranges for hematological parameters (for adults) is provided in the table below:

**Table 26 - Generic ranges for hematological parameters**

| LBTESTCD | LBTEST | GENDER | LBSTRESU | LBGNNRLO- LBGNNRHI |
|---|---|---|---|---|
| HGB | Hemoglobin | F | g/L | 120-160 |
| HGB | Hemoglobin | M | g/L | 135-175 |
| LYM | Lymphocytes | | 10^9/L | 1-2 |
| NEUT | Neutrophils | | 10^9/L | 1.8-3.15 |
| PLAT | Platelets | | 10^9/L | 150-350 |
| WBC | Leukocytes | | 10^9/L | 4.5-11 |
| EOS | Eosinophils | | 10^9/L | 0-0.4 |
| BASO | Basophils | | 10^9/L | 0-0.15 |
| MONO | Monocytes | | 10^9/L | 0.18-0.5 |
| HCT | Hematocrit | M | Fraction of 1 | 0.41-0.53 |
| HCT | Hematocrit | F | Fraction of 1 | 0.36-0.46 |
| RBC | Erythrocytes | F | 10^12/L | 4-5.2 |
| RBC | Erythrocytes | M | 10^12/L | 4.5-5.9 |
| INR | INR | | ratio | 0.8-1.2 |

Based on NEJM (N Engl J Med 2004;351:1548-63.): “Laboratory Reference Values”, Alexander Kratz, M.D., Ph.D., M.P.H., Maryjane Ferraro, Ph.D., M.P.H., Patrick M. Sluss, Ph.D., and Kent B. Lewandrowski, M.D.

The current list of generic ranges for biochemistry parameters (for adults) is provided in the table below:

**Table 27 - Generic ranges for biochemistry parameters**

| LBTEST | LBSTRESU | LBGNNRLO - LBGNNRHI |
|---|---|---|
| Albumin | g/L | 35 - 55 |
| Blood Urea Nitrogen (BUN) | mmol/L | 3.6 – 7.1 |
| Calcium | mmol/L | 2.2 – 2.6 |
| Chloride | mmol/L | 80 - 115 |
| Glucose | mmol/L | 3.9 - 7 |
| Bicarbonate (HCO3) | mmol/L | 22 - 29 |
| Carbon dioxide | mmol/L | 21 - 30 |
| Potassium | mmol/L | 3.5 - 5 |
| Magnesium | mmol/L | 0.8 – 1.2 |
| Sodium | mmol/L | –136 - 145 |
| Phosphate | mmol/L | 1 - 1,4 |
| Protein | g/L | 55 - 80 |
| Urea | mmol/L | 3.6 – 7.1 |

## Appendix I Statistical details for predictive power computation

Details are provided in bibliographic reference (8):

- $\theta$: log of HR , $\theta_{int}, \theta_{fin}$ for IAs and final analysis (FA) respectively
- $d$: the number of events, $d_{int}, d_{fin}$ at IAs and FA respectively
- $p$: randomization ratio
- $\alpha$: one-sided type one error
- $\hat{\theta} \sim N(\theta, \sigma^2)$ with $\sigma^2 = \frac{1}{p(1-p)d}$
- $\theta_{suc}$: log of threshold of HR for statistical significance

$\theta_{suc} = -z_{1-\alpha}\sigma, \ (\sigma = \sigma_{int} \ or \ \sigma_{fin})$ with $\sigma_{int}^2 = \frac{1}{p(1-p)d_{int}}$ $and$ $\sigma_{fin}^2 = \frac{1}{p(1-p)d_{fin}}$

$z_\alpha$: $\alpha$-quantile of the standard normal distribution

Once results of the IA become available, conditional power (CP) calculations can be performed by conditioning on the data observed up to the IA. CP computations remain dependent on an assumed true treatment effect $\theta$ that is hypothesized beyond the IA and is given by:

$$CP\ (\theta_{int}, \theta) = P\left(\hat{\theta}_{fin} \leq \theta_{suc} \middle| \hat{\theta}_{int} = \theta_{int}\right) = \Phi\left(\frac{d_{fin}\theta_{suc} - d_{int}\theta_{int} - \left(d_{fin} - d_{int}\right)\theta}{\sqrt{\frac{1}{p(1-p)}\left(d_{fin} - d_{int}\right)}}\right)$$

with $\Phi$: cumulative distribution function of standard normal distribution

The predictive power for final analysis (PP0) at study start is defined as:

$$PP_0 = \int_{-\infty}^{\infty} P_{fin}\ (\theta) q_0(\theta)\, d\theta = \Phi\left(\frac{\theta_{suc} - \theta_0}{\sqrt{\sigma_{fin}^2 + \sigma_0^2}}\right)$$

Where $q_0(\theta)$ is the prior distribution for $\theta$ defined as $N(\theta_0, \sigma_0^2)$ with $\theta_0 = 0$ and $\sigma_0^2 = \frac{1}{p(1-p)d_0}$ with $d_0$=1

and $P_{fin}\ (\theta) = P\left(\hat{\theta}_{fin} \leq \theta_{suc} | \theta\right) = \Phi\left(\frac{\theta_{suc} - \theta}{\sigma_{fin}}\right)$

The predictive power for final analysis at interim analysis ($PP_1$) is computed as follows:

$$PP_1(\theta_{int}) = \int_{-\infty}^{\infty} CP\ (\theta_{int},\ \theta) q_1(\theta)\, d\theta$$

where $q_1(\theta)$is the posterior distribution for $\theta$ defined as $N(\theta_1,\ \sigma_1^2)$ with
$\theta_1 = \frac{\theta_0 \sigma_{int}^2 + \theta_{int} \sigma_0^2}{\sigma_0^2 + \sigma_{int}^2}$ and $\sigma_1^2 = \frac{\sigma_0^2 \sigma_{int}^2}{\sigma_0^2 + \sigma_{int}^2}$

The study parameters applied are described below:

- $p = 0.5$
- $d_{fin} = 363$
- $d_{int1} = 210$ ($58\%\ IA\ OS$ )
- $d_{int2} = 290$ ($80\%\ IA\ OS$)
- overall $\alpha = 0.015$ (if PFS is not statistically significant)

Signature Page for VV-CLIN-0555606 v5.0
efc15858-16-1-9-sap

| | |
|---|---|
| Approve & eSign | [redacted] |

| | |
|---|---|
| Approve & eSign | [redacted] |