Statistical Analysis Plan for Study M22-137

A Phase 2, Open-Label Study in Subjects with Previously Untreated MET Amplified Locally Advanced/Metastatic Non-Squamous Non-Small Cell Lung Cancer (NSCLC)

Date: 25 April 2023

Version 1.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for telisotuzumab vedotin Study M22-137 "A Phase 2, Open-Label Study in Subjects with Previously Untreated MET Amplified Locally Advanced/Metastatic Non-Squamous Non-Small Cell Lung Cancer (NSCLC)." This SAP is based on protocol amendment version 3.0 (dated 07 March 2023).

Study M22-137 examines the efficacy, safety, and pharmacokinetic (PK) of telisotuzumab vedotin in subjects with previously Untreated MET Amplified Locally Advanced/Metastatic Non-Squamous Non-Small Cell Lung Cancer (NSCLC).

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

This SAP includes changes to analyses described in the protocol. Details are outlined in Section 14.0.

2.0 Study Objectives and Design

2.1 Study Objectives

The study is designed to determine the objective response rate (ORR) of telisotuzumab vedotin in previously untreated subjects with *MET* amplified non-squamous NSCLC.

The clinical hypothesis is that telisotuzumab vedotin in previously untreated subjects with *MET* amplified non-squamous NSCLC will be safe, tolerable, and will demonstrate a clinically meaningful benefit of ORR greater than 40%.

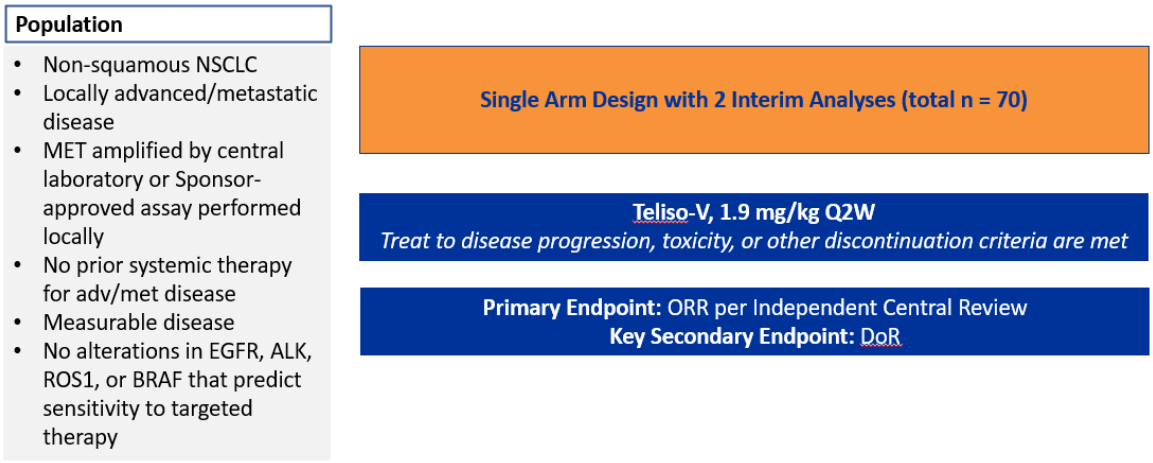
The estimand corresponding to the above primary efficacy objective is the proportion of patients with best confirmed overall response (complete response (CR) or partial response

(PR) confirmed by a repeat assessment ≥ 4 weeks apart defined by Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 as assessed by an independent central review (ICR)) for previously untreated *MET* amplified non-squamous NSCLC subjects receiving telisotuzumab vedotin at 1.9 mg/kg every 2 weeks IV, regardless of premature discontinuation of study drug, prior to or on the date of initiation of new anti-cancer therapy.

2.2 Study Design Overview

The schematic of the study is shown in Figure 1.

Figure 1. Study Schematic



DoR = duration of response; NSCLC = non-small cell lung cancer; ORR = objective response rate; Q2W = Every 2 weeks

2.3 Treatment Assignment and Blinding

This is a non-randomized, single-arm study. Telisotuzumab vedotin will be administered at a dosage of 1.9 mg/kg every 14 days.

2.4 Sample Size Determination

The planned sample size is approximately 70 efficacy evaluable subjects with 2 interim analyses. This sample size will provide approximately 91% power to rule out an ORR of 40% using an exact binomial test assuming the true ORR for telisotuzumab vedotin is 60%. Family-wise type 1 error for testing ORR will be controlled with 1-sided 0.025 significance level. Group sequential design for a single arm study with 2 interim analyses for efficacy at approximately 20 subjects (information fraction of 0.286) and approximately 50 subjects (IF of 0.714), respectively will be carried out using O'Brien-Fleming boundaries (O'Brien 1979) [1] as shown in Table 1 below.

Table 1. Sample Size Calculation

Analysis	Information Fraction (No. of Subjects)	Cumulative Alpha (1-Sided)	Cumulative Power	Expected Analysis Timing (mos)
ORR IA1	28.6% (20)	0.000027	0.4%	20
ORR IA2	71.4% (50)	0.008	67.0%	32
ORR FA	100% (70)	0.025	91.3%	41

3.0 Endpoints

3.1 Primary Endpoint

The primary endpoint is the ORR (assessed by ICR) according to RECIST, version 1.1. ORR will be defined as the proportion of subjects with a confirmed complete response (CR) or confirmed partial response (PR) based on RECIST, version 1.1.

3.2 Secondary Endpoints

3.2.1 Key Secondary Endpoint

1. Duration of Response (DoR) by ICR: Duration of Response will be defined for confirmed responders by ICR as the time from the initial response (CR or PR) to

the first occurrence of radiographic progression per RECIST v1.1, or death from any cause.

3.2.2 Other Secondary Endpoints

1. Disease control rate (DCR) by ICR: Disease control rate will be defined as the percentage of subjects with best overall response of confirmed CR or confirmed PR, or stable disease (SD) by ICR for at least 12 weeks following first dose of study drug, based on RECIST, version 1.1.
2. Progression Free Survival (PFS) per ICR: Progression Free Survival will be defined as the time from the subject's first dose of study drug to the first occurrence of radiographic progression based on RECIST, version 1.1 or death from any cause. Subjects with no PFS event will be censored at the last evaluable radiographic assessment per ICR. Subjects with no event and no evaluable post-baseline assessment will be censored at subject's first dose of study drug.
3. Overall Survival (OS): Overall Survival will be defined as the time from subject's first dose of study drug to the event of death from any cause. Subjects with no documented death will be censored at the last known alive date.
4. Time to deterioration in cough or chest pain or dyspnea as measured respectively by the cough, chest pain, and dyspnea items of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Lung Cancer Module 13 (EORTC QLQ-LC13). *Time to first deterioration will be defined as the time from baseline until a ≥ 10-point change.*
5. Time to deterioration of physical functioning as measured by the physical functioning domain of the EORTC-QLQ-Core 30 (EORTC QLQ-C30). *Time to first deterioration will be defined as the time from baseline until a ≥ 10-point change.*
6. Change from baseline in quality of life as measured by the global health status/quality of life domain of the EORTC QLQ-C30.

3.3 Additional Patient-Reported Efficacy Endpoints

The additional patient-reported outcome endpoints are:

1. Change from baseline in the EuroQoL 5 Dimension 5 Level (EQ-5D-5L) including the Visual Analog Scale (VAS).
2. Patient-Reported Outcome Common Terminology Criteria for Adverse Events (PRO-CTCAE) data.
3. Change from baseline in cough, pain, or dyspnea as measured by the cough, pain, and dyspnea items of the EORTC QLQ-LC13.
4. Change from baseline in remaining scales and sub-scales of the EORTC QLQ-C30.
5. Change from baseline in items of the EORTC QLQ-LC13.
6. Change from baseline in Patient Global Impression of Change (PGIC) and Patient Global Impression of Severity (PGIS) scores.

3.4 Safety Endpoints

Safety and tolerability will be assessed by evaluating AEs, physical examinations, and changes in laboratory data and vital signs, as well as drug discontinuation or dosing modification due to AEs, for the entire study duration.

3.5 Pharmacokinetic Endpoints

Blood samples for assay of telisotuzumab vedotin conjugate in serum and free cytotoxin monomethyl auristatin E (MMAE) toxin levels in plasma, will be collected at specific time-points. Serum samples for assay of telisotuzumab vedotin anti-drug antibodies and neutralizing anti-drug antibodies (ADA/nADA) will also be collected at specified time-points.

4.0 Analysis Populations

The following analysis sets will be used for the analyses.

The Efficacy Evaluable Analysis Set (EEAS) includes all subjects received at least 1 dose of study drug. The EEAS will be used for all efficacy and baseline characteristic analyses. The interim analyses will be conducted based on the number of subjects (i.e., information fraction) specified in Section 2.4.

The Safety Analysis Set (SAS) consists of all subjects who received at least 1 dose of study drug. The SAS will be used for safety analyses.

5.0 Subject Disposition

A summary of subject accountability will be provided where the number and percentage of subjects in each of the following categories will be summarized:

- Subjects who took at least one dose of study drug; (treated)
- Subjects who prematurely discontinued telisotuzumab vedotin (primary reason);
- Subjects who discontinued study (primary reason);
- Subjects in each analysis set, as applicable.

The median follow-up time will be computed using reverse Kaplan-Meier methodology for overall survival. Subjects alive as of the data cut-off date will be considered as events, and deaths on or prior to data cut-off date will be censored in the reverse Kaplan-Meier approach for the estimation of median time on study. Descriptive statistics (min and max) for time on study will be provided as well.

6.0 Study Treatment Duration and Compliance

The following will be applied on Safety Analysis Set. Duration of treatment will be summarized for the treatment group. The planned cycle length for telisotuzumab vedotin is 14 days.

The descriptive statistics (number of non-missing observations, mean, standard deviation, median, and range) will be summarized for each of the following variables:

- The duration of treatment in days for telisotuzumab vedotin is defined as (last dose date – first dose date + 1)
- Number of cycles that subjects received telisotuzumab vedotin

In addition, the number and percentage of subjects will be summarized for the following variables:

- Number of cycles that subjects received telisotuzumab vedotin
- Dose adjustment and dose interruption of telisotuzumab vedotin

The cumulative dose received and relative dose intensity will be summarized for telisotuzumab vedotin as follows:

- Observed dose intensity (ODI, mg/kg/wk): Cumulative dose (mg/kg)/((last dose date – first dose date + 14)/7)
- Calculation for cumulative dose (mg/kg) is further defined as:
  - Actual dose (mg/kg) for each cycle = total dose administered (mg)/ pre-dose weight (kg).
  - Cumulative dose (mg/kg) = sum of actual dose (mg/kg) in study treatment period.
- Planned dose intensity (PDI, mg/kg/wk): Planned total dose over 1 cycle of treatment/planned duration of 1 cycle = 0.95 mg/kg/wk
- Relative dose intensity (RDI): ODI/PDI × 100%

7.0 Subject Characteristics

Demographics, baseline disease characteristics, medical history, and prior and concomitant medications will be summarized for EEAS. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated

based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean and standard deviation, median, minimum and maximum, 1st and 3rd quartiles).

7.1 Demographics and Baseline Characteristics

The following demographic, disease history and baseline disease characteristics will be summarized. Unless otherwise specified, baseline is defined as the last non-missing value prior to the first administration of study treatment.

Demographic variable	**Baseline disease characteristics**
• Age (years) • Age: < 65, ≥ 65 • Height (cm) • Weight (kg) • Sex (Male, Female) • Race (White, Black or African American, Asian, American Indian or Alaskan Native, Native Hawaiian or Pacific Islander, Other) • Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Other) • History of tobacco product use and alcohol use (current, former, never, unknown) • Region of the world (US/Canada, Europe, Pan-Asia (China, Japan, Taiwan, South Korea), Rest-of-World)	• TNM staging at diagnosis • TNM staging at study entry • Histologic type • MET Exon 14 Skipping Mutation • ECOG performance score • KRAS mutation status • RET gene rearrangement • NTRK1/2/3 gene fusion • PD-L1 expression level
Baseline laboratory characteristics	**Baseline disease status**
• Absolute neutrophil count (10^9/L) • Platelets (10^9/L) • Hemoglobin (g/L) • Creatinine clearance (mL/min) (< 30, 30 – < 60, ≥ 60) • Hepatic function per NCI-ODWG classification (normal, mild, moderate, severe)	• Number of metastatic sites, 0 – 1 vs. > 1 • Sum of Longest Diameter for target lesion at baseline <100 mm vs ≥ 100mm • Time (weeks) from initial diagnosis to study entry (continuous)

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category (by MedDRA system organ class and preferred term) will be summarized for EEAS. The system organ class (SOC) will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or preferred term).

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug plus 30 days. The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

For reporting purpose, the following conservative approach will be followed for determination of prior and concomitant medication:

Start date	End date	Classification
Prior to First Dose Date	Prior to First Dose Date	Prior medication
Missing	Prior to First Dose Date	Prior medication
Prior to First Dose Date	First Dose Date to Last dose + 30 days	Both prior medication and concomitant medication
First Dose Date to Last dose + 30 days	First Dose Date to Last dose + 30 days	Concomitant medication
Missing	First Dose Date to Last dose + 30 days	Concomitant medication
First Dose Date to Last dose + 30 day	Missing	Concomitant medication
Missing	Missing	Concomitant medication

7.4 Study Cancer Prior Systemic Therapies and Follow-up Systemic Therapies

Prior systemic therapies and follow-up systemic therapies will be summarized.

7.5 Protocol Deviations

Protocol deviations include, but are not limited to, eligibility criteria violations, receipt of wrong treatment or incorrect dose of study treatment, development of withdrawal criteria without being withdrawn, and use of prohibited concomitant medications. A listing of subjects with protocol deviations will be provided.

8.0 Handling of Potential Intercurrent Events for the Primary and Key Secondary Endpoints

The primary efficacy endpoints of ORR (defined in Section 3.1) will be analyzed based on EEAS and the following methods will be used to address potential intercurrent events:

- Premature discontinuation of study drug: treatment policy strategy- all available disease assessments, including those after the study drug discontinuation, will be used for the analysis of the endpoint.

- Use of new anticancer therapy: while on treatment strategy - All available disease assessment, prior to or on the date of initiation of new anti-cancer therapy, will be used for the analysis of the endpoint.

These methods are further described in Table 3.

The handling of potential intercurrent events for the key secondary endpoint are described in Table 5.

9.0 Efficacy Analyses

9.1 General Considerations

All efficacy analyses will be conducted on the EEAS unless otherwise specified.

All primary and secondary efficacy endpoints of tumor response and progression will be per ICR based on RECIST v1.1. In addition, tumor response and progression endpoints by local investigator will also be summarized as sensitivity analysis.

The first interim analysis of ORR will be based on the first 20 efficacy evaluable subjects approximately, and will occur when all the subjects have the chance been followed for at least 6 months. The second interim analysis of ORR will be based on the first 50 efficacy evaluable subjects approximately and will occur when all the subjects have the chance been followed for at least 6 months. The final analysis of ORR will occur when all the subjects have the chance been followed for at least 6 months.

9.2 Handling of Missing Data

Handling of missing data will be discussed for each endpoint in the following sections.

9.3 Primary Efficacy Endpoint and Analysis

9.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is objective response rate (ORR) by ICR. ORR is defined as the proportion of subjects with a confirmed complete response (CR) or confirmed

partial response (PR) assessed by ICR based on RECIST, version 1.1. Response will need to be confirmed by a repeat assessment no less than 28 days.

9.3.2 Main Analysis of Primary Efficacy Endpoint

Analysis of the primary endpoint of ORR per ICR will be conducted on the EEAS. The ORR per ICR will be summarized, along with 2-sided 95% exact confidence intervals based on the binomial distribution. Subjects without a post-baseline ICR disease assessment will be considered as non-responders in the calculation of ORR. The primary endpoint of ORR per ICR is used to test the clinical hypothesis that the ORR of telisotuzumab vedotin in previously untreated patients with *MET* amplified non-squamous NSCLC exceeds 40%. The exact binomial test will be used for ORR comparison and p value will be provided.

Details of the Type I error rate will be discussed in Section 13.0.

The attributes of the estimand corresponding to the primary efficacy endpoint are summarized in Table 2.

Table 2. Summary of the Estimand Attributes of the Primary Efficacy Endpoints

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Variable**	**Population**	**Intercurrent Events**	**Statistical Summary**
ORR by ICR	Telisotuzumab vedotin at 1.9 mg/kg every 2 weeks IV	Whether patients achieved a confirmed objective response (i.e., complete response or partial response confirmed by a repeat assessment ≥ 4 weeks apart)	- Subjects with previously untreated subjects with MET amplified non-squamous NSCLC.	1. Premature discontinuation of study drug 2. Use of new anti-cancer therapy	1. Proportions of subjects who achieved confirmed objective response

Table 3. Handling Strategies for the Intercurrent Events Related to the Primary Efficacy Endpoint

Handling of Intercurrent Events	**Statistical Summary**
Premature discontinuation of study drug	**Treatment policy strategy – ignore the intercurrent events.** All available disease assessments, including those after the study drug discontinuation, will be used for the analysis of the endpoint.
Use of new anti-cancer therapy	**While on treatment strategy – before the occurrence of the intercurrent events.** All available disease assessment, prior to or on the date of initiation of new anti-cancer therapy, will be used for the analysis of the endpoint

9.3.3 Sensitivity Analyses of the Primary Efficacy Endpoint(s)

ORR per investigator assessment will be conducted as a sensitivity analysis using the same statistical method.

9.4 Secondary Efficacy Endpoints and Analyses

9.4.1 Key Secondary Efficacy Endpoint

The key secondary endpoint of DoR per ICR (as defined in Section 3.2.1) will be analyzed based on the confirmed responders in EEAS.

The attributes of the estimand corresponding to the key secondary efficacy endpoint are summarized in Table 4.

Table 4. Summary of the Estimand Attributes of Key Secondary Efficacy Endpoint

	Attributes of the Estimand				
Estimand Label	**Treatment**	**Variable**	**Population**	**Intercurrent Events**	**Statistical Summary**
DOR by ICR	Telisotuzumab vedotin at 1.9 mg/kg every 2 weeks IV	Time interval from the date of the first response until the first date of progressive disease or death from any cause, whichever occurs first.	Previously untreated subjects with MET amplified non-squamous NSCLC who achieved a confirmed objective response, as determined by the ICR according to RECIST v1.1.	1. Premature discontinuation of study drug: 2. Use of new anti-cancer therapy	1. Distribution for duration of response.

Table 5. Handling Strategies for the Intercurrent Events related to the Key Secondary Efficacy Endpoint

Handling of Intercurrent Events	Statistical Summary
Premature discontinuation of study drug	**Treatment policy strategy – ignore the intercurrent events.** All available disease assessments, including those after the study drug discontinuation, will be used for the analysis of the endpoint.
Use new anti-cancer therapy	**Treatment policy strategy – ignore the intercurrent events.** All available disease assessments, including those after the initiation of the new anticancer therapy, will be used for the analysis of the endpoint.

9.4.2 Main Analysis of Key Secondary Efficacy Endpoint

Duration of Response (DoR) will be estimated using Kaplan-Meier methodology and graphically displayed. The number and percentage of events as well as the earliest contributing event will be summarized. Median DoR along with 2-sided 95% CIs will be estimated.

To assess the impact of missing scheduled tumor assessment on DoR, a sensitivity analysis will be conducted. Any subjects missed 2 or more consecutive scheduled tumor assessments will be censored at the last evaluable radiographic disease assessment before the missing assessments.

Assessment of the impact of new anti-cancer therapy (supplementary analysis): The analysis of DoR will be repeated with such intercurrent event handled using a hypothetical strategy. Subjects will be censored at the last evaluable radiographic disease assessment prior to or at the date of new anti-cancer therapy. DoR per investigator assessment will be conducted similarly as sensitivity analyses.

9.4.3 Other Secondary Efficacy Endpoints and Analyses

Other analyses of secondary efficacy endpoints include:

1. Disease Control Rate (DCR) (per ICR): DCR is defined as the proportion of subjects with best overall response of confirmed complete response, confirmed partial response, or stable disease (SD) lasting for at least 12 weeks from the first dose of study drug. DCR will be summarized along with 2-sided 95% exact confidence intervals based on the binomial distribution. Sensitivity analysis will be conducted using similar methods on DCR by investigator assessment.
2. PFS (per ICR): PFS is defined as the time from the subject's first dose of study drug to either the subject's first objective documented disease progression or death, whichever occurs first. Under the situation that neither event occurs, PFS will be censored at the date of last evaluable tumor assessment. PFS for the treatment cohort will be summarized by Kaplan-Meier estimates and graphically displayed. Median PFS along with 2-sided 95% confidence intervals will be calculated. Sensitivity analysis will be conducted using similar methods on PFS by investigator assessment. As it pertains to the handling of intercurrent event of premature discontinuation of study drug, all available disease assessments, including those after the study drug discontinuation, will be used for the analysis of PFS. As it pertains to the handling of the intercurrent event of the use of new anticancer therapy, the treatment policy will be applied and therefore all available disease assessments, including those after the initiation of new anti-cancer therapy, will be used for the analysis of PFS.
3. Overall Survival (OS): OS is defined as the time from the first dose date to the event of death from any cause. Subjects with no documented death will be censored at the last known alive date.
4. Time to first deterioration in cough, pain or dyspnea as measured by the cough, pain and dyspnea items of the EORTC QLQ-LC13 will be summarized using Kaplan-Meier estimates and associated 95% CI for quartiles (25% - first quartile, 50% - median, 75% - third quartile). Subjects without the specified event will be censored at subject's last EORTC QLQ-LC13 assessment, and all PRO assessments before death will be utilized. Supplementary analysis of this endpoint will consider death before any clinically meaningful deterioration as an event.

5. Time to first deterioration of physical functioning as measured by the physical functioning domain of the EORTC QLQ-C30 will be summarized using Kaplan-Meier estimates and associated 95% CI for quartiles (25% - first quartile, 50% - median, 75% - third quartile). Subjects without the specified event will be censored at subject's last EORTC QLQ-C30 assessment, and all PRO assessments before death will be utilized.
6. Change from baseline in quality of life as measured by the global health status/quality of life domain of the EORTC QLQ-C30 will be summarized by descriptive statistics at each visit.

Additional sensitivity and supplementary analyses of PFS include:

1. Assessment of the impact of missing scheduled tumor assessment on PFS (sensitivity analysis): Any subjects missed 2 or more consecutive scheduled tumor assessment will be censored at the last evaluable radiographic disease assessment before the missing assessments, regardless of progression and survival status after the consecutive missing assessments.
2. Assessment of the impact of new anti-cancer therapy (supplementary analysis): The analysis of PFS will be repeated with such intercurrent event handled using a hypothetical strategy: subjects will be censored at the last adequate radiographic disease assessment before use of new anti-cancer therapy.

9.5 Additional Patient-Reported Outcomes Analyses

For additional PRO efficacy endpoints, descriptive statistics for each visit will be provided.

1. Change from baseline in the EQ-5D-5L including the Visual Analog Scale (VAS).
2. Patient-Reported Outcome Common Terminology Criteria for Adverse Events (PRO-CTCAE) data will be presented using descriptive statistics.

3. Change from baseline in cough, pain, or dyspnea as measured by the cough, pain and dyspnea items of the EORTC QLQ-LC13.
4. Change from baseline in remaining scales and sub-scales of the EORTC QLQ-C30.
5. Change from baseline in remaining items of the EORTC QLQ-LC13.
6. Change from baseline in Global Impression of Change (PGIC) and Patient Global Impression of Severity (PGIS) scores.

9.6 Efficacy Subgroup Analyses

Subgroup analyses for the primary endpoint ORR by ICR will be summarized using descriptive statistics by the following categories:

- MET amplification tests (local + central -/NA vs local + central + vs local - central + vs local NA central +)
- Methodology (FISH vs next generation sequencing (NGS))
- FISH by MET/CEP7 ratio category (< 1.8, 1.8-2.2, > 2.2-5, > 5)
- FISH by Met Gene Copy Number (GCN) (≥ 5, ≥ 10)
- NGS for tissue biopsy by Met Gene Copy Number Variation (CNV) (as defined by analytically validated cut off by FMI, i.e., diploid + 4 copy or Thermo ≥ 5, ≥ 6, ≥ 8)
- NGS for liquid biopsy by Met Gene Copy Number Variation (CNV) (as defined by analytically cut off by Guardant, i.e., ≥ 2.1, ≥ 2.5 or diploid + 4 by F1L)
- Received the allowed cycle of systemic chemo prior to study drug (Yes vs No)

10.0 Safety Analyses

10.1 General Considerations

Safety data will be summarized for the Safety Analysis Set. The statistical test will not be performed for safety analysis.

10.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (PTs) according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

10.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs are defined as any AE with the onset or worsening that is after the first dose of study treatment and no more than 30 days after the last dose of study treatment. Events where the onset date or worsening date is the same as the study drug start date are assumed to be treatment-emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

10.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any treatment-emergent AE
- Any treatment-emergent AE related to study drug according to the investigator
- Any treatment-emergent AE with NCI CTCAE (V5.0) toxicity grades 3, 4 or 5
- Any treatment-emergent AE leading to discontinuation of study drug
- Any treatment-emergent AE related to study drug according to the investigator leading to discontinuation of study drug
- Any treatment-emergent AE leading to interruption of study drug
- Any treatment-emergent AE leading to dose reduction of study drug
- Any serious treatment-emergent AE
- Any serious treatment-emergent AE with NCI CTCAE (V5.0) toxicity grades 3, 4 or 5
- Any serious treatment-emergent AE related to study drug according to the investigator
- Any serious treatment-emergent AE leading to discontinuation of study drug
- Any serious treatment-emergent AE leading to interruption of study drug
- Any serious treatment-emergent AE leading to dose reduction of study drug
- Any treatment-emergent AE leading to death
- Any treatment-emergent AE related to disease progression
- Any treatment-emergent AE not related to disease progression
- Any serious treatment-emergent AE leading to death
- Any treatment-emergent AE leading to death related to study drug according to the investigator
- Any treatment-emergent AEs of special interest including: peripheral neuropathy, pneumonitis/interstitial lung disease (ILD), bone marrow suppression (neutropenia, anemia and thrombocytopenia), and corneal toxicities
- All deaths
- Deaths occurring ≤ 30 days after last dose of study drug
- Deaths occurring > 30 days after last dose of study drug.

10.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

Treatment-emergent adverse events (TEAEs) will be summarized by SOC and PT and by maximum severity and SOC and PT. Specific adverse events will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the treatment group.

10.2.4 Deaths, Serious Adverse Events, and Adverse Events Leading to Study Treatment Discontinuation

Treatment-emergent serious adverse events (SAEs), TEAEs leading to discontinuation of study treatment, and TEAEs leading to death will be summarized by SOC and PT.

Tabular listings will be provided for all deaths, all SAEs, TEAEs leading to death, TEAEs leading to discontinuation of study treatment, and TEAEs leading to study treatment interruptions and dose reduction.

10.2.5 Adverse Events of Special Interest

Adverse events of special interest will be summarized by SOC and PT and will be based on standardized or company MedDRA queries (SMQs or CMQs), or based on adjudication results. Adverse events of special interest are categorized as follows:

- Peripheral neuropathy
- Pneumonitis/ILD
- Corneal toxicity
- Hematological toxicities related to bone marrow suppression

Detailed information about the search criteria is provided in Appendix D.

Time to onset for AESI will be defined as the time from the first dose date to the start date of the first occurrence of an AESI for subjects with events. Descriptive statistics will be provided and summarized (mean, min, max, std, N).

Tabular listings of selected adverse events of special interest will be provided.

10.3 Analysis of Laboratory Data

Unless otherwise specified, the clinical laboratory tests defined in the protocol operations manual (e.g., hematology and clinical chemistry) will be summarized. Post-baseline assessments collected more than 37 days after the last dose of study treatment will not be included in the analyses of laboratory and vital signs values. Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded. Unscheduled assessments will not be included in the summary of change from baseline but will be included in producing shift tables and summary of lab abnormalities.

Each laboratory variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% confidence interval will be presented for the mean change from baseline.

Changes in laboratory parameters will be tabulated using shift tables by NCI CTC criteria CTCAE v4.03. A shift table from baseline to the worse value (based on NCI CTC criteria - CTCAE v4.03) during treatment will be created. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value.

Laboratory abnormalities meeting CTC criteria grade 3 and 4 will be summarized.

Laboratory abnormalities will be evaluated based on Potentially Clinically Important (PCI) criteria (Appendix E). For each laboratory PCI criterion, the number and percentage of subjects who have a laboratory value meeting the criteria will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCI criteria.

Laboratory values will be considered as potentially clinically significant if they are Grade 3 or higher as shown in Table E-1 and Table E-2, with the following exceptions:

- GGT: Grade 4
- Total bilirubin: > 2 × ULN
- Concurrent ALT or AST > 3 × ULN and total bilirubin ≥ 2 × ULN
- Concurrent ALT or AST > 3 × ULN, ALP < 2 × ULN and total bilirubin ≥ 2 × ULN

The maximum ratio relative to the ULN will be used to determine if subjects met the criteria listed above. The ALT, AST, and total bilirubin values need to be concurrent (i.e., collected on same date) and should be collected within 30 days following the last dose date of study drug in order to meet the defined criteria.

A listing of possible Hy's Law cases, defined as those who meet all of the following conditions simultaneously, will be provided: ALT > 3 × ULN or AST > 3 × ULN that is associated with an increase in bilirubin ≥ 2 × ULN and alkaline phosphatase < 2 × ULN.

10.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, pulse oximetry (oxygen saturation) and body temperature will be summarized.

Each vital sign variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from baseline to each applicable post-baseline

visit will be summarized for each vital sign variable, with the number of observations, baseline mean, and visit mean.

Vital sign variables will be evaluated based on potentially clinically important (PCI) criteria (Appendix E). For each vital sign PCI criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria.

10.5 Other Safety Analyses

A listing of subjects with clinically significant ECG abnormalities per local reader evaluation may be provided.

11.0 Pharmacokinetic Analyses

Serum concentrations of telisotuzimab vedotin conjugate, free MMAE toxin in plasma will be listed and summarized by timepoint.

Telisotuzumab vedotin anti-drug antibody (ADA) and neutralizing ADA (if performed) results will be listed and summarized.

Exploratory analyses may be added as deemed appropriate and will not be included in a separate report.

12.0 Interim Analyses

There will be 2 interim analyses.

The first interim analysis (IA1) for efficacy will take place after approximately 20 efficacy evaluable subjects all have the chance been followed for at least 6 months. A second interim analysis (IA2) for efficacy will take place after approximately 50 efficacy evaluable subjects all have the chance been followed for at least 6 months.

12.1 Data Monitoring Committee

There is no independent data monitoring committee for this study.

13.0 Overall Type I Error Control

Family-wise type 1 error (FWER) for testing ORR will be controlled with 1-sided 0.025 significance level. Group sequential design for a single arm study with 2 interim analyses for efficacy at approximately 20 subjects (information fraction of 0.286) and approximately 50 subjects (IF of 0.714), respectively will be carried out using O'Brien-Fleming boundaries.[1]

To control the Type I error for ORR, the stopping boundaries for the 2 interim analyses and final analysis of ORR are to be computed based on the Lan-DeMets approximation to the O'Brien-Fleming boundary[2] using exact binomial testing procedure as shown in Table 6. The statistical boundary (Number of responders) may need to be recalculated, taking into account the actual information fraction at the respective analysis interim time points.

Table 6. Planned Efficacy Stopping Boundaries at IA1, IA2, and Final Analysis (FA) for ORR

Analysis	Information Fraction (No. of Subjects)	Expected Analysis Timing (mos)	Efficacy Stopping Boundaries
			No. of Responders
ORR IA1	28.6% (20)	20	18
ORR IA2	71.4% (50)	32	29
ORR FA	100% (70)	41	37

FA = final analysis; ORR = Objective Response Rate; IA1 = interim analysis 1; IA2= interim analysis 2; mos = months

14.0 Version History

Version	Date	Summary
1.0	25 April 2023	Original version

15.0 References

1. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549-56.
2. DeMets DL, Lan KKG. Interim analysis: the alpha spending function approach. Statistics in Medicine. 1994;13:1341–52.

Appendix A. List of SAP Signatories

Name	Title	Role/Functional Area
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

Appendix B. Protocol Deviations

A listing of subjects who reported at least one of the following protocol deviation categories, or other CSR reportable issues, will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

Appendix C. PRO Assessments

Detailed scoring algorithms will be provided in the statistical programming plan (SPP). The PROs to be analyzed are as follows:

EORTC QLQ-LC13

The EORTC Quality of Life Questionnaire - Lung Cancer Module (EORTC QLQ-LC13) is composed of one multi-item scale to assess dyspnea, and a series of single items assessing coughing, sore mouth, dysphagia, peripheral neuropathy, alopecia, hemoptysis, and pain. The *EORTC QLQ-LC13 scoring manual* will be used to transform raw scores into domain scores.

For the multi-item dyspnea scale, the scale score will be computed if all three items are non-missing. The scale score will be computed using the following formula for computing symptom scale: SS = {(RS – 1)/range} ×100.

If item 5 is missing (question on climbing stairs) then the other two items can be calculated as single-item scores if answered, otherwise the scores will not be calculated. No single-item measure (coughing, hemoptysis, sore mouth, dysphagia, peripheral neuropathy, alopecia, and pain) will be computed if unanswered. If a scale score cannot be computed, the outcome for that score is left blank.

EORTC QLQ-C30

HRQoL, functioning, and symptoms will be assessed with the EORTC QLQ-C30 version 3.[9] The QLQ-C30 is a 30-item subject self-report questionnaire composed of both multi-item and single scales, including five functional scales (physical, role, emotional, social, and cognitive), three symptom scales (fatigue, nausea and vomiting, and pain), a global health status/quality of life scale, and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). The subjects rate items on a four point scale, with 1 as "not at all" and 4 as "very much." The QLQ-C30 was developed and validated for use in a cancer patient population, and its reliability and validity is highly

consistent across different language cultural groups. The items and scales of the QLQ-C30 are as in Table C-1.

Table C-1. EORTC QLQ-C30 Items and Scales

Scale		Number of Items	Item Range	Item Numbers
Global Health Status/Quality of Life				
Global Health Status/Quality of Life	QL2	2	6	29, 30
Functional Scales				
Physical Functioning*	PF2	5	3	1 – 5
Role Function*	RF2	2	3	6 – 7
Emotional Functioning*	EF	4	3	21 – 24
Cognitive Functioning*	CF	2	3	20, 25
Social Functioning*	SF	2	3	26 – 27
Symptom Scales				
Fatigue	FA	3	3	10, 12, 18
Nausea and Vomiting	NV	2	3	14, 15
Pain	PA	2	3	9, 19
Dyspnoea	DY	1	3	8
Insomnia	SL	1	3	11
Appetite Loss	AP	1	3	13
Constipation	CO	1	3	10
Diarrhoea	DI	1	3	17
Financial Difficulties	FI	1	3	28

Scoring algorithms for scales are as follows:

For all Scales, the Raw Score (RS) is the mean of the component items (I):

If items I_1, I_2..., I_n are included in a scale, the procedures are as follows:

$$\text{Raw Score (RS)} = (I_1 + I_2 + ... + I_n) / (\text{number of non-missing items}),$$

Then the Functional Scales:

$$\text{Score} = (1 - (\text{RS} - 1) / \text{range}) \times 100$$

where range is provided in Table 5.

For example, to calculate Emotional Functioning (EF),

$$RS_{EF} = (I_{21} + I_{22} + I_{23} + I_{24}) / 4$$

$$Score_{EF} = (1 - (R_{SEF} - 1) / 3) \times 100$$

If a subject completed more than or equal to 50% of the items in a scale, then the raw score of that subject will contribute to the summary statistics of that scale. If a subject completed less than 50% of the items in a scale, then the raw score of that subject will be dropped from the calculation of the summary statistics of that scale.

EQ-5D-5L

The EuroQol 5 Dimensions 5 Level (EQ-5D-5L) is a preference-based measure of health status that consists of EQ-5D descriptive system and EQ visual analogue scale (VAS). The *EQ-5D-5L* user guide and the published weights will be used to convert the individual items to the utility scores.

The EQ-5D descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each of the dimensions is divided into 5 levels of perceived problems: Level 1 (indicating no problem), Level 2 (indicating slight problems), Level 3 (indicating moderate problems), Level 4 (indicating severe problems), and Level 5 (indicating extreme problems). EQ-5D-5L health status, defined by the EQ-5D-5L descriptive system, will be converted into a single preference – weighted health status or "utility" index score by applying country-specific weights (if available) or U.S. weights (if not available). If country is not available in the look-up table, then weights based on the United States will be applied.

The EQ VAS records the respondent's self-rated health on a visual analogue scale (0 to 100). The VAS score will be measured separately.

Patient Global Impression of Change (PGIC) and Patient Global Impression of Severity (PGIS)

This study uses separate Patient Global Impression of Change (PGIC) and Patient Global Impression of Severity (PGIS) for patients' rating of their severity and overall improvement in their symptoms, cough, pain, dyspnea, physical functioning and quality of life. The present study utilizes six PGIC and PGIS items. The self-reported PGIC assesses the patient's belief about the efficacy of the treatment. The PGIC in this study is a 5-point scale depicting a patient's rating of overall improvement in their symptoms, cough, pain, dyspnea, physical functioning and overall quality of life. Patients rate their change as "much better," "a little better," "no change," "a little worse" and "much worse." The PGIC will be used in the analysis of score quality, as well as to generate within-patient anchor-based estimates for meaningful change for the EORTC QLQ-C30 and EORTC QLQ-LC13 scales of interest (symptoms, cough, pain, dyspnea, physical functioning and overall quality of life).

Similar to PGIC, this study utilizes six PGIS items to assess patients' perspective of their severity of their symptoms, cough, pain, dyspnea, physical functioning and quality of life over the past 7 days. The PGIS for these core concepts ask participants to "Please choose the response below that best describes the severity of your symptoms (or cough, pain, dyspnea, physical functioning and overall QoL)." Patients endorse a five-level rating scale ranging from "none" to "severe." The PGIS questions will be used in the analysis of score quality, as well as to generate within-patient anchor-based estimates of meaningful change for the EORTC QLQ-C30 and EORTC QLQ-LC13 scales and items.

Patient Global Impression of Change (PGIC)

1. Since the start of the treatment received in this study, difficulty with **physical activities** (e.g., walking, moving around, carrying a heavy item, taking care of yourself) due to your **Non-Small Cell Lung Cancer (NSCLC)** is:

   □ Much better
   □ A little better
   □ No change
   □ A little worse
   □ Much worse

2. Since the start of the treatment received in this study, the **cough** you have experienced due to your **NSCLC** is:

   □ Much better
   □ A little better
   □ No change
   □ A little worse
   □ Much worse

3. Since the start of the treatment received in this study, the **pain** you have experienced due to your **NSCLC** is:

   □ Much better
   □ A little better
   □ No change
   □ A little worse
   □ Much worse

4. Since the start of the treatment received in this study, the **shortness of breath** you have experienced due to your **NSCLC** is:

   □ Much better
   □ A little better
   □ No change
   □ A little worse
   □ Much worse

5. Since the start of the treatment received in this study, your **NSCLC symptoms** are:

   □ Much better
   □ A little better
   □ No change
   □ A little worse
   □ Much worse

6. Please rate the change in your **overall quality of life** due to **NSCLC** since you started taking the study medication.

   □ Much better
   □ A little better
   □ No change
   □ A little worse
   □ Much worse

Patient Global Impression of Severity (PGIS)

1. Please choose the response below that best describes how difficult it was to do **physical activities** (e.g., walking, moving around, carrying a heavy item, taking care of yourself) due to your **Non-Small Cell Lung Cancer (NSCLC)** over the past 7 days:

   □ Not at all
   □ Slightly
   □ Moderately
   □ Very much
   □ Extremely

2. Please choose the response below that best describes the severity of your **cough** due to your **NSCLC** over the past 7 days:

   □ None
   □ Mild
   □ Moderate
   □ Severe
   □ Very severe

3. Please choose the response below that best describes the severity of your **pain** due to your **NSCLC** over the past 7 days:

   □ None
   □ Mild
   □ Moderate
   □ Severe
   □ Very severe

4. Please choose the response below that best describes the severity of your **shortness of breath** due to your **NSCLC** over the past 7 days:

   □ None
   □ Mild
   □ Moderate
   □ Severe
   □ Very severe

5. Please choose the response below that best describes the severity of your **NSCLC symptoms** over the past 7 days:

   □ None
   □ Mild
   □ Moderate
   □ Severe
   □ Very severe

6. How much did your **NSCLC** interfere with your **daily life** over the past 7 days?

   □ Not at all
   □ Slightly
   □ Moderately
   □ Quite a bit
   □ Extremely

Appendix D. Definition of Adverse Events of Special Interest

Adverse Events of Special Interest (AESI) will be identified using the following search criteria:

Adverse Events of Special Interest	MedDRA Search Criteria/Codes
Peripheral Neuropathy	SMQ 20000034: Peripheral Neuropathy (narrow)
Anemia	SMQ 20000029: Haematopoietic Erythropenia (broad)
Neutropenia	SMQ 20000030: Haematopoietic Leukopenia (broad)
Thrombocytopenia	SMQ 20000031 Haematopoietic Thrombocytopenia (broad)
Corneal Toxicities	CMQ 10000008: Corneal Epitheliopathy
Pneumonitis	SMQ 20000042 Interstitial Lung Disease (ILD)

Appendix E. Potentially Clinically Important Criteria for Safety Endpoints

The criteria for Potentially Clinically Important (PCI) laboratory findings are described in Table E-1 and Table E-2, and the PCI criteria for vital sign findings are described in Table E-3.

Table E-1. Criteria for Potentially Clinically Important Hematology Values

		Definition of Potentially Clinically Important			
		NCI CTCAE Grade ≥ 3 (CTCAE v4.03)		NCI CTCAE Grade ≥ 4 (CTCAE v4.03)	
Hematology Variables	**Units**	**Low**	**High**	**Low**	**High**
Hemoglobin	g/L	< 80	> ULN+40		
White blood cell count	10^9/L	< 2		< 1	
Neutrophil count	10^9/L	< 1		< 0.5	
Lymphocyte count	10^9/L	< 0.5		< 0.2	
Absolute Platelet count	10^9/L	< 50		< 25	

Note: A post-baseline value must be more extreme than the baseline value to be considered a potentially clinically important finding.

Table E-2. Criteria for Potentially Clinically Important Chemistry Values

		Definition of Potentially Clinically Important			
		NCI CTCAE Grade ≥ 3 (CTCAE v4.03)		NCI CTCAE Grade ≥ 4 (CTCAE v4.03)	
Chemistry Variables	**Units**	**Low**	**High**	**Low**	**High**
Creatinine	mcmol/L		> 3 × ULN		> 6 × ULN
Total bilirubin	mcmol/L		> 3 × ULN		> 10 × ULN
Alanine aminotransferase (ALT/SGPT)	U/L		> 5 × ULN		> 20 × ULN
Aspartate aminotransferase (AST/SGOT)	U/L		> 5 × ULN		> 20 × ULN
Alkaline phosphatase (ALP)	U/L		> 5 × ULN		> 20 × ULN
Gamma glutamyl transferase (GGT)			> 5 × ULN		> 20 × ULN
Sodium	mmol/L	< 130	> 155	< 120	> 160
Potassium	mmol/L	< 3	> 6	< 2.5	> 7
Calcium	mmol/L	< 1.75	> 3.1	< 1.5	> 3.4
Glucose	mmol/L	< 2.2	> 13.9	< 1.7	> 27.8
Cholesterol	mmol/L		> 10.34		> 12.92
Triglycerides	mmol/L		> 5.7		> 11.4
Albumin	g/L	< 20			
Magnesium	mmol/L	< 0.4	> 1.23	< 0.3	> 3.3

Note: A post-baseline value must be more extreme than the baseline value to be considered a potentially clinically important finding.

Table E-3. Criteria for Potentially Clinically Important Vital Sign Values

Vital Signs Variables	Criterion	Definition of Potentially Clinically Important
Systolic Blood Pressure (mmHg)	Low	< 70 mmHg and a decrease of ≥ 30 mmHg from baseline
	High	> 150 mmHg and > 20 mmHg higher than baseline
Diastolic Blood Pressure (mmHg)	Low	< 50 mmHg and a decrease of ≥ 20 mmHg from baseline
	High	> 100 mmHg and higher than baseline
Temperature	Low	≤ 35.6°C
	High	≥ 38.9°C
Pulse (bpm)	Low	< 50 bpm and a decrease of ≥ 30 bpm from baseline
	High	> 120 bpm and an increase of ≥ 30 bpm from baseline
Oxygen saturation (%)	Low	< 95% and a decrease of ≥ 3% from baseline.