Document Type:	Statistical Analysis Plan
Official Title:	A randomized, open-label, active-controlled, Phase II study of intravenous anetumab ravtansine (BAY 94-9343) or vinorelbine in patients with advanced or metastatic malignant pleural mesothelioma overexpressing mesothelin and progressed on first line platinum/pemetrexed-based chemotherapy
NCT Number:	NCT02610140
Document Date:	18-Apr-2018

A randomized, open-label, active-controlled, Phase II study of intravenous anetumab ravtansine (BAY 94-9343) or vinorelbine in patients with advanced or metastatic malignant pleural mesothelioma overexpressing mesothelin and progressed on first line platinum/pemetrexed-based chemotherapy

[Phase II anetumab ravtansine as 2nd line treatment for malignant pleural mesothelioma (MPM)]

Bayer study drug	BAY 94-9343/anetumab ravtansine		
[Study purpose:]	Efficacy and safety		
Clinical study phase:	II	**Date:**	18 Apr 2018
Study No.:	15743	**Version:**	1.0
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Abbreviations

PN	Peripheral Neuropathy
SDE CRF	Study Drug Exposure Case Report Form
SMQ	Standardized MedDRA Query

For additional abbreviations, refer to the main SAP v3.0 dated 9 May 2017.

1. Introduction

This Supplemental Statistical Analysis Plan (SAP) describes analyses that were not included in the main SAP but may be used for Clinical Study Report (CSR) and submission purposes.

This Supplemental SAP version 1.0 supplements SAP version 3.0 dated 9 May 2017.

2. Study Objectives

The primary objective of this study is to:

- Test the superiority of anetumab ravtansine monotherapy over vinorelbine in progression-free survival (PFS)

The secondary objectives of this study are to:

- Test overall survival (OS)
- Evaluate patient-reported outcomes (PROs) – symptom burden and health-related quality of life (QoL)
- Evaluate other indicators of treatment efficacy (indicators of tumor response)
- Evaluate safety

The other objectives of this study are to evaluate the:

- Pharmacokinetics (PKs)
- Immunogenicity
- Biomarkers
    - Further biomarkers to investigate the drug (i.e. mode-of-action-related effect and / or safety) and / or the pathomechanism of the disease (exploratory).

3. Study Design

Refer to the main SAP v3.0 dated 9 May 2017.

4. General Statistical Considerations

Refer to the main SAP v3.0 dated 9 May 2017.

If there are fewer than 10 patients included in a supplemental summary table or figure, listings only may be produced.

Except as otherwise provided in this supplemental SAP, all tables, listings, and figures described in this supplemental SAP will be produced at the time of the OS analysis.

Unless otherwise specified in this supplemental SAP, all endpoints introduced and defined in this SAP will be designated as exploratory.

5. Analysis Sets

Refer to the main SAP v3.0 dated 9 May 2017.

6. Statistical Methodology

6.1 Population characteristics

6.2 Efficacy

Unless otherwise specified in this supplemental SAP, efficacy tables, listings, and figures (TLFs) will be provided for the ITT analysis set.

6.2.1 Waterfall plot of greatest reduction in sum of lesion measurements

The **reduction in sum of the lesion measurements** is defined for each included patient at each post-baseline timepoint, as the difference between the reference baseline sum of the lesion measurements (pleural and non-pleural target lesion diameters) and the sum of the lesion measurements (pleural and non-pleural target lesions diameters) at the timepoint, based on the lesion measurements determined by the central reviewer.

The **greatest reduction in sum of the lesion measurements** is defined for each included patient as the maximum reduction in the sum of the lesion measurements at any post-baseline timepoint.

ITT patients with both a baseline and at least one post-baseline lesion measurement assessment will be included in this analysis.

Confirmation of reduction is not required for this analysis.

A waterfall plot of the greatest reduction in sum of the lesion measurements will be produced in the form of a bar chart. Each bar will represent one subject. Bars will be ordered from maximum increase (left) to greatest reduction (right). Increases will be shown as bars upwards from the baseline, decreases as bars downwards, and 0 change as spaces on the baseline (gap between upward and downward bars). Each patient's treatment group will be depicted as a pattern or grey scale color on the bar.

This analysis will be produced for both the PFS lock data and the OS lock data.

6.2.2 Summary of investigator vs. central review response

A frequency tabulation will be provided of best response on-study per investigator (CR, PR, SD, PD, UE) versus best response on-study per central review (CR, PR, SD, PD, UE).

Best responses of UE (Unevaluable) may be reported as NE (Not Evaluable).

This frequency tabulation will be provided both for the pooled population, and separately for each treatment arm.

This analysis will be produced for both the PFS lock data and the OS lock data.

6.3 Pharmacokinetics/pharmacodynamics

Not applicable.

6.4 Safety

Unless otherwise specified in this supplemental SAP, safety tables, listings, and figures (TLFs) will be provided for the SAF analysis set.

6.4.1 Summary of serious adverse events with incidence ≥ 5%

A summary of treatment-emergent SAEs with incidence (any grade) at least 5% in either treatment arm will be provided by treatment arm and overall, ordered by highest incidence, in a manner otherwise similar to the SAP summary of SAEs.

6.4.2 Time to first incidence of corneal epitheliopathy

The **time to first incidence of corneal epitheliopathy** (CE) is defined in each anetumab ravtansine-treated patient as the time (in weeks) from first anetumab ravtansine treatment to first incidence of treatment-emergent corneal epitheliopathy, i.e. start date of the first CE AE. Patients not experiencing treatment-emergent CE will be censored at the date of their last slit lamp exam.

Time to first incidence of corneal epitheliopathy will be summarized for the anetumab ravtansine treatment arm using Kaplan-Meier estimates. Plots will be produced. Medians and Brookmeyer-Crowley confidence intervals with complementary log-log transformation will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported. For analysis confidence intervals, 95% 2-sided intervals will be calculated. A listing will be provided.

6.4.3 Summary of Peripheral Neuropathy

Peripheral neuropathy (PN) adverse events will be defined and identified in accordance with the Standardized MedDRA Query (SMQ) for peripheral neuropathy as published by MedDRA for the applicable MedDRA version.

The following summary tables of treatment-emergent peripheral neuropathy will be reported in a manner similar to the respective SAP summaries of adverse events:

- any treatment-emergent PN AEs
- any study treatment related PN TEAEs
- any study procedure related PN TEAEs
- any serious PN TEAEs
- any serious study treatment related PN TEAEs
- any serious study procedure related PN TEAEs

- any CTCAE Grade ≥ 3 PN TEAEs
- any PN TEAEs leading to discontinuation
- any PN TEAEs leading to dose interruption
- any PN TEAEs leading to dose reduction

6.4.4 Time to first incidence of peripheral neuropathy

The **time to first incidence of peripheral neuropathy** is defined retrospectively in patients experiencing peripheral neuropathy as the time from first study treatment to first incidence of treatment-emergent peripheral neuropathy, i.e. the start date of the first AE identified as PN by the SMQ.

Time to first incidence of PN will be summarized using summary statistics. A listing will be provided.

6.4.5 Time from first dose modification to resolution of corneal epitheliopathy

The **time from first dose modification to resolution of corneal epithelipathy** is defined in anetumab ravtansine-treated subjects who received a dose modification (reduction, interruption, or discontinuation) due to a CE AE, as the time (in weeks) from the date of dose modification to the date to the date of resolution of the CE AE associated with the dose modification. If the associated CE AE did not resolve, the subject will be censored at the date of the last slit lamp exam.

A patient experienced a dose modification due to a CE AE if the patient:

- Experienced a dose reduction, interruption, or discontinuation, as recorded on the Study Drug Exposure (SDE) CRF.
- The reason for this dose modification per the SDE CRF was adverse event
- At least one of the AEs associated with the dose modification on the SDE CRF was a CE AE.

The date of dose modification is defined as the date of the last dosing visit prior to the visit on which the modification was recorded. If a subject received more than one dose modification due to a CE AE, the first-dated one is selected.

The CE AE associated with a dose modification is defined as the AE on the AE CRF which:

- Is treatment emergent
- Has the same preferred term as the dose modification AE on the SDE CRF.
- Has a start date (or date of grade change if the AE represents a grade change) on or before the date of dose modification.
- Has an action taken with study drug corresponding to the modification (reduction, interruption, or discontinuation) on the SDE CRF.

- If more than one CE TEAE meets the above criteria the latest one with start or grade change date on or before the date of dose modification is selected.

Time from first dose modification to resolution of CE will be summarized using Kaplan-Meier estimates. Plots will be produced. Medians and Brookmeyer-Crowley confidence intervals with complementary log-log transformation will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported. For analysis confidence intervals, 95% 2-sided intervals will be calculated. A listing will be provided.

6.4.6 Time from first dose modification to resolution of peripheral neuropathy

The **time from first dose modification to resolution of peripheral neuropathy** is defined retrospectively in subjects who received a dose modification (reduction, interruption, or discontinuation) due to a PN AE and whose peripheral neuropathy resolved, as the time (in weeks) from the date of dose modification to the date to the date of resolution of the PN AE associated with the dose modification.

See Section 6.4.5 for definitions of patient experiencing dose modification due to PN AE, date of dose modification, the PN AE associated with the dose modification, substituting PN for CE events. No censoring is performed.

Time from first dose modification to resolution of PN will be summarized using summary statistics. A listing of all patients with dose modifications due to a PN AE will be provided. If the patient's selected PN AE did not resolve, a time to resolution will not be calculated.

7. Document history and changes in the planned statistical analysis

- Supplemental SAP FINAL version 1.0 dated 01 April 2018

8. References

Refer to the main SAP v3.0 dated 9 May 2017.

A randomized, open-label, active-controlled, Phase II study of intravenous anetumab ravtansine (BAY 94-9343) or vinorelbine in patients with advanced or metastatic malignant pleural mesothelioma overexpressing mesothelin and progressed on first line platinum/pemetrexed-based chemotherapy

[Phase II anetumab ravtansine as 2nd line treatment for malignant pleural mesothelioma (MPM)]

Bayer study drug	BAY 94-9343/anetumab ravtansine		
Clinical study phase:	II	**Date:**	9 MAY 2017
Study No.:	15743	**Version:**	3.0
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AD	As-Diagnosed
ADAs	Anti-Drug Antibodies
AE	Adverse Event
AMR	Application Maintenance Request
AUC	Area Under the Curve
CI	Confidence Interval
CR	Complete Response
CRF	Case Report Form
CSR	Clinical Study Report
CSS	Composite Symptom Score
DCR	Disease Control Rate
DOR	Duration of Response
DRR	Durable Response Rate
eCRF	Electronic Case Report Form
HRQoL	Health Related Quality of Life
IBPRC	Independent Blinded PRO (patient reported outcomes) Review Committee
ICF	Informed Consent Form
IHC	Immunohistochemistry
IMM	Immunogenicity
IRR	Infusion Related Reaction
ITT	Intent-to-treat
IxRS	Interactive (voice or web) Response system
LCSS-Meso	Lung Cancer Symptom Scale-Mesothelioma
LKAD	Last Known Alive Date
MDASI-MPM	MD Anderson Symptom Inventory-Malignant Pleural Mesothelioma
MedDRA	Medical Dictionary for Regulatory Activities
mRECIST	Modified Response Evaluation Criteria in Solid Tumors
M&S	Modelling and Simulation
N/A	Not Applicable
NABs	Neutralizing Antibodies
NCI-CTCAE	National Cancer Institute's Common Terminology Criteria for Adverse Events
OS	Overall Survival
PD	Progression Disease
PFS	Progression Free Survival
PIR	Pain Improvement Rate
PK	Pharmacokinetics
PR	Partial Response
PRO	Patient Report Outcome
QoL	Quality of Life
RoW	Rest of the World
SAE	Serious Adverse Event
SAF	Safety
SAP	Statistical Analysis Plan
SD	Stable Disease
SIR	Symptom Improvement Rate
SOC	System Organ Class
TEAE	Treatment-Emergent AE
TTP	Time To Progression
TTWP	Time to Worsening of Pain
TTWS	Time to Worsening of Symptoms
UE	Un-Evaluable
WHO-DD	World Health Organization Drug Dictionary

1. Introduction

Study 15743 is a randomized, open-label, active-controlled Phase II registrational study to compare the efficacy and safety of intravenous anetumab ravtansine (BAY 94-9343) against vinorelbine in patients with advanced or unresectable locally metastatic malignant pleural mesothelioma overexpressing mesothelin who progressed on first line platinum/pemetrexed-based chemotherapy.

Two analyses are planned, the final analysis for primary endpoint progression-free survival (PFS), which will also serve as an interim analysis for secondary endpoint overall survival (OS), and the final OS analysis.

This Statistical Analysis Plan (SAP) specifies the analyses and data presentations planned for both the final primary endpoint and the final OS analysis.

This SAP was written based on the following documentation:

Document	Date	Version
Protocol	17 August 2015	1.0
Protocol Amendment no. 1, local amendment for UK only	04 December 2015	Integrated CSP version 1.0
Protocol Amendment no. 2, global amendment	9 February 2016	Integrated CSP version 2.0
Protocol Amendment no. 3, local amendment for Finland only	23 May 2016	Integrated CSP version 2.0
Protocol Amendment no. 4, global amendment	11 Aug 2016	Integrated CSP version 3.0
Protocol Amendment no. 5, global amendment	18 April 2017	Integrated CSP version 4.0
eCRF - Subject Case Report Forms	30 November 2015	1.0
eCRF – AMR1	26 Feb 2016	2.0
eCRF – AMR2	15 Mar 2016	3.0
eCRF – AMR3	18 Apr 2016	4.0
eCRF – AMR4	30 Aug 2016	5.0
eCRF – AMR5	19 Sep 2016	6.0
eCRF – AMR6	7 Dec 2016	7.0
Bayer Therapeutic Area Oncology Standards (TAS)	2 August 2016	July 2016
Bayer Oncology Standard Tables	08 Dec 2016	3.3
Anetumab Ravtansine Standards	11 August 2016	August 2016
Bayer Global Medical Standards (GMS)	9 May 2016	March-April 2016
15743 Data Monitoring Committee (DMC) Charter	2 Feb 2016	1.0
15743 Data Monitoring Committee (DMC) Charter	24 May 2016	2.0
15743 Independent Radiological Review Imaging Charter	16 December 2015	1.0
15743 Independent Radiological Review Imaging Charter	21 March 2016	2.0

2. Study Objectives

The primary objective of this study is to:

- Test the superiority of anetumab ravtansine monotherapy over vinorelbine in progression-free survival (PFS)

The secondary objectives of this study are to:

- Test overall survival (OS)
- Evaluate patient-reported outcomes (PROs) – symptom burden and health-related quality of life (QoL)
- Evaluate other indicators of treatment efficacy (indicators of tumor response)
- Evaluate safety

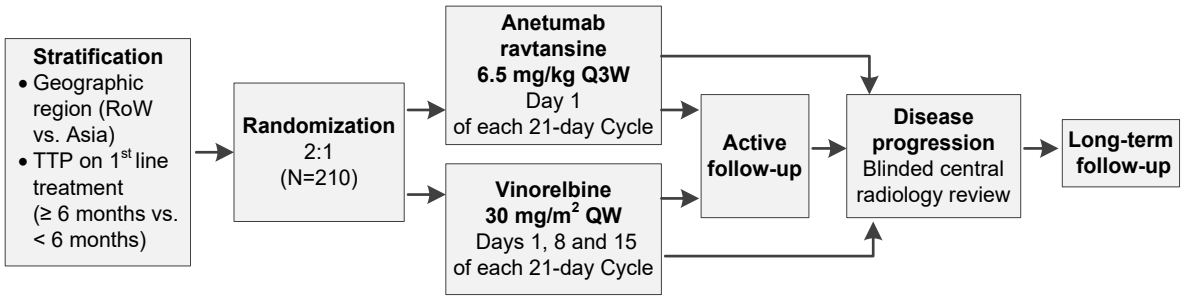
The other objectives of this study are to evaluate the:

- Pharmacokinetics (PKs)

- Immunogenicity
- Biomarkers
    - Further biomarkers to investigate the drug (i.e. mode-of-action-related effect and / or safety) and / or the pathomechanism of the disease (exploratory).

3. Study Design

Figure 3–1 Overall study design



This is a randomized, open-label, active-controlled, 2-arm, multicenter, Phase II trial to evaluate the safety and efficacy of anetumab ravtansine as a single agent administered as an IV infusion Q3W in comparison to IV vinorelbine given according to the usual QW schedule. Approximately 210 patients will be randomized in this study.

At the time of the start of study treatment, the patients will have unresectable locally advanced or metastatic MPM recurrent/relapsing after a 1st line treatment with platinum in combination with pemetrexed with or without bevacizumab, and overexpressing mesothelin as determined by immunohistochemistry (IHC). Only patients who demonstrate mesothelin overexpression at the moderate and stronger level by IHC in at least 30% of tumor cells can be randomized into the study. A prescreening step, including the mesothelin expression level testing, can be performed without evidence of disease progression after the initial treatment cycles with platinum/pemetrexed (with or without bevacizumab) at the investigator's discretion.

The start of the study is defined by signing of the informed consent form (ICF) for prescreening. After meeting the eligibility criteria for prescreening and signing the ICF for full study), approximately 210 patients who meet all of the inclusion and none of the exclusion criteria will be randomly assigned in a 2:1 ratio to receive anetumab ravtansine or vinorelbine, respectively. The anetumab ravtansine arm will consist of approximately 140 patients and the vinorelbine comparator arm of approximately 70 patients.

Patients will be stratified at randomization according to geographic region (Rest of the world [RoW] versus Asia) and per time to progression (TTP) on 1st line treatment (≥ 6 month versus < 6 months). An approximately 54% screen fail rate is anticipated (25% at

prescreening and a subsequent 33% among biomarker expressers. Approximately 420 patients are estimated to be required for prescreening to yield approximately 263 biomarker-positive patients, resulting in 210 randomized eligible patients).

The start of the treatment period is defined for efficacy purposes by randomization to study drug (anetumab ravtansine or vinorelbine), and for safety purposes by first administration of study drug. Start of treatment has to be within 24 hours after the randomization call. Patients in the anetumab ravtansine arm will receive anetumab ravtansine IV infusion at a dose of 6.5 mg/kg (recommended Phase II dose [RPIID]) on Day 1 of a 21-day cycle. Patients in the comparator arm will receive vinorelbine 30 mg/m^2 IV on Days 1, 8 and 15 of a 21-day cycle. Treatment will be continued until death or occurrence of PD as defined by Modified Response Evaluation Criteria in Solid Tumors (mRECIST) for mesothelioma [1] and assessed by blinded central radiology review, or clinical progression, or until another criterion for withdrawal from the study is met. In case of radiological progression, however, treatment may be continued provided that the patient derives clinical benefit as determined by the treating physician.

All patients who discontinue study treatment for any reason will enter the safety follow-up period. Safety follow-up visit will be performed 30 (+7) days after the last administration of study treatment.

Patients who discontinue study treatment for any reason other than centrally confirmed radiological PD will be followed for progression during active follow-up (which includes the safety follow-up period) until centrally confirmed PD for this patient is observed.

All patients who end study treatment for any reason will be followed for OS and any new anti-cancer treatment every 3 months during the long-term follow-up period until data maturation for the OS final analysis is reached, or until death, consent withdrawal or end of study, whichever occurs first.

Primary efficacy will be assessed based on radiological tumor evaluation by contrast enhanced CT or contrast-enhanced magnetic resonance imaging (MRI) of chest/abdomen/pelvis. The first tumor images will be obtained during full screening and will be sent to blinded central review to confirm radiological eligibility prior to randomization. During treatment as well as active follow-up, tumor imaging will be performed with the same modality every 6 weeks during the first 6 months after the start of study treatment, every 9 weeks until the end of year 2, and every 12 weeks thereafter until centrally confirmed radiological disease progression or end of study. Primary analysis results will be based on central review.

Patients will be contacted to assess survival status every 3 months during long-term follow-up. In addition, extra survival sweep contacts will be conducted at the time of PFS final analysis and prior to OS final analysis to ensure that long-term follow-up data is current.

The effect of treatment on disease-specific symptoms and disease-specific health-related QoL will be assessed using the MD Anderson Symptom Inventory-Malignant Pleural Mesothelioma (MDASI-MPM) and the Lung Cancer Symptom Scale-Mesothelioma (LCSS-Meso), respectively, at full screening, at each cycle during treatment, at the safety follow-up

visit, and during active follow-up. An independent blinded PRO review committee of PRO, statistical, and psychometric experts, will conduct analyses to support validation of the MDASI-MPM and determine endpoint definitional details.

Safety evaluations will be done at full screening, at each clinic visit during the treatment, and at the safety follow-up visit. The National Cancer Institute's Common Terminology Criteria for Adverse Events (NCI-CTCAE) version 4.03 will be used to grade severity of adverse events (AEs). In addition, a Bayer grading system (see Protocol Table 7–5 and Table 7–6) will be used to assess corneal epitheliopathy.

Sparse plasma sampling for PK will be performed on all patients.

Immunogenicity assessment will be performed for patients in the anetumab ravtansine arm only.

Obligatory biomarker sampling will be performed on all patients to measure mesothelin expression levels in tumor material at prescreening. In addition, plasma levels of soluble mesothelin will be studied to evaluate whether plasma mesothelin levels may correlate with response rate and be of predictive value. Biomarker plasma will be collected to analyze circulating tumor DNA, too. Exploratory biomarker analysis may also be performed using additional fresh or archival tumor tissue to determine alterations in tumor-associated genes and to perform gene expression analysis.

3.1 Determination of Sample Size

The sample size is primarily designed to support hypothesis test of the primary endpoint PFS, and to provide a limited formal evaluation of secondary endpoint OS.

Power calculations were performed using East 6.3 software.

3.1.1 Primary endpoint progression-free survival

The sample size is designed to test the following hypotheses for primary endpoint PFS:

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, PFS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to PFS under treatment with 30 mg/m^2 QW vinorelbine.

versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, PFS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to PFS under treatment with 30 mg/m^2 QW vinorelbine.

These hypotheses will be tested using a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per TTP on 1st line treatment (≥ 6 months versus < 6 months) at a significance level of 0.0125.

Assuming median PFS of 3.6 months under vinorelbine treatment and constant hazards and a 2:1 treatment: comparator randomization, a 100% prolongation of PFS (hazard ratio 0.5) in the anetumab ravtansine arm in comparison to the comparator arm can be detected at a 1-sided significance level of 0.0125 with 90% power, with a single-stage trial with approximately 117 PFS events. Assuming a maximum accrual rate of 12.5 patients/month (20.83 patients/month screened with 40% overall screening failure rate) with 6-month linear accrual ramp-up, and a 3.4%/month dropout (loss to follow-up and unevaluable for tumor assessment) rate, 210 patients be will accrued in approximately 19.8 months and reach endpoint maturation of 117 events in approximately 22.0 months. The total number of unevaluable/dropout patients over the duration of the study through final PFS analysis is estimated at 33 (15.7%).

3.1.2 Secondary endpoint overall survival

The sample size is also designed to test the following hypotheses in secondary endpoint OS:

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, OS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to OS under treatment with 30 mg/m^2 QW vinorelbine.

versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, OS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to OS under treatment with 30 mg/m^2 QW vinorelbine.

These hypotheses will be tested using a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per TTP on 1st line treatment (≥ 6 months versus < 6 months), at a significance level of 0.025.

Assuming true median OS of 9.6 months under vinorelbine treatment and constant hazards, a 60% prolongation of true OS (median 15.4 months) in the anetumab ravtansine arm in comparison to the comparator arm (median 9.6 months) can be detected with an overall 80% power and an overall 1-sided significance level of 0.025 (hazard ratio 0.625), with a 2-stage group sequential test with a total of 159 events (modified by amendment 2).

The OS analysis assumes the same accrual as for primary endpoint PFS, with 210 patients accrued in approximately 19.8 months. A 0.3% per month loss to OS follow-up rate is assumed. An interim OS analysis will be performed at the time of the final primary endpoint (PFS) analysis at an estimated 22 months from first patient randomized, when an estimated 80 OS events will have been observed. If the study is not stopped for superiority at the OS interim analysis, the final OS analysis will occur after approximately159 OS events have been observed, at approximately 41.3 months from first patient randomized. A Lan-Demets alpha spending function [10] with O'Brien-Fleming boundaries [11] will be used, based on actual events at the time of the interim analysis. Under the null hypothesis of no anetumab ravtansine OS superiority, approximately 0.00158 alpha is estimated to be spent at the interim

analysis and 0.02342 at the final (0.025 alpha overall). Under the alternative hypothesis of 60% OS improvement under anetumab ravtansine treatment, the chance of finding superiority is estimated at 16.9% at the interim analysis and 63.2% at the final analysis (80% power overall). The total number of patients lost to OS follow-up over the duration of the study through final OS analysis is estimated at 10 (4.8%).

3.2 Blinding

This is an open-label study except for:

- The independent and central radiological review for the assessment of disease progression and other radiological imaging based endpoints, which will be conducted in a blinded fashion.
- The independent blinded PRO review committee which will conduct analyses to support validation of the MDASI-MPM instrument and determine endpoint definitional details.

An independent data monitoring committee (DMC) will monitor safety in this study, its role is further described in the Data Monitoring Committee Charter. An interim OS analysis will be performed by the sponsor at the time of final PFS analysis. The sponsor will not perform interim efficacy assessments. No further interim efficacy analyses are planned.

4. General Statistical Considerations

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS release 9.2 or higher (SAS Institute Inc., Cary, NC, USA). Calculations of significance level adjustment for group sequential designs based on alpha spending functions will be performed using East 6.3 or higher (Cytel Software Corporation, Cambridge, MA, USA).

All variables will be analyzed by descriptive statistical methods.

In general, continuous variables will be summarized using number of non-missing values (n), mean, standard deviation, median, maximum, minimum, and interquartile range.

Ordinal variables will be summarized using n, median, maximum, minimum, and interquartile range.

Categorical variables will be summarized using n, number of missing values, and percentages.

Time-to-event variables will be summarized using Kaplan-Meier estimates [5].

Frequency tables will be generated for categorical data.

Unless otherwise specified, stratified analyses will be performed based on IxRS stratification factors.In the event that a per-IXRS stratification factor enrollment distribution results in sparse cells preventing convergence or numerically reliable estimates, the stratification factor will be dropped from the analysis.

4.2 Handling of Dropouts

A "dropout" is defined as a randomized patient who discontinues study treatment prior to start of study treatment (i.e. without having received any study treatment) for any reason. Dropouts are asked to continue in active follow-up and/or long-term follow-up.

Once randomized, patients (e.g. dropouts and patients withdrawn from study treatment) will not be replaced. Refer to Section 6.3 in the study protocol for withdrawal of patients from study.

In addition, the following terms identify patients who discontinue study participation at stages prior to randomization:

A patient whose tumor tissue is centrally tested by IHC for mesothelin overexpression and whose result is not moderate and stronger mesothelin overexpression in at least 30% of the tumor cells, or who fails to meet any of the other eligibility criteria for prescreening, is regarded as a "prescreening failure";

A patient who passes the prescreening, including the mesothelin overexpression testing, enters full screening, but for any other reason (e.g. failure to satisfy the remaining selection criteria) terminates the study before randomization is regarded as a "full screening failure".

4.3 Handling of Missing Data

All missing or partial data will be presented in the patient data listing as they are recorded on the Case Report Form (CRF).

Additional descriptive analyses in the presence of missing data

The number of patients who prematurely discontinue the study and study treatment for any reason, as well as the reasons for premature discontinuation of study and study treatment, will be reported. Kaplan-Meier [5] plots for "Time to end of study" and "Time to end of study treatment" will be provided.

Patients withdrawing from treatment, active follow-up, and long-term follow-up will be evaluated with respect to

- baseline characteristics
- potential differences between the treatment groups in the proportion of patient withdrawals or in the timing of withdrawals, and
- the reasons for premature discontinuation of study and/or study treatment

and potential withdrawal patterns will be described.

The number, timing, pattern, and reasons for missing values of all relevant efficacy and safety variables will be displayed by means of descriptive statistics and visualized if applicable. Data exploration will include investigation of

- potential missing data imbalances
- baseline characteristics of patients with and without missing values

- the direction of any change over time with time courses of patients with and without missing data

Treatment withdrawal criteria are specified in Figure 6-1 in the Integrated Clinical Study Protocol.

Unless otherwise specified, where values are required to be "known" or "valid" as a criterion for inclusion in an analysis set., the values must be complete without imputation (e.g. no partial dates).

4.3.1 Imputation rules

Imputation rules for specific types of data are described in Appendix 9.1.

4.4 Interim Analyses and Data Monitoring

No formal interim analysis will be performed for primary endpoint PFS.

A formal interim analysis for secondary endpoint OS will be performed by the sponsor at the same time as the primary endpoint analysis. A Lan-Demets [10] alpha spending function with O'Brien-Fleming [11] boundaries will be used, based on actual events at the time of the interim analysis. The actual alpha level will be based on the actual number of OS events included in the interim analysis, i.e. in the database as of the PFS primary endpoint lock. The information fraction for interim OS analysis used as input into the alpha spending function will be calculated as the number of OS events occurring as of PFS primary endpoint lock, divided by the total number of OS events (159) planned for the final OS analysis per protocol.

In addition, an interim descriptive analysis will be performed for duration of response and durable response rate at the time of the final PFS analysis.

Accordingly, two analyses of this study are planned: the PFS analysis; which will be reported in the Clinical Study Report and include interim analyses of OS, DOR, and DRR; and the final OS analysis, which will include the final analysis of DOR and DRR and an updated analysis of other endpoints.

Secondary OS will be evaluated separately in a 2-stage group sequential procedure as further described below, with final analysis expected to occur after the primary endpoint analysis. Final analysis for other secondary variables, except for DOR and DRR, will occur at the same time as the primary endpoint analysis. Final analysis of DOR and DRR will occur at the time of final OS analysis. To preserve the validity of the secondary endpoint hierarchy and control of overall study-wide secondary Type I error, superiority for secondary endpoint hypothesis tests ranking below OS in the hierarchy cannot be declared for regulatory label claim purposes until OS hypothesis testing succeeds. Accordingly, unless OS superiority is found at the interim OS analysis, the hypothesis testing outcome of endpoints ranking below it will not be fully effective, and any superiority for these endpoints cannot be declared for

regulatory label claim purposes, until the outcome of the OS hypothesis test is determined and OS superiority found at the final OS analysis.

4.5 Data Rules

Generally, for each date stored in the database, a set of organizational variables will be derived in order to describe the temporal context of that date in the specific study: phase of treatment (pre, during or post study treatment), day relative to the start of study treatment, day relative to the end of study treatment.

Additional contextual variables may be created in analysis datasets.

Unless otherwise specified, the baseline assessment will be the last valid value prior to first treatment. Unless otherwise specified, assessments or procedures which are performed prior to first treatment per protocol will be assumed to occur prior to first treatment. Any unscheduled assessments will be included in the determination of the baseline value.

4.6 Blinded Review, Validity Findings, and Protocol Deviations

Protocol deviations and validity review will be performed as described in the sponsor's operational instruction on conducting blinded review meetings, cleaning and reviewing study data, in addition to detailed guidance on important deviations and validity findings.

4.6.1 Blinded Review

Blinded review will be conducted prior to database lock for each analysis, as described in the sponsor's operational instruction on conducting blinded review meetings. Reviewers will be tasked with evaluating the overall quality and reliability of the study data and its suitability for the planned statistical analysis. Blinded review will also assess protocol deviations and validity findings.

4.6.2 Validity Findings

Criteria for validity, i.e. eligibility for each analysis set, are described in Section 5. Details may be further described in a separate document. Validity findings will be obtained as described in both the sponsor's operational instruction on cleaning and reviewing study data, and the sponsor's detailed guidance on important deviations and validity findings, and include major protocol deviations as described in Section 4.6.3 and the separate Protocol Deviations Document. The blinded review will review and finalize the validity findings prior to the database lock for each analysis.

4.6.3 Protocol Deviations

Protocol deviations will be classified as major, important or other. Major protocol deviations are deviations affecting analysis set eligibility or treatment group assignment and constitute validity findings as described in Section 4.6.2.

Important protocol deviations will be identified based on the criteria described in the sponsor's operational instruction on cleaning and viewing study data. Assessment criteria, deviation definitions, methods of evaluation, and further details will be described in a separate Protocol Deviations Document.

5. Analysis Sets

- **Intent-to-treat set (ITT)**. All randomized patients will be included in the ITT set. Patients in this set will be reported by treatment arm as randomized. This set will be used for patient characteristics, demographic, and efficacy evaluations. Patients will be included in ITT analyses according to the treatment to which they were randomized per IxRS.
- **Safety set (SAF)**. All treated patients, that is, all randomized patients receiving any amount of any study treatment, will be included in the SAF. Patients in this set will be reported by treatment arm as treated. The SAF will be used for safety evaluations.
- **Pharmacokinetic set (PK)**. All patients with valid at least one valid PK sample for anetumab ravtansine will be included in the analysis set for the anetumab ravtansine PK analysis. All patients with at least one valid PK sample for vinorelbine will be included in the analysis set for the PK analysis of vinorelbine. PK samples will be considered valid under the following conditions: known dose, known duration of treatment, known time of sample collection.
- **Immunogenicity set (IMM)**. All patients who have received at least one dose of anetumab ravtansine and have at least one valid post-baseline measurement of anti-drug antibodies (ADAs) or neutralizing antibodies (NABs) will be included in the immunogenicity set. The IMM analysis set will be used for the immunogenicity analysis of anetumab ravtansine.
- **Quality of life set (QoL)**. All randomized patients with at least one non-missing LCSS-Meso evaluation at baseline will be included in the QoL dataset. Patients in this set will be reported by treatment arm as randomized. The QoL set will be used for QoL evaluations.
- **Biomarker set (BIO),** also called the **As-Diagosed set (AD)**. All patients evaluated for mesothelin expression at pre-screening with a valid mesothelin expression test result) will be included in the biomarker set. The biomarker set will be used for biomarker analyses, for patient disposition, and for comparability to Ventana's biomarker analysis.
- **Enrolled set (ENR)**. All patients who signed the informed consent for any portion of the study including pre-screening. Enrolled patients will be used for patient disposition.

6. Statistical Methodology

6.1 Population characteristics

6.1.1 Disposition of patients

The number and percentage of patients entering pre-screening, completing pre-screening, entering screening, randomized, treated, entering safety follow-up, active follow-up and long-term follow-up will be presented by treatment group and overall. The number of patients discontinuing from pre-screening, screening, treatment, safety follow-up, active follow-up, and long-term follow-up will also be summarized by treatment group. The reasons for patients discontinuing from pre-screening, screening, treatment, safety follow-up, active follow-up, and long-term follow-up, will be summarized by treatment group. In addition, the number of patients screened and included in each analysis set will be displayed overall. and, the number of patients in the ITT and SAF analysis sets will also be displayed by region and country .

6.1.2 Demographic and Baseline Characteristics

Descriptive summaries of demographics and baseline characteristics will be presented by treatment group and overall for the ITT analysis set. Comparability of the treatment groups with respect to demographics and baseline characteristics will be assessed using descriptive summaries.

The following demographic data will be summarized:

- Age (years)
- Age category (< 65, ≥65 years)
- Sex (Male, Female)
- Race (White, Black, Asian Other, NA)
- Ethnicity (Hispanic, not Hispanic. NA)
- Height (cm)
- Weight (kg)

The following baseline characteristics will be summarized:

- Mesothelin overexpression level (percent moderate or stronger cells) (distribution of frequencies by 10% levels counting A+5% as A; ordinal summary statistics)
- ECOG PS (0, 1)
- Stage of malignant pleural mesothelioma at diagnosis
- Histology (epithelioid, biphasic, sarcomatoid)
- Time since initial diagnosis (months), defined as time from initial diagnosis as recorded on the CRF to randomization date

- Time since first progression (months), defined as time from date of first progression as recorded on the CRF to randomization date.
- Number of mRECIST pleural lesion measurements
- Sum of mRECIST pleural lesion measurements
- Number of non-pleural target lesions
- Sum of diameters of non-pleural target lesions
- Sum of pleural lesion measurements and non-pleural target lesion diameters
- MDASI-MPM symptom Composite Symptom Score (CSS) (summary statistics; categories0-<4, ≥4)
- MDASI-MPM pain score (summary statistics; categories0--<4, ≥4)
- LCSS-Meso total score (summary statistics)

Ophthalmologic characteristics will be reported separately by ocular laterality (Left eye, right eye, highest visual acuity of both eyes):

- Slit Lamp Exam Bayer Grade (0, 1)
- Snellen Acuity Equivalent (summary statistics)
- Intra-ocular pressure (mmHg) (summary statistics)
- Schirmer test (mm/5 min) (categories normal/abnormal, abnormal: ≤ 10 mm/5 min)

Note that the date of MPM diagnosis will be described in a listing.

Categorical summaries of each randomization stratification factor per IXRS:

- Region (Rest of World, Asia)
- Time to progression on first line treatment (≥ 6 months, < 6 months)

, Categorical summaries the randomization stratification factor per CRF (time to progression on first-line treatment) will also be performed.

6.1.3 Medical history

Medical history will be summarized by Medical Dictionary for Regulatory Activities (MedDRA) body system organ class (SOC) and preferred term for the ITT population, by treatment group.

6.1.4 Prior and Concomitant Medications

All investigator-reported non-study medications taken before and/or during the study will be coded using the World Health Organization Drug Dictionary (WHO-DD) 2005 Q3 and the Anatomical Therapeutic Chemical (ATC) classification system. Coding will include the drug class and generic drug name.

Non-study medications taken before and/or during the study will be categorized as prior medications, concomitant medications, and post treatment medications.

Post treatment medications are defined as non-study medications taken after the treatment period (including treatment and safety follow-up).

Prior medications are defined as all medication used before the treatment period.

Concomitant medications are defined as all medications used during the treatment period.

All non-study medications will be listed, including verbatim descriptions and coded terms, and flags for prior and post-treatment medications. Priorand concomitant medications will be summarized using frequencies of patients reporting each drug category and generic drug name. For each patient, multiple records of the same concomitant medication will be counted once within a drug class and generic drug name.

For the analysis of medications, the treatment period begins on the date of first treatment and ends 30 days following the date of last study treatment. A medication may be classified in more than one category. Classification details are shown in Table 6-1.

Table 6-1 Medication classification

	Prior to study drug	Start of study drug	Treatment with study drug	Safety Follow-Up	Post Safety Follow up	Prior Medication?	Concomitant Medication?	Post-treatment Medication?
C1						Yes	No	No
C2						Yes	Yes	No
C3						No	Yes	Yes
C4						Yes	Yes	Yes
C5						No	Yes	No
C6						No	No	Yes
C7						No	Yes	No

C1= medication started and ended before study drug
C2= medication began before study drug and ended during study drug administration
C3= medication started at or after study drug administration and ended after end of safety follow-up
C4= medication started before study drug administration and ended after end of safety follow-up
C5= medication started after safety follow-up started and ended before safety follow-up ended
C6= medication started after safety follow-up ended
C7= medication started at or after study drug administration and ended before safety follow-up ended

6.1.5 Prior Therapy for Cancer

Prior therapies for cancer (MPM) will be summarized for the ITT population, by treatment group. The following summaries will be reported:

- Prior Bevacizumab therapy (yes, no)
- Prior pleurodesis (yes, no)
- Number of prior systemic therapy regimens (0, 1, >1)
- Number of prior surgeries (0, 1, 2, 3, >3)
- Number of prior radiotherapies (0, 1, 2, 3, >3)
- Number of prior other diagnostic and therapeutic procedures (i.e. other than systemic therapy regimens, surgeries, and radiotherapies) (0, 1, 2, 3, >3)

In addition, prior systemic anti-cancer therapies will be summarized using frequencies of patients reporting each drug category and generic drug name. For each patient, multiple records of the same therapy will be counted once within a drug class and generic drug name. Prior anti-cancer therapy for study cancer will also be reported through listings. Pleurodesis represents either pleurodesis surgery or pleurodesis or a pleurodectomy procedure.

6.1.6 Systemic anti-cancer therapy

Subsequent and any concomitant systemic anti-cancer therapy reported on the CRF will be reported through listings, and summarized using frequencies of patients reporting each drug category and generic drug name. For each patient, multiple records of the same therapy will be counted once within a drug class and generic drug name.

6.1.7 Additional radiotherapies and diagnostic and therapeutic procedures

- Numbner of concurrent radiotherapies (0, 1, 2, 3, >3)
- Concurrent diagnostic procedures (yes, no)
- Concurrent therapeutic procedures (yes, no)
- Number of radiotherapies during follow-up (0, 1, 2, 3, >3)

6.2 Efficacy

6.2.1 Primary Efficacy Analysis

The primary efficacy variable is progression-free survival (PFS), as determined by the central radiological reviewer. The primary efficacy analysis will be performed in the ITT analysis set.

Endpoint Definition

Progression-free survival (PFS) is defined as time from randomization until disease progression according to modified RECIST criteria for malignant pleural mesothelioma (mRECIST) per blinded central radiology review, or death. Patients not experiencing death or progression will be censored at the last tumor assessment.

Modified RECIST criteria for malignant pleural mesothelioma (mRECIST) ([1],[2]) are further described in Protocol Section 16.3 and in the Image Review Charter for Study 15743.

Censoring Rules

Censoring rules for primary endpoint PFS are described in Table 6-2.

Table 6-2 Censoring rules for PFS

Censoring for progression-free survival for Situation	End Date	Censored	Reason for Censoring
No baseline radiological tumor assessment	Date of Randomization	Yes	No baseline assessment
No post-baseline radiological assessment or death	Date of randomization	Yes	No post-baseline radiological assessment or death
No post-baseline radiological assessment, death following window for 2 consecutive missed tumor assessments following randomization	Date of randomization	Yes	No post-baseline radiological assessment
No post-baseline radiological assessment, death within window for 2 missed tumor assessments following randomization	Date of death	No	N/A
Patient had a radiological assessment of PD following randomization (no 2 consecutive missing radiological assessments)	Date of first radiological progression event	No	N/A
Death or PD immediately following two or more consecutive missed radiological assessments	Date of last radiological assessment before missed assessments	Yes	Missed two or more consecutive tumor assessments
Patient discontinued from study without radiological PD or death	Last radiological assessment date	Yes	Patient discontinued from study without PD or death
Death during the study (no 2 consecutive missing radiological assessments) or before first radiological PD assessment	Date of Death	No	N/A
Patient still on study at the time of data cut-off without PD	Last radiological assessment before data cut-off	Yes	Patient is still alive without PD

Censoring rules based on 2 consecutive missing tumor assessments prior to an event will be implemented programmatically based on tumor assessment results provided by the central radiological reviewer. These rules are further described in Table 6-3.

Table 6-3 Censoring for 2 consecutive missed assessments

Last Non-Missed Assessment	Event Time from Randomization	Interval Required to Censor*
No post-baseline assessments	Any	15 weeks
Prior to 25 Weeks	Prior to 25 Weeks	15 weeks
Prior to 25 Weeks	25 Weeks to 106 weeks	19 weeks
25 Weeks to 106 Weeks	25 Weeks to 106 Weeks	23 weeks
25 Weeks to 106 Weeks	More than 106 Weeks	27 weeks
More than 106 Weeks	More than 106 Weeks	30 weeks

*Event is censored if occurring more than indicated interval beyond last non-missing assessment. Interval represents 2.5 scheduled assessments (2 missing in between) appropriate to assessment period.

Hypotheses Tested

The analysis tests the following hypotheses in primary endpoint PFS.

Null Hypothesis H0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, PFS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to PFS under treatment with 30 mg/m^2 QW vinorelbine.

Versus

Alternative Hypothesis HA: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, PFS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to PFS under treatment with 30 mg/m^2 QW vinorelbine.

Primary Analysis

These hypotheses will be tested using a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per TTP on 1st line treatment (≥ 6 months versus < 6 months), with a 1-sided significance level of 0.0125.

The final primary endpoint analysis will be performed after approximately 117 PFS events have been observed.

In addition to the stratified log-rank test described above, PFS will be summarized using Kaplan-Meier estimates [5]. Plots will be produced by treatment arm. Medians and Brookmeyer-Crowley confidence intervals [6] with complementary log-log transformation [7] will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported. The hazard ratio (anetumab ravtansine regimen/vinorelbine) will be estimated using Cox proportional hazards models [8] with Wald confidence intervals [9] stratified by the same factors. For primary endpoint analysis

confidence intervals, both 95% 2-sided intervals and 97.5% 2-sided intervals (reflecting primary endpoint significance level) will be calculated.

Components Analysis

A frequency table will be provided by treatment arm identifying the component events of first progression or death in each arm, as follows:

- Censored,
- Death,
- Progression of pleural or non-pleural target lesions,
- Progression of non-target lesions,
- New lesion(s)

For patients with multiple first progression component events on the same day, all such components will be reported.

For censored patients, a frequency table will be provided for censoring in each arm, as described in Table 6-2.

Checking of Assumptions

To assess the assumption of consistent observation across the treatment arms, summary statistics on the difference between actual and target tumor assessment study day for each arm will be calculated for each target visit study day by treatment arm. Each target tumor assessment visit date per the PFS visit schedule (every 6 weeks during the first 6 months (24 weeks) after the start of study treatment, every 9 weeks until the end of year 2 (105 weeks), and every 12 weeks thereafter) will be associated with its closest actual tumor assessment visit date. Ties will be resolved in favor of the earlier visit. For each patient, this assessment will be performed through the last included assessment prior to an event; events and assessments excluded from the primary analysis (e.g. due to the missing assessment censoring rule) will be excluded from this analysis. Absolute differences will be reported similarly.

To assess the assumption of non-informative censoring, Kaplan-Meier [5] survival curves, and corresponding exponential regression hazard estimates will be calculated for censoring times by 3-month intervals, characterizing PFS censorings as events (events as censorings) for purposes of this analysis. Similarity and proportionality of the censoring times will be evaluated. All primary analysis censorings, including censorings due to the missing assessment and other rules, will be included in this analysis.

Sensitivity Endpoints and Analyses

To evaluate proportionality of the hazards, hazards by time will be estimated in each arm using exponential regression models by 3-month intervals. Tables and plots of hazards and corresponding estimated survival functions, with interval model estimates overlaying Kaplan-Meier [5] estimate for each arm, will be produced.

A non-stratified log-rank test and a test stratified by stratification factors per CRF will be performed. In addition, stratification factors per IXRS and per CRF will be included in subgroup analyses. (Note: region is per site identifier, only time to progression on first line-treatment differs per CRF).

Sensitivity endpoints for primary endpoint PFS will be calculated programmatically based on the tumor assessments performed by the central reviewer or investigator as applicable, and applicable additional CRF data.

The following sensitivity endpoints will be defined:

Time to Progression (TTP), a component of PFS, is defined as time from randomization until disease progression according to modified RECIST criteria for malignant pleural mesothelioma (mRECIST) per blinded central radiology review. Patients who died, were lost to follow-up, or were alive at data cut-off without radiological progression will be censored at the last tumor assessment. Censoring for TTP is described in Table 6-4. Windows for missed assessments are described in Table 6-3.

Table 6-4 Censoring for time to progression

Situation	End Date	Censored	Reason for Censoring
No baseline radiological tumor assessment	Date of Randomization	Yes	No baseline assessment
No post-baseline radiological assessment	Date of randomization	Yes	No post-baseline radiological assessment
Patient had a radiological assessment of PD following randomization (no 2 consecutive missing radiological assessments)	Date of first radiological progression event	No	N/A
PD immediately following two or more consecutive missed radiological assessments	Date of last radiological assessment before missed assessments	Yes	Missed two or more consecutive tumor assessments
Patient discontinued from study without radiological PD	Last radiological assessment date	Yes	Patient discontinued from study without PD
Death without prior progression	Last radiological assessment date	Yes	Patient died without PD
Patient still on study at the time of data cut-off without PD	Last radiological assessment before data cut-off	Yes	Patient still alive without PD

Progression-free survival censored for subsequent therapy (PFS1) is defined in the same manner as primary PFS, except that patients receiving non-protocol anti-cancer therapy prior to centrally confirmed radiological progression will be censored at the last tumor assessment prior to first non-protocol anti-cancer therapy.

Progression-free survival not censoring for missed assessments (PFS2) is defined in the same manner as primary PFS, except that events following 2 or more missed assessments are included, i.e. rule censoring events following 2 or more missed assessments does not apply.

Investigator progression-free survival (PFS3) is defined in the same manner as primary PFS, except based on investigator assessments.

Clinical progression-free survival per investigator (PFS4) is defined in the same manner as PFS3, except that clinical progression documented on the CRF will also be counted as an event.

Sensitivity endpoint analyses (excluding TTP) will be performed in the same manner as the primary endpoint analyses, using a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per time to progression on 1st line treatment (≥ 6 months versus < 6 months), with a 1-sided significance level of 0.0125. Sensitivity endpoints (including TTP) will be summarized using Kaplan-Meier estimates [5]. Plots will be produced by treatment arm. Medians and Brookmeyer-Crowley confidence intervals [6] with complementary log-log transformation [7] will be reported. will be reported. The hazard ratio (anetumab ravtansine regimen/vinorelbine) will be estimated using Cox proportional hazards models [8] with Wald confidence intervals [9] stratified by the same factors. For confidence intervals, both 95% 2-sided intervals and 97.5% 2-sided intervals will be calculated.

6.2.2 Secondary Efficacy Analysis

All secondary efficacy endpoints will be analyzed for the ITT analysis set.

6.2.2.1 Hierarchy and timing of analyses

The following secondary endpoints will be evaluated:

- Overall Survival (OS)
- Objective Response Rate (ORR)
- Disease Control Rate (DCR)
- Duration of Response (DOR)
- Durable Response Rate (DRR)
- Time to Worsening of Symptoms Characteristic of Mesothelioma (TTWS)
- Time to Worsening of Pain (TTWP)
- Improvement Rate of Symptoms Characteristic of Mesothelioma (SIR)
- Improvement Rate of Pain (PIR)

For drug label claim purposes, to preserve the overall Type I error rate, secondary endpoint hypothesis testing will be performed for the following key secondary endpoints, only in the event of primary endpoint superiority, according to the following hierarchy:

- Overall survival
- Time to Worsening of Symptoms Characteristic of Mesothelioma
- Time to Worsening of Pain

- Improvement Rate of Symptoms Characteristic of Mesothelioma
- Improvement Rate of Pain

A secondary endpoint may be claimed in the drug label only if the 1-sided p-values for this endpoint and all that are ranked before it are 0.025 or less. Secondary OS will be evaluated separately in a 2-stage group sequential procedure as further described in Section 6.2.2.2. Final analysis of DOR and DRR will occur at the same time as the final OS analysis. Final analysis for other secondary endpoints will occur at the same time as the primary endpoint analysis. However, to preserve the validity of the secondary endpoint hierarchy and control of overall study-wide secondary Type I error for regulatory label claim purposes, secondary endpoint hypothesis tests ranking below OS in the hierarchy tests will be applied after OS hypothesis testing. Accordingly, unless OS is found superior at the interim OS analysis, the hypothesis testing outcome of endpoints ranking below it will not be fully effective for regulatory label claim purposes, and any superiority for these endpoints cannot be declared for such purposes, until the outcome of the OS hypothesis test is determined and OS superiority found at the final OS analysis.

6.2.2.2 Overall Survival (OS)

Endpoint Definition

Overall survival (OS) is defined as time from randomization until death from any cause. Patients lost to follow-up or alive at the time of analysis will be censored at the last known alive date.

Censoring Rules

Censoring rules for overall survival are described in Table 6-5.

Hypotheses Tested

The secondary efficacy OS analysis tests the following hypotheses:

Null Hypothesis H_0: In patients with advanced or metastatic malignant pleural mesothelioma overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, OS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to OS under treatment with 30 mg/m^2 QW vinorelbine.

versus

Alternative Hypothesis H_A: In patients with advanced or metastatic malignant pleural mesothelioma overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, OS under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to OS under treatment with 30 mg/m^2 QW vinorelbine.

These hypotheses will be tested using a 2-stage group sequential hypothesis test procedure, with an overall significance level preserved at 0.025 one-sided. Each stage will utilize a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per TTP on 1st line treatment ($\geq$ 6 months versus < 6 months).

Table 6-5 Censoring for Overall Survival

Situation	End Date	Censored	Reason for Censoring
Death between randomization date and the date of analysis / database cut-off date (i.e. there is a death date and this date is before or on the date of analysis / database cut-off date)	Date of Death	No	N/A
Alive at the date of analysis / database cut-off date (i.e. there is a clear assessment that the patient was alive this day, e.g. from information collected on or after the date of analysis / database cut-off date)	Last known alive date (LKAD) prior to database cut-off	Yes	No Death
No information of survival status as of the data cut-off date used for an analysis database (e.g., lost to follow-up)	Last known alive date (LKAD) prior to database cut-off	Yes	No Death and no survival status
Lost to follow-up with no contact information at all after the randomization date	Date of randomization	Yes	Lost to follow-up after randomization date
If month and/or year are missing for death date	The nearest prior time point: Last known alive date (LKAD) prior to database cut-off	Yes	Missing month and/or year for death date

The interim OS analysis will be performed at the time of the final primary endpoint (PFS) analysis. If the study is not stopped for superiority at the interim analysis, the final OS analysis will occur after 159 OS events have been observed. A Lan-DeMets alpha spending function [10] with one-sided O'Brien-Fleming superiority boundary [11] will be used to determine the alpha level applicable to the interim and final analyses, based on the actual number of OS events at the time of the interim analysis, i.e. as of the database cut-off for the final PFS analysis. For this calculation, the information fraction for the interim OS analysis will be calculated as the actual number of OS events included in the final PFS analysis, divided by the per-protocol total number of OS events (159).

The alpha level for the final analysis will be determined by subtracting the interim analysis alpha level from the total overall significance level of 0.025.

In addition to the stratified log-rank test described above, OS will be summarized at each analysis using Kaplan-Meier estimates [5]. Plots will be produced by treatment arm. Medians and Brookmeyer-Crowley confidence intervals [6] with complementary log-log transformation [7] will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported. The hazard ratio (anetumab ravtansine regimen/vinorelbine) will be estimated using Cox proportional hazards models [8] with Wald confidence intervals [9] stratified by the same factors. For confidence intervals,

both 95% 2-sided CIs, and CIs adjusted for the group sequential design (2-sided confidence intervals at 100*(1-2*αi)%, where αi is the calculated 1-sided significance level for the applicable analysis) will be reported.

Assumption Checking

To assess the assumption of non-informative censoring, Kaplan-Meier [5] survival curves, and corresponding exponential regression hazard estimates by 3-month intervals will be calculated for censoring times, characterizing OS censorings as events (events as censorings) for purposes of this analysis. Similarity and proportionality of the censoring times will be evaluated. All primary analysis censorings, including censorings due to the missing assessment and other rules, will be included in this analysis.

Sensitivity Endpoints and Analyses

To evaluate proportionality of the hazards, hazards by time will be estimated in each arm using exponential regression modelsby 3-month intervals. Tables and plots of hazards and corresponding estimated survival functions, with interval model estimates overlaying Kaplan-Meier estimate [5] for each arm, will be produced.

A non-stratified log-rank test and a test stratified by stratification factor per CRF will be performed. In addition, stratification factors per IXRS and per CRF will be included in subgroup analyses.

6.2.2.3 Objective Response Rate (ORR)

Each patient's **best tumor response on-study** will be calculated based on central review tumor assessments and CRF data for non-protocol anti-cancer therapy. For this calculation assessments of CR, PR, and SD following PD, and/or following non-protocol anti-cancer therapy, will be excluded. A best response of CR requires confirmation by a 2nd subsequent CR assessment at least 4 weeks later with no assessments of PR, SD, or PD in between. A best response of PR requires confirmation by a 2nd subsequent CR or PR assessment at least 4 weeks later with no assessments of PD in between. Responses of SD must be recorded at least 5 weeks (6 weeks less 1-week per protocol window) following randomization. Responses will be ordered as follows: (CR, PR, SD, PD, UE).

Endpoint Definition

A patient is a **responder** if the patient has a confirmed best tumor response on-study of CR or PR, as determined by the central radiological reviewer per mRECIST criteria.

The **objective response rate (ORR)** in each arm is the number of responders divided by the number of randomized patients.

Summary statistics for objective response rates will be reported in each treatment arm with 95% 2-sided exact binomial confidence intervals. Summary statistics for best tumor response on-study with 95% 2-sided exact binomial confidence intervals will also be calculated.

As sensitivity analysis, the same statistics will be performed based on investigator assessments.

6.2.2.4 Disease Control Rate (DCR)

A patient has **disease control** if the patient has a best tumor response on-study (See Section 6.2.2.5) of CR, PR, or SD.

The disease control rate (DCR), defined as proportion of patients achieving CR, PR, or SD per mRECIST criteria, as determined by the central radiological reviewer. DCR is calculated in each arm as the number of patients with disease control divided by the number of randomized patients.

Summary statistics for disease control rates will be reported in each treatment arm with 95% 2-sided exact binomial confidence intervals. Summary statistics and 95% 2-sided exact binomial confidence intervals will also be reported for patients' best response observed on-study (confirmed CR, confirmed PR, SD, PD, UE).

As sensitivity analysis, the same statistics will be performed based on investigator assessments.

6.2.2.5 Duration of Response (DOR)

Duration of response (DOR) is defined in responders (See Section 6.2.2.3) as the time from central documentation of tumor response (date of first response in the confirmation sequence) to the earlier of disease progression as determined by the central radiological reviewer, or death without centrally documented progression.

Censoring rules for DOR are described in Table 6-6.

Duration of response will be summarized at both the interim and final analyses using Kaplan-Meier estimates [5]. Plots will be produced by treatment arm. Medians and Brookmeyer-Crowley confidence intervals [6] will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported.

Table 6-6 Censoring for duration of response

Situation	End Date	Censored	Reason for Censoring
Patient had a radiological assessment of PD following response (no 2 consecutive missing radiological assessments)	Date of first radiological progression event	No	N/A
Death or PD immediately following two or more consecutive missed radiological assessments	Date of last radiological assessment before missed assessments	Yes	Missed two or more consecutive tumor assessments
Patient discontinued from study without radiological PD or death	Last radiological assessment date	Yes	Patient discontinued from study without PD or death
Death during the study (no 2 consecutive missing radiological assessments)	Date of Death	No	N/A
Patient still on study at the time of data cut-off without PD	Last radiological assessment before data cut-off	Yes	Patient is still alive without PD

Sensitivity Endpoints and Analyses

The above statistics for duration of response will also be performed based on investigator assessments.

6.2.2.6 Durable Response Rate (DRR)

A **durable responder** is a responder (See Section 6.2.2.3 with duration of response of 180 days or more.

The **durable response rate (DRR)** in each arm is the number of durable responders divided by the number of randomized patients.

Summary statistics for durable response rates will be reported in each treatment arm with 95% 2-sided exact binomial confidence intervals.

Sensitivity Endpoints and Analyses

The above statistics for durable response rate will also be performed based on investigator assessments.

6.2.2.7 Time to worsening of symptoms characteristic of mesothelioma

Time to worsening of symptoms characteristic of mesothelioma will be evaluated based on the MD Anderson Symptom Inventory-Malignant Pleural Mesothelioma (MDASI-MPM), and is defined as follows:

Time to worsening of symptoms characteristic of mesothelioma is defined in patients **evaluable for assessing worsening of symptoms**, as the time from randomization until the first **worsening of symptoms** characteristic of mesothelioma. Worsening must be confirmed by a second MDASI-MPM assessment. Patients who died, were lost to follow-up, or ended MDASI-MPM assessments without confirmed worsening of symptoms will be censored at the date of their last MDASI-MPM assessment with a non-missing **composite symptom score CSS)**.

Confirmation of this endpoint requires two consecutive MDASI-MPM assessments with worsened symptoms, with no more than one missing CSS assessment in between.

An independent PRO review committee of PRO, statistical, and psychometric experts, blinded to study treatment, will conduct analyses based on pooled study data following database lock at primary analysis to support validation of the MDASI-MPM instrument and determine the key definitions and details described below. The PRO review committee and its analyses will be further described in the PRO review committee charter and/or the independent PRO review analysis plan.

For analyses of this endpoint, the following terms will be defined by the independent blinded PRO review committee:

- **Worsening of symptoms** characteristic of mesothelioma as measured by the MDASI-MPM questionnaire composite symptom score (CSS).
- **Composite Symptom Score (CSS)** applicable to worsening of symptoms
- **Subset of Symptoms** applicable to CSS
- **Evaluability** for assessing worsening of symptoms

The blinded PRO review committee will determine the specific change from baseline constituting worsening of symptoms characteristic of mesothelioma, and define the minimum and/or maximum baseline score and other criteria constituting evaluability to determine worsening.

The blinded PRO review committee will also define the applicable subset of MDASI-MPM items constituting the relevant subset of symptoms, and the calculation algorithm to calculate the CSS on that subset at each assessment, including handling of missing or partially missing data.

Censoring rules are described in Table 6-7.

Hypotheses Tested

The analysis tests the following hypotheses.

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, time to worsening of symptoms characteristic of mesothelioma under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to time to worsening of symptoms characteristic of mesothelioma under treatment with 30 mg/m^2 QW vinorelbine.

Versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, time to worsening of symptoms characteristic of mesothelioma under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to time to worsening of symptoms characteristic of mesothelioma under treatment with 30 mg/m^2 QW vinorelbine.

Table 6-7 Censoring for time to worsening of symptoms characteristic of mesothelioma

Situation	End Date	Censored	Reason for Censoring
No baseline CSS	Date of Randomization	Yes	No baseline assessment
No post-baseline CSS	Date of randomization	Yes	No post-baseline assessment
Patient had confirmed* worsening of symptoms (no 2 consecutive missing MDASI-MPM CSS assessments)	Date of first worsening of symptoms in confirmation sequence	No	N/A
Confirmed* worsening of symptoms immediately following two or more consecutive missed MDASI-MPM CSS assessments	Date of last non-missing MDASI-MPM CSS assessment following missing CSS assessments	Yes	Missed two or more consecutive assessment
Patient discontinued MDASI-MPM assessments without confirmed* worsening of symptoms	Date of last non-missing MDASI-MPM CSS assessment	Yes	Patient discontinued MDASI-MPM assessments without worsening
Death without prior confirmed* worsening	Date of last non-missing MDASI-MPM CSS assessment	Yes	Patient died without worsening
Patient still on study at the time of data cut-off without confirmed* worsening	Date of last non-missing MDASI-MPM CSS assessment	Yes	Patient still alive without worsening

*Confirmed worsening of symptoms requires 2 consecutive non-missing MDASI-MPM worsened CSS assessments separated by no more than 1 missing MDASI-MPM CSS assessment.

Analysis

These hypotheses will be tested using a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per time to progression on 1st line treatment (≥ 6 months versus < 6 months), with a 1-sided significance level of 0.025, at the same time as the primary endpoint analysis.

In addition to the stratified log-rank test described above, time to worsening of symptoms characteristic of mesothelioma will be summarized using Kaplan-Meier estimates [5]. Plots will be produced by treatment arm. Medians and Brookmeyer-Crowley confidence intervals [6] with complementary log-log transformation [7] will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported. The hazard ratio (anetumab ravtansine regimen/vinorelbine) will be estimated using Cox proportional hazards models [8] with Wald confidence intervals [9] stratified by the same factors. For analysis confidence intervals, 95% 2-sided intervals will be calculated.

Symptoms characteristic of mesothelioma will also be analyzed descriptively. Plots of mean CSS over time will be produced based on dosing-independent assessment points including screening and following randomization (every 6 weeks during the first 6 months (24 weeks) after the start of study treatment, every 9 weeks until the end of year 2 (105 weeks), and every 12 weeks thereafter).. For these plots, patients will be included in the analysis for a time point if their applicable assessment was taken within 1 week of target on or before 24 weeks, within 3 week through 105 weeks, and every 4 weeks thereafter. In addition, plots by scheduled visit per CRF will be provided. For both sets of plots, data will be plotted until the last time point where >6 patients/arm have a non-missing data in both arms. Plots will show 95% confidence intervals for the mean, and the number of patients each mean is based on.

Summary statistics will be provided, associated with each plot, for all timepoints plotted for both sets, including timepoints where < 6 patients/arm have non-missing data.

6.2.2.8 Time to worsening of pain

Time to worsening of pain will be evaluated based on the “pain at its worst” (worst pain score) item of MD Anderson Symptom Inventory-Malignant Pleural Mesothelioma (MDASI-MPM), and is defined as follows:

Time to worsening of pain is defined in patients **evaluable for assessing worsening of pain**, as time from randomization until the first **worsening of pain**. Worsening must be confirmed by a second MDASI-MPM assessment. Patients who died, were lost to follow-up, or ended MDASI-MPM assessments without confirmed worsening of pain will be censored at the date of their last MDASI-MPM assessment with a non-missing pain score.

Confirmation of this endpoint requires two consecutive MDASI-MPM assessments with worsened pain, with no more than one missing assessment in between.

An independent PRO review committee of PRO, statistical, and psychometric experts, blinded to study treatment, will conduct analyses based on pooled study data following database lock at primary analysis to support validation of the MDASI-MPM instrument and

determine the key definitions and details described below. The PRO review committee and its analyses will be further described in the PRO review committee charter and/or the independent PRO review analysis plan.

For analyses of this endpoint, the following terms will be defined by the independent blinded PRO review committee:

- Worsening of pain as measured by the MDASI-MPM questionnaire "pain at its worst" item
- Evaluability for assessing worsening of pain.

The blinded PRO review committee will determine the specific change from baseline in the "pain at its worst" MDASI-MPM item constituting worsening of pain, and define the minimum and/or maximum baseline score and other criteria constituting evaluability to determine worsening.

Censoring rules are described in Table 6-8.

Hypotheses Tested

The analysis tests the following hypotheses.

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, time to worsening of pain under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to time to worsening of pain under treatment with 30 mg/m^2 QW vinorelbine.

Versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, time to worsening of pain under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to time to worsening of pain under treatment with 30 mg/m^2 QW vinorelbine.

Analysis

These hypotheses will be tested using a 1-sided log-rank test, stratified by geographic region (RoW versus Asia) and per time to progression on 1st line treatment ($\geq$ 6 months versus < 6 months), with a 1-sided significance level of 0.025, at the same time as the primary endpoint analysis.

In addition to the stratified log-rank test described above, time to worsening of pain will be summarized using Kaplan-Meier estimates [5]. Plots will be produced by treatment arm. Medians and Brookmeyer-Crowley confidence intervals [6] with complementary log-log transformation [7] will be reported. The number of patients at risk, number of events, and number of censored patients per month will also be reported. The hazard ratio (anetumab ravtansine regimen/vinorelbine) will be estimated using Cox proportional hazards models [8] with Wald confidence intervals [9] stratified by the same factors. For analysis confidence intervals, 95% 2-sided intervals will be calculated.

Pain will also be analyzed descriptively. Plots of mean "pain at its worst" score will be produced based on dosing-independent assessment points including screening and following randomization (every 6 weeks during the first 6 months (24 weeks) after the start of study treatment, every 9 weeks until the end of year 2 (105 weeks), and every 12 weeks thereafter). For these plots, patients will be included in the analysis for a time point if their applicable assessment was taken within 1 weeks of target on or before 24 weeks, within 3 weeks through 105 weeks, and every 4 weeks thereafter. In addition, plots by visit per CRF will be provided. For both sets of plots, data will be plotted until the last time point where >6 patients/arm have a non-missing data in both arms. Plots will show 95% confidence intervals for the mean, and the number of patients each mean is based on.

Summary statistics will be provided, associated with each plot, for all timepoints plotted for both sets, including timepoints where < 6 patients/arm have non-missing data.

The MDASI-MPM worst pain score analysis will also be performed subgrouped by baseline worst pain score (<4 vs. ≥ 4).

Table 6-8 Censoring for time to worsening of pain

Situation	End Date	Censored	Reason for Censoring
No baseline "pain at its worst" score	Date of Randomization	Yes	No baseline assessment
No post-baseline "pain at its worst" score	Date of randomization	Yes	No post-baseline assessment
Patient had confirmed* worsening of pain (no 2 consecutive missing MDASI-MPM "pain at its worst" assessments)	Date of first worsening of pain in confirmation sequence	No	N/A
Confirmed* worsening of pain immediately following two or more consecutive missed MDASI-MPM "pain at its worst" assessments	Date of last non-missing MDASI-MPM "pain at its worst" assessment following missing pain assessments	Yes	Missed two or more consecutive assessments
Patient discontinued MDASI-MPM assessments without confirmed* worsening of pain	Date of last non-missing MDASI-MPM "pain at its worst" assessment	Yes	Patient discontinued MDASI-MPM assessments without worsening
Death without prior confirmed* worsening of pain	Date of last non-missing MDASI-MPM "pain at its worst" assessment	Yes	Patient died without worsening
Patient still on study at the time of data cut-off without confirmed* worsening of pain	Date of last non-missing MDASI-MPM "pain at its worst" assessment	Yes	Patient still alive without worsening

*Confirmed worsening of pain requires 2 consecutive non-missing MDASI-MPM "pain at its worst assessments separated by no more than 1 missing MDASI-MPM "pain at its worst" assessment.

6.2.2.9 Improvement rate of symptoms characteristic of mesothelioma

Improvement rate of symptoms characteristic of mesothelioma will be evaluated based on the MD Anderson Symptom Inventory-Malignant Pleural Mesothelioma (MDASI-MPM), as follows:

Improvement rate of symptoms characteristic of mesothelioma is defined in each randomized study arm as the number of patients with confirmed **improvement of symptoms** characteristic of mesothelioma, divided by the number of patients **evaluable for improvement** of symptoms characteristic of mesothelioma. Improvement in each patient must be confirmed by a second MDASI-MPM assessment.

Confirmation of this endpoint requires two consecutive MDASI-MPM assessments with improved symptoms, with no more than one missing CSS assessment in between.

An independent PRO review committee of PRO, statistical, and psychometric experts, blinded to study treatment, will conduct analyses based on pooled study data following database lock at primary analysis to support validation of the MDASI-MPM instrument and

determine the key definitions and details described below. The PRO review committee and its analyses will be further described in the PRO review committee charter and/or the independent PRO review analysis plan.

For analyses of this endpoint, the following terms will be defined by the independent blinded PRO review committee:

- **Improvement of symptoms characteristic of mesothelioma** as measured by the MDASI-MPM questionnaire composite symptom score (CSS).
- **Composite Symptom Score (CSS)** applicable to improvement of symptoms (Same as in Section 6.2.2.7 above)
- **Subset of Symptoms** applicable to CSS (Same as in Section 6.2.2.7 above)
- **Evaluability** for assessing improvement of symptoms.

The blinded PRO review committee will determine the specific change from baseline constituting improvement of symptoms characteristic of mesothelioma, and define the minimum and/or maximum baseline score and other criteria constituting evaluability to determine improvement.

The blinded PRO review committee will also define the applicable subset of MDASI-MPM items constituting the relevant subset of symptoms, and the calculation algorithm to calculate the CSS on that subset at each assessment, including handling of missing or partially missing data.

Hypotheses Tested

The analysis tests the following hypotheses.

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, improvement rate of symptoms characteristic of mesothelioma under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to improvement rate of symptoms characteristic of mesothelioma under treatment with 30 mg/m^2 QW vinorelbine.

Versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, improvement rate of symptoms characteristic of mesothelioma under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to improvement rate of symptoms characteristic of mesothelioma under treatment with 30 mg/m^2 QW vinorelbine.

Analysis

These hypotheses will be tested using a 1-sided Cochran-Mantel-Haenszel test [12], stratified by geographic region (RoW versus Asia) and per time to progression on 1st line treatment ($\geq$ 6 months versus < 6 months), with a 1-sided significance level of 0.025, at the same time as the primary endpoint analysis.

Summary statistics on improvement rate of symptoms characteristic of mesothelioma in each arm will be reported with 95% 2-sided exact binomial confidence intervals.

6.2.2.10 Improvement rate of pain

Improvement rate of pain will be evaluated based on the "pain at its worst" item of MD Anderson Symptom Inventory-Malignant Pleural Mesothelioma (MDASI-MPM), and is defined as follows:

Improvement rate of pain is defined in each randomized study arm as the number of patients with confirmed **improvement of pain**, divided by the number of **patients evaluable for improvement of pain**. Improvement in each patient must be confirmed by a second MDASI-MPM assessment.

Confirmation of this endpoint requires two consecutive MDASI-MPM assessments with improved pain, with no more than one missing assessment in between. An independent PRO review committee of PRO, statistical, and psychometric experts, blinded to study treatment, will conduct analyses based on pooled study data following database lock at primary analysis to support validation of the MDASI-MPM instrument and determine the key definitions and details described below. The PRO review committee and its analyses will be further described in the PRO review committee charter and/or the independent PRO review analysis plan.

For analyses of this endpoint, the following terms will be defined by the independent blinded PRO review committee:

- **Improvement of pain** as measured by the MDASI-MPM questionnaire "pain at its worst" item
- **Evaluability** for assessing improvement of pain.

The blinded PRO review committee will determine the specific change from baseline in the "pain at its worst" MDASI-MPM item constituting improvement of pain, and define the minimum and/or maximum baseline score and other criteria constituting evaluability to determine improvement.

Hypotheses Tested

The analysis tests the following hypotheses.

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, improvement rate of pain under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to improvement rate of pain under treatment with 30 mg/m^2 QW vinorelbine.

Versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, improvement rate of pain under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to improvement rate of pain under treatment with 30 mg/m^2 QW vinorelbine.

Analysis

These hypotheses will be tested using a 1-sided Cochran-Mantel-Haenszel test [12], stratified by geographic region (RoW versus Asia) and per time to progression on 1st line treatment (≥ 6 months versus < 6 months), with a 1-sided significance level of 0.025, at the same time as the primary endpoint analysis.

Summary statistics on improvement rate in each arm will be reported with 95% 2-sided exact binomial confidence intervals.

6.2.3 Other efficacy endpoint analysis

Other efficacy variables include:

- MDASI-MPM CSS
- MDASI-MPM Worst Pain Score
- Time-Averaged AUC MDASI-MPM CSS
- Time-Averaged AUC MDASI-MPM Pain Score
- MDASI-MPM symptom interference score
- LCSS-Meso total score
- LCSS-Meso 3 summary item score

6.2.3.1 Time-Averaged AUC MDASI-MPM CSS and Pain Score

Time-Averaged area under the curve (AUC) MDASI-MPM composite symptom score is defined in each subject with a baseline and at least 1 post-baseline MDASI-MPM Composite Symptom Score (CSS) (as defined in Section 6.2.2.7), as the time-averaged MDASI-MPM CSS over the period during which MDASI-MPM was assessed for that patient, after imputing intermediate missing assessments as described below.

Time-Averaged area under the curve (AUC) MDASI-MPM ain Score is defined in each subject with a baseline and at least 1 post-baseline MDASI-MPM worst pain score (as defined in Section 6.2.2.8, as the time-averaged MDASI-MPM worst pain score over the period during which MDASI-MPM was assessed for that patient, after imputing intermediate missing assessments as described below.

Calculation details for time-averaged AUC MDASI-MPM CSS and pain score

Time-averaged scores are calculated as follows:

$$\frac{\sum_{i=1}^{m}\left(Score_i + Score_{(i-1)}\right) \times \left(Date_i - Date_{(i-1)}\right)}{2 \times \left(Date_m - Date_0\right)}$$

Parameters are described in Table 6-9.

Table 6-9 Time-averaged score parameters

m	is the total number of MDASI-MPM score assessments for the subject (CSS or pain as applicable) including actual non-missing-value and interpolated intermediate missing-value protocol-scheduled assessments Note: In the event e.g. the subject's last pain score is missing but last CSS score is non-missing, or vice versa, m for CSS and pain scores may be different.
$Score_i$ for $0 \leq i \leq m$	is the subject ith-dated MDASI-MPM score (including interpolated intermediate missing-value scores) for CSS or pain as applicable, with baseline score as $Score_0$
$Date_0$	Is the subject's randomization date
$Date_i$ for $0 < i \leq m$	Is the date of the ith MDAS-MPM CSS or worst pain score as applicable (including interpolated intermediate missing-value scores) with date of subject's last applicable non-missing MDASI-MPM score as $Date_m$

For each subject, a protocol-scheduled MDASI-MPM assessment, whether or not missing, is intermediate if it is not the baseline assessment and not the subject's last assessment with a non-missing applicable score. If an intermediate MDASI-MPM assessment has a missing total symptom score, it is imputed by linear interpolation (with respect to date) from the proceeding and following non-missing (non-imputed) scores.

Hypotheses Tested

The analysis tests the following hypotheses.

Null Hypothesis H_0: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, the population mean time-averaged AUC CSS for symptoms characteristic of mesothelioma (respectively, time-averaged AUC pain score) under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is not superior to the mean time-averaged AUC CSS for symptoms characteristic of mesothelioma (respectively, time-averaged AUC pain score) under treatment with 30 mg/m^2 QW vinorelbine.

Versus

Alternative Hypothesis H_A: In patients with advanced or metastatic MPM overexpressing mesothelin and progressed on 1st line platinum/pemetrexed-based chemotherapy, the population mean time-averaged AUC CSS for symptoms characteristic of mesothelioma (respectively, time-averaged AUC pain score) under treatment with anetumab ravtansine at 6.5 mg/kg Q3W is superior to the mean ime-averaged AUC CSS for symptoms characteristic of mesothelioma (respectively, time-averaged AUC pain score) under treatment with 30 mg/m^2 QW vinorelbine.

Analysis

These hypotheses will be tested using a 1-sided t-test with a 1-sided significance level of 0.025, at the same time as the primary endpoint analysis. The normality assumption will be checked with the Shapiro-Wilk test [17] and homoskedasticiy will be tested with the Folded-F test [18]. If the data are not approximately normal, a Mann-Whitney-Wilcoxin test [19] will be used. If the data are normal but heteroskedastic, the Satterthwaite method will be used [18].

Summary statistics for time-averaged AUC CSS and pain score will be reported in each arm with 2-sided 95% (normal) confidence intervals.

The MDASI-MPM pain score will also be evaluated subgrouped by baseline worst pain score (<4 vs. ≥ 4).

6.2.3.2 MDASI-MPM symptom interference score and LCSS-MESO scores

Plots of mean MDASI-MPM symptom interference score, LCSS-Meso total score, and LCSS-Meso 3 summary item score will be produced over time, by treatment arm with 95% confidence intervals for the mean.

For these endpoints, plots over time will be produced by visit (per CRF) at every scheduled assessment visit (CxD1, CxD8, and CxD15 visits for the first 3 treatment cycles and CxD1 visit of each cycle thereafter). Plots will show 95% confidence intervals for the mean, and the number of patients each mean is based on. Data will be plotted until the last time point where ≥ 6 patients/arm have non-missing data in both arms.

Summary statistics will be provided, associated with each plot, for all timepoints plotted, including timepoints where < 6 patients/arm have non-missing data.

6.3 Pharmacokinetics/pharmacodynamics

The details of the PK modeling analysis will be described in a separate Modelling and Simulation (M&S) Analysis Plan and the results will be reported in a separate M&S Report.

Plasma concentration data for all analytes will be listed in the CSR.

6.4 Safety

No formal statistical tests will be done for the safety endpoints.

A safety analysis will be performed at the same time as the final primary endpoint analysis. This analysis will include treatment-emergent safety events as of database closure for primary endpoint completion, and will form the basis of the safety evaluation in the Clinical Study Report.

An updated analysis of safety events will be performed as of the final OS analysis. Results of this analysis will be reported separately in an update or an addendum for the OS analysis.

6.4.1 Extent of exposure

Extent of exposure will be summarized for the Safety set (SAF) by treatment group, using descriptive statistics.

- Duration of study treatment for each patient will be calculated in days, depending upon study treatment received, as follows:
  - Anetumab ravtansine: the date of the last dose of study treatment – date of the first dose of study treatment + 21
  - Vinorelbine: the date of the last dose of study treatment – date of the first dose of study treatment + 7
- Number of study treatment infusions received.
- Cumulative dose of study treatment will be calculated for each patients as the sum of all doses received (mg/kg anetumab ravtansine and mg/m^2 vinorelbine)
- Dose intensity of study treatment will be calculated for each patient as
- (cumulative dose)/(duration of study treatment) (mg/kg/day anetumab ravtansine and mg/m^2/day vinorelbine) Patients with dose modifications will be summarized by dose modification category (dose omission, infusion interruption, change in infusion rate, dose delay, dose reduction, or permanent discontinuation of study drug). In addition, descriptive statistics on number of dose modifications will be reported in patients experiencing at least one modification.

6.4.2 Adverse events

All adverse events will be coded according to the latest MedDRA Version as of database lock for the applicable analysis. The MedDRA version used will be documented in the CSR. The intensity of an AE will be documented using the NCI-CTCAE v4.03. Corneal epitheliopathy events will be graded according to the Bayer Grading System for corneal epitheliopathy, as described in Protocol Tables 7-5 and 7-6. Corneal epitheliopathy adverse events will be reported under the MedDRA classification category corneal disorders, and will also be reported separately classified according to Bayer Grading System classification.

Adverse Events of Special Interest

Adverse events of special interest for anetumab ravtansine are:

- Any grade corneal epitheliopathy

Treatment Period

The treatment period for this study, for purposes of safety analyses, extends from the initiation of study treatment until 30 days after the last administration of study treatment.

Pre-treatment AEs

Pre-treatment AEs will be defined as AEs that started and either stopped before the first dose of study treatment or continued after and did not worsen in intensity (i.e. increase in CTCAE toxicity grade or became serious) during the treatment period.

Treatment-emergent AEs (TEAEs)

All AEs starting or worsening within the treatment period will be considered TEAEs, for example:

- Events that started on or after the first dose and within the treatment period and are not a continuation of a pre-treatment event.
- Events that started before the first dose, and worsened on or after the first dose and on or before the end of the treatment period.

Post-treatment AEs

Post-treatment follow-up AEs will be defined as AEs that started or worsened after the treatment period.

An overall summary of AEs will be provided to present the number and percentage of patients with

- any pre-treatment AEs
- any TEAEs
- any study treatment related TEAEs
- any study procedure related TEAEs
- any serious TEAEs
- any serious study treatment related TEAEs
- any serious study procedure related TEAEs
- any CTCAE or Bayer Grade ≥3 TEAEs
- any TEAEs leading to discontinuation
- any TEAEs leading to dose interruption
- any TEAEs leading to dose reduction
- any infusion related reactions (IRR)
- any TEAEs leading to death
- any TEAEs of special interest

TEAEs will be summarized by MedDRA system organ class and preferred term. For each patient, multiple occurrences of the same event will be counted once within a system organ class and preferred term.

The same summaries will be repeated for study treatment related TEAE, study procedure related TEAEs, serious TEAEs, serious study treatment related TEAEs, CTCAE or Bayer Grade ≥3 TEAEs, TEAEs leading to discontinuation, and TEAEs leading to dose interruption, TEAEs leading to dose reduction, and TEAEs leading to death.

The maximum severity of the TEAEs will be summarized according to the NCI-CTCAE toxicity criteria, except for corneal epitheliopathy, whose severity will be summarized according to the Bayer Grading system. For each patient, multiple occurrences of the same event will be counted once at their maximum severity within a system organ class and preferred term. TEAEs will also be summarized by NCI-CTCAE toxicity or Bayer Grading grade (as applicable) and relationship to study medication. Corneal epitheliopathy will be summarized classified both according to MedDRA (as corneal disorders) and according to the Bayer Grading System. In addition, corneal disorder (per MedDRA) AEs other than corneal epitheliopathies (identified per Bayer Grading System) will be reported separately.

The analysis of TEAEs will be reported for the Safety set (SAF), by treatment group as treated.

IRR listings will include the IRR diagnosis and its associated symptoms.

AE Listings

Data listings will be produced for all AEs recorded in the study. Verbatim descriptions and coded terms will be listed for all AEs.

Serious adverse events (SAEs), deaths, AEs leading to discontinuation and AEs of special safety interest will each have a separate listing.

6.4.3 Deaths

Deaths reported during the study period will be tabulated by treatment group.

6.4.4 Clinical laboratory data

The following laboratory parameters will be summarized:

- Complete blood count : Hemoglobin, hematocrit, platelet count, white blood cell count (WBC) including differential: neutrophil, lymphocyte, monocyte, basophil, and eosinophil counts, red blood cell count (RBC)
- Electrolyte and chemistry panel: sodium, potassium, chloride, calcium (total, corrected, or ionized), phosphorus, glucose (fasting or random/unspecified), AST, ALT, gamma-glutamyl transferase (GGT), bilirubin (total and direct), ALP, uric acid, total protein, albumin, lipase, amylase, lactic dehydrogenase (LDH), blood urea nitrogen (BUN) or urea, and creatinine
- Coagulation panel: PTT or aPTT or PTT ratio, and PT or PT-INR or PT ratio.
- eGFR

Urinalysis and serum pregnancy test results will be reported through listings.

Where possible, worst grades for hematological and biochemical toxicities will be calculated according to CTCAE, version 4.03 based on laboratory measurements. Where clinical data is required by CTCAE criteria to provide a valid grade (e.g. fasting status for glucose), this requirement will be noted on the report, and the worst grade will be imputed.

Baseline laboratory grades will be determined based on the last non-missing (gradable) assessment taken between 28 days prior to first treatment and first treatment day (inclusive).

For graded labs, worst grade on study and a shift table (worst grade on study by worst grade at baseline) will be reported. The number of grades imputed due to missing clinical data, if applicable, will be reported. For ungraded labs with normal ranges, low/normal/high status will be reported as applicable.

Descriptive statistics (number of observations, mean, standard deviation, minimum, median and maximum values) will be presented, including mean change changefrom baseline by treatment arm. Box plots of mean and quartile values over time will be generated for laboratory parameters comprising adverse events of special interest (e.g. hematology, eGFR) to investigate trends over time and outliers in the data.

If more than one assessment occurred at any visit (i.e. repeat samples taken), the last valid (non-missing) value will be used in the summaries. Where clinical data is required for grading, assessments with both laboratory and clinical data non-missing will be selected over assessments with non-missing laboratory and missing clinical data. Unscheduled laboratory data will be listed but will not be included in the summary tables.

6.4.5 Slit-Lamp examination findings

Slit-lamp biomicroscopy findings will be classified and graded according to the Bayer Grading System, described in Protocol Tables 7-6 and 7-7. For each laterality, the number and percent of patients with each abnormality term and worst Bayer grade on-study will be reported by treatment arm. In addition, the number and percentage with the worst Bayer grade for any laterality will be reported.

6.4.6 Other ophthalmologic examination findings

The following ophthalmologic safety endpoints will be reported:

- Snellen acuity equivalent
- Intraocular pressure (mmHg)
- Schirmer test (mm/5 min)

For these endpoints, the baseline value for each patient and laterality is the latest-dated non-missing value taken between 28 days before first study treatment and first study treatment (inclusive).

Snellen acuity equivalent

For each patient and laterality, the worst (lowest) on-study Snellen acuity equivalent will be calculated by applying the Snellen fraction (e.g. 20/50 = 0.4). The percent worst change from baseline for any post-baseline assessment for that patient and laterality will be calculated.

Snellen acuity equivalent will be evaluated by three lateralities: left eye, right eye, and the eye with the best Snellen acuity equivalent at baseline.

Descriptive statistics on percent change from baseline will be reported for each laterality by arm (n, mean, std, median, interquartile range). For each arm and laterality, the number and percent of patients with a ≥50% and ≥75% decrease in Snellen acuity from baseline at any time on-study will be reported.

Intraocular Pressure (IOP)

For each patient and eye laterality, the highest and lowest intra-ocular pressure on-study will be determined. Summary statistics will be provided for worst change from baseline in intraocular pressure. The percent worst changes from baseline (greatest increase and decrease) will be calculated and summarized respectively). The number and percentage of patients who develop IOP > 21 mmHg at any time on study, and the number and percentage of patients with an increase of 10 mmHg or more from baseline at time on study, will also be reported.

Descriptive statistics on percent changes from baseline will be reported for each laterality by arm (n, mean, std, median, interquartile range). For each arm and laterality, the number and percent of patients with a ≥50% and ≥75% decrease (respectively increase) in intraocular pressure from baseline at any time on-study will be reported.

Schirmer Test

On-study Schirmer test findings will be reported through listings.

6.4.7 ECOG performance score

The baseline ECOG PS value for each patient is the latest-dated non-missing value taken between 28 days before first study treatment and first study treatment (inclusive).

A shift table will be reported with the number and percent of patients in each baseline and worst-on-study ECOG performance score category.

6.4.8 12-Lead ECG

Each patient's 12-lead ECG results will be classified by the investigator into one of the following categories: Normal or normal variant; abnormal, clinically insignificant findings; abnormal, clinically significant findings.

For each patient, the worst on-study findings will be determined. The number and percentage of patients with worst on-study findings will be reported by category. A shift table of baseline findings by worst on-study findings will be reported.

Other ECG parameters and findings will be reported through listings.

6.4.9 Cardiac function

Each patient's EchoCG or MUGA scan results will be classified by the investigator into one of the following categories: Normal or normal variant; abnormal, clinically insignificant findings; abnormal, clinically significant findings.

For each patient, the worst on-study findings will be determined. The number and percentage of patients with worst on-study findings will be reported by category. A shift table of baseline findings by worst on-study findings will be reported.

Summary statistics will be reported for percent left ejection fraction, by treatment arm.

6.4.10 Vital signs

Summary statistics on vital signs (weight, blood pressure, heart rate, respiratory rate, temperature) will be reported for each visit. Vitals signs collected at unscheduled visits will not be summarized but will be included in listings using both change from baseline and by visit.

6.4.11 Physical examinations

Not applicable

6.5 Other endpoints

6.5.1 Tumor assessments

Endpoints based on tumor assessments are described in Section 6.2 Efficacy. Tumor assessment data supporting such endpoints will be provided as listings. For each patient, assessment and lesion (where applicable), the listings will comprise for each time point:

- Pleural target lesion measurements (Date of procedure, measurement number, cut number, method of assessment, thickness (mm), reason not measured, other/specify)
- Non-pleural target lesions (Lesion number, anatomical description, assessment period, date of procedure, method of assessment, longest diameter or short axis, reason not measured, other/specify)
- Non-target lesions (Lesion number, anatomical location, anatomical description, assessment period, date of procedure, method of assessment, status, short axis [if present and lymph node], contributed to unequivocal progression, reason not assessed)
- New lesions (Lesion number, anatomical location, anatomical description, assessment period, date of procedure, method of assessment, status, short axis [if present and lymph node], reason not assessed)

6.5.2 Patient reported outcomes

6.5.2.1 Disease Symptoms

MDASI-MPM Data

The MDASI-MPM (See Protocol Appendix 16.5.1) asks patients to rate a list of symptoms (at the worst) and activities (amount of interference from symptoms) on a 0 (symptom not present or no interference) to 10 (symptom as bad as can be imagined or complete interference) numeric scale in the last 24 hours. The MDASI-MPM contains the 13 symptom items (those found to have the highest frequency and/or severity in patients with various cancers and treatment types) and 6 interference items of the validated MDASI-Core and an additional 6 symptom items developed by qualitative interviews of 20 patients with pleural mesothelioma and 6 patients with different types of cancer receiving anetumab ravtansine.

The MDASI-MPM will be used to assess the severity of multiple symptoms and the impact of symptoms on daily functioning and is based on the ITT population.

6.5.2.2 Health Related Quality of Life (HRQoL)

LCSS-Meso

The Lung Cancer Symptom Scale for Mesothelioma (LCSS-Meso) is designed as a site-specific measure of QoL, particularly for use in clinical trials. (Protocol Appendix 16.5.2; [14]; [15]). It evaluates 5 major symptoms associated with mesothelioma and their effect on overall symptomatic distress, functional activities, and global QoL. It captures in detail those dimensions most likely to be influenced by therapeutic interventions and evaluates other dimensions globally.

The patient questionnaire consists of visual analogue scales (VAS) (100 mm horizontal line). The patient is instructed to put a mark on the line to indicate intensity of response to the items in question (0 = lowest rating; 100 = highest rating). Completion of the patient scale takes approximately 8 min initially, including demonstration of the VAS and 3-5 min for subsequent administrations

HRQoL data from the LCSS-Meso, whether or not included in efficacy analysis endpoints, will be reported through listings. Data from the LCSS-Meso information sheet will also be listed. The analysis of HRQoL data will be based on QoL set.

6.5.3 Biomarkers

A listing will be provided showing the individual percentages of patients whose tumor cells express mesothelin at staining intensity of moderate and stronger in at least 30% of tumor cells. Exploratory biomarker analyses will be described in a separate document.

6.5.4 Immunogenicity

Summary statistics will be reported on number and percent of patients with ADAs and NABs.

Titers will be reported through listings.

7. Document history and changes in the planned statistical analysis

- Final Version 1.0 dated 30 September 2016
- Final Version 2.0 dated 20 April 2017
- Final Version 3.0 dated 9 May 2017

7.1 Examination of subgroups

Subgroup analyses will be conducted for the primary efficacy endpoints PFS and secondary efficacy endpoint OS on the ITT population. Subgroup analyses will be provided by randomized treatment arm. Descriptive statistics and hazard ratio estimates with 95% CI will be provided at least for the subgroups listed below, provided there is a sufficient number of events in total (at least 10) within the subgroup.

The hazard ratio (anetumab ravtansine regimen/vinorelbine) will be estimated using Cox proportional hazards models with Wald confidence intervals stratified by the same factors, and will be reported in tables and using forest plots. 95% 2-sided CIs will be used. In addition, the following adjusted confidence intervals will also be reported in tables and forest plots:

- For primary endpoint PFS, confidence intervals based on the 0.0125 primary PFS significance level, i.e. 97.5% confidence intervals.
- For secondary endpoint OS, confidence intervals adjusted for the group sequential design, i.e. 2-sided confidence intervals at 100*(1-2*αi)%, where αi is the calculated 1-sided significance level for the applicable analysis.

Subgroup analyses will be performed based on the following baseline categories:

- Age category (< 65, ≥65 years)
- Sex (Male, Female)
- Race (White, Black, Asian and Other)
- Ethnicity (Hispanic, not Hispanic)
- Prior bevacizumab therapy (yes, no)
- ECOG PS (0, 1)
- MDASI-MPM symptom Composite Symptom Score (CSS) (<4, ≥4)
- MDASI-MPM pain score (<4, ≥4)
- Slit Lamp Exam Bayer Grade (0 both eyes, > 0 any eye)

Randomization stratification factors:

- Region (Rest of World, Asia)
- Time to progression on first line treatment (≥ 6 months, < 6 months)

CRF stratification factor:

- Time to progression on first line treatment (≥ 6 months, < 6 months)

8. References

1. Byrne, MJ and Nowak, AK. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Annals of Oncology* 15: 257–260. (2004)
2. Labby, ZE et al. Optimization of response classification criteria for patients with malignant pleural mesothelioma. *J Thorac Oncol*. 7: 1728–1734 (2012)
3. Fleming, TR, Rothmann, MD, and Lu, HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol*. 27:2874-2880 (2009).
4. United States Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: Clinical trial endpoints for the approval of cancer drugs and biologics. May 2007.
5. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 53:457-81 (1958)
6. Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics. 38:29-41 (1982).
7. Collett, D. *Modelling Survival Data in Medical Research: Third Edition*. CRC Press. (2015) p. 29.
8. Cox DR, Oakes D. *Analysis of Survival Data*. New York: Chapman & Hall. (1984).
9. SAS Institute, Inc. *SAS/STAT 9.2 Users Guide*. SAS Institute, Inc. p. 4593 (2008)
10. Lan, KG and DeMets, KL. Discrete sequential boundaries for clinical trials. *Biometrika* 70:659-63 (1983)
11. O'Brien, PC and Fleming, TR. A multiple testing procedure for clinical trials. *Biometrics* 35:549-556 (1979).
12. Agresti, A. *Categorical Data Analysis*. Wiley, 227-30 (2012)
13. Lung Cancer Symptom Scale (LCSS) http: //www.lcss-ql.com/index.htm.
14. Hollen PJ, et al. Adapting the Lung Cancer Symptom Scale (LCSS) to mesothelioma: using the LCSS-Meso conceptual model for validation. *Cancer*. 2004;101:587-95 (2004).
15. Hollen PJ, et al. Measuring quality of life in patients with pleural mesothelioma using a modified version of the Lung Cancer Symptom Scale (LCSS): psychometric properties of the LCSS-Meso. *Support Care Cancer*. 14:11-21 (2006).
16. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons. (1980)

17. Shapiro, SS and Wilk, MB. An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611 (1965).
18. SAS Institute, Inc. *SAS/STAT 9.2 Users Guide*. SAS Institute, Inc. pp. 7408-7410 (2008)
19. Hettmansperger, TW and McKean, JW. *Robust nonparametric statistical methods, 2nd Edition*. Chapman & Hall/CRC Press. (2010).

9. Appendices

9.1 Imputation rules

9.1.1 Imputation rules for adverse event and non-study medication dates

Adverse event and non-study medication start and stop dates will not be imputed.

9.1.2 Imputation rules for systemic anti-cancer medication during treatment and follow-up

Dates for prior systemic anti-cancer medications will not be imputed.

Stop dates for systemic anti-cancer medications will not be imputed.

For systemic anti-cancer therapy start and stop dates during active follow-up,

Table 9-1 Imputation rules for systemic anti-cancer medications during follow-up

Imputation Rules for start and stop dates	
Start Date Scenario	**Imputation rule**
Partially Missing	Imputation of start date If partial date has day and month missing: Date will not be imputed.H. If partial date has missing day only: If the month and year are same as the year and month of last dose date, then the day of last dose will be assigned to the missing day. Otherwise, the first day of the month will be assigned to the missing day.
Completely Missing	Date is not imputed.

9.1.3 Imputation rules for last known alive date

The last known alive date (LKAD) is derived from the key data listed in Table 9-2 below. The last available date across all key data listed below will be picked as the LKAD by patient.

To determine the LKAD from the key data panels in Table 9-2 with partial missing dates:

- Impute the partial missing dates in each key data item according to the imputation rules below in Table 9-3,
- Pick the last available date, following imputations, as the LKAD for each patient.

Note: The imputation method applies only to dates imputed only as part of the LKAD calculation. If partial dates are imputed as part of the LKAD calculation but not imputed or imputed differently in analyses of the underlying data, LKAD may not match the underlying reported data date(s).

Table 9-2 Key Data

Data Item
Start Date of Treatment
End Date of Treatment
Demographics Patient End Date
Demographics Reference Start Date
Start Date of Patient Disposition Event*
Date/Time of Laboratory Specimen Collection
End Date/Time of Patient Visit
Start Date/Time of Patient Visit
Date/Time of Tumor Measurement
PRO Questionnaire Entry Date
Survival Assessment Contact (Alive) Date

*Note: Disposition events not indicating a patient contact, (LOST TO FOLLOW-UP) should not be used.

Table 9-3 Imputation rules for LKAD derivation

Scenario: Partially Missing	If partial date has day and month missing, date will not be imputed. If partial date has missing day only, then the first day of the month will be assigned to the missing day. C If the first of the month is before the *randomization date,* then the date will be imputed as randomization date.
Scenario: Completely Missing	No imputation.

9.1.4 Imputation rules for death date

Every effort should be made to resolve incomplete or missing dates during the course of the study (i.e. edit checks, data cleaning/monitoring etc.). However, in rare circumstances, missing parts of either date of last contact or the date of death may occur where an imputation algorithm has to be defined. In general the following rule should be followed: Missing month or year is not acceptable. Dates will be imputed as the day after LKAD if LKAD is in the same month as the death date, otherwise as first of the month. Details are described in Table 9-4 below.

Table 9-4 Imputation rules for death date derivation

Scenario: Partially Missing	A. If partial death date has day missing, and year and month are the same as LKAD year and month, then impute the missing day as the day after LKAD. B. If partial death date has day missing, and year and month are after the LKAD year and month, then impute the missing day as the first pf the month. C. If month and/or year are missing, death will be classified as missing and the patient will be censored according to the censoring rules.
Scenario: Completely Missing	No imputation.

9.1.5 Imputation rules for time-to-event endpoints

Unless otherwise specified, imputation rules for partial PFS and OS assessment dates will be as described in Table 9-5. Unless otherwise specified, other partial time-to-event endpoint assessment dates will not be imputed (calculated as missing).

Table 9-5 Imputation rules for time-to-event assessment date

Scenario: Partially Missing	If partial date in the specific time-to-event endpoint data has day and month missing, then the date will be classified as missing (not imputed) If partial date in the specific time-to-event endpoint data has missing day only, then the first day of the month will be assigned to the missing day. If first of the month is before the *[randomization date* then the date will be imputed as the *randomization date*..
Scenario: Completely Missing	No imputation.