STATISTICAL ANALYSIS PLAN

Current Title:	SHERLOC: A Phase 2 Study of MM-121 in Combination with Docetaxel versus Docetaxel Alone in Patients with Heregulin Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer"
Original Title:	A Phase 2 Study of MM-121 in Combination with Docetaxel or Pemetrexed versus Docetaxel or Pemetrexed Alone in Patients with Heregulin Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer
Author:	Ron Knickerbocker, Statistical Consultant
Date and Version	28 Aug 2018, Version 2.0

Sponsor:

merrimack
One Kendall Square
Suite B7201
Cambridge, MA 02139

Approval Page

Statistical Analysis Plan for Seribantumab (MM-121) in Patients with Heregulin Positive, Locally Advanced or Metastatic Non-Small Cell Lung Cancer

The undersigned approve of the plan.

Ron Knickerbocker, Ph.D
Statistical Consultant

27 Aug 2018
Date

J. Marc Pipas, MD.
Senior Medical Director

8/30/18
Date

Table of Contents

List of Tables

Abbreviations

Terms	Definition
AE	Adverse event
AESI	Adverse Event of Special Interest
CI	Confidence Interval
CLIA	Clinical Laboratory Improvement Amendments
CR	Complete Response
CRF	Case report form
CSR	Clinical study report
CTCAE	Common terminology criteria for adverse events
DMC	Data Monitoring Committee
DOR	Duration of Response
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
ECHO	Echocardiogram
eCRF	Electronic case report forms
HRG	Heregulin
IA	Interim Analysis
ICH	International conference on harmonization
ITT	Intent-to-Treat
IWRS	Interactive web response system
LVEF	Left ventricular ejection fraction
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat
msec	Milliseconds
MUGA	Multiple gated accession scan
NC	Not Computable
NCI	National Cancer Institute
NE	Not Evaluable
ORR	Objective response rate
OS	Overall Survival
PD	Progressive disease
PFS	Progression free survival
PK	Pharmacokinetic
PP	Per Protocol
PR	Partial response
PT	Preferred term
RECIST	Response Evaluation Criteria In Solid Tumors
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable Disease
SOC	System Organ Class
SRB	Safety Review Board
TFL	Tables, Figures, Listings
Wk	Week

1. Introduction

This is the Statistical Analysis Plan (SAP) for protocol MM-121-01-02-09 version 5.0 02FEB2018.

1.1. Protocol Changes and Clarifications

This is the first amendment after the previous signed off SAP version 16NOV2015, which was based on protocol MM-121-01-02-09 version 2.0. Changes related to statistical analysis from Protocol version 2.0 through version 5.0 are updated in this SAP.

1.2. Study Design

This study is a randomized, open-label, international, multi-center, Phase 2 study in patients with locally advanced or metastatic NSCLC, histologically classified as adenocarcinoma (or squamous cell carcinoma for patients enrolled prior to protocol version 4.0) that have progressed following one or two lines (or more than two lines for patients enrolled prior to protocol version 4.0) of systemic therapy for locally advanced and/or metastatic disease, of which one must have been a platinum containing regimen (for patients enrolled under protocol versions 4.0 and 5.0). The purpose of the study is to assess patients with HRG-positive tumors, to understand if the addition of MM-121 to docetaxel (or pemetrexed for patients enrolled prior to version 4.0) is more effective than docetaxel (or pemetrexed for patients enrolled prior to version 4.0) alone, in prolonging progression-free survival (PFS).

Eligible patients will be randomized in a 2:1 ratio (experimental arm versus comparator arm) using an Interactive Web Response System (IWRS). Randomization will be stratified based on number of prior systemic therapies for locally advanced and/or metastatic disease: 1 or 2) 1 or ≥2 for patients enrolled prior to protocol version 4.0). One line of therapy is equivalent to one prior systemic therapy. One line of therapy may consist of one or more drugs.

Patients will be randomized to Arm A or Arm B:

Arm A (Experimental Arm):

Safety Population

- Patients receiving MM-121: fixed dose of 3000 mg IV on day 1 of each 21-day cycle, in combination with docetaxel (75 mg/m^2 IV) or pemetrexed (500 mg/m^2 IV): on day 1 of each 21-day cycle (for all protocol versions.)

mITT Population

- Patients receiving MM-121: fixed dose of 3000 mg IV on day 1 of each 21-day cycle, in combination with docetaxel (75 mg/m^2 IV) on day 1 of each 21-day cycle (for protocol versions 4.0 & 5.0, including patients enrolled under earlier versions also meeting the inclusion/exclusion criteria of versions 4.0 and 5.0.)

Arm B (Comparator Arm):

Safety Population

- Patients receiving docetaxel (75 mg/m^2 IV) or pemetrexed (500 mg/m2 IV) on day 1 of each 21-day cycle (for all protocol versions.)

mITT Population

- Patients receiving docetaxel: (75 mg/m2 IV) on day 1 of each 21-day cycle (for protocol versions 4.0, 5.0, including patients enrolled under earlier versions also meeting the inclusion/exclusion criteria of versions 4.0 and 5.0.)

1.3. Planned Interim Analysis

A planned interim analysis on the mITT population will be conducted when approximately 75% of final PFS events (57 PFS events) have been observed in the primary efficacy population and will be tested using a one-sided significance level of 0.0089. This interim analysis plan controls the overall type I error at a one-sided 0.025 level. The purpose of the interim analysis is to assess the safety and efficacy at an early stage.

1.4. Study Objectives

1.4.1. Primary Objective

The primary objective of this study is to determine whether the combination of MM-121 and docetaxel is more effective than docetaxel alone based on investigator assessed Progression-Free Survival (PFS) according to RECIST 1.1 in HRG-positive patients (defined as HRG ISH score of ≥1+).

1.4.2. Secondary Objectives

- To determine whether the combination of MM-121 and docetaxel or pemetrexed is more effective than docetaxel or pemetrexed alone in HRG-positive NSCLC patients (defined as HRG ISH score of ≥ 1+) for the following clinical outcome parameters:
  - Overall Survival (OS)
  - Objective Response Rate (ORR) based on RECISTv1.1
  - Time to Progression (TTP)
- To describe the safety profile of MM-121 in combination with docetaxel
- To characterize the pharmacokinetic (PK) profile of MM-121 when given in combination with docetaxel and of docetaxel when given in combination with MM-121

1.4.3. Exploratory Objective

- To evaluate if mechanistically linked exploratory biomarkers from tumor tissue or blood samples correlate with clinical outcomes.

2. Analysis Sets

2.1. Modified Intent-to-Treat (mITT) Population

This population includes all randomized patients who received at least one dose of assigned therapy and met the inclusion/exclusion criteria of protocol version 4.0 and 5.0, also including patients enrolled under earlier versions meeting the following inclusion criteria of versions 4.0 and 5.0.

1. Diagnosed with NSCLC with adenocarcinoma histology

2. Received only 1-2 prior lines of therapy in metastatic setting
3. Received docetaxel (but not Pemetrexed) as monotherapy or in combination with seribantumab
4. Received one prior platinum-based regimen for management of primary or recurrent disease
5. HRG positive tissue using archived or fresh biopsy is required for enrollment.
6. Received nivolumab, pembrolizumab, or other anti-PD-1 or anti-PD-L1 therapy, where available and clinically indicated.

2.2. Per Protocol (PP) population

This population is a subset of the mITT. The PP population will only include patients that are confirmed HRG-positive, based on centralized analysis of tissue collected after progression on the most recent line of therapy (pre-treatment tissue sample). The PP population will exclude those patients with critical protocol violations. Primary considerations for exclusion will be those patients with no measurable baseline tumor, non-evaluated baseline non-target tumor assessments, and incomplete tumor evaluations. Not all circumstances will result in exclusion; however, a continual review of patient data will be performed to address data inconsistencies. Data will be reviewed without knowledge of progression/non-progression. This population will be used for a sensitivity analysis of the primary endpoint in the event that there is a substantial number of exclusions.

2.3. Safety (SAF) population

The safety population includes patients receiving at least one dose of study medication. Patients will be analyzed by the treatment received and not by the treatment to which they were randomized. All safety analyses will be performed on the mITT population or this population.

2.4. PK population

All treated patients with at least one PK assessment.

3. Primary and Secondary Endpoints

3.1. Progression-Free Survival (PFS) – Primary Endpoint

The investigator will make tumor assessments. All PFS calculations (months) will be calculated using the following formula:

$$\frac{\left[\left(\frac{Progression}{\frac{Death}{Censor\ Date}}\right) - Randomization\ Date + 1\right]}{30.4375}$$

PFS duration is defined as the time from the date of randomization to the date of death or first documented (objective) disease progression using RECISTv1.1, whichever occurs earliest. Patients that do not experience progression or death at the time of analysis will have time to progression censored at the date of last valid tumor assessment. Tumor assessments are to be performed every 6 weeks (± 7 days) from day of first dose until treatment discontinuation.

Censorship: In general, patients are censored at the date of the last objective disease assessment that verified lack of disease progression if they are last known to be alive, on-treatment or within 30 days following treatment discontinuation, and progression-free. For patients who die without assessment, a wider window of 60 days will be used. Censoring is explicitly described below in Table 1 and in Table 4.

- Patients with inadequate/incomplete baseline disease assessment will be censored at the date of randomization
- Patients not reassessed after randomization will be censored at the date of randomization, unless death occurred within the time window for the first two assessment visits, provided the patient is still on-treatment, or ≤ 60 days after treatment discontinuation
- Patients with at least one on-study disease assessment who discontinue treatment without documented disease progression or death will be censored at the date of the last objective disease assessment documenting no progression

There are three exceptions:

- If objective progression or death is documented ≤ 30 days after treatment discontinuation, then progression or death is an event
- If death is documented ≤ 60 days after treatment discontinuation in a patient with no on-study disease assessment, then death is an event
- If a new anti-cancer treatment is started prior to objective progression and ≤ 30 days after discontinuation, then censorship will be at the date of the last objective disease assessment that verified lack of disease progression prior to the new treatment unless objective progression occurs within 30 days after discontinuation

Patients with documentation of progression or death after an unacceptably long interval (i.e., 2 or more missed or indeterminate/not evaluable assessments) since the last tumor assessment will be censored at the time of last objective assessment prior to the missed/indeterminate/not evaluable assessments.

Table 1: Primary Endpoint Evaluation of Documented Disease Progression and Censoring

Situation	Date of	Outcome
Inadequate baseline tumor assessment	Randomization	Censored
No post-baseline assessment and no death prior to first scheduled assessment or within 60 days of treatment discontinuation		
Death or documented PD after ≥ 2 consecutively missed and/or not evaluable scheduled tumor assessments	Last tumor assessment documenting no PD (prior to the missed/not evaluable assessments)	
Alive, on-treatment, and no documented PD	Last tumor assessment documenting no PD	

Documented PD (on, prior to, or ≤ 30 days after treatment discontinuation)	First tumor assessment documenting PD	Progression (Event)
New anticancer treatment prior to PD	Last tumor assessment documenting no PD prior to new anticancer treatment (unless if within 30 days of treatment discontinuation in which case it is an event on date of documented PD per rule above)	Censored
Treatment discontinuation due to toxicity, undocumented progression, or other reason and no documented PD	Last tumor assessment documenting no PD prior to discontinuation	Censored
Treatment discontinuation due to toxicity, undocumented progression, or other reason and PD documented ≤ 30 days after treatment discontinuation	First tumor assessment documenting PD	Progression (Event)
No post-randomization scans, death ≤ 60 days after treatment discontinuation	Death	Death (Event)
Death without documented PD and ≤ 30 days after treatment discontinuation or while on treatment	Death	Death (Event)

Note: Tumor assessments are planned every 6 weeks, with a window of +/-7 days for each assessment. Thus, a patient will be considered to have missed ≥2 consecutive assessments if the time between two evaluable assessments is > 14 weeks (98 days). Follow-up scans can be performed up to 4 weeks from last dose date according to the protocol.

3.2. Overall Survival (OS) – Secondary Endpoint

OS is defined as the time from randomization to the date of death from any cause. In the absence of confirmation of death, survival time will be censored at the last date the patient is known to be alive. Patients with no post-randomization information will be censored on the date of randomization. A sensitivity analysis will be performed censoring patients who leave the study at the date of discontinuation.

OS calculations (months) will be the following:

$$\frac{\left(\frac{Death}{Censor\ Date} - Randomization\ Date + 1\right)}{30.4375}$$

3.3. Objective Response Rate (ORR) – Secondary Endpoint

Best Overall Response (BOR) is recorded from randomization until documented disease progression or death. ORR is defined as the proportion of patients with a BOR characterized as either a Complete Response (CR) or Partial Response (PR), as defined according to RECISTv1.1 guidelines, relative to the total number of evaluable patients. Only patients with measurable disease at baseline will be included in the analysis of the objective response. In accordance with RECISTv1.1 guidelines, patients with documented PR or CR will not need confirmatory tumor assessments to document tumor response, although the subset of PRs or CRs that are confirmed will also be noted in the summaries. Patients without baseline or post-baseline tumor assessments will be considered non-evaluable for this analysis.

The following table describes the responder/non-responders.

Table 2: ORR Determination

Best Overall Response	Reason for Treatment Discontinuation	Responder/Non-responder
Not Evaluable/None	Due to PD, symptomatic deterioration, toxicity, death, or new anticancer treatment	Non-responder
PD, SD	Any	Non-responder
PR, CR	Any	Responder
SD	None	Non-responder
Not Evaluable/None	None and > 9 weeks from 1st dose	Non-responder
Not Evaluable/None	None and ≤ 9 weeks from 1st dose	Not included in analysis

3.4. Duration of Response (DoR)

Duration of response is defined as the time from the date of first documented response (CR or PR) to the earliest date of disease progression or death due to any cause, whichever occurs first. Censorship will be determined using the primary endpoint definitions in Section 3.1. DoR will be calculated for the subgroup of responder patients.

$$\frac{\left[\left(\frac{Progression}{Death}\right) - Date\ of\ First\ CR/PR + 1\right]}{30.4375}$$

3.5. Time to Progression (TTP)

TTP is defined as the time from the date of randomization to the date of objective tumor progression. Those patients without objective tumor progression will be censored at the date of last tumor assessment documenting no objective progression. Patients who died prior to first scan will be censored on the date of death.

3.6. Safety Endpoints

Safety profile will be characterized by type, frequency, and severity of adverse events, coded using the MedDRA coding system, as graded by NCI Common Toxicity Criteria for Adverse Events version 4.03 (NCI CTCAE v4.03), timing and relationship to treatment, laboratory abnormalities, vital signs, and left ventricular imaging data.

3.6.1. Treatment-Emergent Adverse Events (TEAE)

Adverse events will be considered treatment-emergent if:

- The event occurs for the first time after the start of study treatment and on or before 30 days after final dose of study treatment and was not observed prior to start of treatment
- The event was observed prior to the start of treatment but increased in NCI CTCAE v4.03 grade during study treatment

The eCRF will capture TEAEs on the adverse event page, while those events occurring pre-treatment will be captured on the medical history page.

AEs that start on the same day as the first dose of study drug will be considered treatment emergent.

Additionally, a subset of events noted as adverse events of special interest (AESI) will be summarized including pulmonary embolism, deep vein thrombosis, and infusion related

reactions will also be summarized. A detailed list of the preferred terms to be included will be documented separately prior to the final database lock for the study.

3.6.2. Laboratory, ECG, MUGA/ECHO, and Vital Signs Data

Assessments will be assigned to cycles based on the collection date of the sample relative to the start dates of the cycles from the study drug administration page.

3.7. Stratification Factors

Randomization is stratified based on number of prior systemic therapies for locally advanced and/or metastatic disease (1 or 2). The primary analysis will include this stratification factor in the log-rank test and the estimation of hazard ratio.

4. Handling of Missing Data

4.1. General Conventions

Dates will be presented in listings as DDMONYYYY. Listings will be presented as the collected or raw data with missing information presented as such: - - MONYYYY for missing day, DD - - - YYYY for missing month, and DDMON - - - - for missing year. Partial start dates will be imputed as the earliest possible, while partial stop dates will be imputed as the latest possible based on the available information. Imputation of start / stop dates for AEs will follow the general rule, with the additional condition that if the available information on the start / stop dates is compatible with the AE being treatment emergent, then the imputation will reflect this. A similar approach will be taken for medications (i.e. these will be considered concomitant wherever possible based on partial dates).

4.2. Safety Data

Missing adverse event, vital sign, MUGA or ECHO, ECG, or laboratory data will be considered missing in summary analyses.

4.3. Pharmacokinetic Data

PK data will be analyzed according to a separate PK SAP.

5. Statistical Methods

5.1. Statistical Hypothesis and Sample Size Estimate

The primary hypothesis is to test for improvement of experimental drug MM-121 in progression-free survival (PFS), i.e.,

- H_0: Hazard Ratio MM-121 / Control = 1

  vs

- H_1: Hazard Ratio MM-121 / Control < 1

Approximately 76 PFS events from ~100 eligible enrolled patients are required to have at least 80% power to detect a 3-month improvement in median PFS over 3 months when comparing patients receiving the MM-121 combination (Arm A) versus Comparator Arm (Arm B, i.e., hazard ratio ≤ 0.50), using a one-sided, stratified log-rank test at a significance level of 0.0225 with patients randomized in a 2:1 ratio. An interim analysis is planned after 57 events (3/4 of the planned 76 events) and will be tested at a one-sided 0.0089 level. This interim plan maintains

the overall study type I error rate of 0.025 (one-sided). The hazard ratio will be estimated using a Cox regression model including treatment and the stratification factor as covariates. If the number of events that have occurred in the database interim lock is different than 57, then these significance levels will be modified using the prespecified spending function (see Section 5.3)

5.2. General Considerations

Every effort will be made to maintain the data integrity of the study, including not to look at the aggregated efficacy data before database lock. All analyses will be done using SAS version 9.4 or higher.

5.3. Interim Analysis

A planned interim analysis will be conducted when approximately 75% of final PFS events (57 PFS events) have been observed in the primary efficacy population. The purpose of the interim analysis is to assess the potential efficacy at an early stage.

Analyses will be focused on the primary, key secondary efficacy endpoints and safety endpoints.

The primary endpoint PFS will be analyzed with methods specified in 5.6.1. The one-sided P-value will be compared against significance level 0.0089 according to Hwang-Shih-DeCani α-spending function ($\gamma = -4$). If the number of events that have occurred in the database interim lock is different than 57, then these significance levels for the interim and final analysis will be modified using the true information time in the spending function.

Appendix 7.4 provides further detail on these significance levels as well as the planned analyses to be conducted as part of the interim analysis.

5.4. Standards

The information presented in this section describes the standard displays/mocks and analysis to be performed. Standard table displays will include columns for treatment groups called: Arm A and Arm B. The group titles can be abbreviated as Arm A: MM-121, Arm B: Control, and Overall. Abbreviations should be provided in the footnote.

Continuous variables, as deemed appropriate, will be summarized using the number of non-missing observations, mean, standard deviation, median, minimum and maximum values. Missing data values will be treated as missing in summarization. No imputation of missing data will be performed.

Categorical variables will be summarized using the number of non-missing observations, frequency, and percentage out of the number of non-missing observations in each category. Percentages will be presented to one decimal place (e.g., xx.x). Percentages will not be presented when zero counts occur.

Final p-values will be presented for the primary analysis as 1-sided, measured against significance level 0.0225. All secondary analyses will be tested against a one-sided, alpha at 0.025.

5.4.1. Data Cut-off

Two database cut-off dates for primary efficacy analyses will be determined at the time of the confirmed number of PFS events: 57 for the interim analysis (IA) and at least 76 for the final

analysis. All data will be cleaned as much as possible up to cut-off dates, although further data entry is allowed during the cleaning period.

5.4.2. Analysis of Time-to-Event Endpoints

Time-to-event endpoints between two treatment arms will be compared with a stratified log-rank test. Region is the only stratification factor. Cox proportional hazards models will be constructed with stratification factor and treatment in the model. The estimated hazard ratio and 2-sided confidence interval (CI) will be provided from the respective Cox proportional hazards model. For the primary endpoint (PFS), the significance levels will be 0.0089 at interim and 0.0225 at final, both one-sided. The CI for the hazard ratio is adjusted accordingly (two-sided 95.5%).

Kaplan-Meier methods will be used in estimating the probability of event and displayed graphically. Time-to-event calculations will be performed in months unless specified otherwise. Median event times and 2-sided 95% CI will be provided based on Brookmeyer-Crowley methods. Probability of event estimates at 3, 6, 9, and 12-month intervals will be reported in tables.

5.4.3. Analysis of Binary Endpoints

Frequencies and percentages will be reported for binary endpoints. The rate of binary endpoints between 2 treatment arms will be compared using a Cochran Mantel-Haenszel (CMH) test stratified according to the stratification factor and a Pearson χ^2 test, both at a 2-sided, $\alpha = 0.05$ level. The relative risk ratio estimate will be used to contrast treatment effects on response rates. Point estimates of the rates for each treatment arm will be provided along with the corresponding exact 2-sided 95% CI using exact methods based on the binomial distribution.

5.4.4. Analysis of Continuous Endpoints

Mean, standard deviation, median, minimum, maximum will be reported for continuous endpoints. Least-squares mean and standard error with 95% confidence intervals will be reported for models. Models will be tested using a 2-sided, $\alpha = 0.05$ level.

5.4.5. Baseline Definition

For efficacy endpoints, baseline is defined as the last assessment prior to randomization. According to the protocol, screening disease evaluation procedures have to be completed within 28 days prior to the first dose of study drug.

For safety endpoints, baseline is defined as the last assessment prior to first dose of study drug, provided that it is within the 28 days prior to the first dose.

5.5. Baseline Analysis

Baseline analysis will be performed on the mITT and safety population.

5.5.1. Disposition, Populations, and Patients Excluded from Population

Treatment termination and study termination will be summarized overall and by reason. Included in the summarization will be deaths, defined as overall number of deaths, deaths on-treatment (up to 30 days post last dose date), deaths after treatment.

A listing of patients and those excluded from a population along with reason will be provided, if available.

Protocol deviations will be listed, and major protocol deviations will be summarized.

5.5.2. Demographics, Cancer Diagnosis, Medical History

Descriptive statistics will be reported where available for Age (years), sex (male, female), race (white/Caucasian, Black or African American, Alaska native or American Indian, Asian/Oriental, Native Hawaiian or Pacific Islander, other), height (cm), weight (kg), ethnicity, and ECOG at baseline. Other cancer diagnosis factors such as frequency of adenocarcinoma, squamous cell; TNM Stage; Bone disease, brain metastasis, visceral metastasis, number of prior chemotherapies, KRAS status, best response will be reported. Time from cytological or histo-pathologic diagnosis to randomization will be analyzed as a continuous parameter.

The stratification factor will be summarized by frequencies and percentages. Frequency of number of prior systemic treatments for locally advanced and/or metastatic disease (1, 2) will be summarized.

Prior surgical therapy (yes/no), prior radiation therapy (yes/no) will be characterized.

Medical history will be coded using the latest version of MedDRA coding. Medical history will be summarized by System Organ Class and preferred term.

5.5.3. Extent of Disease

Number of patients with target lesions only, non-target lesions only, and both target and non-target lesions will be summarized. Individual disease sites for target or non-target lesions will be summarized.

5.5.4. Prior Cancer Treatment

Medical review will determine the assignment of prior anti-cancer treatment categories and regimen. The final list will be maintained in the TMF and signed by the study Medical Director. The following will be summarized categorically unless stated as a continuous parameter:

- Type of prior anti-cancer treatment
- Number of lines of prior anti-cancer treatment
- Time since last anti-cancer treatment prior to study enrollment will be summarized as a continuous parameter

5.6. Efficacy Analysis

Efficacy analysis for tumor assessment related endpoints (PFS, ORR, DoR, TTP) will be done on the mITT population unless stated otherwise.

5.6.1. Primary Efficacy PFS Analysis on mITT population

PFS as determined by Investigator assessment will be analyzed using a stratified log-rank test on mITT population. A stratified Cox proportional hazard models will be used to calculate hazard ratio and its confidence interval. The stratification factor will be number of prior systemic therapies for locally advanced and/or metastatic disease (1 or 2).

5.6.2. PFS (Sensitivity) Analysis on PP population and safety population

PFS as determined by Investigator assessment will be analyzed using a stratified log-rank test on above mentioned populations. A stratified Cox proportional hazard models will be used to

calculate hazard ratio and its confidence interval. The stratification factor will be number of prior systemic therapies for locally advanced and/or metastatic disease (1 or 2).

Additionally, if scans are reviewed centrally by an Independent Review Committee then the primary analysis and relevant sensitivity and secondary analyses will be repeated using this assessment.

5.6.3. Subgroup Analysis on Primary Efficacy Endpoint PFS

The subgroup analysis on primary efficacy endpoints PFS will be performed on the subgroups defined by the following baseline variables. Forest plots of the hazard ratio estimates, and confidence intervals will be displayed.

- Age (< 65 vs ≥ 65)
- Sex
- Region (North America *vs* the rest)
- Received nivolumab or other approved anti-PD-1 or anti-PD-L1 therapy *vs* none
- Number of prior systemic therapies (1 *vs* 2)
- NSCLC disease stage (IV *vs* III)
- ECOG PS (1 *vs* 2)
- Heregulin results from fresh tissue (positive vs negative vs missing/insufficient/not evaluable)
- Baseline number of metastatic sites

Additional relevant subgroups in addition to these may be summarized for PFS in the safety population (eg, adenocarcinoma vs. squamous).

5.6.4. DCR, OS, ORR, DOR Secondary Efficacy Analysis

OS and DOR will be analyzed using stratified and unstratified log-rank tests. Forest plots of the hazard ratio estimates, and confidence intervals will be displayed for OS and DoR Cox proportional hazards models with covariates and biomarker factors in the model. Other analyses, as specified for time-to-event endpoints in section 5.4 earlier, will also be performed.

Frequencies and percentages of DCR and ORR responders will be summarized. CMH and Pearson's chi-square test will be performed for responders.

All secondary analyses will be performed on mITT population except OS and ORR, which will be performed on both mITT and Safety populations.

OS Analysis may be updated and followed beyond the timing of the primary analysis of PFS..

5.6.5. Exploratory Biomarker Analysis

All exploratory biomarker analysis will be performed post-hoc.

5.7. Safety Analysis

All safety analysis will be performed using the mITT population. All safety tables will be based on comparison between Arm A *vs* Arm B.
Selected key safety analyses will also be reported on Safety Population.

5.7.1. Study Drug Exposure

Extent of treatment exposure will be summarized for MM-121 and docetaxel as follows:

- Number and percentage of patients beginning 1, 2, 3,4,5 and 6 and > 6 cycles
- Number of cycles started will be treated as a continuous variable
- Actual Dose received defined as total actual dose (mg)
- Relative Actual Dose = [Total actual dose (mg) divided by Total intended dose (mg)] x 100, where total intended dose is defined as the dose as prescribed which is summed from first to last intake date irrespective of missing, interruption, or reduction.
- Total Exposure (weeks) derived as (last dose date – first dose date + 1)/7
- Dose intensity (mg/wk) will be summarized as Actual Dose received divided by (duration in days multiplied by 7)

Treatment Delays and Dose Modifications

Definitions:

A **dose reduction** is the actual dose being less than the dose specified per protocol on any given day for any reason with the exception that a day with total dose administered of 0 mg is not considered a dose reduction.

A **dose interruptions/delay/missed dose** is defined as a planned dosing day with 0 mg administered

Patients will be summarized by frequency and percentage according to:

- Patients with at least one dose reduction overall and by cycle
- Patients with at least one dose reduction due to an adverse event
- Total number of dose reductions per patient
- Number of dose interruptions/missed overall and by cycle

5.7.2. Adverse Events

Adverse events will be coded using the MedDRA dictionary (Version 19.1) at the time of coding. NCI CTCAE v4.0 will be used to grade severity. Frequencies and percentages will be reported by System Organ Class (SOC) and preferred term (PT). Patients with multiple events for a category will be counted once in the highest CTCAE classification. Summary tables will include treatment-emergent AEs, unless specified otherwise. Treatment-emergent adverse events (TEAEs) will be listed by the event onset date relative to first dose of study drug in the study. Missing onset date for an event will be considered as TEAE. Adverse events with a starting date 30 or more days after the last dose date are not considered TEAEs. Relatedness will be considered: Not related, Possibly, Unlikely, or Definitely related using the case report form classification. If relatedness is missing, the adverse event will be considered TEAE. The order of reporting the events in summary tables will be according to the following, where applicable:

- SOC will be alphabetically reported
- Descending order of frequencies according to the number of patients in the MM-121 treatment group experiencing an adverse event
- PT will be alphabetically reported

The following tables will be reported:

- TEAE
- Any TEAE by SOC
    - Any TEAE by SOC and PT*
    - Any TEAE by PT
    - Any TEAE by PT by NCI CTCAE Grade
    - Any TEAE by PT by cycle
    - Any TEAE by PT by NCI CTCAE Grade by cycle
- Related TEAE
    - Any TEAE by SOC and PT*
    - Any TEAE by PT
    - Any TEAE by PT by NCI CTCAE Grade
- TEAE - NCI CTCAE Grade 3 or higher
    - Any TEAE by SOC and PT
    - Any TEAE by PT
- Related TEAE - NCI CTCAE Grade 3 or higher
    - Any TEAE by SOC and PT
    - Any TEAE by PT
- SAE – TEAEs only
    - Any SAE by SOC and PT*
    - Any SAE by PT
- AESI – TEAEs only
    - Any AESI by SOC and PT*
    - Any AESI by PT
    - Any AESI by PT by cycle
    - Any AESI by PT by NCI CTCAE Grade
- TEAE - Various
    - Any TEAE Resulting in Dose Reduction
    - Any TEAE Resulting in Treatment Discontinuation
    - Any TEAE Resulting in Death

In the AE preferred term summary tables, a patient having multiple events within a System Organ Class is counted only once for that SOC. Similarly, a patient having multiple events under a preferred term is counted only once for that PT. The part I safety analysis will include tables

indicated with an asterisk (*) above. If data dictates, additional tables may be generated. Listings will be provided for all adverse event data.

5.7.3. Laboratory Evaluation

Laboratory analytes will be summarized according to panel (i.e., hematology, chemistry, coagulation).

Shift in laboratory analyte grade will be characterized by worst from baseline and at end of treatment. Shift categories will include normal grade 1 to grade 5 and totals. Percentages will be based on the number with baseline and postbaseline values for that particular analyte.

5.7.4. Vital Signs

Vital sign parameters: systolic and diastolic blood pressure (mmHg), respiratory rate (breaths per minute), body temperature (C°), weight (kg) and pulse rate (beats per minute) will be summarized by observed and change from baseline according to minimum and maximum values. Additionally, the frequency and percentage of patients for each observed parameter outside the normal range during treatment (up to 30 days post last dose date) will be reported.

Table 3: Vital Signs Normal Range (Resting)

Parameter	Units	Normal Range
Systolic Blood Pressure	mmHg	80 - 130
Diastolic Blood Pressure	mmHg	60 - 90
Pulse Rate	Beats per minute	60 - 100
Respiratory Rate	Breaths per minute	12 – 25

5.7.5. Electrocardiogram

The number and percentage of patients with worst ECG result as order: "abnormal", "clinically significant"; "abnormal, not clinically significant"; "normal" will be summarized by treatment at scheduled visits. Results will be counted once for each patient and scheduled visit.

Descriptive summaries of observed and change from baseline values will be presented for ECG measures of HR, PR interval, QRS duration, QT interval, and QTc interval by scheduled visit. QTc interval is derived using Fridericia's correction method. These summaries will be presented by scheduled visit.

Prolongation of QTc interval [as defined by International Conference on Harmonization] for each on-treatment ECG reading will be summarized by number and percentage with a maximum QTc value, as well as maximum increases from baseline according to the following:

- Observed
  - $450 \leq$ Maximum QTc (msec) < 480
  - $480 \leq$ Maximum QTc (msec) < 500
  - $500 \leq$ Maximum QTc (msec)
- Increase from Baseline
  - $30 \leq$ Maximum QTc (msec) < 60
  - $60 \leq$ Maximum QTc (msec)

5.7.6. Echocardiogram (ECHO)/MUGA

Observed, change, and percent change from baseline values will be displayed by visit. Percent change from baseline in LVEF will be displayed in groups of 10% (e.g., 0 to ≤ 10%, 10% < to ≤ 20%, etc.) by visit and overall.

5.7.7. Pharmacokinetic (PK)

An analytical plan will be developed separately for the summary of PK parameters for the final analysis.

6. References

I. K. Hwang, W. J. Shih, and J. S. DeCani. Group sequential designs using a family of type I error probability spending functions. Statistics in Medicine, 9:1439–1445, 1990.

7. Appendices

7.1. Categorizing Lesions at Baseline

Measurable Lesions - Defined as lesions accurately measured in at least one dimension.

- Lesions with longest diameter twice the slice thickness and at least 10 mm or greater when assessed by CT or MRI (slice thickness 5-8 mm)
- Lesions with longest diameter at least 20 mm when assessed by Chest X-ray
- Superficial lesions with longest diameter 10 mm or greater when assessed by caliper
- Malignant lymph nodes with the short axis 15 mm or greater when assessed by CT

NOTE: Shortest axis is used as the diameter for malignant lymph nodes, longest axis for all other measurable lesions.

Non-measurable Disease - Defined as lesions too small to be considered measurable (including nodes with short axis between 10 and 14.9 mm) and truly non-measurable disease such as pleural or pericardial effusions, ascites, inflammatory breast disease, leptomeningeal disease, lymphangitic involvement of skin or lung, clinical lesions that cannot be accurately measured with calipers, abdominal masses identified by physical exam that are not measurable by reproducible imaging techniques.

- Bone disease: Not measurable with the exception of soft tissue components that can be evaluated by CT or MRI and meet the definition of measurability at baseline
- Previous local treatment: A previously irradiated lesion (or lesion subjected to other local treatment) is non-measurable unless it has progressed since completion of treatment.

Normal Sites

- Cystic lesions: Simple cysts should not be considered as malignant lesions and should not be recorded either as target or non-target disease. Cystic lesions thought to represent cystic metastases can be measurable lesions, if they meet the specific definition above. If non-cystic lesions are also present, these are preferred as target lesions.
- Normal nodes: Nodes with short axis <10 mm are considered normal and should not be recorded or followed either as measurable or non-measurable disease.

7.2. Recording Tumor Assessments

All sites of disease must be assessed at baseline. Baseline assessments should be done as close as possible prior to study start. For an adequate baseline assessment, all required scans must be done within 28 days prior to treatment and all disease must be documented appropriately. If baseline assessment is inadequate, subsequent statuses generally should be not evaluable.

Target Lesions

- All measurable lesions up to a maximum of 2 lesions per organ, 5 lesions in total, representative of all involved organs, should be identified as target lesions at baseline. Target lesions should be selected on the basis of size (longest lesions) and suitability for accurate repeated measurements. Record the longest diameter for each lesion, except in the case of pathological lymph nodes for which the short axis should be recorded. The sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all

target lesions at baseline will be the basis for comparison to assessments performed on study.

- If two target lesions coalesce the measurement of the coalesced mass is used. If a large target lesion splits, the sum of the parts is used.
- Measurements for target lesions that become small should continue to be recorded. If a target lesion becomes too small to measure, 0 mm should be recorded if the lesion is considered to have disappeared; otherwise a default value of 5 mm should be recorded.

NOTE: When nodal lesions decrease to <10 mm (normal), the actual measurement should still be recorded.

Non-target Disease

All non-measurable disease is non-target. All measurable lesions not identified as target lesions are also included as non-target disease. Measurements are not required but rather assessments will be expressed as ABSENT, UNEVALUATED, PRESENT, UNEQUIVOCAL PROGRESSION. Multiple non-target lesions in one organ may be recorded as a single item on the case report form (e.g., 'multiple enlarged pelvic lymph nodes' or 'multiple liver metastases').

7.3. Objective Response Status at Each Evaluation

Disease sites must be assessed using the same technique as baseline, including consistent administration of contrast and timing of scanning. If a change needs to be made the case must be discussed with the radiologist to determine if substitution is possible. If not, subsequent objective statuses are not evaluable.

Target Disease

- Complete Response (CR): Complete disappearance of all target lesions with the exception of nodal disease. All target nodes must decrease to normal size (short axis < 10 mm). All target lesions must be assessed.
- Partial Response (PR): Greater than or equal to 30% decrease under baseline of the sum of diameters of all target measurable lesions. The short diameter is used in the sum for target nodes, while the longest diameter is used in the sum for all other target lesions. All target lesions must be assessed.
- Stable: Does not qualify for CR, PR or Progression. All target lesions must be assessed. Stable can follow PR only in the rare case that the sum increases by less than 20% from the nadir, but enough that a previously documented 30% decrease no longer holds.
- Documented Progression Disease (PD): 20% increase in the sum of diameters of target measurable lesions above the smallest sum observed (over baseline if no decrease in the sum is observed during therapy), with a minimum absolute increase of 5 mm.
- Not Evaluable: Progression has not been documented, and
- one or more measurable target lesions have not been assessed
- or assessment methods used were inconsistent with those used at baseline
- or one or more target lesions cannot be measured accurately (e.g., poorly visible unless due to being too small to measure)
- or one or more target lesions were excised or irradiated and have not reappeared or increased

Non-target Disease

- CR: Disappearance of all non-target lesions and normalization of tumor marker levels. All lymph nodes must be 'normal' in size (<10 mm short axis).
- Non-CR/Non-PD: Persistence of any non-target lesions and/or tumor marker level above the normal limits.
- PD: Unequivocal progression of pre-existing lesions. Generally, the overall tumor burden must increase sufficiently to merit discontinuation of therapy. In the presence of SD or PR in target disease, progression due to unequivocal increase in non-target disease should be rare.
- Not Evaluable: Progression has not been determined and one or more non-target sites were not assessed or assessment methods were inconsistent with those used at baseline.

New Lesions

The appearance of any new unequivocal malignant lesion indicates PD. If a new lesion is equivocal, for example due to its small size, continued assessment will clarify the etiology. If repeat assessments confirm the lesion, then progression should be recorded on the date of the initial assessment. A lesion identified in an area not previously scanned will be considered a new lesion.

Supplemental Investigations

If CR determination depends on a residual lesion that decreased in size but did not disappear completely, it is recommended the residual lesion be investigated with biopsy or fine needle aspirate. If no disease is identified, objective status is CR.

If progression determination depends on a lesion with an increase possibly due to necrosis, the lesion may be investigated with biopsy or fine needle aspirate to clarify status.

Symptomatic/clinical Progression

Subjects requiring discontinuation of treatment without objective evidence of disease progression should not be reported as PD on tumor assessment CRFs. Every effort should be made to document objective progression even after discontinuation of treatment.

Table 4: Objective Response Status at Each Evaluation

Target Lesions	Non-target Lesions	New Lesions	Objective Status
CR	CR	No	CR
CR	Non-CR/Non-PD	No	PR
CR	NE, or Missing	No	PR
PR	Non-CR/Non-PD, NE, or Missing	No	PR
SD	Non-CR/Non-PD, NE, or Missing	No	SD
NE or Missing	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

7.4. Interim Analysis Plan

The following analyses are planned as part of the interim analysis:

Core Analyses for MM-121 IA

Demography and Baseline Characteristics – mITT population
Baseline Disease Characteristics - mITT population
Prior Anti-cancer Therapy - mITT population
Progression Free Survival - mITT population (with K-M plot)
Forest Plot of Progression Free Survival Hazard -mITT (per SAP subgroups)
Overall Survival - mITT population (with K-M plot)
Best Overall Response - mITT population
Duration of Response (subset of mITT with response)
Overall Summary of Adverse Events (mITT)
Summary of Treatment Emergent Adverse Events by System Organ Class (mITT)
Summary of Treatment Emergent Adverse Events by System Organ Class and Preferred Term (mITT)
Summary of Treatment Emergent Adverse Events by Preferred Term and Maximum NCI-CTCAE Grade (mITT)
Summary of Related Treatment Emergent Adverse Events by Preferred Term (mITT)
Summary of Study Drug Exposure (mITT)
Summary of Treatment Emergent Adverse Events of Special Interest by Preferred Term and Maximum NCI-CTCAE Grade (mITT)
Summary of Serious Treatment Emergent Adverse Events by System Organ Class and Preferred Term (mITT)
Summary of Serious Treatment Emergent Adverse Events by Preferred Term (mITT)

Sensitivity Analyses

Progression Free Survival - mITT population(with K-M plot) for patients in mITT who took immunotherapies
Overall Survival - mITT population (with K-M plot) for patients in mITT who took immunotherapies

Best Overall Response - mITT population for patients in mITT who took immunotherapies
Progression Free Survival, OS - (with K-M plots) for any randomized patients who took immunotherapies, including those not in mITT
PFS for the subset of mITT randomized on or after Amendment 4 (excludes retrospectively added patients)

A spending function was predefined, which allows for modifications in case the number of events at the interim analysis is different than the planned 57 events. The following table summarizes the one-sided significance levels that will be used if the number of events in the data snap shot differs from 57.

Significance Levels at Interim and Final Analysis Using Pre-Defined Alpha Spending Function

Number of Events at Interim	Information Time	1-sided significance level required at interim	1-sided significance level required at final
56	56/76=73.7%	0.0084	0.0226
57	57/76=75%	0.0089	0.0225
58	58/76=76.3%	0.0094	0.0224
59	59/76=77.6%	0.0099	0.0223
60	60/76=78.9%	0.0105	0.0222
61	61/76=80.3%	0.0111	0.0221
62	62/76=81.6%	0.0117	0.022