STATISTICAL ANALYSIS PLAN (SAP):

Nintedanib Plus Standard of Care Immunosuppression Versus Standard of Care Immunosuppression Alone in Patients with Progressive Fibrotic Myositis Associated - Interstitial Lung Disease: A Randomized, Double-Blind, Exploratory Trial

Short title: Myositis Interstitial Lung Disease Nintedanib Trial (MINT)

Clinical Phase: Exploratory (Pilot) Trial

Coordinating Center: University of Pittsburgh

Funding/Sponsor: Boehringer Ingelheim

Protocol number: 1199-0508 NCT number: NCT05799755

VERSION NUMBER/DATE: Version 2.0, March 4, 2024

PRINCIPAL INVESTIGATOR:

Rohit Aggarwal, MD, MS (University of Pittsburgh)
3601 Fifth Avenue, Suite 2B
Pittsburgh, PA 15213
Phone: (412) 647-8708
aggarwalr@upmc.edu

CO-INVESTIGATORS

Chester V. Oddis, MD (University of Pittsburgh)
200 Lothrop Street
Pittsburgh, PA 15261
Phone: (412) 383-8861
cvo5@pitt.edu

Siamak Moghadam-Kia, MD, MPH (University of Pittsburgh)
200 Lothrop Street
Pittsburgh, PA 15261
Phone: (412) 648-9782
moghadamkias@upmc.edu

Dana P. Ascherman, MD (University of Pittsburgh)
200 Lothrop Street
Pittsburgh, PA 15261
Phone: (412) 383-8734
Dascher@pitt.edu

Vladimir Liarski, MD (University of Pittsburgh)
200 Lothrop Street
Pittsburgh, PA 15261
Phone: (412) 383-8000
Vliarski@pitt.edu

Daniel Sullivan, MD
3469 Fifth Avenue
Pittsburgh, PA 15213
Phone: (412) 692-2210
sullivandi@upmc.edu

Daniel Kass, MD
200 Lothrop Street
Pittsburgh, PA 15261
Phone: (412) 624-7225
kassd2@upmc.edu

Kevin Gibson, MD
3469 Fifth Avenue
Pittsburgh, PA 15213
Phone: (412) 624-7225
gibsonkf@upmc.edu

Tanya Chandra, MD
3601 Fifth Avenue, Suite 2B
Pittsburgh, PA 15213
Phone: (412) 383-8000
Chandrat3@upmc.edu

1. General Analysis Definitions

1.1 Subject Sets Analyzed

The efficacy analyses will include all randomized subjects and will be performed according to their randomly assigned treatment regardless of the treatment they actually received. The treatment policy strategy defined in ICH E9 (R1) will be adopted in the analyses to handle intercurrent events (ICEs). Specifications for ICEs and the corresponding handling rules are provided in Section 1.2.

The safety analyses will include all randomized and treated subjects who have received at least one dose of study medication. The analyses will be performed according to the treatment they actually received.

1.2 Intercurrent Events

Intercurrent Events (ICEs) are events occurring after randomization which affect the existence or interpretation of the outcomes associated with the clinical objectives. The expected intercurrent events in this study are listed as follow:

- Treatment discontinuation or interruption (any cause)
- Start of a restricted medication
- Death
- Lung transplant
- Important protocol deviation

All these ICEs will be handled using treatment policy / intension-to-treat strategy, and data after these events, if available, will be included in the analyses.

According to the ICE guidance, summary of the number and timings of each ICE will be reported by treatment groups. Timing of an ICE is defined as the time from a patient's treatment initiation to the start date of the ICE.

1.3 Handling of Missing Data

For primary and secondary efficacy analyses, missing data will remain missing and will not be imputed. Sensitivity analyses may be performed to assess the impact of missing data utilizing multiple imputation methods..

For safety analyses, data will be summarized without any data imputation.

1.4 Study Sample Size

Approximately 134 patients are expected to be randomized in the study (67 per arm) and with 20% dropout rates we expect to have 54 subjects in each arm.

In INBUILD, patients in the nintedanib and placebo arms had a mean (SD) absolute change from baseline at 12 weeks in the L-PF dyspnea score of 0.59 (13.44) and 1.45 (13.04), respectively. The minimal clinically important difference in the L-PF dyspnea score is 6.

The table below lists the width of a 2-sided 95% confidence interval for the L-PF for a variety of subgroup sizes and assumes a standard deviation of 13.44. Assuming 134 patients are enrolled and 20%

of patients drop out, there will be 108 patients with complete follow up (with 54 in each arm). In each case, the confidence intervals are appropriate for an exploratory study.

Subgroup size	27	54	81	108	134
95% CI width	±5.07	±3.59	±2.93	±2.54	±2.28

2. Efficacy Evaluation

2.1 Primary Efficacy Analyses

Endpoint: Change in L-PF Dyspnea score from baseline to Week 12.

Descriptive statistics will be used for a summary of the primary endpoint between the treatments. The endpoint will be listed descriptively as median with interquartile and other percentages (including minimum and maximum values depending on the underlying distribution), and as mean value with standard deviation. The endpoint will be analyzed using a restricted maximum likelihood (REML) based approach using a mixed model with repeated measurements (MMRM) comparing the change from baseline in L-PF Dyspnea at Week 12 between treatment groups (controlled and uncontrolled for covariates). The main analysis uncontrolled for covariates will include the fixed categorical effect of treatment at each visit, fixed continuous effect of baseline L-PF score at each visit, and the random effect of subject. The sensitivity analysis controlled for covariates will additionally include terms for baseline predicted FVC%, age, gender, BMI. Continuous endpoints will be evaluated for outliers, and a secondary sensitivity analysis excluding outliers will be performed.

The endpoint will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

Subgroup analysis of the primary endpoint will include stratification by usual interstitial pneumonia diagnosis, mycophenolate medication, study site type, anti-synthetase status, and baseline FVC. A Bonferroni correction will be applied to subgroup analysis p-values based on the number of levels within each stratification variable.

2.2 Key Secondary Efficacy Analyses

a. Other L-PF scores including:
    - Change in L-PF Dyspnea score from baseline to Week 24
    - Change in L-PF Total score from baseline to Week 12 and Week 24
    - Change in L-PF Symptoms score from baseline to Week 12 and Week 24
    - Change in L-PF Cough score from baseline to Week 12 and Week 24
    - Change in L-PF Energy score from baseline to Week 12 and Week 24
    - Change in L-PF Impacts score from baseline to Week 12 and Week 24

    These endpoints will be analyzed descriptively and using a repeated measures mixed model similar to the primary endpoint analysis and with the same set of covariates. The results will be presented graphically as well as numerically.

b. Change in immunosuppressive regimen will be summarized by proportion of patients requiring an increased dose or a change in their glucocorticoid (GC) / immunosuppression (IS) agent for clinical worsening / flare of MA-ILD. The results will be presented by treatment groups. Comparison in the changes across different treatment group will also be performed by using two-group T-test or Analysis of Variance (ANOVA) models with significance level of 0.05. P-values will be presented in the tables together with the other summary statistics. To present the dose changes in immunosuppressive regimen, mean and median dose of GC/IS regimen (over time will be presented in figures using dose-over-time curves. Area under the curves will be calculated to summarize the cumulative dose of GC/IS regimen over the whole study period.
c. Changes in lung function assessments from baseline at Week 12 (+/- 2 weeks) and at Week 24 (+/- 2 weeks) including
    - Absolute and relative change in FVC (mL) and FVC (%) predicted from baseline at Week 12 and Week 24
    - Proportion of patients with a relative decline from baseline in FVC (mL) and FVC (%) of >= 10%, >= 7.5%, and >=5% within Week 12 and within Week 24.
    - Proportion of patients with stable (within 5%) or improved (>= 5%, >= 7.5%, >= 10%) FVC (mL) and FVC (%) predicted within Week 12 and within Week 24.
    - Time to FVC (mL) and FVC (%) predicted improvement and decline from baseline by (>= 5%, >= 7.5%, >= 10%) over 24 weeks.
    - Time to progression (defined in the protocol) or death or lung transplant over 24 weeks.

The continuous secondary endpoints will be analyzed using a MMRM model similar to the primary analysis. Proportions will be analyzed using a logistic regression model with treatment as the main effect and with a covariate set similar to the primary analysis. Time-to-event endpoints will be analyzed in a descriptive manner first using Kaplan-Meier (KM) estimates and KM curves and will be presented by treatment groups. A Cox-proportional model with covariates similar to the primary analyses will also be applied to the time-to-event endpoints to evaluate the hazard ratio. The point estimate as well as the 95% confidence interval for the hazard ratio will be presented in tables.

2.3 Other Secondary and further endpoints

a. Rate of patient recruitment, enrollment, screen failure, and dropout rates between the local sites (average per site) and the remote site will be summarized and presented in tables. The rate of recruitment/enrollment is defined and will be calculated as number of patients enrolled in the study per month. It will be calculated at the end of recruitment for both local sites and remote sites for comparison. Two-group proportion test will be adopted to compare the screen failure and drop out rates between local sites and remote sites with significance level of 0.05.
b. Psychometric properties of home spirometry as compared to clinical / lab-based spirometry. Assessments may be compared descriptively and using testing hypothesis depending on the data availability.
c. Absolute and relative changes in DLCO, FEV1, and threshold from baseline to Week 12 and Week 24 will be analyzed using the same approaches as the primary analyses.
d. Progression free survival, which includes

i. Free of CTD-ILD progression outcome as per OMERACT, which is defined as the occurrence of death, lung transplant, FVC decline ≥ 10%, or FVC decline ≥ 5% together with DLCO decline ≥ 15% during 12-week or 24-week periods
ii. Free of clinical worsening which is defined as death, lung transplant, new/worsening O2 requirement at rest (> 2L), non-elective hospitalization for ILD clinical worsening/flare, or ILD worsening definition (per protocol) or out of protocol rescue medication for ILD clinical worsening/flare.

For each of the progression-free definition, survival analysis will be performed using KM estimates and KM curves to summarize the probability of event-free/survival during the 12-week and 24-week periods. A Cox-proportional model with covariates similar to the primary analyses will also be applied to the time-to-event endpoints to evaluate the hazard ratio. The point estimate as well as the 95% confidence interval for the hazard ratio will be presented in tables.

a. Steroid use at Week 12 and Week 24 will by summarized by treatment groups, which includes
    i. Mean and median decrease in glucocorticoid dose
    ii. Absolute and relative changes in glucocorticoid dose
    iii. Area under the curve for cumulative total glucocorticoid exposure (cumulative dose)
    iv. Proportion of patients able to decrease glucocorticoid to 5, 7.5, and 10 mg or less of prednisone equivalent (calculated among patients on doses higher than 5, 7.5, and 10 mg respectively)
    v. Proportion of patients requiring GC rescue therapy (specified in the protocol) for ILD clinical worsening/flare
    vi. Proportion of patients requiring IV GC rescue therapy for ILD clinical worsening/flare (results in withdrawal of study drug)

    The comparison across different treatment groups in steroid use will be performed as follows: decrease in dose, absolute change in dose, as well as relative change in dose will be compared using two-group T-test or ANOVA models at significance level of 0.05. The proportions mentioned above will be compared using Chi-square test between two treatment groups at significance level of 0.05.

For the following endpoints, the summary statistics mean, SD, Q1, Q3, and median will be calculated for each assessment. Two-group T-test and ANVOA models will be used for comparison with significance of 0.05.

- Change in supplemental oxygen needs from baseline to 12 and 24 weeks will be analyzed by summarizing the following variables (collected at the end of L-PF symptoms module) at baseline and 12/24 weeks. Changes in the continuous variables (indicated below) will be summarized as well. The variables include:
    - Rest
        - Increased, decreased or same or new start of supplemental oxygen
        - Flow rate as a continuous variable

    - Activity
        - Increased, decreased or same or new start of supplemental oxygen Flow rate as a continuous variable
- Overall assessment of change at Week 12 and 24 which include patient overall assessment of change (by the patient) and physician assessment of change (by the investigator). The changes will by summarized as continuous variables by treatment group.
- Changes in physical activity monitor (step-based and cadence measures)
- Changes in Sit to Stand test
- Change in HRCT chest from baseline to 24 weeks, which are derived using semi-quantitative and quantitative image-based scores, will be summarized by treatment groups. HRCT chest quantitative scoring will be performed as a separate project.

In addition, there are few exploratory endpoints for drug efficacy may also be explored in the study. For the numerical endpoints, the summary statistics mean, SD, Q1, Q3, and median will be calculated for each assessment. Two-group T-test and ANVOA models will be used for comparison with significance of 0.05. For categorical variables, frequencies of each category will be summarized. Comparison will be made using Chi-square or Fisher exact tests at significant level of 0.05.

a. Lung function assessments collected via home spirometry. The measurements as well as the analysis methods will be similar to the analyses for lung function assessments listed under the key secondary endpoints.
b. Changes in patient reported outcomes (PROs) including SGRQ, PROMIS, Myositis, as well as muscle enzymes will be summarized numerically by treatment groups.
c. Changes in pulse oximetry will by summarized by treatment groups.
d. Patient's perspective of the trial may by summarized using the feedback from patients. Feedback will be summarized as categorical variables and presented by treatment groups.

3. Safety Evaluation

3.1 Adverse Events

The safety evaluation for this study will mainly include analyses on Adverse Events (AEs) and laboratory parameters. The analysis period for the evaluation is from the first study drug intake to the end of treatment plus 28 days (residual effect periods). All the AEs and laboratory parameters collected during this period will be included for the analyses.

The analysis for AEs will be on subject level and will include summary of AEs by treatment groups. Multiple tables will be presented to summarize the following:

- Patients with any AEs
- Patients with Serious Adverse Events (SAEs)
- Patients with AESI (Adverse Events of Special Interest), which includes death or transplant, gastrointestinal perforation, liver function abnormalities, and hepatic injury

In addition to the summary tables, listings of AEs with AE onset dates, end dates, severity will be provided to report all the AE incidences during the period.

For laboratory parameters collected by CRF, summary of parameters at each visit as well as the change from baseline in these values over the period will be provided in tables.

3.2 Safety Exclusions

A participant presenting with any AESI (death or transplant, gastrointestinal perforation, liver function abnormalities, or hepatic injury) at any visit will immediately discontinue study drug and be removed from the study due to safety concerns.