TRAIL1 Statistical Analysis Plan

TRIAL FULL TITLE	A Randomized, Double-Blind, Placebo-Controlled, Phase 2 Study of Safety, Tolerability and Efficacy of Pirfenidone in Patients with Rheumatoid Arthritis Interstitial Lung Disease
SAP VERSION	2.0
SAP VERSION DATE	May 10, 2021
PROTOCOL VERSION (associated with SAP)	4.0
TRIAL PRINCIPAL INVESTIGATOR	Ivan Rosas, MD
SAP AUTHOR(s)	S Hurwitz, R Maurer

SAP Signatures

I approve for the attached SAP entitled "A Randomized, Double-Blind, Placebo-Controlled, Phase 2 Study of Safety, Tolerability and Efficacy of Pirfenidone in Patients with Rheumatoid Arthritis Interstitial Lung Disease" dated May 10, 2021

Statistician

Name: Shelley Hurwitz, PhD

Signature: _____

Date: 5/10/21

Statistician

Name: Rie Maurer, MA

Signature: _____

Date: 05/10/2021

Principal Investigator

Name: Ivan Rosas, MD

Signature: _____

Date: 05/10/2021

Contents

Abbreviations and Definitions

AE	Adverse event
ANCOVA	Analysis of covariance
CCC	Clinical Coordinating Center
CI	Confidence interval
COVID-19	Coronavirus Disease 19
CRF	Case report form
CRP	C-reactive protein
DAS	Disease activity score

DCC	Data Coordinating Center
DLCO	Diffusing capacity of lung for carbon monoxide
% DLCO	Percent predicted diffusing capacity of lung for carbon monoxide
DSMB	Data Safety Monitoring Board
eCRF	Electronic case report form
ESR	Erythrocyte sedimentation rate
FEV1	Forced expiratory volume in one second
FVC	Forced vital capacity
%FVC	Percent predicted forced vital capacity
HRCT	High resolution computed tomography
HRQOL	Health Related Quality of Life
ILD	Interstitial Lung Disease
IMP	Investigational Medicinal Product
IPF	Idiopathic pulmonary fibrosis
ITT	Intention to treat
LCQ	Leicester Cough Questionnaire
ln	Natural logarithm
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
mg/d	Milligrams per day
mL	Milliliter
mm Hg	Millimeters of mercury
msec	Millisecond
OMERACT	Outcome Measures in Rheumatology
PF	Pulmonary fibrosis
PFS	Progression-free survival
PRO	Patient reported outcomes
RA	Rheumatoid arthritis

RA-ILD	Rheumatoid arthritis-associated interstitial lung disease
RAPID3	Routine Assessment of Patient Index Data 3
SAE	Serious adverse event (AE)
SAS	Statistical Analysis Software
SGRQ	St. George's Respiratory Questionnaire
SpO2	Peripheral capillary oxygen saturation
TLCO	Transfer capacity of lung for carbon monoxide
ULN	Upper limit of normal
wk	Week

Introduction

Preface

Rheumatoid arthritis-associated interstitial lung disease (RA-ILD) is a life-threatening and severely debilitating disease characterized by poor survival and a lack of efficacious therapeutic options. Results from previous controlled trials suggest that pirfenidone treatment is safe and well tolerated in idiopathic pulmonary fibrosis (IPF) with efficacy in a variety of domains that include fibrosing ILD. Pirfenidone may be of benefit in RA-ILD, which shares many histopathological and clinic-epidemiological features with IPF. Given the loss of pulmonary function over time in patients with RA-ILD, this protocol is designed to evaluate whether pirfenidone 2403 mg/d reduces decline in forced vital capacity (FVC) or mortality over 52 weeks, compared with placebo, in patients with RA-ILD.

TRAIL1 is a Phase 2, randomized, double blind, placebo controlled clinical trial designed to evaluate the efficacy and safety of administering pirfenidone or placebo for 52 weeks to patients with RA-ILD. Patients meeting the eligibility criteria for the study have been randomized to receive either pirfenidone 2403 mg/d or placebo. Efficacy will be evaluated through pulmonary function tests, patient reported outcomes (PROs), adverse events and survival. Safety will be assessed by

determining differences between the treatment arms for the rate of adverse events, serious adverse events, rates of acute exacerbation, hospitalization and all- cause mortality.

Scope of the Statistical Analysis Plan (SAP)

These analyses will assess the efficacy and safety of pirfenidone 2403 mg/day as compared with placebo in patients with RA-ILD. This SAP assumes familiarity with the TRAIL1 Clinical Trial Protocol, including amendments. In particular, the SAP is based on the planned analysis specification as written in the protocol Section 10 “Statistical Considerations.” Readers may consult the protocol for more background information on the design and conduct of the trial, and panned analyses.

Study Objectives and Endpoints

Study Objectives

To assess the efficacy and safety of pirfenidone 2403 mg/day as compared to placebo in patients with RA-ILD.

To explore the role of peripheral blood biomarkers in predicting disease progression and survival in patients with RA-ILD.

To explore a spectrum of validated questionnaires to assess disease specific PROs including overall health, and perspectives on symptoms, performance and quality of life.

Endpoints

Primary Endpoint

The primary outcome variable of this study is the incidence of the composite endpoint of decline from baseline in percent predicted FVC of 10% or greater or death during the study period. The primary endpoint will be analyzed using the ITT population

Secondary Endpoints

Secondary outcome for this study will be assessed from baseline to the end of study for each patient. The following secondary outcomes will be analyzed using the ITT population:

1. Incidence of a decline from baseline in percent predicted FVC of 10% or greater during the study period.
2. Frequency of progressive disease as defined by Outcome Measures in Rheumatology (OMERACT): Relative decline from baseline in percent predicted FVC of ≥10%, or relative change from baseline in percent predicted FVC ≥ 5% and < 10%, and ≥15% relative decline in diffusing capacity of lung for carbon monoxide (DL_{CO}) (Khanna, Mittoo et al. 2015)
3. Change from baseline to end of study in absolute value of FVC
4. Change from baseline to end of study of percent predicted FVC
5. Slope of percent predicted FVC over 52-week treatment period
6. Slope of absolute value of FVC over study period
7. Time to decline of 10% or greater in percent predicted FVC or death while on study
8. Change from Baseline to end of study in dyspnea, as measured by the Dyspnea 12 questionnaire

The following secondary outcomes will be analyzed using the as-treated population:

9. Proportion of participants with all-cause mortality
10. Proportion of participants with all-cause hospitalization
11. Proportion of participants with hospitalization for respiratory cause
12. Number of acute exacerbations requiring hospitalizations
13. Proportion of participants with and number of treatment-emergent adverse events (AEs)
14. Proportion of participants with and number of treatment-emergent serious adverse events (SAEs)
15. Proportion of participants with and number of treatment-emergent/treatment-related AEs
16. Proportion of participants with and number of treatment-emergent/treatment-related SAEs
17. Proportion of participants with and number of AEs leading to early discontinuation of

study treatment

18. Proportion of participants with and number of treatment-emergent death or transplant
19. Proportion of participants with and number of treatment-emergent RA-ILD-related mortality

Exploratory Endpoints

The following exploratory outcomes will be analyzed using the ITT population:

1. Change from Baseline to end of study in Disease Activity Score (DAS)
2. Change from baseline to end of study in Routine Assessment of Patient Index Data 3 (RAPID3) score
3. Change from Baseline to end of study in Erythrocyte Sedimentation Rate (ESR)
4. Change from Baseline to end of study in C-Reactive Protein (CRP)
5. Candidate biomarker expression in the peripheral blood of patients with RA-ILD over the 52 weeks of treatment.
6. Changes from Baseline to end of study in high resolution computed tomography (HRCT) parameters evaluated by quantitative functional imaging
7. Changes from Baseline to Week 13, 26, 39 and final visit in the St. George's Respiratory Questionnaire (SGRQ)
8. Changes from Baseline to Week 13, 26, 39 and final visit in Dyspnea 12 questionnaire
9. Changes from Baseline to Week 13, 26, 39 and final visit in Leicester Cough Questionnaire (LCQ)
10. Changes from Baseline to Week 13, 26, 39 and final visit in the Patient global assessment
11. Changes from Baseline to Week 13, 26, 39 and final visit in the Health assessment questionnaire

Study Methods

General Study Design and Plan

This parallel-group double-blind randomized longitudinal study was designed to assess the superiority of Pirfenidone versus placebo, over a 52-week follow-up period. Patients are randomized with equal probability to receive Pirfenidone or placebo. After the screening visit, randomization occurs at visit 2. The week 52 visit is visit 11. Adverse events are additionally queried by phone 28 days after visit 11.

Inclusion-Exclusion Criteria and General Study Population

Eligible participants include patients age 18 through 85 years, with RA and ILD, without evidence or suspicion of an alternative diagnosis that may contribute to their ILD. Additional details to be found in the protocol.

Randomization and Blinding

After consent was obtained, participants were screened to determine eligibility. After participants were determined to be eligible, they were randomized with equal probability to receive either Pirfenidone or placebo in a double-blinded fashion. Treatment assignment slots were not re-used; each consecutive randomized participant received the next consecutive unused treatment assignment slot. Randomization was stratified by center with random block sizes. The password protected computer text file that contained the translation of treatment labels A and B to either Pirfenidone or placebo was maintained by the unblinded DCC statistician. Details will be provided in the final study report.

Study Assessments

The schedule of all study assessments is in **Appendix A** of the protocol. Eligibility and randomization take place at Visit 2. Time windows for visits are below. There is an additional phone visit that occurs 28 days after the last study drug dose to assess vital status and adverse events.

Visit Number	Visit Week	Target Day of Visit	Visit Window
1	-8 to 0	-56 to 0	-56 to 0 days
2	0	1	0
3	1	7	± 2 days
4	2	14	± 2 days
5	4	28	± 1 week
6	8	56	± 1 week
7	13	91	± 2 weeks
8	19	133	± 1 week
9	26	182	± 2 weeks
10	39	273	± 2 weeks
11*	52	364	± 1 week

* Visit 11 marks the end of the follow-up period. Protocol amendment V8 (US/CAN/AUS) /V10 (UK) extends the visit 11 window to ± 4 weeks due to COVID-19 laboratory closures.

Definition of Baseline

Baseline values for spirometry are defined as the screening values. Spirometry is repeated at randomization to ensure stability of values, as defined in the study protocol, prior to randomization. Baseline is defined as screening because the screening spirometry must meet established criteria for quality and is done before and after bronchodilator administration. The underlying true spirometry values are assumed to be the same at randomization. The absolute differences between the measured screening and randomization spirometry values will be summarized descriptively.

Baseline values for patient reported outcomes are defined as the randomization visit values, and when unavailable, the screening or rescreening values will be used. Baseline for other variables is defined as the randomization values for variables collected at randomization, otherwise it is defined as the screening values.

Endpoints and Relevant Variables

Refer to separate file for list of relevant variables for each endpoint, along with relevant datasets and endpoint description ("TRAIL1 endpoints and dataset variable names.xlsx" and "Primary & Secondary Outcomes" tab.)

Primary and Secondary Endpoints and Relevant Variables

Composite endpoint of decline from baseline in percent predicted FVC of 10% or greater or death during study period.

- FVC (percent predicted) at baseline
- Last FVC (percent predicted) while on study

  Decline from baseline in percent predicted FVC is (last percent predicted FVC while on study) - (percent predicted FVC at baseline)
- Death (yes/no) while on study

Decline from baseline in percent predicted FVC of 10% or greater during study period.

- FVC (% predicted) at baseline
- Last FVC (% predicted) while on study

  Decline from baseline in percent predicted FVC is (last percent predicted FVC while on study) - (percent predicted FVC at baseline)

Frequency of progressive disease as defined by OMERACT: Relative decline from baseline in percent predicted FVC of ≥10%, or relative decline from baseline in percent predicted FVC ≥ 5% and < 10%, and ≥15% relative decline from baseline in percent predicted DLCO (Khanna, Mittoo et al. 2015)

- FVC (percent predicted) at baseline
- Last FVC (percent predicted) while on study

  Relative decline from baseline in percent predicted FVC = [(Last FVC(percent predicted) – Baseline FVC (percent predicted))/Baseline FVC (percent predicted)]*100
- DLCO (percent predicted) at baseline
- Last DLCO (percent predicted) while on study

Relative decline in percent predicted DLCO = [(Last DLCO (percent predicted) – Baseline DLCO (percent predicted))/Baseline DLCO (percent predicted)]*100

If DLCO (percent predicted) is missing, TLCO (percent predicted) will be substituted.

Change from baseline over study period in FVC (L)

- FVC (L) at baseline
- Last FVC (L) while on study

Change from baseline over study period of percent predicted FVC

- FVC (percent predicted) at baseline
- Last FVC (percent predicted) while on study

Slope of percent predicted FVC over study period

- All available percent predicted FVC measurements and dates while on study

Slope of absolute FVC over study period

- All available FVC measurements and dates while on study

Time to decline of 10% or greater in percent predicted FVC at death over study period

- Baseline percent predicted FVC and randomization date
- All measurements of percent predicted FVC while on study and date
- Death date, death indicator
- Last study visit date

Proportion of participants with all-cause mortality

- Death date present or absent

Proportion of participants with all-cause hospitalization

- Hospital admission date present or absent

Proportion of participants with hospitalization for respiratory cause

- Hospital admission date
- Hospitalization reason

Number of participants with acute exacerbations requiring hospitalizations

- Hospital admission date(s)
- Hospitalization reason for each admission

Proportion of participants with and number of participants with treatment-emergent/treatment-related AEs

- Participants with at least one treatment emergent AE (yes/no for each participant) where treatment emergent is defined as the AE onset date being greater than or equal to the date of the first dose of study dedication, or the randomization date if the date of the first dose of study medication is missing
- Number of treatment-related AEs for each participant while on study

Proportion of participants with and number of treatment-emergent/treatment-related SAE (a subset of AEs)

- Participants with at least one treatment emergent SAE (yes or no for each participant)
- Number of treatment-related SAEs for each participant while on study

Proportion of participants with and number of AEs leading to early discontinuation of study treatment

- Early treatment discontinuation (yes or no for each participant)

- Reason for each early treatment discontinuation
- Number of AEs per participant where the action is "permanent discontinuation or withdrawn from study medication

Proportion of participants with and number of participants with treatment-emergent death, transplant, and transplant then death

- Cause of death
- Cause of transplant

Proportion of participants with and number of participants with treatment-emergent RA-ILD-related mortality

- Cause of death

Change from Baseline to Week 52 in dyspnea, as measured by the Dyspnea 12 questionnaire

- Baseline Total Dyspnea 12 questionnaire score and date
- Last Total Dyspnea 12 questionnaire score on study and date

Exploratory Endpoints and Relevant Variables

Change from Baseline to end of study period in Disease Activity Score (DAS28) and RAPID3 score

- Baseline DAS (continuous, 0 to 10)
- Last DAS on study (continuous, 0 to 10) on study
- Baseline RAPID3 score (continuous)
- Last RAPID3 score (continuous) on study

Note: DAS28 ranges from 2 to 10, and is calculated as DAS28= (0.56*√(Tender Joint Count)+0.28*√(Swollen Joint Count)+0.7*ln(ESR)+0.014*(Global Health)). The screening or rescreening ESR will be used for the calculation. Cumulative and weighted RAPID3 scores will be used.

Change from Baseline to end of study period in Erythrocyte Sedimentation Rate (ESR)

- Baseline ESR
- Last ESR on study

Change from Baseline to end of study period in C-Reactive Protein (CRP)

- Baseline CRP
- Last CRP on study

Candidate biomarker expression in the peripheral blood of patients with RA-ILD over the 52 weeks of treatment and the study follow-up period

- The exploratory analysis of candidate biomarkers will be a substudy performed at a later date.

Changes from Baseline to end of study period in HRCT parameters evaluated by quantitative functional imaging

- The quantitative parameters will be defined after the end of the trial, as exploratory analyses, and will include the following. These scores are separate from the radiologist -derived scores; they are computer interpretation of the scans, and range from 0 to 100.
    - DTA fibrosis score (total fibrosis score, dta_fibrosis)
    - Honeycombing score (hc_score)
    - Reticulation (reticular abnormality, ra_score)
    - Ground glass abnormality score (ggo_score)
- The data values that will be used:
    - Baseline HRCT quantitative measures
    - Last HRCT quantitative measures on study

Changes from Baseline to Week 13, 26, 39 and final visit in the SGRQ

- All measurements of SGRQ while on study, total score

- The SGRQ has 2 parts and 3 components (symptoms, activity and impacts). Scores for each component range from 0 to 100 with higher scores meaning more disability. A total score is calculated from component scores with a range from 0 to 100 with higher scores meaning more disability.

Changes from Baseline to Week 13, 26, 39 and final visit in Dyspnea 12 questionnaire

- All measurements of Dyspnea 12 while on study, total score
- The Dyspnea 12 has 12 questions scored 0 (none) to 3 (severe). Score are added up for total (0 to 36) with higher scores indicating greater severity of symptoms.

Changes from Baseline to Week 13, 26, 39 and final visit in LCQ

- All measurements of LCQ while on study, total score
- The scores are calculated as a mean for each domain and total score is found by adding the domain scores.

Changes from Baseline to Week 13, 26, 39 and final visit in the Patient Global Assessment

- All measurements of Patient Global Assessment while on study, total score
- The Patient Global Assessment has 3 questions that are compared separately for change. One question is scaled 0 to 10 and two questions are scaled on a likert scale.

Changes from Baseline to Week 13, 26, 39 and final visit in the Health Assessment Questionnaire

- All measurements of Stanford Health Assessment Questionnaire while on study and date.

  The Health Assessment Questionnaire has 8 sections scored 0 – 3. Using the highest score in each section, the Health Assessment Questionnaire Score is the average across all sections with data (see Fries, et al., 1980)

Planned Sample Size

The planned sample size for this trial was 270 participants randomized with equal probability to one of two treatment arms. The considerations delineated in the protocol show that 254 participants with complete data provide at least 85% power to demonstrate the treatment arm difference in terms of the combined primary endpoint reached by week 52, defined as either death or a decline from baseline of $\geq$10 percentage points in the percent predicted FVC. The robustness of this determination under a variety of feasible scenarios is also demonstrated in the protocol. The planned number of participants randomized was 270 because of possible loss of data due to participants dropping out before week 52 or other reasons.

Actual Sample Size

Enrolment at centers was interrupted by the COVID-19 pandemic during March, 2020. After all centers were shut down for approximately three months, and after an interim DSMB meeting that focussed on the interruption of recruitment, a decision was made to permanently close enrolment. The final sample size is 123 randomized participants. The last participant was randomized on March 24, 2020.

General Analysis Considerations

Timing of Analyses

Data cleaning will be performed throughout the enrolment period, and will continue and through the 12-month study period of the last participant randomized. Data cleaning will be conducted in conjunction with a blind data review while the study is ongoing. The database will be locked and the final analysis will be performed after the last participant completes visit 12 at the end of the follow-up period, and after all outstanding data queries and discrepancies are resolved. This statistical analysis plan will be finalized before closure of the database and before any final analyses are conducted. Independent study monitoring will be conducted by SABER in adherence to the Good Clinical Practice guidelines.

Analysis Populations

Each participant's inclusion or exclusion status with regard to each analysis population will be determined prior to breaking the blind. For relevant variables, please refer to the accompanying file (TRAIL1 endpoints and dataset variable names.xlsx, Analysis Population type tab.)

Intention to Treat (ITT) population (full analysis set):

Defined as all participants who were randomized. The ITT population will be used for all primary, secondary, and exploratory efficacy analyses.

Safety population (as-treated population)

Defined as all randomized participants who received any study medication and who received at least one study medication dose. The Safety population will be used for safety-related analyses.

Per Protocol population

Defined as all randomized participants who did not deviate from the protocol, who satisfy all eligibility criteria, and who are compliant with study medication, determined on a per-participant basis by trial leadership before database lock. The Per Protocol population will be used for sensitivity analyses to demonstrate robustness of the results, and for selected sub-studies to be determined. Out of visit window visits do not automatically disqualify participants from the per protocol population. If a visit is less than four weeks out of the window (either before or after,) then it is included in the Per Protocol population. Prohibited medications will exclude participants from the Per Protocol population. Medication compliance is required for inclusion in the Per Protocol population, and is defined elsewhere in this document as having taken at least 80% and not more than 120% of dispensed pills according to pill diaries. Other protocol deviations will be discussed on a case-by-case basis prior to the final database lock and while all team members are blinded to treatment assignments.

Sensitivity Analyses

To assess the robustness of the primary analysis of the primary endpoint, the following two sensitivity analyses will be performed. (1) The analysis of the primary composite endpoint will be repeated using the Per-Protocol population. (2) The analysis of the primary composite endpoint will be repeated using an adaptation of the Per-Protocol population where the lower threshold for medication compliance will be reduced from 80% to 65%.

Coronavirus Disease 19 (COVID-19)

The effect of COVID-19 on TRAIL1 will be investigated as follows: Protocol deviations due to COVID-19 will be presented and visits out of window due to COVID-19 will also be identified. Sensitivity analyses will be determined if necessary prior to database lock, and may include the following: Excluding all affected participants, using time-varying covariates to identify measurements taken during the affected period, increasing the duration of visit windows beyond those specified in the protocol, setting all data in a participant's affected period to missing and imputing values. (reference: Ongoing trials during COVID-19: Quantitative strategies and methodological aspects: Cytel, Inc. 2020.)

COVID-19-related factors that have the potential to impact this trial:

- COVID-19 infection resulting in voluntary or mandatory dropout
- Different investigator(s) filling in
- Study center closures resulting in loss to follow-up
- Stopped enrollment resulting in study termination or delay
- Delayed assessments and expanded protocol visit windows
- Missed visits and/or assessments leading to partial data collection
- Alternative data collection, phone visits, exchangeability of methods, validation
- Stopping drug due to safety concerns
- Stopping drug due to lack of lab monitoring

- Concomitant meds due to COVID-19, and interaction with the Investigational Medicinal Product (IMP)

Covariates and Subgroups

Covariates

Pre-specified baseline covariates known to be related to specific endpoints will be included as predictors in the primary analyses. One covariate has been pre-specified: the HRCT pattern, categorized as usual interstitial pneumonia (UIP)-like fibrotic pattern versus other fibrotic patterns.

Subgroups

Subgroup analyses are considered exploratory. Subgroupings were selected for pre-specification where there is an expectation of differential treatment effects. The p-value for the interaction between the subgroup variable and treatment arm will be reported. Sufficient power to reject a null hypothesis for the interaction is not anticipated. If the interaction (difference in the treatment effects among the subgroups) is judged to be clinically important, then subgroup-specific estimates of the treatment effect and 95% confidence intervals will be reported. Pre-specified subgroups will include sex, race (white, black or African American, Asian, native Hawaiian or other pacific islander, American Indian or Alaska native,) and ethnicity (Hispanic or Latino.)

Multi-center Studies

In this multi-center clinical trial, to allow for center-specific effect sizes, center will serve as a random effect in all primary analyses comparing treatment arms. Model fitting problems from sparse data may be encountered since enrolment stopped early due to COVID-19. If that occurs, then network (US, Canada, UK, Australia) will serve as random effect. As exploratory analyses, results within center or network will be described.

Missing Data

Adhering to the definition of the primary outcome as either death or a decline from baseline of >10 percentage points in percent predicted FVC, reached by week 52, all participants will be included in the ITT population. Outcomes involving repeated measurements will be analyzed using mixed models, accommodating varying amounts of follow-up data among participants. The statistical models allow for missing data, assuming they are missing at random. Even participants with only one post-baseline assessment can be included, and can therefore contribute to variance estimation. It is reasonable to assume that missing data due to COVID-19 center closures are missing at random. The statistical model assumes that participants who prematurely discontinue study participation would have behaved similarly to those who remained in the study. The mixed effect model for repeated measures analyses allows for form-level missing data, assuming they are missing at random. Item-level data for PROs will be handled according to the instructions provided by the instrument developer.

Sensitivity analyses using alternative assumptions will be conducted to investigate the potential effect of missing data on the results of the primary analysis. One such analysis will exclude participants with less than six months of follow-up. It is conceivable that one arm, such as the active drug arm, may have more early discontinuations and/or more missing data than the other arm. The extent of missing data will be reported overall and by treatment arm, and quantified for key primary and secondary endpoints, with a graphical representation. Reasons for treatment discontinuation and for study discontinuation will be reported. Because missing data arising from the temporary closure of centers due to the COVID-19 pandemic is assumed to be missing at random, treatment arm comparisons will be valid. Where covariates are missing, a value of 0 will be imputed, along with a missing-indicator variable for the covariate as an additional covariate in the regression model (White, Thompson, 2005).

Time-to-event missing data

In the analyses of the time-to-event endpoints, missing or incomplete data will be managed by censoring at the last date recorded without the event. In the case of partially missing dates, the following imputation will be done:

1. If day is missing, then imputed day will be the 15th of the month.
2. If day and month are missing, they will be imputed by inference from the participant's other visit dates.
3. If year is missing, it will be imputed by inference from participant's other visit dates.

Interim Analyses and Data Monitoring

This phase II clinical trial was not designed to stop early for efficacy in order to preserve the 5% probability of a Type I error for the final analysis. Throughout the study, an external Data Safety Monitoring Board (DSMB) reviewed safety and administrative data periodically, according to the schedule specified in the DSMB Charter, to determine the safety and feasibility of continuing the trial. Open session reports summarized the following types of data: accrual, protocol deviations, data completeness, demographics and baseline characteristics, clinical and laboratory data, adverse events and other safety parameters, treatment compliance and dose modifications, reasons for treatment discontinuations, and causes of death. The open report included analyses with all participants combined (i.e., pooled across treatment group). The closed report presented the data according to treatment group in a blinded fashion. The DSMB could request additional participant-level data and/or recommend that the study stop temporarily or permanently for safety concerns that are set forth in the DSMB Charter, or for administrative reasons such as slow accrual. An interim DSMB meeting was convened on June 1, 2020 that focussed on the ongoing COVID-19 pandemic, and the potential permanent discontinuation of enrolment.

Scope of Adaptations

Serious safety concerns could lead to a Data Safety Monitoring Board (DSMB) recommendation to temporarily stop the trial pending additional information, or to permanently stop the trial. Issues with

slow accrual could lead to a recommendation to stop the trial or to expand recruitment with marketing efforts and/or by adding centers, for example. The unprecedented COVID-19 pandemic precipitated the permanent discontinuation of enrolment.

Practical Measures to Minimize Bias

Any level of unblinding, either of individual participants or of treatment estimates, could induce biases. Therefore, all study personnel were blinded throughout the study, seeing only pooled summaries, with the exception of the Data Coordinating Center (DCC) statisticians producing the DSMB closed reports and the Clinical Coordinating Center (CCC) statisticians producing the final analysis after the data lock. To minimize bias during the trial before the data lock, the treatment arms had masked labels such as "A" and "B." Individual participant treatment assignments were not known unless there was a safety concern where knowledge of treatment assignment determined patient care.

Plan for Interim and Final Analyses, and Documentation

- The DCC statisticians will perform interim analyses for DSMB periodic reports
- The DCC statisticians and the DSMB members will see the DSMB analyses broken down by masked treatment arm
- A decision to stop or modify the trial will be provided to the sponsor and investigators
- The CCC statisticians will perform final analyses.

Snapshots of the data available at each interim analysis will be preserved, as will all documentation of analysis plans, programming code and reporting provided at each interim.

Multiple Testing

As this is a Phase II trial with only one single pre-specified primary endpoint, adjustment for multiple testing in the analyses is not planned for the primary endpoint or for the hypothesis-generating

secondary endpoints. In the exploration of the pre-specified subgroups, estimates of effect and 95% confidence intervals will be reported with an accompanying statement regarding multiplicity and the exploratory nature of the subgroup analyses.

Summary of Study Data

All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, minimum, and maximum. The frequency and percentages (based on participants' non-missing values for each particular variable) of observed levels will be reported for all categorical measures. In general, all data will be listed, treatment, and participant, and when appropriate by visit number within participant. Age will additionally be summarized by pre-specified age groups. Selected other continuous variables, that are specified before the database lock, will be additionally summarized using categories.

All summary tables will be structured with a column for each treatment in the order (Placebo, Pirfenidone) and will be annotated with the total population size relevant to that table/treatment. The number of missing observations will be reported.

Summary tables of baseline characteristics will be based on the ITT population. Summary tables of adverse events will be based on the As Treated population. Summary tables of outcomes will be presented for the ITT population, with the primary endpoint also being presented using the Per Protocol population.

Subject Disposition

Please refer to the accompanying spreadsheet for forms and variable names (TRAIL1 endpoints and dataset variable names.xlsx, Subject Disposition tab.)
The consort diagram will be presented.

http://www.consort-statement.org/consort-statement/flow-diagram

The number of participants reaching the milestones in the trial will show the number screened, randomized, and reaching each visit, or study close, and the number dropped out and for what reasons (death, toxicity, treatment failure, withdrew consent). A crosstabulation will show the accrual by quarter and by network and center. A listing of participant discontinuation from study medication will be presented for all randomized participants. Frequency counts and percentages of all participants providing informed consent, randomized/enrolled, completing and discontinuing from the study and study treatment will be presented according to treatment arm and overall. A summary of discontinuations will also be presented by visit. The proportion who discontinued the study early will be compared using chi-square test in the ITT population. The proportion who stopped treatment early on each arm will be compared using chi-square test in the as-treated population.

Protocol Deviations

Important protocol deviations will be listed for all participant providing consent.

1. Those who entered the study even though they did not satisfy the entry criteria.
2. Those who developed withdrawal criteria during the study but were not withdrawn: Patients with visit dates out of visit window will be included in the ITT population and the per protocol population.
3. Randomized participants who report taking excluded medications will be included in the ITT population only.

Protocol deviations will be summarised by center and grouped into the following categories:

- Informed consent deviation
- Study procedure not completed (laboratory, serology, liver enzymes, ecg, biomarkers, patient reported outcome questionnaire, physician global assessment, hrct, diffusion capacity/lung volumes, physical exam, oxygen saturation)
- Study procedure performed incorrectly
- Eligibility violation (FVC ≤40%, change in FVC >10%, FEV1/FVC ratio <0.7, spirometry quality not met, eGFR out of range)

- Study medication dose error/not taken as prescribed error
- Prohibited medication (Nintedanib)
- Visit or procedure performed out of window
- Missed visit
- Other (incorrect equipment, incorrect/missing drug diary, additional/duplicate procedure, procedure/visit done over multiple days, study medication dispensed at incorrect visit, wrong medication (clofazimine), extra blood drawn)

Demographic and Baseline Variables

Summarization by treatment arm will include mean, standard deviation, median, minimum, maximum for continuous variables, and the number and percent for binary and ordinal variables. The age at randomization will be use in analyses. The following is a list of the demographic and baseline variables, recorded at, or shortly, before Randomization or first treatment administration:

- Age (Years)
- Age group
    - 18 to 35 years
    - >35 to 55 years
    - >55 to 75 years
    - >75 years
- Sex
- Ethnicity
- Percent predicted FVC at Screening (Pre-Bronchodilator) as 100*FVC/$FVC_{predicted}$
- Relative Change in Pre-Bronchodilator FVC between Screening and Baseline
- Percent predicted DLCO at Screening (Note that TLCO in SI units × 2.987 = DLCO)
    - DLCO corrected for Hg
- Number (%) with Either DLCO or TLCO Provided at Screening

Treatment Compliance

Pill diaries will be used to determine compliance. The large number of unreturned pill bottles due to COVID-19 and participant non-compliance is the reason that pill counts will not be used. Compliance is defined as the number of pills taken (based on participant reporting in diaries) divided by the number prescribed. Compliance will be summarized and presented in descriptive statistics that include the sample size, mean, standard deviation, median, min, and max. For missing medication time period dates, we will assume the time period is contiguous with the next or the previous time period. Time periods with missing pill numbers prescribed and/or taken will be excluded from the calculation. Participants will be categorized as compliant if at least 80% and not more than 120% of prescribed pills are reported to have been taken. The proportion of participants who are compliant will also be presented for each treatment arm.

Treatment Exposure

Treatment exposure is defined as the time measured in days from when the participant receives the first study drug until the participant either discontinues from treatment or completes the treatment period as planned. The duration of treatment exposure will be listed and summarized by treatment group for the ITT population. Summary statistics will include mean per participant exposure in days, standard deviation, median, min, and max.

Efficacy Analyses

All efficacy variables will be listed by participant within study center. Data will be summarized by treatment arm using descriptive statistics (N, Mean, Standard Deviation, Minimum, and Maximum) for continuous efficacy variables. Number and percent will be presented for categorical efficacy variables. Columns will be Placebo and Pirfenidone. The primary efficacy population will be the ITT population, where all randomized participants are included according to their randomized treatment arm.

The primary efficacy analysis for this trial has a binary composite outcome. (A) Patients will be

categorized based on their decline since baseline in percent predicted FVC of 10% or greater during study period. (B) Patients will also be categorized based on death during study period. The composite primary endpoint is the occurrence of either (A) or (B) during the study period. Comparisons between treatment arms will be performed using a mixed effect logistic regression model, adjusting for the continuous fixed covariate baseline percent predicted FVC and the fixed binary covariate HRCT pattern. Treatment arm will be an independent fixed effect and study center will be a random effect in the analysis (Kahn, 2014). The likelihood-ratio test will be used to test for differences between treatment arms. Treatment arm will be tested at a 2-sided 5% significance level. The components of the primary endpoint will be similarly analyzed.

The adjusted odd ratios together with 95% confidence interval and p-value will be used to quantify the effect of treatment, comparing pirfenidone with placebo as the reference. The data will be coded such that an odds ratio less than 1.0 will indicate that pirfenidone is protective. Assumptions for models will be assessed by examining the plots for the residual values. If data sparseness causes model fitting problems, the model will be simplified by using study network instead of study center as a random effect and/or eliminating the random center (or network) effect.

The longitudinal continuous efficacy endpoints will be presented graphically and numerically. The analysis will use a restricted maximum likelihood-based repeated measures approach with a random slope and intercept model. Visit will be analyzed as weeks since baseline. The analysis will include the fixed, categorical effect of treatment arm, fixed continuous effect of visit, and the treatment arm-by-visit interaction. The baseline value will be included as a fixed covariate. Random effects will be included for participant response and for both visit and intercept. An unstructured covariance structure will be used to model the within-participant errors. If this analysis fails to converge, the other structures that will be tested will include spatial power because visits are unevenly spaced, and compound symmetry. The covariance structure converging to the best fit, as determined by Akaike's information criterion, will be used as the

primary analysis. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom and adjust standard errors. The analysis will use all measurements obtained over the 52-week study period, including those from patients who had discontinued study medication. The model allows for missing data assuming missing at random. If model optimization does not converge, then the random effect will be removed. Significance tests will be based on least squares means using a two-sided $\alpha = 0.05$ with two-sided 95% confidence interval. The primary treatment comparison of slopes will be assessed through the treatment-by-visit interaction coefficient. Assumptions for models will be assessed by examining plots of the residual values.

Analyses of time-to-event outcomes will be performed using mixed effect proportional hazards regression modelling with treatment arm as main effect, and with study center as random effect. The analysis will include the fixed categorical effects of treatment arm and baseline HRCT pattern category. For time-to-event analyses involving spirometry, the baseline value will also be a covariate. The censoring date will be the earlier of the end of study period date or the date the participant was last observed without the event. The actual dates, rather than visit number will be used to calculate time to event, and time will be expressed in weeks. The proportional hazards assumption will be assessed using Kaplan Meier plots and Schoenfeld residuals. A time dependent covariate will be added if proportionality is not supported. The treatment arm difference will be tested at the 2-sided 5% significance level. The adjusted hazard ratio and 95% confidence interval for the hazard ratio will be reported. The data will be coded such that for death and for endpoints that indicate a decline in function, a hazard ratio less than 1.0 will indicate that pirfenidone is protective. Breslow's method for handling ties will be used.

Safety Analyses

The analysis population for the safety analyses will be the As-Treated population. Safety data will be described as follows.

- The sort order of listings will be the participant number within randomized treatment arm.
- Summary information will be grouped according to randomized treatment arm.
- Continuous data will be summarized using descriptive statistics (median, minimum, maximum, mean, standard deviation and number with non-missing data.)
- Categorical data that is not ordered will be summarized with frequencies, percents, and number with non-missing data.
- When calculating the incidence of adverse events, or any sub-classification thereof by treatment, time period, severity, etc., each participant will only be counted once and any repetitions will be ignored; the denominator will be the total population size.

Per ClinicalTrial.gov reporting requirements, three tables summarizing adverse events are required:

- All-Cause Mortality: A table of all anticipated and unanticipated deaths due to any cause, with number and frequency of such events in each arm/group of the trial.
- SAEs: A table of all anticipated and unanticipated serious adverse events, grouped by organ system, with number and frequency of such events in each arm of the trial.
- Other (Not Including Serious) AEs: A table of anticipated and unanticipated events (not included in the serious adverse event table) that exceed an incidence rate of 5% on either arm of the trial, grouped by organ system, with number of such events in each arm/ of the trial.

Adverse Events

An AE is any untoward medical event associated with the use of a drug in humans, whether or not it is considered related to a drug. Adverse events will be coded from the actual term described by the investigator using the Medical Dictionary for Regulatory Activities (MedDRA). Treatment-emergent AEs are events that first occur or increase in severity after the first study drug dose. In case of repeated events in participants, the proportions on each treatment arm will use the number of participants as the denominator, and the number of participants with at least one event as the numerator, so each participants will be counted once. Incidence of AEs by treatment arm will

be compared using chi square tests.

AEs will be summarized by treatment arm, overall and by system organ class, severity, and relationship to study drug.

The following body systems will be summarized according to treatment arm:

- Any Treatment-Emergent AEs
- Blood and Lymphatic System Disorders
- Cardiac Disorders
- Ear and Labyrinth Disorders
- Eye Disorders
- Gastrointestinal Disorders
- General Disorders
- Hepatobiliary Disorders
- Immune System Disorders
- Infections and Infestations
- Injury, Poisoning and Procedural Complications
- Metabolism and Nutrition Disorders
- Musculoskeletal and Connective Tissue Disorders
- Nervous System Disorders
- Psychiatric Disorders
- Renal and Urinary Disorders
- Reproductive System and Breast Disorders
- Respiratory, Thoracic and Mediastinal Disorders
- Skin and Subcutaneous Tissue Disorders
- Surgical and Medical Procedures

AE listings will be sorted by treatment arm, then participant ID, and will include the following:

- Treatment Emergent (yes/no)
- AE # for the participant
- AE Primary Category
- AE Event Term
- AE Start Date
- AE Stop Date
- Days from Start of Study Medication
- Severity Grade
- Expected?
- Relationship to Study Drug
- Study Drug Action Taken
- Outcome

Deaths, Serious AEs and other Significant AEs

Separate summaries will be provided for deaths, serious AEs, and other AEs by treatment arm. Tables will include the number and percent. Summaries will be shown for the number of events and the number of patients experiencing the event. Frequency of SAEs by treatment arm will be compared using chi square tests.

Clinical Laboratory Evaluations

Normal ranges for each study center will be tabulated. Scheduled laboratory tests will be summarized. Unscheduled follow-up tests performed for medical or safety concerns, will be listed only. Test results will be presented using shift tables, showing the change in laboratory values from baseline to each subsequent visit. Accompanying scatter plots and line plots will have a different symbol or color for each treatment arm. All summary analyses will be conducted by treatment group using the as-treated population.

Descriptive statistics will be presented by treatment group and visit week for the laboratory measurements. Continuous laboratory measures will also be compared to reference ranges to determine whether they are abnormally high, low, or normal. The incidence and percent of high, low and normal values will be listed for each of the treatment arms. For subjective (qualitative) laboratory assessments, count and percent of normal and abnormal values will be presented. For laboratory values that are entered as outside the detectible range for a laboratory, the midpoint of the undetectable range will be used for-analyses and nonparametric analyses will be performed.

Prior and Concurrent Medications

Concomitant medication is defined as any medication a participant has received concurrently with study treatment. Prior Medication is defined as any medication or therapies initiated prior to date of first dose of study drug. Medications that are started prior to the date of first dose of study drug and continue after the first dose of drug are considered to be both prior and concomitant medications.

Reporting Conventions

P-values ≥ 0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters, not on the same scale as raw observations (e.g. regression coefficients) will be reported to 3 significant digits.

Unblinding

Study Center Level Unblinding: A minimum number of personnel will see the randomization table and treatment assignments before the study is complete. The treatment assignments will be

blinded to participants and investigators until the end of the study. Emergency un-blinding for AEs may be performed ONLY if the participant's well-being requires knowledge of the participant's treatment assignment. The investigator should make every effort to contact the study team prior to un-blinding a participant's treatment assignment.

Sponsor/Trial Level Unblinding: The study team will remain blinded to treatment assignments until all participants have completed the study and the database has been finalized and locked for analysis.

Quality Assurance of Statistical Programming

Programs and anonymized datasets will be stored on a server at Brigham and Women's Hospital that has nightly back-up. To provide high quality code that is understandable, and allows reproduction of the analysis the following will be explicitly stated:

- Author
- Date of writing
- Description of the code's purpose, e.g. to create the analytic dataset, to analyze the primary outcome variable, etc.
- List of steps or strategy to achieve purpose of code
- References to inputs and outputs
- Reference to any parent code file that runs the child code file
- Population to be used in a table or figure will be explicitly set at the start of a block of code that computes the output.

Statistical software

Statistical analyses will be performed using SAS version 9.4 (Cary, North Carolina.) Supplemental graphics may be generated using JMP, Stata, and/or Excel.

References

1. Liu GF, Zhan X. Comparisons of methods for analysis of repeated binary responses with missing data. Journal of Biopharmaceutical Statistics, 21(3); 371-392, 2011
2. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? BMC Medical Research Methodology 2014, 14:20
3. Agresti A, Hartzel J: Strategies for comparing treatments on a binary response with multi-centre data. Stat Med 2000, 19(8):1115–39.
4. White, Thompson, 2005. Adjusting for partially missing baseline measurements in randomized trials. Statistics in Medicine, Volume 24, pp. 993-1007.
5. Ongoing trials during COVID-19: Quantitative strategies and methodological aspects: Cytel, Inc. (2020)
6. Committee for Human Medicinal Products (CHMP), European Medicines Agency. Points to consider on implications of Coronavirus disease (COVID-19) on methodological aspects of ongoing clinical trials. Draft 25 March 2020
7. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, James P. Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Therapeutic Innovation and Regulatory Science 42(4):303-319. 2008
8. Hernández AV, Steyerberg EW, Habbema JDF. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J Clin Epidemiol. 2004;57(5):454-60.
9. Meyer RD, Ratitch B, Wolbers M, Marchenko O, et al. Statistical issues and recommendations for clinical trials conducted during the COVID-19 pandemic. Statistics in Biopharmaceutical Research. Epub ahead of print.
10. Fries JF, Spitz P, Kraines RG, & Holman HR, Measurement of patient outcomes in arthritis. Arthritis and Rheumatism, 23, 1980, pp.137-145.