KHAA SAP v4

Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Phase 3 Study of Baricitinib in Patients with COVID-19 Infection

NCT04421027

Approval Date: 19-Mar-2021

1. Statistical Analysis Plan: I4V-MC-KHAA(e): A Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Phase 3 Study of Baricitinib in Patients with COVID-19 Infection

Baricitinib (LY3009104) COVID-19 Infection

Study I4V-MC-KHAA (KHAA) is Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group study designed to evaluate the efficacy and safety of 4-mg baricitinib given once daily (QD).

Eli Lilly and Company
Indianapolis, Indiana USA 46285

Protocol I4V-MC-KHAA
Phase 3

Statistical Analysis Plan Version 1 electronically signed and approved by Lilly: 11 July 2020
Statistical Analysis Plan Version 2 electronically signed and approved by Lilly: 30 August 2020
Statistical Analysis Plan Version 3 electronically signed and approved by Lilly: 13 January 2021
Statistical Analysis Plan Version 4 electronically signed and approved by Lilly on date provided below.

Approval Date: 19-Mar-2021 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical analysis plan (SAP) Version 1 was approved prior to first unblinding.

Statistical analysis plan Version 2 was approved prior to Data Monitoring Committee 3 (DMC3).

Statistical analysis plan Version 3 was approved prior to Data Monitoring Committee Meeting 6.

Statistical analysis plan Version 4 was approved prior to unblinding of the study team for the primary outcome database lock.

The main changes incorporated in SAP Version 4 are as follows:

Change	Section	Summary of Changes
Minor changes to increase clarity and correct spelling and grammatical errors	Throughout	To increase accuracy and clarity.
General Considerations	Table KHAA.6.1	Indicated that p-values for primary and key secondary analyses will be reported to as many decimal places as needed.
Analysis Populations	6.1.1	Removed the As Treated Population, as it very similar to the Safety Population.
Definition of Study Baseline	6.1.2	Changed the definition of baseline for efficacy and health outcomes to be relative to the first date that clinical status is assessed rather than first study drug dose. This is consistent with basing these analyses on the ITT population.
Definition of Study Period Time Intervals	Table KHAA.6.2	Removed 2 periods from the Definition of Study Period Time Intervals. Instead, these will be described in analyses, as needed.
Baseline Stratification Variables	6.1.4	Added a section on baseline stratification variables so that these can be referred to throughout the document.
Logistic Regression Model	6.2.1	Changed text regarding logistic regression model, mainly to accommodate multiple imputation.
Analysis of Variance/Covariance	6.2.2	Clarified models for ANOVA
Time-to-Event Analysis	6.2.3	Clarified models for time-to-event analyses.
Proportional Odds Model	6.2.4	Clarified the model for proportional odds analyses

Change	Section	Summary of Changes
Last Observation Carried Forward (LOCF)	6.4.2	Changed mLOCF to LOCF, as it more closely matches the planned analyses. There aren't separate rules in for different intercurrent events.
Multiple Imputation	6.4.3	Filled in details of planned multiple imputation analyses.
Multiple Comparisons/Multiplicity	6.6	Corrected language regarding the multiplicity testing.
Description and Derivation of Primary, Key Secondary, and Associated Supportive Analyses	Table KHAA.6.4	Added derivation details for several variables
Description of Primary, Key Secondary, and Associated Supportive Analysis	Table KHAA.6.5	Minor wording changes for clarity. Changed mLOCF to LOCF. Changed analysis for Oxygen Saturation from MI to LOCF.
Description and Derivation of Other Secondary Outcomes	Table KHAA.6.6	Added analytic details for several analyses.
Description of Other Secondary Analyses	Table KHAA.6.7	Removed analysis of time to recovery by disease duration at baseline. Removed Fine-Gray proportional hazards regression with death as a competing risk, as death is already handled in a manner that considers it a competing risk (i.e., via IETI).
Primary Outcome and Methodology	6.11.1	Changed the final allocation of alpha for Population 1 to 99%. This change was made based on a review of blinded data that showed that Population 2 had a much smaller sample size than originally anticipated.
Key Secondary Efficacy Analyses	6.11.2	Added an analysis of all-cause mortality in the safety population.
Subgroup Analyses	6.14	Updated subgroup analyses to reflect current plans.
Interim Analyses and Data Monitoring	6.16	Removed the paragraph that suggested that a limited number of preidentified individuals may gain access to the limited unblinded data, as there is no plan to do early unblinding for PK/PD analyses.
Sample Size Re-Estimation	6.16.2	Removed the clause "(in addition to the final analysis)" as the 90% allocation of α to Population 1 only applied to the

Change	Section	Summary of Changes
		sample size re-estimation. For the final primary outcome analyses, 99% will be allocated to Population 1.
References	8	Added new references
Multiple Imputation	Appendix 2	Added appendix with details of the multiple imputation methodology
Exploratory Analyses	Appendix 3	Added appendix with additional exploratory analyses

Abbreviations: ANOVA = analysis of variance; DMC = Data Monitoring Committee; DOH = duration of hospitalization; IETI = infinite event time imputation; VFD = ventilator-free days.

4. Study Objectives

4.1. Primary Objective

Objectives	Endpoints
Primary	
To evaluate the effect of baricitinib 4-mg once daily (QD) compared to placebo on disease progression in patients with COVID-19 infection	Proportion of patients who die or require non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including extracorporeal membrane oxygenation [ECMO]) by Day 28 in the following populations: • Population 1 - all randomized patients • Population 2 – patients who, at baseline, require oxygen supplementation (OS 5 and 6) and are not receiving dexamethasone or other systemic corticosteroids for the primary study condition.

The associated estimand for this objective is to measure the effect of treatment with baricitinib as assessed by proportion of patients who die or progress to requiring non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including extracorporeal membrane oxygenation [ECMO]) by Day 28.

4.2. Secondary Objectives

Note that all secondary objectives apply to Population 1.

Objectives	Endpoints
Key Secondary	
To evaluate the effect of baricitinib 4-mg QD compared to placebo on clinical outcomes in patients with COVID-19 infection	Proportion of patients with at least 1-point improvement on NIAID-OS or live discharge from hospital at Day 4, Day 7, Day 10, Day 14
	Number of ventilator-free days (Day 1 to Day 28)
	Time to recovery (NIAID-OS) (Day 1 to Day 28)
	Overall improvement on the NIAID-OS evaluated at Day 4, Day 7, Day 10, Day 14
	Duration of hospitalization (Day 1 to Day 28)
	Proportion of patients with a change in oxygen saturation from <94% to ≥94% from baseline to Day 4, Day 7, Day 10, Day 14
	All-cause mortality (Day 1 through Day 28)
Other Secondary	
To evaluate the effect of baricitinib 4-mg QD compared to placebo on other clinical outcomes in patients with COVID-19 infection	Treatment Period – (Day 1 to Day 28, unless otherwise specified) • Time to recovery (NIAID-OS) by disease duration of <7 days or ≥7 days • Duration of stay in the intensive care unit (ICU) in days • Time to clinical deterioration (one-category increase on the NIAID-OS)

Objectives	Endpoints
	• Time to clinical improvement in one category of the NIAID-OS • Time to resolution of fever, in patients with fever at baseline • Overall improvement on the NIAID-OS evaluated at Day 21, Day 28 • Mean change in National Early Warning Score (NEWS) • Time to definitive extubation • Time to independence from non-invasive mechanical ventilation • Time to independence from oxygen therapy in days • Time to oxygen saturation of ≥94% on room air in days • Number of days with supplemental oxygen use • Number of days of resting respiratory rate <24 breaths per minute
Other Secondary	
	Landmark analyses – Day 4, Day 7, Day 10, Day 14, Day 28 • Proportion of patients in each severity category on the NIAID-OS • Proportion of patients with at least 2-point improvement on the NIAID-OS or live discharge from the hospital • Proportion of patients with at least 1-point improvement on NIAID-OS or live discharge from hospital

Abbreviations: NIAID-OS = National Institute of Allergy and Infectious Diseases ordinal scale; QD = once daily.

4.3. Exploratory Objectives

Exploratory
Exploratory objectives and endpoints may include the following: Serum cytokines, hs-CRP, D-dimer, lactate dehydrogenase (LDH), ferritin (baseline, and during treatment up to Day 14) • Virologic measures • To characterize PK of baricitinib in intubated patients with COVID-19 infection • Long-term (at least Day 60) clinical outcomes.

Abbreviations: hs-CRP = high-sensitivity C-reactive protein; PK = pharmacokinetics.

Planned analyses for long-term clinical outcomes and for patients participating in Addendum 5, who are OS 7 at baseline, are included in Appendix 3.

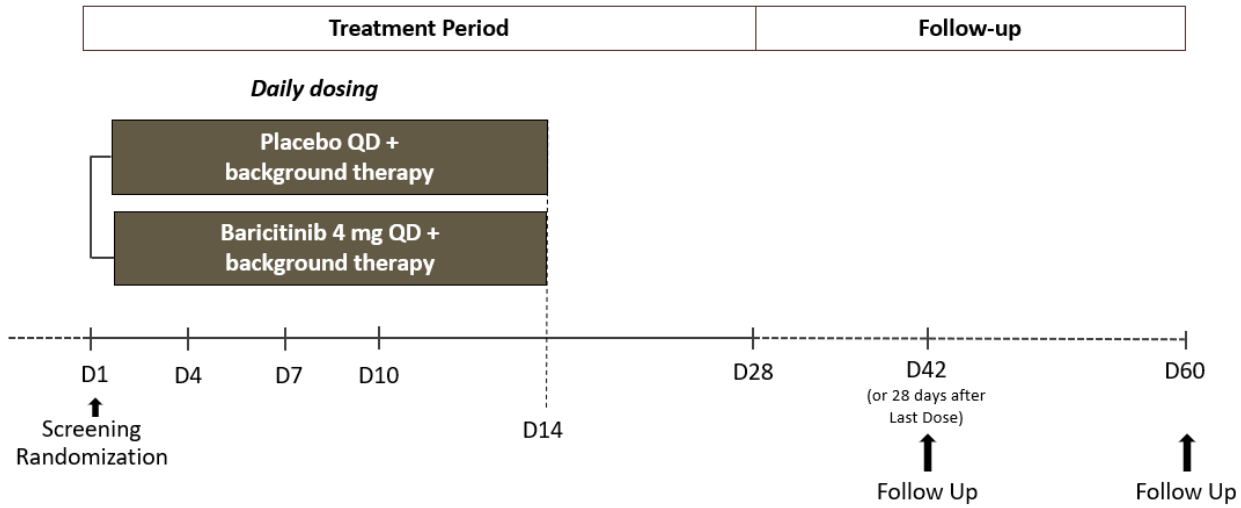
5. Study Design

5.1. Summary of Study Design

Study I4V-MC-KHAA (KHAA) is Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group study designed to evaluate the efficacy and safety of baricitinib 4-mg given once daily (QD). The primary endpoint is the proportion of patients who die or require non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) by Day 28.

The study duration will be up to approximately 60 days over 3 study periods (see Figure KHAA.5.1):

- Screening: on Day 1 prior to dosing
- Treatment period: treatment is administered for up to 14 days, or up to the day of hospital discharge, whichever comes first, followed by treatment evaluations up to Day 28; and
- Follow-up: period starting after treatment evaluation, with a follow-up visit approximately 28 days after last dose of study drug and another follow-up visit at approximately Day 60.



Abbreviations: D = day; QD = once daily.
Note: Dosing occurs from the day of randomization until Day 14, or until hospital discharge, whichever comes first. Placebo or baricitinib are given with background therapy in keeping with local clinical practice for management of COVID-19, as defined in the protocol.

Figure KHAA.5.1. Schema of Study I4V-MC-KHAA.

Patients will be enrolled if they are hospitalized with coronavirus (SARS-CoV-2) infection and meet other study entry criteria. Patients requiring invasive mechanical ventilation (including ECMO) at the time of study entry are not eligible (unless they are enrolled in Addendum 5.0, which allows such patients).

While hospitalized, enrolled patients will receive either baricitinib or placebo until Day 14 or until the day of hospital discharge, whichever comes first.

A follow-up visit approximately 28 days after last dose is required for all randomized patients, including those discharged from the hospital before Day 14. Another follow-up visit occurs at approximately Day 60. The follow-up visits can be conducted as telephone visits.

Discharge from the hospital prior to Day 14 is not considered early discontinuation from the study drug or from the study. All randomized patients, including patients meeting criteria for early discontinuation of study drug, as specified in KHAA Protocol (Section 5.1) should be encouraged to remain in the study for the scheduled study assessments specified in the Schedule of Activities (SoA) (Protocol Section 1.3). Patients who prematurely discontinue from the study should have an early termination visit (ETV) and final follow up visit, if possible, as shown in the SoA.

5.2. Determination of Sample Size

In protocol amendment e, the final amendment, the sample size was updated to approximately 1400 patients based on blinded review of the proportion of patients requiring oxygen supplementation without the use of dexamethasone or systemic corticosteroids at baseline and the potential that concomitant use of systemic corticosteroids may reduce the magnitude of the treatment effect.

The table below describes the power calculations for various scenarios with a total sample size of 1400. This assumes, for illustration, that α_1 for Population 1 is 75% of the total alpha and that 60% of the patients were taking dexamethasone or other corticosteroids at baseline.

Treatment Effect Size in Patients Who are at OS 5 or OS 6 at Baseline		**Combined Effect Size**	**Power for at Least One of the Two Primaries to Succeed**
Patients using dexamethasone or a systemic corticosteroid	**Patients not using dexamethasone or a systemic corticosteroid**		
0.075	0.075	0.075	81%
0.040	0.075	0.054	54%

Abbreviations: NIAID = National Institute of allergy and Infectious Diseases; OS 5 = #5 on the 8-point NIAID ordinal scale - Hospitalized, requiring supplemental oxygen; OS 6 = #6 on the 8-point NIAID ordinal scale – Hospitalized, on noninvasive ventilation or high-flow oxygen devices.

Note: Power estimates were obtained from a custom simulation program.

Amendment e also allowed for the sample size to be increased using an unblinded sample size re-estimation (Gao et al. 2008) during an interim analysis (that occurred in January 2021).

5.3. Method of Assignment to Treatment

Blinding will be maintained in the Phase 3 study.

Method of treatment assignment

Patients who meet all criteria for enrollment will be randomized in a 1:1 ratio (baricitinib 4-mg: placebo) at Day 1.

Randomization will be stratified by these factors:

- disease severity by ordinal scale (OS):
    - hospitalized not requiring supplemental oxygen, requiring ongoing medical care (OS 4)
    - hospitalized requiring supplemental oxygen by prongs or mask (OS 5)
    - hospitalized requiring non-invasive ventilation or high-flow oxygen (OS 6)
- age (<65 years; ≥65 years)
- region (US, Europe, rest of world), and
- dexamethasone and/or other systemic corticosteroid used at baseline for primary study condition (Yes/No). (Note that the original protocol included symptom onset <7 days or ≥7 days prior to randomization as a stratification factor. It was changed to this one in Amendment (c).)

Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web-response system (IWRS).

6. A Priori Statistical Methods

6.1. General Considerations

Unless otherwise specified, efficacy analyses will be conducted on the Intent-to-Treat (ITT) Population. Patients will be analyzed according to the treatment to which they were assigned. All tests of treatment effects will be conducted at a 2-sided alpha level of 0.05, unless otherwise stated.

Table KHAA.6.1. General Methods for Statistical Analysis

Data Type	Purpose	Analysis
Categorical/discrete data	Summary and descriptive analysis	Number of patients with available data (n), either observed or imputed, at the relevant time point, and will be presented as frequency counts and percentages. Percentages will be calculated using the total number of patients in the analysis population included as the denominator. Percentages will generally be presented to 1 decimal place but will not be presented for zero counts.
	Treatment comparison	Logistic regression analysis
Continuous data	Summary and descriptive analysis	Number of observations, mean, standard deviation (SD), standard error of the mean (SEM), median, 1st quartile, 3rd quartile, minimum, and maximum. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, first quartile, median, and third quartile will be reported to 1 more decimal place than the raw data recorded in the database. The SD and SEM may be reported to 2 more decimal places than the raw data recorded in the database. Any exceptions will be noted in the programming specifications. P-values for primary and key secondary analyses will be reported to as many decimal places as needed (≥5).
	Treatment comparison	Analysis of variance (ANOVA), mixed model repeated measures (MMRM), Wilcoxon rank-sum test
Ordinal data	Treatment comparison	Proportional Odds model, Wilcoxon rank-sum test
Time to event data	Treatment comparison	Log-rank test, Kaplan-Meier curves, Cox proportional hazards. As needed: Fine-Gray model, Max-combo test

Note: the detail of the analysis method defined in Section 6.2.

Change from baseline will be calculated as the visit value of interest minus the baseline value. If the baseline value is missing for a particular variable, then it will not be imputed and the change from baseline and percent change (or percent improvement) from baseline will not be calculated.

For all models, patients will be excluded who do not have the necessary covariates to run the model.

For the primary and key secondary analyses that rely on OS values, the patient's worst ordinal score for the day will be used.

The Per Protocol Set (PPS) will be used if needed for post hoc analyses.

Throughout this document, "Day x" refers to Study Day x.

Additional supportive analyses will be performed as needed.

6.1.1. Analysis Populations

The following populations are defined for this study:

Population	Description
Entered	All participants who sign the informed consent form
Intent-to-Treat (ITT)	All participants randomly assigned to study intervention. Participants will be analyzed according to the intervention to which they were assigned.
Per Protocol Set (PPS)	The PPS will include those participants in the ITT population who do not have any identified important protocol violations considered to impact efficacy analyses. Qualifications for, and identification of, significant or important protocol violations will be determined while the study remains blinded, prior to database lock.
Safety	All participants randomly assigned to study intervention and who receive at least 1 dose of study intervention and who did not discontinue from the study for the reason 'Lost to Follow-up' at the first postbaseline visit. Participants will be analyzed according to the intervention they actually received.
Follow-up	All randomized participants who received at least 1 dose of investigational product and have entered the post-treatment follow-up period. Participants will be analyzed according to the intervention to which they received.

A listing of patients who were randomized but did not receive study drug will be prepared. This listing will include baseline information and their last known status.

A sensitivity analysis for the primary efficacy analysis will exclude patients who die within 24 hours of randomization and have Do Not Resuscitate (DNR) or Do Not Intubate (DNI) in ITT analysis populations will be conducted.

6.1.2. Definition of Study Baseline

Unless otherwise specified, for efficacy and health outcomes, baseline is defined as the last nonmissing assessment recorded on or prior to, the first date that clinical status is assessed.

Baseline for safety analyses is described in Table KHAA.6.8.

Change from baseline will be calculated as the visit value of interest minus the baseline value. If a baseline value or the value at the visit is missing for a variable, then the change from baseline is defined as missing.

The baseline of vital signs is defined as the vital signs collated at Day 1 from the case report form (CRF) Assessment for the National Early Warning Score (NEWS) (AVPU1001_VS1001_F1). The post baseline vital signs for analyses other than NEWS is defined as vital signs collated from CRF Vital Signs: Minimum and Maximum (VS1001).

6.1.3. Study Time Intervals

To calculate the length of any time interval or time period in this study, the following formula will be used:

$$Length\ of\ interval\ (days) = End\ Date - Interval\ Start\ Date + 1$$

To convert any time length from days to weeks, the following formula will be used:

$$Length\ of\ interval\ (weeks) = Length\ of\ interval\ (days)/7$$

Only for the purpose of calculating the length of study period time intervals, the words "prior to" in Table KHAA.6.2 should be understood to mean "the day before" while the words "after" should be understood to mean "the day after." For the purpose of determining whether a date/time lies within an interval, these words are intended to convey whether the start or end of the period is inclusive of the specified date.

Table KHAA.6.2. Definition of Study Period Time Intervals

Study Period	Interval Start Definition	Interval End Definition
Screening: All participants who sign informed consent are considered as entering the Screening Period.	Informed consent date	Prior to the start of Treatment Period.
Post-Treatment Follow-Up: All participants who had a follow up visit are considered as entering follow-up period.	After the Treatment Period ends.	The maximum of the last study visit date or study disposition date.

6.1.4. Baseline Stratification Variables

Throughout the SAP, when baseline stratification variables are referred to, they include the following categorical variables:

- baseline disease severity (ordinal scale [OS] 4, OS 5, OS 6),
- age (<65 years, ≥65 years),
- region (US, Europe, rest of world) and
- dexamethasone and/or other systemic corticosteroid used at baseline for primary study condition (Yes/No).

If any of these are redundant for a particular model they will be dropped.

6.2. Statistical Methods

The following will be applied to the ITT population as described in Section 6.1.1.

6.2.1. Logistic Regression Model

Treatment comparisons of discrete/binary efficacy variables between treatment groups will be made using a logistic regression analysis adjusted for baseline stratification variables (Section 6.1.4). The p-value and odds ratio with 100(1-α)% confidence interval (CI) from the logistic model will be reported.

Point estimates for proportions and difference in proportions will be presented. Confidence intervals and p-values for the estimates of the differences of the proportions and will be presented. For last-observation-carried-forward (LOCF) analyses the CIs will be based on the Newcombe-Wilson method without continuity correction. For multiple imputation analyses, percentages will be based on the observed percentages; CIs will be based on the Wald method without continuity correction.

6.2.2. Analysis of Variance

Treatment comparisons of quantitative efficacy and health outcome variables will be made using analysis of variance (ANOVA) adjusted for baseline stratification variables (Section 6.1.4). Where appropriate, baseline value for the endpoint of interest will also be included in the model. Type 3 tests for least squares (LS) means will be used for statistical comparison between treatment groups. The LS mean difference, standard error, p-value, and 95% CI will also be reported.

6.2.3. Time-to-Event Analysis

The primary analysis for time-to-event analysis will be the unstratified log-rank test. Treatment comparisons for time-to-event analysis may also be analyzed using a Cox proportional hazards model adjusted for baseline stratification variables (Section 6.1.4). The hazard ratio with 95% CIs will be reported. Kaplan-Meier curves may also be produced. Diagnostic tests for checking the validity of the proportional hazards assumption may be performed. If the assumption of proportional hazards is not justified, a statistical model capable of handling nonproportional hazards will be explored to assess treatment effect, such as a max-combo test (Lee 1996), restricted mean survival time model (Royston and Parmar 2013), and win ratio analysis (Pocock et al. 2012).

6.2.4. Proportional Odds Model

The proportional odds model will be used to assess the overall improvement in the OS. Treatment comparisons will be adjusted for baseline stratification variables (Section 6.1.4). The treatment odds ratio estimated from the model will be presented along with the CI and p-value.

6.2.5. Wilcoxon Rank-Sum Test

The Wilcoxon rank-sum test, without continuity correction, will be used for analyzing endpoints associated with number of days, such as number of ventilator free days. The p-value for each treatment group will be provided.

6.2.6. Mixed-Effects Model of Repeated Measures

Mixed model repeated measures analyses were performed to mitigate the impact of missing data. This approach assumed that missing observations were missing-at-random (missingness is related to observed data) during the study and borrowed information from patients in the same treatment arm taking into account both the missingness of data and the correlation of the repeated measurements.

MMRM model will be used a restricted maximum likelihood (REML) estimation. The model will be adjusted with baseline stratification variables (as appropriate), and treatment-by-landmark days-interaction as fixed categorical effects and baseline and baseline-by-landmark days-interaction as fixed continuous effects. An unstructured (co)variance structure will be used to model the between- and within-patient errors. If this analysis failed to converge, the heterogeneous autoregressive [ARH(1)], followed by the heterogeneous compound symmetry (CSH), followed by the heterogeneous Toeplitz (TOEPH), followed by autoregressive [AR(1)], followed by compound symmetry (CS), was used. The Kenward–Roger method was used to estimate the degrees of freedom. Treatment LS mean were estimated within the framework of the MMRM using type 3 sums of squares. Differences in LS mean between each dose of baricitinib and placebo (and associated p-values, standard errors, and 95% CI) were used for statistical inference.

6.3. Adjustments for Covariates

Unless otherwise specified, the statistical analysis models will control for baseline stratification variables (see Section 6.1.4).

6.4. Handling of Dropouts or Missing Data

The following imputation rules will be used for subjects who are lost to follow-up, withdrew from the study early, or do not have further outcome data available after discharge for any reason or death.

Efforts to use all available data and minimize missing data imputation will be considered. For clinical outcomes related to National Institute of Allergy and Infectious Diseases ordinal scale (NIAID-OS), the outcome may be derived using pre-specified relevant clinical data before missing data imputation approaches applied.

6.4.1. Infinite Event Time Imputation (IETI)

For some time-to-event endpoints, if there are competing risks to the event of interest, then the event times censored due to the competing risk will be imputed as infinite.

This imputation method will be applied if the event of interest is in the opposite direction of death (e.g., recovery or improvement). For time to recovery or time to improvement, all deaths

within 28 days will be considered censored at Day 28 with respect to time to event of interest. Conceptually, a death corresponds to an infinite time to event of interest, but censoring at any time greater than or equal to Day 28 gives the same answer as censoring at Day 28; both correspond to giving death the worst rank.

6.4.2. Last Observation Carried Forward (LOCF)

Some analyses (in particular for quantitative or ordinal scale measures) will use the LOCF approach. Intermittent or terminally missing data will be filled in by carrying forward the last available measurement prior to the missing data. This methodology will only be utilized for patients who had both a baseline and a postbaseline measurement.

6.4.3. Multiple Imputation

A multiple imputation method will be used to impute the missing NIAID-OS scores (Rubin, 1996). The multiply-imputed datasets will be used for the primary analyses of several endpoints involving the NIAID-OS scores. A total of 100 multiply-imputed datasets will be generated. For random number generation, the seed will be set 3012021. The multiple imputation will be performed in a stratified manner with the imputation performed separately for each of the following levels: (1) baricitinib patients with baseline NIAID-OS of 5 or 6 and baseline steroid use, (2) baricitinib patients with baseline NIAID-OS of 5 or 6 and no baseline steroid use, (3) other baricitinib patients, (4) comparator patients with baseline NIAID-OS of 5 or 6 and baseline steroid use, (5) comparator patients with baseline NIAID-OS of 5 or 6 and no baseline steroid use, and (6) other comparator patients. In all these strata, baseline steroid use is defined as in Section 6.1.4. These strata are slightly reduced from what might be expected in order to keep a sufficient number of patients in each stratum. Imputation will be performed using a Markov model where each transition to a future state is dependent on only the previous state. This approach is described in detail in Appendix 2.

The intended estimand for the multiple imputation approach is based on the treatment policy strategy for handling intercurrent events (ICH E9R [ICH 2017]). In this strategy the value of the NIAID-OS score is the value of interest regardless of any intercurrent events that occurred. The NIAID-OS includes a state for death, and thus it is meaningful even for patients who have died.

If the between-imputation variance calculated while combining some analysis results of the multiply-imputed datasets is zero then that specific analysis will not be performed using multiple imputation. Instead, the observed data will be used to perform the same analysis.

6.5. Multicenter Study

Study KHAA is a multicenter study; the following countries will conduct trials:

Countries		
ARGENTINA	ITALY	RUSSIAN FEDERATION
BRAZIL	JAPAN	SPAIN
GERMANY	KOREA	UNITED KINGDOM
INDIA	MEXICO	UNITED STATES

6.6. Multiple Comparisons/Multiplicity

Multiplicity controlled analyses will be performed on the primary and key secondary endpoints to control the overall family-wise Type I error rate at a 1-sided α level of 0.025. The graphical multiple testing procedure described in Bretz et al. (2011) will be used. The graphical approach is a closed testing procedure; hence it strongly controls the family-wise error rate (FWER) across all endpoints (Alosh et al. 2014).

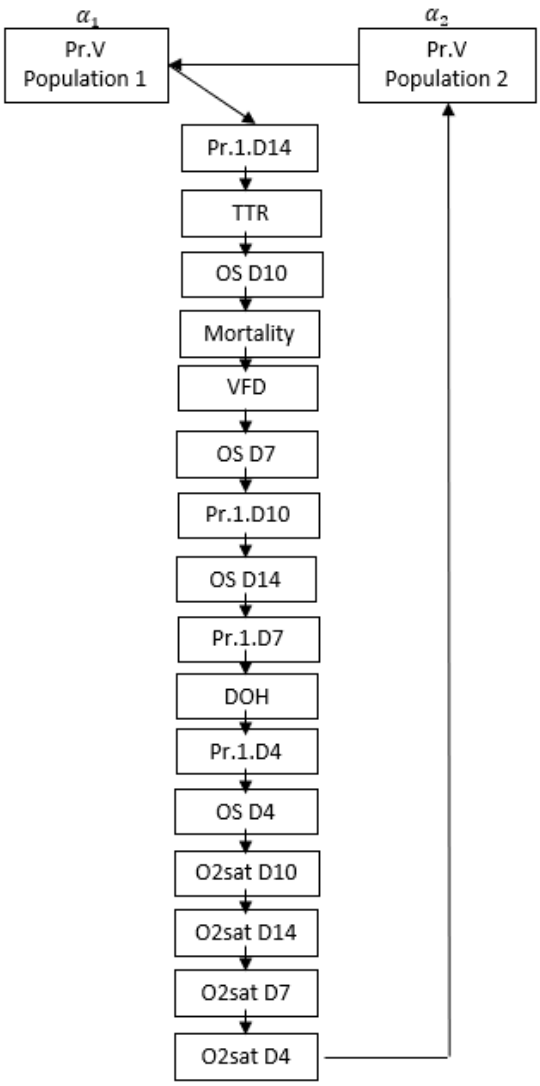
See Figure KHAA.6.1 for the graphical testing scheme for this study. This scheme allows for some α to be sent to the key secondary analyses if Population 1 achieves statistical significance (see next paragraph). Note that for the 2 primary populations the scheme is slightly different from what is in the protocol. As the testing scheme only impacts interpretation of the results and not the actual analysis, there is no plan to update the protocol.

If Population 2 is rejected at α_2, Population 1 will be tested at level α, otherwise Population 1 will be tested at level α_1. If Population 1 is rejected, each of the gated secondaries will be tested in order until one is not rejected as depicted in Figure KHAA.6.1. The key secondaries will be tested at level α_1 if Population 1 was rejected or at level α if both Population 1 and Population 2 were rejected. Finally, if Population 1 and all secondaries are rejected at level α_1, but Population 2 was not rejected at level α_2, Population 2 will be tested again at level α.

The testing scheme will provide a strong control of Type-I error for the study at a 1-sided 0.025 level and is based on the closed testing principle (Marcus et al. 1976). The testing scheme is parametric, where the known correlation is accounted for between the 2 test statistics z_1 and z_2. The initial significance levels for the Primary endpoints Population 1 (α_1) and Population 2 (α_2) will be computed using the fact that their test statistics have a correlation of $\sqrt{p}$ where p is the percentage of Population 2 within Population 1 observed in the final analysis data set and a 0.99 weight on Population 1 and 0.01 weight on Population 2. The two significance levels will be computed using the R package gMCP.

The primary and key secondary endpoints to be tested are listed in Section 6.11.1 and Section 6.11.2, respectively.

The graphical testing scheme is considered final.



Abbreviations: Pr.V: Proportion of patients who die or require non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) by Day 28.
TTR: Time to recovery, where recovery is defined as clinical status in states 1, 2, or 3 of the 8-point ordinal scale, censored at Day 28.
Pr.1.D4/7/10/14: proportion of patients with at least 1-point improvement on the NIAID-OS or live discharge from the hospital at Day 4, Day 7, Day 10, and Day 14.
VFD: number of ventilator-free days (Day 1 to Day 28). OS D4/7/10/14: overall improvement on the NIAID-OS evaluated at Day 4, Day 7, Day 10, and Day 14. DOH: duration of hospitalization (Day 1 to Day 28).
O2sat D4/7/10/14 proportion of patients with a change in oxygen saturation from <94% to ≥94% from baseline to Day 4, Day 7, Day 10, Day 14. Mortality: All-cause mortality (Day 1 to Day 28).
Additional details are included in Sections 6.11.1 and 6.11.2.

Figure KHAA.6.1. Graphical testing scheme for Study KHAA.

6.7. Patient Disposition

An overview of patient populations will be summarized by treatment group. Frequency counts and percentages of patients excluded prior to randomization by primary reason for exclusion will be provided for patients who failed to meet study entry requirements during screening.

Patient disposition will be summarized using the ITT population. Frequency counts and percentages of patients will be summarized by treatment group by the following dispositions:

- dosing (study drug) period disposition:
    - ongoing dosing (study drug) period
    - discontinued dosing (study drug) period
    - completed dosing (study drug) period
- treatment period disposition:
    - ongoing treatment period
    - discontinued treatment period (reason will be summarized)
    - completed treatment period
- study disposition:
    - ongoing
    - discontinued (reason will be summarized)
    - completed

A listing of patient disposition will be provided for all randomized patients, with treatment assignment, the extent of their participation in the study, and the reason for discontinuation.

6.8. Patient Characteristics

6.8.1. Demographics

Patient demographics will be summarized as described above. The following demographic information will be included:

- age
- age group (<65 vs. ≥65)
- age group (<65, ≥65 to <75, ≥75 to <85, ≥85)
- gender (male, female)
- race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)
- geographic region (US, Europe, Rest of World)
- country

- weight (kg)
- weight category (<60 kg, ≥60 to <100 kg, ≥100 kg)
- height (cm)
- body mass index (BMI) (kg/m^2)
- body mass index category (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2), and
- baseline renal function status: impaired (estimated glomerular filtration rate [eGFR] <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2).

A listing of patient demographics will also be provided for the ITT population.

6.8.2. Baseline Disease Characteristics

The below baseline disease information (although not inclusive) will be categorized and presented for baseline coronavirus disease 2019 (COVID-19) clinical characteristics, baseline health outcome measures, and other baseline demographic and disease characteristics as described above:

- duration of symptoms prior to enrollment (≥7 days or <7 days)
- World Health Organization (WHO) ordinal scale
- baseline disease severity:
    - hospitalized not requiring supplemental oxygen, requiring ongoing medical care
    - hospitalized requiring supplemental oxygen by prongs or mask
    - hospitalized requiring noninvasive ventilation or high-flow oxygen
- inflammatory biomarkers:
    - C-reactive protein, high-sensitivity (hs-CRP)
    - Ferritin
    - D-Dimer
    - Lactate dehydrogenase (LDH)
- Prior therapy of interest: nonsteroidal anti-inflammatory drugs (NSAIDs), Antivirals, Antibiotics, Immunosuppressants (anti-malarials, corticosteroids and others)
- Symptom onset (<7 days, ≥7 days)
- Disease duration of symptoms prior to enrollment (<7 days, ≥7 days)
- Renal function status (impaired, no impaired)
- Preexisting comorbid conditions (None, Any; obesity; diabetes, chronic respiratory disease, hypertension)

- dexamethasone and/or other systemic corticosteroid used at baseline for primary study condition (Yes/No)
- remdesivir (Yes/No)

6.8.3. Historical Illness and Preexisting Conditions

Historical illnesses are defined as those conditions recorded in the Preexisting Conditions and Medical History electronic case report form (eCRF) or from the Prespecified Medical History: Comorbidities eCRF with an end date prior to the informed consent date. The number and percentage of patients with selected historical diagnoses will be summarized by treatment group using the ITT population. Historical diagnoses will be categorized using the Medical Dictionary for Regulatory Activities (MedDRA; most current available version) algorithmic Standardized MedDRA Queries (SMQs) or similar pre-defined lists of Preferred Terms (PTs) of interest.

Preexisting conditions are defined as those conditions recorded in the Pre-existing Conditions and Medical History eCRF, or the Prespecified Medical History: Comorbidities eCRF with a start date and time prior to the informed consent and with a stop date that is after the informed consent date or have no stop date (ongoing). Adverse events (AEs) are recorded in the eCRFs. For events recorded on the AE page, we considered it as a preexisting event if its onset date was before the first dose date. For events occurring on the day of the first dose of study treatment, the date and time of the onset of the event will both be used to determine if the event was preexisting. Conditions with a partial or missing start date (or time if needed) will be assumed to be 'not preexisting' unless there is evidence, through comparison of partial dates, to suggest otherwise. Preexisting conditions will be categorized using the SMQs or similar predefined lists of PTs of interest. Frequency counts and percentages of patients with selected preexisting conditions will be summarized by treatment group using the ITT population.

The number and percentage of participants using preferred terms of prior medications will be presented.

The following historical illness and preexisting conditions will be presented:

- preexisting comorbid conditions:
    - none, any
- preexisting comorbid conditions of interest
    - obesity
    - diabetes (Type I and Type II)
    - chronic respiratory disease
    - hypertension

6.9. Treatment Compliance

As all study drug doses will be administered at the study site, treatment compliance will not be reported.

6.10. Concomitant Therapy

Medications will be classified into anatomical therapeutic chemical (ATC) drug classes using the latest version of the WHO drug dictionary. Medication start and stop dates will be compared to the date informed consent is obtained to allow medications to be classified as concomitant.

Prior medications are those medications that start and stop prior to the date of the first dose of study treatment. *Concomitant medications* are those medications that start before, on, or after the first day of study treatment and continue into the treatment period. For all summary tables of concomitant medications, PTs of concomitant medication will be sorted by descending frequency in the LY total arm.

If dose-related corticosteroid analyses are conducted, the information in Appendix 1 will be used as reference for a conversion factors for each corticosteroid medication identified during the study, instructions for selecting corticosteroids, and the manual review process.

Table KHAA.6.3. Summary Tables Related to Concomitant Medications

Analysis	Details
Prior medications	Number and percentage of participants using Preferred Terms of prior medication • Ordered by decreasing frequency No inferential statistics
Concomitant medications	Number and percentage of participants using Preferred Terms of concomitant medication • Ordered by decreasing frequency No inferential statistics

6.11. Efficacy Analyses

The general methods used to summarize efficacy data, including the definition of baseline value for assessments are described in Section 6.1.

Table KHAA.6.4 includes the descriptions and derivations of the primary, key secondary, and associated supportive analysis efficacy outcomes.

Table KHAA.6.5 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for efficacy analyses.

Table KHAA.6.6 includes the descriptions and derivations of the other secondary and associated supportive analysis efficacy outcomes.

Table KHAA.6.7 provides the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for efficacy analyses.

6.11.1. Primary Outcome and Methodology

The primary comparison of interest is the proportion of patients who die or require non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) by Day 28. The primary comparison will be performed on two different populations:

- Population 1 – all randomized patients
- Population 2 – patients who, at baseline, require oxygen supplementation and are not receiving dexamethasone or other systemic corticosteroids for the primary study condition.

Patients on non-invasive ventilation/high-flow oxygen at baseline will be counted toward this endpoint if they progressed to invasive mechanical ventilation. Treatment comparisons between baricitinib 4-mg QD + background therapy and placebo + background therapy will be made using logistic regression with baseline stratification factors and treatment group in the model.

- Pr.V: Proportion of patients who died or require non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) by Day 28. For patients who start with ventilation at baseline, the patients need to be worsening in symptom (at least 1-point worsening in NIAID-OS) to be counted.

For the primary comparison involving two different populations, the alpha will be split between the two populations such that 99% of alpha is assigned to Population 1 and the rest to Population 2. The primary endpoint will be met if any one or both of these two populations show a significant treatment effect.

If the sample size is increased as a result of the interim analysis, the Cui, Hung, and Wang (CHW) procedure (Cui et al. 1999) will be applied to the primary endpoints for the two populations and the secondary endpoints to control the type I error at a one-sided α=0.025 significance level. The CHW method ensures strong control of type I error when the sample size is increased in a data dependent manner. Details about the CHW statistics are provided in Section 6.16.2.4.

If the sample size is increased as a result of the interim analysis, an unadjusted point estimate for both of the primary efficacy analysis will be calculated and reported. A median unbiased point estimate and a stage-wise adjusted confidence interval for the primary efficacy analysis will be calculated and reported based on the approach described by Brannath and colleagues (Brannath et al. 2009) to assess sensitivity of the point estimate.

No CHW adjustments will be applied to the endpoints if the recommended increased number of patients from the sample size re-estimation had already been enrolled before the completion of the interim.

Table KHAA.6.4 and Table KHAA.6.6 include the descriptions and derivations of the primary and secondary outcomes.

Table KHAA.6.5 and Table KHAA.6.7 provide the detailed analyses including analysis type, method and imputation, population, time point, and comparisons for efficacy analyses.

6.11.2. Key Secondary Efficacy Analyses

Secondary comparisons of interest (key secondaries) include:

- TTR: Time to recovery, where recovery is defined as clinical status in states 1, 2, or 3 of the 8-point ordinal scale, censored at Day 28
- Pr.1.D14: proportion of patients with at least 1-point improvement on the NIAID-OS or live discharge from the hospital at Day 4, Day 7, Day 10, and Day 14
- VFD number of ventilator-free days (Day 1 to Day 28)
- NIAID-OS D4/7/10/14: overall improvement on the NIAID-OS evaluated at Day 4, Day 7, Day 10, and Day 14
- DOH: duration of hospitalization (Day 1 to Day 28)
- O2sat D4/7/10/14 proportion of patients with a change in oxygen saturation from <94% to ≥94% from baseline to Day 4, Day 7, Day 10, Day 14; and
- All-cause mortality (Day 1 to Day 28).

Table KHAA.6.4. Description and Derivation of Primary, Key Secondary, and Associated Supportive Analyses

Measure	Description	Variable	Derivation / Comment	Definition of Missing
NIAID-OS	Using data from the COVID-19 Clinical Status Assessment, results will be calculated for ordinal scales currently being used in other studies, and in this study, to measure clinical outcomes in patients treated for COVID-19, in particular the NIAID-OS. The NIAID-OS is as follows: 1. Not hospitalized, no limitations on activities 2. Not hospitalized, limitation on activities and/or requiring home oxygen 3. Hospitalized, not requiring supplemental oxygen – no longer requires ongoing medical care: (This would include those kept in hospital for quarantine/infection control, awaiting bed in rehabilitation facility or homecare, etc.) 4. Hospitalized, not requiring supplemental oxygen – requiring ongoing medical care (COVID-19 related or otherwise) 5. Hospitalized, requiring supplemental oxygen 6. Hospitalized, on noninvasive ventilation or high-flow oxygen devices 7. Hospitalized, on invasive mechanical ventilation or ECMO 8. Death	Patients who die or progress to requiring ventilation	Proportion of patients who progress to ventilation or death. In order to be counted as having this event: • Patients who start at OS 4 or 5 must progress to at least OS 6. • Patients who start at OS 6 must progress to at least OS 7.	Missing if NIAID-OS is missing and the variable status cannot be identified using remaining data.
		Ventilator-free days (days free of invasive mechanical ventilation)	Total number of days patients are alive and have a NIAID-OS less than 7	Missing if patient's NIAID-OS is missing for any day.
		1-point improvement in OS (at days 4, 7, 10 and 14)	Proportion of patients with baseline NIAID-OS minus NIAID-OS ≥ 1	Missing if either baseline NIAID-OS is missing or if the observed NIAID-OS is missing on the day of assessment of 1-point improvement
		Time to recovery	Time to reach NIAID-OS 1, 2, or 3 for the first time. The date reached is the first full day that OS 1, 2, or 3 is the patient's maximum OS for the day.	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data
		Overall improvement (at days 4, 7, 10, and 14)	NIAID OS	Missing if observed NIAID-OS is missing
		Duration of hospitalization	Total number of days patients have a NIAID-OS of 4, 5, 6, or 7. If the patient died on or prior to Day 28, their duration of hospitalization is considered to be 28 days. On Day 1 the patient's worst ordinal scale between baseline and the rest of Day 1 will be used to define their hospitalization status.	Missing if patient's NIAID-OS is missing for any day

Measure	Description	Variable	Derivation / Comment	Definition of Missing
		All-cause mortality	Time to death by Day 28. Patients included: ITT population Start date: Baseline date Event date: date of death if death is within 28 days of their baseline date. (That is, if baseline is Day 1, any deaths up to and including Day 28 are included). Censor date: Date of last non-missing OS or visit that is on or prior to Day 28. If there is other information that makes it clear that they were alive at Day 28, they will be censored at Day 28.	Missing if the patient's maximum NIAID-OS is not 8 and patient's NIAID-OS is missing for all days after a particular day
Oxygen Saturation	Measure of the oxygen level of the blood measured by pulse oximetry	Oxygen saturation from <94% at baseline to ≥94% (at days 4, 7, 10 and 14)	Patients included: randomized patients whose oxygen saturation (based on NEWS) is < 94% at baseline. A patient is a responder at Day x if their oxygen saturation is ≥ 94% and their OS is not higher than their baseline OS. Otherwise, they are a non-responder. Patients who are missing data at Day x due to having died are considered to be non-responders.	Missing if either baseline or observed oxygen saturation measurement is missing

Abbreviations: ECMO = extracorporeal membrane oxygenation; NIAID-OS = National Institute of Allergy and Infectious Diseases ordinal scale.

Table KHAA.6.5. Description of Primary, Key Secondary, and Associated Supportive Analyses

Measure	Variable	Population	Analysis Method	Time Point	Analysis Type
NIAID-OS	Proportion of patients who died or progressed to non-invasive ventilation or high-flow oxygen (OS ≥ 6)	ITT: Population 1 - all randomized patients	Logistic regression using MI	By Day 28	Primary analysis
		ITT: Population 2 – patients who, at baseline, require oxygen supplementation and are not receiving dexamethasone or other systemic corticosteroids for the primary study condition.	Logistic regression using MI	By Day 28	Primary analysis
		ITT	Logistic regression using LOCF	By Day 28	Supportive analysis
		ITT: Population 2	Logistic regression using LOCF	By Day 28	Supportive analysis
		Safety Population	Logistic regression using MI	By Day 28	Supportive Analysis
		ITT Population, excluding patients who die within 24 hours of baseline and have Do Not Resuscitate (DNR) or Do Not Intubate (DNI) orders	Logistic regression using MI	By Day 28	Sensitivity Analysis
	Ventilator free days (VFD)	ITT	ANOVA using MI	By Day 28	Key secondary analysis
			Wilcoxon rank-sum test using LOCF	By Day 28	Supportive analysis
	Proportion of patients with 1-point improvement	ITT	Logistic regression using MI	Day 4, 7, 10, 14	Key secondary analysis
			Logistic regression using LOCF	Day 4, 7, 10, 14	Supportive analysis

Measure	Variable	Population	Analysis Method	Time Point	Analysis Type
	Time to recovery	ITT	Log-rank test using IETI	By Day 28	Key secondary analysis
			Cox proportional hazards model using IETI	By Day 28	Supportive analysis
			Max-combo test using IETI (if needed)	By Day 28	Supportive analysis
NIAID-OS	Overall improvement	ITT	Proportional odds model using MI	Day 4, 7, 10, 14	Key secondary analysis
			Proportional odds model using LOCF	Day 4, 7, 10, 14	Supportive analysis
	Duration of hospitalization	ITT	ANOVA using MI	By Day 28	Key secondary analysis
			Wilcoxon rank-sum test LOCF	By Day 28	Supportive analysis
	All-cause mortality	ITT	Log-rank test	By Day 28	Key secondary analysis
			Cox proportional hazards model	By Day 28	Supportive analysis
		Safety	Log-rank test	By Day 28	Supportive analysis
			Cox proportional hazards model	By Day 28	Supportive analysis
Oxygen Saturation	Proportion of patients with change in oxygen saturation from <94% at baseline to ≥94% at the observed time point	ITT	Logistic regression using LOCF	Day 4, 7, 10, 14	Key secondary analysis

Abbreviations: ANOVA = analysis of variance; IETI = infinite event time imputation; ITT = Intent-to-Treat; MI = multiple imputation; LOCF = last observation carried forward; NIAID-OS = National Institute of Allergy and Infectious Diseases ordinal scale; PPS = Per Protocol Set; VFD = ventilator-free days.

6.11.3. Other Secondary Efficacy Analyses

Table KHAA.6.6 summarizes other secondary efficacy analyses.

Table KHAA.6.6. Description and Derivation of Other Secondary Outcomes

Measure	Variable	Derivation / Comment	Definition of Missing
NIAID-OS	Time to recovery by disease duration of <7 days or ≥7 days prior to enrollment?	Time to reach NIAID-OS 1, 2, or 3 for the first time by disease duration of <7 days or ≥7 days.	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data or the baseline disease duration is missing.
	Time to clinical deterioration	First time patient's observed NIAID-OS – baseline NIAID-OS ≥ 1	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data.
	Time to clinical improvement in one category	First time patient's baseline NIAID-OS - observed NIAID-OS ≥ 1 for each NIAID-OS baseline category.	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data.
	Overall improvement	NIAID-OS	Missing if either baseline or observed NIAID-OS is missing.
	Time to independence from noninvasive mechanical ventilation	Patients included: patients whose baseline OS is 6. Start date: Baseline date Event date: First date patient achieves OS 5 or less Censor date (for patients who haven't died but never achieve OS 5 or less by Day 28): Date of last non-missing OS that is on or prior to Day 28. Censor date for patients who die on or prior to Day 28: Date of what would have been their Day 28.	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data.
	Time to independence from oxygen therapy in days	Patients included: patients whose baseline OS is 5 or 6. Start date: baseline date Event date: first date patient achieves OS 4 or less Censor date (for patients who never achieve OS 4 or less by Day 28): Date of last non-missing OS that is on or prior to Day 28. Censor date for patients who die on or prior to Day 28: Date of what would have been their Day 28 visit.	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data.

Description and Derivation of Other Secondary Outcomes

Measure	Variable	Derivation / Comment	Definition of Missing
	Number of days with supplemental oxygen use	Total number of days patients' NIAID-OS is ≥5	Missing if any observed NIAID-OS is missing.
	Proportion of patients in each severity category	Total proportion of patients in each NIAID-OS category at particular time points	Missing if observed NIAID-OS is missing at particular time point.
	Patients with at least 2-point improvement on the NIAID-OS or live discharge from the hospital	Proportion of patients whose baseline NIAID-OS - observed NIAID-OS ≥ 2	Missing if observed NIAID-OS is missing or if observed NIAID-OS is ≥3 and baseline NIAID-OS is missing
	Patients with at least 1-point improvement on the NIAID-OS or live discharge from the hospital	Proportion of patients whose observed NIAID-OS is ≤ 2 or whose baseline NIAID-OS - observed NIAID-OS ≥ 1	Missing if observed NIAID-OS is missing or if observed NIAID-OS is ≥3 and baseline NIAID-OS is missing
Intensive Care Unit (ICU) stay	Duration of stay in the intensive care unit (ICU)	Total number of days spent in the ICU plus days from death up to and including Day 28.	Missing if status of ICU stay is missing for any day for the patients and the patient is alive.
Fever	Time to resolution of fever	Time to patients being free of fever for the first time for patients who had fever at baseline. Fever resolution is defined by: • ≤36.6°C (axilla, forehead) • ≤37.2°C (oral cavity), or • ≤37.8°C (rectum, ear, temporal artery) If temperature is assessed multiple times in a day, fever will be considered to be resolved if all of the measurements fall in the "fever resolution" ranges defined above	Missing if fever status is missing or baseline fever status is missing and time to event of interest cannot be identified using remaining data.
National Early Warning Score (NEWS)	Mean change	Mean of observed NEWS – NEWS at baseline NEWS will be calculated using aggregate score approach in Royal College of Physicians [WWW].	Missing if either baseline or observed NEWS component is missing.

Description and Derivation of Other Secondary Outcomes

Measure	Variable	Derivation / Comment	Definition of Missing
Extubation	Time to definitive extubation	Patients included: randomized patients who progress to OS 7 at any time on or prior to Day 27. Start date: first date that patient progressed to OS 7 (this could be their baseline date if they progressed the first day on study, after baseline) Event date: First date when patient is removed from invasive mechanical ventilation (i.e., OS 6 or less) *for the last time* on or prior to Day 28. Censor date for patients who are OS 7 at their last observation or by Day 28: Date of last observation. Censor date for patients who die after their Start Date: Date of what would have been their Day 28.	Missing if NIAID-OS is missing and time to event of interest cannot be identified using remaining data.
Oxygen Saturation	Time to oxygen saturation ≥ 94% on room air in days	Patients included: randomized patients whose oxygen saturation (based on NEWS) is < 94% at baseline. Event Date: the first date that their oxygen saturation is ≥ 94% and their OS is ≤ 4.	Missing if NIAID-OS is missing or oxygen saturation status is missing and time to event of interest cannot be identified using remaining data.
Resting respiratory rate	Number of days of resting respiratory rate <24 breaths per minute	Total number of days when patients' resting respiratory rate <24 breaths per minute. This will be based on values from NEWS.	Missing if Patient's resting respiratory rate is missing for at least 1 day.

Abbreviations: ICU = intensive care unit; NIAID-OS = National Institute of Allergy and Infectious Diseases ordinal scale.

Table KHAA.6.7. Description of Other Secondary Analyses

Measure	Variable	Population	Analysis Method	Time Point	Analysis Type
NIAID-OS	Time to recovery by disease duration at baseline: < 7 days and ≥7 days	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
	Time to clinical deterioration	ITT	Log rank test	By Day 28	Secondary analysis
			Cox Proportional Hazards model	By Day 28	Supportive analysis
	Time to clinical improvement in one category	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
	Overall improvement	ITT	Proportional Odds model using MI	Day 21 and Day 28	Secondary analysis
			Proportional Odds model using LOCF	Day 21 and Day 28	Supportive analysis
	Time to independence from noninvasive mechanical ventilation	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
	Time to independence from oxygen therapy in days	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
	Number of days with supplemental oxygen use	ITT	ANOVA with MI	By Day 28	Secondary analysis
			Wilcoxon Rank Sum Test using LOCF	By Day 28	Supportive analysis
	Proportion of patients in each severity category	ITT	Observed	Day 4,7,10,14	Secondary analysis
			Descriptive using LOCF	Day 4,7,10,14	Supportive analysis
	Patients with at least 2-point improvement on the NIAID-OS or live discharge from the hospital	ITT	Logistic regression using MI	Day 4, 7, 10, 14, 21, 28	Secondary analysis
			Logistic regression using LOCF	Day 4,7,10,14, 28	Supportive analysis
	Patients with at least 1-point improvement on the NIAID-OS or live discharge from the hospital	ITT	Logistic regression using MI	Day 28	Secondary analysis
			Logistic regression using LOCF	Day 28	Supportive analysis

Measure	Variable	Population	Analysis Method	Time Point	Analysis Type
Intensive Care Unit (ICU) stay	Duration of stay in the intensive care unit (ICU)	ITT	Wilcoxon Rank Sum Test using LOCF	By Day 28	Secondary analysis
Fever	Time to resolution of fever	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
National Early Warning Score (NEWS)	Mean change from baseline in NEWS	ITT	MMRM	Day 4,7,10,14	Secondary analysis
base	Time to definitive extubation	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
Oxygen Saturation	Time to oxygen saturation ≥94% on room air in days	ITT	Log rank test using IETI	By Day 28	Secondary analysis
			Cox Proportional Hazards model using IETI	By Day 28	Supportive analysis
Resting respiratory rate	Number of days of resting respiratory rate <24 breaths per minute	ITT	Wilcoxon Rank Sum Test using observed data.	By Day 28	Secondary analysis

Abbreviations: ANOVA = analysis of variance; IETI = infinite event time imputation; ITT = Intent-to-Treat; MI = multiple imputation; LOCF = last observation carried forward; MMRM = mixed-effect model repeated measure; NIAID-OS = National Institute of Allergy and Infectious Diseases ordinal scale.

6.11.4. Sensitivity Analyses

A sensitivity analysis excluding patients who die within 48 hours of screening and have DNR or DNI in ITT analyses population will be conducted. The sensitivity analyses may include the following endpoints:

- time to recovery,
- proportion of patients who die or require non-invasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) by Day 28, and
- all-cause mortality (Day 1 to Day 28).

6.12. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

Plasma concentration data will be collected in patients who progress to intubation in ICU. The concentration-time data for baricitinib in these patients after NG tube administration will be primarily evaluated via graphical comparison to known pharmacokinetic (PK) profiles at 4-mg QD dosing that have been characterized for other populations such as healthy subjects, patients with rheumatoid arthritis, atopic dermatitis, etc. The observed individual data points and/or median and corresponding 90% CIs of plasma concentrations from patients with COVID-19 infection will be overlaid onto the estimated median and corresponding 90% prediction intervals of concentration-time profiles at 4-mg QD in the previous populations.

The PK data may also be analyzed using a population modeling approach via a nonlinear mixed-effects modeling (NONMEM) program, if the observed 4-mg exposure in the COVID-19 patients dramatically deviates from that in other patient populations and if adequate numbers of PK samples are available from patients who progress to intubation. The population PK model previously developed for baricitinib in aforementioned patient populations will be used for the modeling analysis. No PK/PD analyses will be conducted for this study.

6.13. Safety Analyses

The planned safety analyses are consistent with compound level safety standards, which are based on various sources, including company standards, internal and external subject matter experts, and cross-industry initiatives (for example, white papers produced by a Pharmaceutical Users Software Exchange [PhUSE] Computational Science Working Group [a collaboration with Food and Drug Administration (FDA) and PhUSE], published in the PhUSE Deliverables Catalog [PhUSE [WWW]). Descriptions of the safety analyses are provided in this SAP; however, some details are in compound level safety standards.

A treatment-emergent adverse event (TEAE) during the analysis period is defined as an event that either first occurred or worsened in severity after the first dose of study treatment and on or prior to the end of the postbaseline period (defined as Day 28 for the primary outcome datalock).

A follow-up-emergent adverse event (FEAE) is defined as an event that either first occurred or worsened in severity after the treatment period and on or prior to the last visit date during the post-treatment follow-up period (up to the Day 60 visit) of the study.

Table KHAA.6.8. Baseline and Postbaseline Definitions

Analysis Type	Baseline	Postbaseline
1.1) Treatment-Emergent Adverse Events	The baseline period is defined as the start of screening and ends prior to the first dose of study treatment.	Starts after the first dose of study treatment and ends up to data cut date (for DMC interim analyses), or up to Day 28 for the primary outcome datalock.
1.2) Adverse Events including serious adverse events	**NA**	
1.3) Treatment-Emergent Abnormal Labs, Vital Signs and shift summaries in labs	Baseline will be all scheduled and unscheduled measurements recorded during the baseline period as defined above (1.1).	Postbaseline will be defined as above (1.1). All scheduled and unscheduled measurements will be included.
1.4) Change from Baseline for Labs, Vital Signs	The last scheduled nonmissing assessment recorded prior to the date of first dose of study treatment during the baseline period defined above (1.1).	Postbaseline will be defined as above (1.1). Only scheduled visits will be included. The early termination visits (ETV) are considered scheduled visits.

6.13.1. Extent of Exposure

Mean and median modal dose and total patient days of exposure will be reported for all the treatment arms. Descriptive statistics will be provided for subject days of exposure and the frequency of subjects falling into the following different exposure ranges will also be summarized:

- ≥4 days, ≥7days, and ≥10 days, and >=14 days
- >0 to <4 days, ≥4 days to <7 days, ≥7 days to <10 days, ≥10 days to <14 days and ≥14 days.

In addition, study duration will be analyzed in a similar fashion to what was planned for exposure duration but expanding ranges to 28 days.

6.13.2. Adverse Events

The planned summaries for AEs are provided in Table KHAA.6.9 and are described more fully in compound level safety standards and in the AE-related PhUSE white paper (Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Documents [PhUSE 2017]).

Table KHAA.6.9. Summary Tables Related to Adverse Events

Analysis	Population or Analysis Set
An overview table, with the number and percentage of subjects who experienced a TEAE, serious adverse event, death, discontinued from study treatment due to an adverse event	Safety population
An overview table, with the number and percentage of subjects who experienced an event, serious adverse event, death, discontinued from study due to an adverse event	Follow-up analysis set Safety analysis set
The number and percentage of subjects with TEAEs using MedDRA Preferred Term nested within System Organ Class	
The number and percentage of subjects with FEAEs using MedDRA Preferred Term nested within System Organ Class	Follow-up analysis set
The number and percentage of subjects with TEAEs using MedDRA Preferred Term	Safety population
The number and percentage of subjects with TEAEs by maximum severity using MedDRA Preferred Term	Safety population
The number and percentage of subjects with TEAEs using MedDRA Preferred Term for the common TEAEs (occurred in ≥2% before rounding of treated subjects)	Safety population
The number and percentage of subjects who experienced a serious adverse event (including deaths and SAEs temporally associated or preceding deaths) using MedDRA Preferred Term nested within System Organ Class	Safety population
The number and percentage of subjects who experienced a serious adverse event (including deaths and SAEs temporally associated or preceding deaths) using MedDRA Preferred Term nested within System Organ Class	Follow-up analysis set
Listing of All SAEs	Safety population
The number and percentage of subjects who permanently discontinued from study treatment due to an adverse event (including adverse events that led to death) using MedDRA Preferred Term nested within System Organ Class	Safety population
Listing of AEs which led to permanent discontinuation from the study treatment and from the study	Safety population

Abbreviations: AEs = adverse events; ITT = Intent-to-Treat; MedDRA = Medical Dictionary for Regulatory Activities; SAE = serious adverse event; TEAE = treatment-emergent adverse event.

6.13.3. Clinical Laboratory Evaluation

The planned summaries for clinical laboratory evaluations are provided in Table KHAA.6.10 and are described more fully in compound level safety standards and in the laboratory-related PhUSE white papers (Analyses and Displays Associated with Measures of Central Tendency – Focus on Vital Sign, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Submission Documents [PhUSE 2013] and Analyses and Displays Associated with Outliers or Shifts from Normal to Abnormal: Focus on Vital Signs, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents [PhUSE 2015]).

Table KHAA.6.10. Summary Tables Related to Clinical Laboratory Evaluations

Analysis	Population or Analysis Set
Box plots for observed and/or change from baseline values	Safety population
Common Terminology Criteria for Adverse Events (CTCAEs) shifts tables for labs	Safety population
Tables with the number and percentage of subjects who shift from normal/high to low (that is, treatment-emergent low) and the number and percentage of subjects who shift from normal/low to high (that is, treatment-emergent high) may be analyzed using appropriate and validated reference ranges	Safety population
Listing of abnormal findings for laboratory analyte measurements, including qualitative measures	Safety population

6.13.4. Vital Signs and Other Physical Findings

The planned summaries for vital signs (systolic blood pressure [BP], diastolic BP, pulse, weight, BMI, temperature) are provided in Table KHAA.6.11 and are described more fully in compound level safety standards and in the vitals-related PhUSE white papers (Analyses and Displays Associated with Measures of Central Tendency – Focus on Vital Sign, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Submission Documents [PhUSE 2013] and Analyses and Displays Associated with Outliers or Shifts from Normal to Abnormal: Focus on Vital Signs, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents [PhUSE 2015]).

Table KHAA.6.11. Summary Tables Related to Vital Signs

Analysis	Population or Analysis Set
Box plots for observed and/or change from baseline values	Safety population
Tables with the number and percentage of subjects who shift from normal/high to low (that is, treatment-emergent low) and the number and percentage of subjects who shift from normal/low to high (that is, treatment-emergent high); the limits are defined in the compound level safety standards and are based on literature.	Safety population

6.13.5. Special Safety Topics

This section includes safety topics of interest, whether due to observed safety findings, potential findings based on drug class, or safety topics anticipated to be requested by a regulatory agency for any reason. In general, safety topics of interest will be identified by 1 or more SMQs, by an Eli Lilly and Company- (Lilly-) defined MedDRA PT listing based upon the review of the most current MedDRA version, or by treatment-emergent relevant laboratory changes. The topics of interest may include, but may not be limited to:

- abnormal hepatic tests
- hematologic changes
- renal function effects
- elevations in creatinine phosphokinase

- serious infections, herpes zoster, and opportunistic infection,
- major adverse cardiovascular events and other cardiovascular events, and
- venous thromboembolic and arterial thromboembolic events.

Refer to the compound level safety standards for details of analyses of the special safety topics.

6.14. Subgroup Analyses

For primary and gated secondary endpoints, following subgroup analysis will be conducted:

- baseline disease severity:
    - hospitalized not requiring supplemental oxygen, requiring ongoing medical care (OS 4)
    - hospitalized requiring supplemental oxygen by prongs or mask (OS 5)
    - hospitalized requiring noninvasive ventilation or high-flow oxygen (OS 6)
    - OS 5 and OS 6 combined
- baseline steroid use: dexamethasone and/or other systemic corticosteroid used at baseline for primary study condition (Yes/No)
- baseline remdesivir use (Yes/No)

The following subgroups categorized into disease-related characteristics and demographic characteristics will be evaluated for primary endpoint. These analyses may also be conducted selectively on key secondary endpoints if applicable:

- patient demographic and characteristics subgroups:
    - gender (male, female)
    - age group (<65, ≥65)
    - baseline weight (<60 kg, ≥60 to <100 kg, ≥100 kg)
    - baseline BMI (<25 kg/m^2, ≥25 to <30 kg/m^2, ≥30 kg/m^2)
    - race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Multiple)
    - baseline renal function status: impaired (eGFR <60 mL/min/1.73 m^2) or not impaired (eGFR ≥60 mL/min/1.73 m^2)
- geographic region subgroups:
    - region (US, Europe, Rest of World)
- baseline disease-related characteristics subgroups:
    - baseline disease duration of symptoms prior to enrollment:<7 days, ≥7 days

- preexisting comorbid conditions of interest: none, any

For subgroup analyses of categorical endpoints, a logistic regression model or proportional odds model using the Firth correction (Firth 1993) will be used to obtain a p-value for the treatment-by-subgroup interaction. The response variable will be each corresponding endpoint. The explanatory variables will be treatment, subgroup, treatment-by-subgroup interaction, and the baseline stratification variables (Section 6.1.4). Within each subgroup category, odds ratios (bari 4-mg over placebo) and associated CIs and p-values will be provided. Point estimates and CIs of the proportions and difference in proportions will be presented. For LOCF analyses the CIs will be based on the Newcombe-Wilson method without continuity correction. For multiple imputation analyses, proportions/percentages will be based on percentages in the multiply-imputed datasets; CIs will be based on the Wald method without continuity correction.

For ANOVA analyses of continuous endpoints, the following explanatory variables will be included for the test of interaction: treatment, subgroup, treatment-by-subgroup interaction and the baseline stratification factors (Section 6.1.4). The F test will be used to obtain a p value for treatment-by-subgroup interaction, type III sums of squares. Within each subgroup category, LS mean and associated p-values from ANOVA (using models given in Section 6.2.2) will be provided.

For the time-to-event endpoints, an unstratified log-rank test will be performed within each level of specified subgroups if sample sizes allow. Analysis of individual levels of subgroups will follow what is specified in Section 6.2.3, possibly with the use of the Firth correction.

Descriptive statistics will be provided for each treatment and stratum of a subgroup, regardless of sample size. Inferential statistics will be provided if sample sizes and event numbers are sufficiently large (e.g., >100 patients total and at least 1 event in each arm).

Additional subgroup analyses on efficacy may be performed as deemed appropriate and necessary.

Subgroup analyses for safety parameters will be specified in the List of Analyses for this study.

Table KHAA.6.12 lists the analysis methods for the different subgroup analysis.

Table KHAA.6.12. Description of Subgroup Analyses for Primary and Key Secondary Endpoints

Measure	Variable	Subgroup	Analysis Method	Time Point
NIAID-OS	Proportion of patients who died or progressed to non-invasive ventilation or high-flow oxygen (OS ≥ 6)	ITT: Population 1 – baseline disease severity	Logistic regression using MI	By Day 28
		ITT: Population 2 – baseline disease severity	Logistic regression using MI	By Day 28
		ITT: Population 1 – baseline disease severity	Logistic regression using LOCF	By Day 28
		ITT: Population 2 – baseline disease severity	Logistic regression using LOCF	By Day 28
		All subgroups other than baseline disease severity	Logistic regression using LOCF	By Day 28
	Ventilator-free days (VFD)	All subgroups	ANOVA using LOCF	By Day 28
	Proportion of patients with 1-point improvement	All subgroups	Logistic regression using LOCF	Day 4, 7, 10, 14
	Time to recovery	All subgroups	Log-rank test using IETI	By Day 28
	Overall improvement	All subgroups	Proportional odds model using LOCF	Day 4, 7, 10, 14
	Duration of hospitalization	All subgroups	ANOVA using LOCF	By Day 28
	All-cause mortality	All subgroups	Log-rank test	By Day 28
Oxygen Saturation	Proportion of patients with change in oxygen saturation from <94% at baseline to ≥94% at the observed time point	All subgroups	Logistic regression using LOCF	Day 4, 7, 10, 14

Abbreviations: ANOVA = analysis of variance; IETI = infinite event time imputation; ITT = Intent-to-Treat; MI = multiple imputation; LOCF = last observation carried forward; NIAID-OS = National Institute of Allergy and Infectious Diseases ordinal scale; VFD = ventilator-free days.

6.15. Protocol Violations

Protocol deviations will be tracked by the clinical team and their importance will be assessed by key team members during protocol deviation review meetings. All important protocol deviations (IPDs) identified with the potential to affect efficacy analyses will result in exclusion from the PPS population.

Potential examples of deviations include patients who receive excluded concomitant therapy, significant noncompliance with study medication (<80% of assigned doses taken, failure to take study medication, and taking incorrect study medication), patients incorrectly enrolled in the study, and patients whose data are questionable due to significant site quality or compliance issues. Refer to a separate document for the important protocol deviations.

Trial Issue Management Plan includes the categories and subcategories of IPDs and whether or not these deviations will result in the exclusion of patients from per protocol set.

The number and percentage of patients having IPD(s) will be summarized within category and subcategory of deviation by treatment group using the ITT population. Individual patient listings of IPDs will be provided. A summary of reasons patients were excluded from the PPS population will be provided by treatment group.

6.16. Interim Analyses and Data Monitoring

A Data Monitoring Committee (DMC) will oversee the conduct of this trial. The DMC will consist of members external to Lilly. This DMC will follow the rules defined in the DMC Charter, focusing on potential and identified risks for this molecule and for this class of compounds. The DMC will include at least 2 physicians with experience as a DMC member and/or clinical trial experience with a specialty in acute care, pulmonary medicine, or infectious diseases, as well as a biostatistician with DMC and clinical trial experience.

The DMC will be authorized to review unblinded results of analyses by treatment group prior to the final database lock, including, but not limited to, study discontinuation data, AEs including serious adverse events (SAEs), selective clinical laboratory data, and vital sign data. The DMC may recommend continuation of the study, as designed; modify study protocol as specified; temporary suspension of enrollment; the discontinuation of the entire study; or recommend increasing the sample size. Details of the DMC, including its operating characteristics are documented in the Data Monitoring Committee Charter for Phase 3 Study of Baricitinib in COVID-19 Program and further details are given in Section 6.16.1 of the Interim Analysis Plan.

6.16.1. Interim Analysis Plan

The following are the planned interim analyses for the KHAA Phase 3 study:

- interim analysis: for futility and safety monitoring, unblinded, selective efficacy/safety data will be reviewed, and
- safety and mortality review: for safety and mortality monitoring purposes only, unblinded mortality data and selective safety data will be reviewed.

- Sample size re-estimation (SSR): results of an interim analysis for the purposes of sample size re-estimation will be reviewed. Criteria will be provided, and the DMC may recommend increasing the sample size if the criteria for increase are met.

Additional details of interim analyses are summarized in the DMC charter.

The primary endpoint for the interim analysis for futility is:

- proportion of patients that require noninvasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) or die by Day 28.

The interim futility analysis methods are described in detail in the DMC charter. Beside the futility interim analysis, the DMC will also review safety data as specified in the DMC charter provide their recommendation for SSR. The details of SSR are provided in Section 6.16.2.

Summaries will include TEAEs, SAEs, special topic AEs, and treatment-emergent high and low laboratory and vital signs in terms of counts, percentages, and incidence rates (IRs), where applicable. For continuous analyses, box plots of laboratory analytes will be provided by time point and summaries will include descriptive statistics.

For interim futility review, the summary of proportion of patients that require noninvasive ventilation/high-flow oxygen or invasive mechanical ventilation (including ECMO) or die by Day 28 will be included.

The DMC may request efficacy data if they feel there is value and to confirm a reasonable benefit/risk profile for ongoing patients in the studies. Further details are given in the DMC Charter.

6.16.2. Sample Size Re-Estimation

This section provides an overview of design parameters for this sample size adaptive trial. It includes a detailed description of the unblinded sample size procedure and associated formulas for executing the design.

For purposes of sample size re-estimation, the overall α will be split such that 90% is allocated to Population 1 (α_1).

The proposed sample size re-estimation may result in overall sample sizes from approximately 1400 patients to 1700 patients.

6.16.2.1. Execution Details

An interim analysis will be performed when approximately 1000 patients have randomized in the trial and had the opportunity to contribute data to the primary endpoint (i.e., had the opportunity to complete the study or discontinued). This analysis will be used to determine the final sample size of the trial, in accordance with adaptive SSR methodology described in Mehta and Pocock (2011). Discontinuing the study due to safety concerns at the interim is also possible. Due to the uncertainty in enrollment projections, the exact timing of the interim may occur at a sample size smaller or larger than the planned number of 1000 patients.

At the interim analysis, the conditional power (CP) for statistical significance of the primary endpoint will be calculated for each of the two population separately using the CP formula as given in Mehta and Pocock (2011) for each dose-placebo comparison. The maximum of the two conditional powers will be considered. This CP could fall into one of the three zones: "Unfavorable" (CP < 30%), "Promising" (30%≤CP<90%), and "Favorable" (CP ≥ 90%). Sample size will be increased if the maximum of two CPs falls in the promising zone but will remain at the minimum otherwise.

6.16.2.2. Calculation of Conditional Power at Interim

Table KHAA.6.13 defines the values that are needed for the conditional power calculation at interim analysis.

Table KHAA.6.13. Defining the Values for the Conditional Power Calculation

Quantity	Input Value or Derivation/Calculation
z_1	This is the chi-square test statistic for testing difference between of baricitinib versus placebo for the primary efficacy analysis conducted *using only patients included in the interim analysis*.
n_1	This is the planned total number of randomized patients included in the interim analysis including both treatment groups. The interim is scheduled to occur when approximately 71% of randomized patients have had the opportunity to complete the treatment phase at $n_1 = 1000$ for Population 1. Due to the uncertainty in enrollment projections, the exact timing of the interim may occur at a sample size smaller or larger than the planned number of 1000 patients for Population 1. For Population 2, n_1 will be the number of patients observed at the time of the interim analysis in Population 2
n_2	This is the increment from n_1 for the total number of patients in both treatment groups should the study remain at the planned *minimum* sample size based on interim results. The planned minimum sample size is 1400 and if the interim occurs at the planned $n_1 = 1000$, the increment needed to get to 1400 patients is $n_2 = 400$ for Population 1. For Population 2, the proportion of patients p not on dexamethasone or other systemic corticosteroids will be estimated from the available data for patients requiring supplemental oxygen at baseline. n_2 for Population 2 will then be estimated by multip7lying n_2 for the overall population by the observed proportion, i.e., (n_2 for Population 2) = (n_2 for Population 1) × p

As provided in Mehta and Pocock (2011), the conditional power for the primary efficacy analysis is given by

$$CP(z_1, n_2) = 1 - \Phi\left(\frac{z_\alpha\sqrt{n_1 + n_2} - z_1\sqrt{n_1}}{\sqrt{n_2}} - \frac{z_1\sqrt{n_2}}{\sqrt{n_1}}\right)$$

where $\Phi(\,)$ is the cumulative distribution function of a standard normal variable. The CP will be calculated for Population 1 and for Population 2 with the prespecified split type-I errors α_1 and α_2 respectively. Note that this conditional power calculation assumes that the data post-interim follows a distribution with an effect size equal to the effect size observed at interim. The preceding conditional power formula is written in terms of the test statistic for the baricitinib versus placebo dose comparison.

6.16.2.3. New Sample Size Calculation

If the sample size is increased as a result of the interim analysis, in which the maximum conditional power of the two populations is in the "promising zone," the sample size will be increased for both treatment groups to maintain the 1:1 ratio. The formula for the updated sample size to maintain a targeted conditional power is provided in Mehta and Pocock (2011) and is given below. The formula below represents the recommended updated total sample size to maintain 90% conditional power. The value z_1 in the formula is the interim test statistic corresponding to the population with the maximum conditional power.

$$N_{new} = n_1 + \left[\frac{n_1}{z_1^2}\right]\left[\frac{z_\alpha*\sqrt{(n_1+n_2)}-z_1*\sqrt{n_1}}{\sqrt{n_2}} + (\Phi^{-1}(0.90))\right]^2$$

If the maximum conditional power is observed corresponding to Population 2, then the increment in sample size obtained from the formula will be divided by the proportion of patients not on dexamethasone or other systemic corticosteroids obtained from the data as described before to get the increment in the overall sample size.

If the results of the interim analysis suggest an increase in the sample size, the final sample size will be the minimum of N_{new} and 1700.

6.16.2.4. Calculation of CHW Statistics

The final CHW test statistic for the primary efficacy analysis (at the completion of the treatment phase after the sample size re-estimation) can be written as a weighted combination of the independent increments comprising the interim Wald test statistic and the post-interim Wald test statistic (Cui et al. 1999; Mehta and Pocock 2011):

$$z_{chw} = \frac{\sqrt{n_1}}{\sqrt{n_1+n_2}}z_1 + \frac{\sqrt{n_2}}{\sqrt{n_1+n_2}}z_2$$

Here z_1 is the test statistic for the test conducted using the interim population and z_2 is the test statistic for the test conducted using only the set of patients included in the post-interim assessment.

6.17. Planned Exploratory Analyses

The planned exploratory analyses are described in Section 4.3 and Appendix 3.

6.18. Annual Report Analyses

Annual report analyses, such as the Development Update Safety Report, will be documented in a separate analysis plan.

6.19. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include a summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AEs are summarized by treatment group and by MedDRA PT.

- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- Adverse event reporting is consistent with other document disclosures (e.g., clinical study report, manuscripts).

Similar methods will be used to satisfy the European Clinical Trials Database requirements.

7. Unblinding Plan

Refer to the blinding and unblinding plan document for details.

8. References

Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. *Stat Med*. 2014;33(4):693-713. https://doi.org/10.1002/sim.5974.

Brannath W, Mehta CR, Posch M. Exact confidence bounds following adaptive group sequential tests. *Biometrics*. 2009;65(2):539-546. https:// doi.org/10.1111/j.1541-0420.2008.01101.x.

Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913. https://doi.org/10.1002/bimj.201000239.

Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics.* 1999;55(3):853-857. https://doi.org/10.1111/j.0006-341X.1999.00853.x.

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38. https://doi.org/10.1093/biomet/80.1.27.

Gao P, Ware JH, Mehta C. Sample size re-estimation for adaptive sequential design in clinical trials. *J Biopharm Stat*. 2008;18(6):1184-1196. https:// doi.org/10.1080/10543400802369053.

Honaker J, King G. What to do about missing values in time-series cross-section data. *Am J Political Sci*. 2010;54(2):561-581.

[ICH] International Conference on Harmonisation. ICH E9 (R1): Addendum to statistical principles for clinical trials on choosing appropriate estimands and defining sensitivity analyses in clinical trials. 2017. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf. Accessed _March 16, 2021.

Lee JW. Some versatile tests based on the simultaneous use of weighted log-rank statistics. *Biometrics*. 1996;52(2):721-725. https://doi.org/10.2307/2532811.

Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655-660. https://doi.org/10.2307/2335748.

Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med*. 2011;30(28):3267-3284. https://doi.org/10.1002/sim.4102.

[PhUSE] Pharmaceutical Users Software Exchange resources page. PhUSE web site. Available at: http://www.phuse.eu/css-deliverables. Accessed March 16, 2020.

[PhUSE] Pharmaceutical Users Software Exchange. PhUSE Computational Science Standard Analyses and Code Sharing Working Grouyp. Analysis and Dispay White Papers Project Team. Analysis and displays associated with adverse events: focus on adverse events in phase 2-4 clinical trials and integrated summary documents. Version 1.0. 2017. Available at: mhttp://www.phuse.eu/documents//working-groups/cs-whitepaper-adverseevents-v10-4442.pdf. Accessed June 26, 2020.

[PhUSE] Pharmaceutical Users Software Exchange. PhUSE Computational Science Development of Standard Scripts for Analysis and Programming Working Group. Analysis and Dispay White Papers Project Team. Analyses and displays associated with outliers or shifts from normal to abnormal: focus on vital signs, electrocardiogram, and laboratory analyte measurements in phase 2-4 clinical trials and integrated summary documents. Version 1.0. 2015. Available at: http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CS_WhitePaper_OutliersShifts_v1.0.pdf. Accessed June 26, 2020.

[PhUSE] Pharmaceutical Users Software Exchange. Analyses and displays associated with measures of central tendency – focus on vital sign, electrocardiogram, and laboratory analyte measurements in phase 2-4 clinical trials and integrated submission documents. Version 1.0. 2013. Available at: http://www.phusewiki.org/docs/CSS%20White%20Papers%202016/CSS_WhitePaper_CentralTendency_v1.0.pdf. Accessed June 26, 2020.

Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33(2):176-182. https://doi.org/10.1093/eurheartj/ehr352.

Rohmeyer K, Klinglmueller F. *gMCP: Graph Based Multiple Comparison Procedures. R package version 0.8-15.* 2020. Available at: https://cran.r-project.org/web/packages/gMCP/index.html. Accessed March 16, 2021.

Royal College of Physicians. National Early Warning Score (NEWS): Standardising the assessment of acute-illness severity in the NHS. 2012. Available at: https://www.rcplondon.ac.uk/file/32/download?token=5NwjEyTq. Accessed March 31, 2020.

Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13:152. https://doi.org/10.1186/1471-2288-13-152.

Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91(434):473-489. https://doi.org/10.1080/01621459.1996.10476908.

Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592. https://doi.org/10.1093/biomet/63.3.581.

Appendix 1. Dose Conversion for Corticosteroid

In case conversion of dosage of corticosteroids is needed, the following table should be used for converting nonprednisone medications to prednisone equivalent:

> Multiply the dose of the corticosteroid taken by the patient (in milligrams) in Column 1 by the conversion factor in Column 2 to get the equivalent dose of prednisone (in milligrams).
>
> *Example: Patient is taking 16 mg of methylprednisolone po daily. To convert to prednisone: 16 mg methylprednisolone × 1.25 = 20 mg prednisone. 16 mg of methylprednisolone po daily is equivalent to 20 mg of prednisone po daily.*

Column 1	Column 2
Corticosteroid Preferred Term	Conversion factor for converting to an equivalent prednisone dose
Prednisone	1
Prednisone acetate	1
Prednisolone	1
Prednisolone acetate	1
Prednisolone sodium phosphate	1
Methylprednisolone	1.25
Methylprednisolone acetate	1.25
Methylprednisolone sodium succinate	1.25
Triamcinolone	1.25
Triamcinolone acetonide	1.25
Triamcinolone hexacetonide	1.25
Cortisone	0.2
Cortisone acetate	0.2
Hydrocortisone	0.25
Hydrocortisone acetate	0.25
Hydrocortisone sodium succinate	0.25
Betamethasone	6.25
Betamethasone acetate	6.25
Betamethasone dipropionate	6.25
Betamethasone sodium phosphate	6.25
Dexamethasone	6.25
Dexamethasone acetate	6.25
Dexamethasone phosphate	6.25
Dexamethasone sodium phosphate	6.25

Column 1	Column 2
Corticosteroid Preferred Term	Conversion factor for converting to an equivalent prednisone dose
Paramethasone	2.5
Deflazacort	0.83
Celestona bifas	6.25
Depo-Medrol med lidocaine	1.25
Diprospan	6.25
Fluocortolone	1
Meprednisone	1.25

Appendix 2. Multiple Imputation

1) Unique challenges for multiple imputation of the NIAID-OS:

While many methods of multiple imputation (and single imputation) are readily available in standard software, the NIAID-OS in KHAA presents several unique problems. First, it includes an absorbing state of death (NIAID-OS = 8). Also, complete recovery (NIAID-OS = 1) is ordinarily an absorbing state with some exceptions where patients become sick again. Second, change between days is generally slow. Patients often stay in the same state for several days. Thus, the correlation between days is very strong which may cause issues of (near) multicollinearity when fitting regression models for use in imputation. Also, some methods may not converge when many highly correlated days are being imputed for a patient. Third, standard assumptions such as multivariate normality and approximate linearity are highly suspect for the NIAID-OS. Finally, for study KHAA additional information is available for some patients. For example, for many patients the exact NIAID-OS score is missing for some days, but it may be inferred from other available data that the patient is not dead (i.e., NIAID OS $\neq 8$). It would be beneficial to use an imputation technique which can make use of this information.

For these reasons, a novel Markov model multiple imputation (MMMI) method is adopted to impute the NIAID-OS scores. A simulation study was performed to show the operating characteristics of the suggested method. The manuscript for this method is currently under preparation and the results of the simulation study are available upon request.

2) Overview of multiple imputation using a Markov model:

A Markov model assumes that each transition to a future state is dependent on only the previous state. Let $H_{it} \in \{1,2,\ldots 8\}$ be a random variable representing the sequence of NIAID-OS states for individual $i \in \{1,2,\ldots N\}$ on day $t \in \{1,2,\ldots T\}$. The parameters of interest in a Markov model are the initial probabilities and the transition probabilities:

Initial Probabilities: Let π_j represent the probability that a patient will be in state j on Day 1.

Transition Probabilities: Let $P_{b_t}(k|j)$ be the probability of transitioning from ordinal state $k \in \{1,2,\ldots 8\}$ to state $j \in \{1,2,\ldots 8\}$ on day t to day $t+1$. Here b_t represents a bin associated with a transition from day t to day $t+1$. We assume that within each bin the transition probabilities are the same. Daily transitions were divided into 4 bins as follows: (bin 1) includes the first 6 transitions; (bin 2) includes the next 7 transitions; (bin 3) includes the next 7 transitions; and (bin 4) the final 7 transitions.

The parameters $\boldsymbol{\pi}$ and $\boldsymbol{P}$ are collectively referred to as $\boldsymbol{\theta}$. We let $\boldsymbol{h}_{1:T}$ represent a vector of states of length T, and $\boldsymbol{h}_{i,1:T}^{obs}$ represent the observed scored for individual i. Also we define the set S_{iT}={*All possible trajectories* $\boldsymbol{h}$ *s.t. consistent with observed data* $\boldsymbol{h}_i^{obs}$}. The set may be constructed to account for complex types of missingness such as knowing that a patient is not dead.

The likelihood for the Markov model for a single individual i may be written by marginalizing out the missing scores:

$$l_i(\boldsymbol{\pi}, \boldsymbol{P}|\boldsymbol{h}_{i,1:T}^{obs}) = \sum_{\boldsymbol{h}\in S_{iT}} P(H_{i1} = h_1, \cdots, H_{it} = h_t)$$
$$= \sum_{\boldsymbol{h}\in S_{iT}} \pi_{h_1} \prod_{t=2}^{T} P_{b_{t-1}}(h_t|h_{t-1})$$

The overall likelihood may be found by multiplying together the likelihoods from each patient. As described by Rubin (1976) the maximum likelihood estimator after marginalizing out the missing observations is still asymptotically unbiased under a missing at random (MAR) assumption. Maximum likelihood estimates of the model parameters ($\boldsymbol{\theta}$) may be calculated using the expectation maximization (EM) algorithm. The same algorithms commonly used for "hidden Markov models" (HMMs) may be equally applied to the Markov model with missing data. The EM algorithm in this context is known as the Baum-Welsh algorithm in the HMM literature. A standard algorithm can be used to simulate from the distribution of the missing data conditional on the observed data (i.e., simulate from $\boldsymbol{P_\theta}(\boldsymbol{H_i^{mis}}|\boldsymbol{H_{i,1:T}^{obs}})$).

As described by Honaker and King (2010), a bootstrap + EM procedure is used to appropriately account for model uncertainty. The imputation proceeds by performing the following steps for $m \in \{1,2,\ldots M\}$ with $M = 100$:

1. Generate a bootstrapped dataset by resampling patients with replacement.
2. Estimate the model parameters using the maximum likelihood approach as calculated using the EM algorithm. For iteration *m* the bootstrapped parameter estimates are $\tilde{\boldsymbol{\theta}}_m$.
3. Impute the missing data in the original data (i.e., not the bootstrap data) by simulating from the distribution of the missing states conditional on the observed states assuming the estimated bootstrap model parameters (i.e., simulate from $\boldsymbol{P}_{\tilde{\boldsymbol{\theta}}_m}(\boldsymbol{H_i^{mis}}|\boldsymbol{H_{i,1:T}^{obs}})$).

The above steps are applied 100 times to generate 100 multiply imputed datasets. We note that with MMMI covariates can be accounted for by using a stratified imputation approach where imputation is performed separately for each strata. The mathematical details of the MMMI algorithm is provided in the following section.

3) <u>Additional Notation Conventions:</u>

To simplify the notation, we will drop the index for individuals. We used the notation $\boldsymbol{h}_{t_1:t_2}$ to represent a vector of states from day t_1 to t_2 inclusive. We define $\boldsymbol{h}^{obs}$ as the observed states *where the missing values will be coded as belonging to state 9 and patients who are missing but not dead will be defined as being in state 10*. We also define the flowing indicator function:

$$q(h^{obs}, h) = \begin{cases} I[h = h^{obs}] & h^{obs} \leq 8 \\ 1 & h^{obs} = 9 \\ I[h \neq 8] & h^{obs} = 10 \end{cases}.$$

The $q(h^{obs}, h)$ can be flexibly altered to account knowledge about which states a patient with missing data may be. The transition probability in our case has the form:

$$P_{b_j} = \begin{bmatrix} p_{b_j 11} & p_{b_j 12} & \cdots & p_{b_j 17} & p_{b_j 18} \\ p_{b_j 21} & p_{b_j 22} & & p_{b_j 27} & p_{b_j 28} \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix},$$

and initial probability has the form:

$$\pi = (\pi_1, \pi_2, \dots, \pi_7, \pi_8).$$

Using the function above, the likelihood may be rewritten as:

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{h}_{1:T}^{obs}) &= \sum_{\boldsymbol{h}_{1:T}} P(H_1 = h_1, \cdots, H_T = h_T) \prod\prod_{j=1}^{T} q(h_j^{obs}, h_j) \\ &= \sum_{\boldsymbol{h}_{1:T}} \pi_{h_1} q(h_1^{obs}, h_1) \prod\prod_{j=1}^{T-1} P_{b_j}(h_{j+1}|h_j) q(h_{j+1}^{obs}, h_{j+1}) \end{aligned}$$

a) <u>Forward Procedure:</u>

We define the forward variables at $t = 1, \dots, T$ as:

$$\alpha_i(t) = \sum_{\boldsymbol{h}_{1:t-1}} P(H_1 = h_1, \cdots, H_{t-1} = h_{t-1}, H_t = i) \prod_{j=1}^{t-1} q(h_j^{obs}, h_j)\ q(h_t^{obs}, i),\ i = 1, \dots, 8.$$

When the product $\prod_{j=1}^{t-1} q(h_j^{obs}, h_j)$ has the upper bound less than the lower bound we define the product as 1, that is $\alpha_i(1) = P(H_1 = i) q(h_1^{obs}, i) = \pi_i q(h_1^{obs}, i)$.

For $t \geq 2$ we may write the forward variables as:

$$\alpha_i(t) = \sum_{\boldsymbol{h}_{1:t-1}} \pi_{h_1} q(h_1^{obs}, h_1) \left(\prod\prod_{j=1}^{t-2} P_{b_j}(h_{j+1}|h_j) q(h_{j+1}^{obs}, h_{j+1}) \right) P_{b_{t-1}}(i|h_{t-1}) q(h_t^{obs}, i).$$

$\alpha_i(t)$ may be calculated using the recursion formula:

$$\begin{aligned} \alpha_i(1) &= \pi_i q(h_1^{obs}, i) \\ \alpha_i(t) &= q(h_t^{obs}, i) \sum_{j=1}^{8} \alpha_j(t-1) P_{b_{t-1}}(i|j), \qquad t = 2, \dots, T. \end{aligned}$$

To make the algorithm more numerically stable and tackle the underflow problem, we follow the commonly used normalized forward algorithm. The likelihood until time t is:

$$L(h_{1:t}^{obs}) = \sum_{i=1}^{8} \alpha_i(t).$$

Then, we can introduce normalized forward variables as:

$$\alpha_i^*(t) = \frac{\alpha_i(t)}{L(h_{1:t}^{obs})}.$$

From the recursion formula above, we have:

$$\alpha_i^*(t) = \frac{L(h_{1:t-1}^{obs})}{L(h_{1:t}^{obs})} q(h_t^{obs}, i) \sum_{j=1}^{8} \alpha_j^*(t-1) P_{b_{t-1}}(i|j), \qquad t = 2, \dots, T.$$

Define:

$$c_t = \frac{L(h_{1:t}^{obs})}{L(h_{1:t-1}^{obs})}, \qquad t = 1, \dots, T,$$

where: $c_1 = L(h_1^{obs}) = \sum_{i=1}^{8} \alpha_i(t)$. Then,

$$c_t = \sum_{i=1}^{8} q(h_t^{obs}, i) \sum_{j=1}^{8} \alpha_j^*(t-1) P_{b_{t-1}}(i|j), \qquad t = 2, \dots, T,$$

and:

$$L(h_{1:T}^{obs}) = \prod_{t=1}^{T} c_t.$$

b) <u>Backwards Procedure:</u>

We define the backwards variables at $t = 1, \dots, T$ as:

$$\beta_i(t) = \sum_{h_{t+1:T}} P(H_{t+1} = h_{t+1}, \cdots, H_T = h_T | H_t = i) \prod_{j=t+1}^{T} q(h_j^{obs}, h_j).$$

These may be calculated using the recursion formula:

$$\beta_i(T) = 1$$

$$\beta_i(t) = \sum_{j=1}^{8} \beta_j(t+1) P_{b_t}(h_{t+1} = j|i) q(h_{t+1}^{obs}, j).$$

Similarly, we can define normalized backward variables as:

$$\beta_i^*(t) = \frac{L(h_{1:t}^{obs})}{L(h_{1:T}^{obs})} \beta_i(t), \qquad t = 1, \dots, T,$$

with $\beta_i^*(T) = 1$. The normalized recursion formula is:

$$\beta_i^*(t) = \frac{1}{c_{t+1}} \sum_{j=1}^{8} \beta_j^*(t+1) P_{b_t}(h_{t+1} = j|i) q(h_{t+1}^{obs}, j), \qquad t = T-1, \dots, 1.$$

c) **Baum Welch (i.e., EM) Procedure:**

Define the quantities:

$$\gamma_i(t) = P\left(H_t = i \middle| H_{1:T}^{obs}\right)$$
$$= \frac{\beta_i(t)\alpha_i(t)}{\sum_{j=1}^{8}\beta_j(t)\alpha_j(t)} = \frac{\beta_i(t)\alpha_i(t)}{L\left(\theta \middle| H_{1:T}^{obs}\right)} = \beta_i^*(t)\alpha_i^*(t),$$

And:

$$\xi_{ij}(t)$$
$$= P(H_t = i, H_{t+1} = j | H_{1:T}^{obs})$$

$$= \frac{\alpha_i(t)P_{b_t}(j|i)\beta_j(t+1)q\left(h_{t+1}^{obs}, j\right)}{\sum_{s=1}^{8}\sum_{u=1}^{8}\alpha_s(t)P_{b_t}(u|s)\beta_u(t+1)q\left(h_{t+1}^{obs}, u\right)}$$

$$= \frac{\alpha_i(t)P_{b_t}(j|i)\beta_j(t+1)q\left(h_{t+1}^{obs}, j\right)}{L\left(\theta \middle| H_{1:T}^{obs}\right)}$$

$$= \frac{\alpha_i^*(t)P_{b_t}(j|i)\beta_j^*(t+1)q\left(h_{t+1}^{obs}, j\right)}{c_{t+1}}.$$

Then, we can update θ as following:

$$\pi_i^* = \gamma_i(1),$$

$$P_b^*(j|i) = \frac{\sum_{t\in\{t:b_t=b\}}\xi_{ij}(t)}{\sum_{t\in\{t:b_t=b\}}\gamma_i(t)}.$$

The convergency criteria is set by change of likelihood value between two consecutive steps is numerically indistinguishable.

4) Simulating missing states (i.e., simulating from $P_\theta(H^{mis}|H_{1:T}^{obs})$):

First the backward and forward algorithms are calculated for each patient.

Next, for each patient we follow the following procedure:

1. Simulate h_1 from the discrete distribution defined by the vector of probabilities $(\gamma_1(1), \cdots, \gamma_8(1))$.
2. For t>1, simulate h_t from the discrete distribution defined by the vector of probabilities $(\xi_{h_{t-1}1}(t-1), \cdots, \xi_{h_{t-1}8}(t-1))/\zeta$ where ζ is a normalizing constant, $\sum_{i=1}^{8}\xi_{h_{t-1},i}(t-1)$.

Appendix 3. Exploratory Analyses

1) Additional analyses for the main study (through Day 28)

The primary analysis (progression to ventilation or death by Day 28) and time to recovery will be repeated for the following subgroups, if sample sizes allow:

- No baseline steroids, no baseline remdesivir
- No baseline steroids, yes baseline remdesivir
- Yes baseline steroids, no baseline remdesivir
- Yes baseline steroids, yes baseline remdesivir

1.1) Additional analyses for the main study (Day 60)

The following table gives planned analyses for the main study using Day 60 data.

Variable	Population	Analysis Method	Time Point	Groups/Subgroups
Observed OS values at Day 60	ITT population with a baseline NIAID OS and a value at the Day 60 Visit	Descriptive statistics as described in Table KHAA.6.1 for categorical/discrete data.	At the Day 60 visit.	Overall and by baseline • OS 4, 5, 6 and 5, 6 combined. • Corticosteroid status (see Section 6.1.4) • Age group (<65, ≥65)
All-cause mortality:	Patients in the ITT population, excluding randomized patients who discontinued on the day of randomization and have no baseline or postbaseline OS data.	Time to death from all causes, all time on study. Event dates are date of death for patients who died. Censor dates are last known information that indicates patient was alive. Patients who die after the visit window for Day 60, will be censored at the last date of their Day 60 visit window. KM curves and standard summary statistics (see Section 6.2.3) will be generated.	From baseline to end of study participation.	

Additionally, the following analyses will be performed after the Day 60 database lock for relevant subsets of the ITT population.

- Demographics
- Disposition
- Concomitant medications
- Safety analyses will include all adverse events occurring post primary outcome datalock up to an including Day 60. The following are the planned safety analyses:
    - Follow-up-emergent adverse events (FEAEs). FEAE is defined as an event that either first occurred or worsened in severity after Day 28 and on or prior to end of the study.
    - Serious adverse events
    - Adverse events leading to death
    - Overview of infections (includes herpes zoster, herpes simplex, opportunistic infections, and tuberculosis)

2) Addendum 5 Analyses (for Patients who are OS 7 at Baseline)

The following analyses are planned for randomized patients in Addendum 5 with baseline OS = 7. Patients who discontinue on the day of randomization with no baseline or postbaseline OS data will not be included.

1. NIAID-OS D4/7/10/14/28: overall improvement on the NIAID-OS evaluated at Day 4, Day 7, Day 10, 14 and Day 28
2. Proportion of patients with at least 1-point improvement on the NIAID-OS or live discharge from the hospital at Day 4, Day 7, Day 10, 14 and Day 28
3. Number of days alive and free of mechanical ventilation (VFD per main SAP), Day 1 to Day 28
4. Duration of hospitalization (Day 1 to Day 28)
5. Time to recovery (by Day 28)
6. All-cause mortality (Day 1 to Day 28).

Analyses 1-4 will use logistic regression with LOCF methodology. Analyses 5-6 will use time-to event methodology. All will follow the same methodology as the main SAP, except that treatment comparisons will be adjusted for baseline age group (<65, ≥65 years) and region rather than all the baseline stratification variables given in Section 6.1.4.

Patient disposition, patient characteristics and safety analyses will follow the main SAP to the extent possible (see Sections 6.7, 6.8, and 6.13).

The following table gives additional analyses for the Addendum 5 patients, utilizing their Day 60 data.

Variable	Population	Analysis Method	Time Point	Groups/Subgroups
Observed OS values at Day 60.	ITT population with a baseline NIAID OS and a value at the Day 60 Visit	Descriptive statistics.	At the Day 60 visit.	Overall (OS 7) • Baseline corticosteroid status (see Section 6.1.4) • Age group (<65, ≥65)
All-cause mortality	ITT population with a baseline NIAID OS	Time to death from all causes, all time on study. Event dates are date of death for patients who died. Censor dates are last known information that indicates patient was alive. Patients who die after the visit window for Day 60, will be censored at the last date of their Day 60 visit window. KM curves and standard summary statistics (see Section 6.2.3) will be generated.	From baseline to end of study participation.	

Demographics, disposition, and safety analyses for Addendum 5 patients will mirror the main study to the extent possible. The main comparative safety analyses will be based on data from baseline to Day 28. There will be separate analyses of FEAEs, as described in Section 6.13.

Leo Document ID = 983e9bb1-2ca8-4121-b5da-3d4709a78151

Approver: PPD

Approval Date & Time: 19-Mar-2021 14:58:22 GMT

Signature meaning: Approved