Stanford MEDICINE | Quantitative Sciences Unit

Statistical Analysis Plan (SAP)

Title: STOP-PASC

CRU/Department/Division/Center: Division of Infectious Diseases

IRB Number: 66994

Primary Investigators: Linda Geng, Upinder Singh

Biostatisticians: Sa Shen, Jane Liang, Blake Shaw, Haley Hedlin, Wei Deng, Bren Botzheim, Lu Tian

Version: 2.0

Version Date: 2023 October 20

Investigator Agreement

☐ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s).
☐ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract, manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses.
☐ If substantial additional analysis is necessary or the aims of the project change, a new SAP will need to be developed.

☐ If you engaged the CTSA's BERD please ensure you cite UL1TR003142 award when disseminating work. If your study was cancer-related, please ensure you cite P30CA124435 award when disseminating work. If your study was diabetes-related, please ensure you cite P30DK116074 award when disseminating work. These publications should also be submitted to PubMed Central.
☐ I have reviewed the SAP and understand that any changes must be documented.

Acknowledged by:

Primary Investigator:
Linda Geng, MD, PhD

Date: 10/20/2023

Lead Biostatistician:
Haley Hedlin, PhD

Date: 10/20/23

---

1. Study Overview

1.1 Study Aims:

Post-acute sequelae of SARS-CoV-2 (PASC) also known as Long COVID is a major public health problem in the wake of the pandemic, with estimates of over 27 million PASC cases in the United States alone as of August, 2022. PASC encompasses various conditions and symptoms, such as fatigue, brain fog, and dyspnea, that develop in about 10-30% of individuals after their initial COVID-19 infection and can last for months or longer with significant impact on quality of life and function. The underlying mechanisms of disease are unclear and there are currently no known effective therapies to treat PASC. Given the large-scale impact of PASC globally on the individual patient and the broader society and economy, there is an urgency for rigorous and timely evaluation of potential treatments.

The overall goal of this study is to is to investigate the efficacy and safety of ritonavir-boosted SARS-CoV-2 antiviral medication nirmatrelvir in participants with PASC and explore biological and digital wearable biometric markers of disease and disease severity.

1.2 Study Objectives

We hypothesize that adult participants with PASC treated with nirmatrelvir/ritonavir (PAXLOVID) for 15 days will report reduced PASC symptom(s) severity at 10 weeks compared to placebo/ritonavir. The primary PASC symptoms evaluated in this study include fatigue, brain fog, dyspnea, body aches, CV symptoms, and GI symptoms (detailed in eligibility criteria section 2.2).

1.2.1. Primary Objective:

To compare the efficacy of a 15-day course of nirmatrelvir/ritonavir versus placebo/ritonavir in reducing symptom(s) severity of participants with PASC

1.2.2. Secondary Objectives:

- To compare overall symptom burden of participants with PASC treated with Paxlovid versus placebo/ritonavir
- To evaluate other patient-reported outcomes (e.g., functional status, global health status, etc.) in participants with PASC treated with nirmatrelvir/ritonavir versus placebo/ritonavir
- To identify which core PASC symptom(s) are most responsive to Paxlovid treatment, if any

1.2.3. Exploratory Objectives:

- To investigate potential biological biomarkers of PASC in participants treated with nirmatrelvir/ritonavir versus placebo/ritonavir

2. Study Design

This is a phase 2, single-center, randomized, blinded, placebo-controlled pilot trial to evaluate the efficacy and safety of nirmatrelvir-ritonavir (PAXLOVID) in treating PASC in adults and to

explore potential biological and digital wearable biometric markers of disease. A total of 200 participants with PASC who meet all the inclusion criteria will be randomized 2:1 to a 15-day course of twice-daily (a) nirmatrelvir/ritonavir (nirmatrelvir 300 mg – ritonavir 100 mg), or (b) placebo/ritonavir (placebo 0 mg – ritonavir 100 mg). A total of ~133 and ~67 participants will be enrolled in the two arms above, respectively. Randomization will be stratified by the number of moderate or severe core symptoms (2 or 3 vs >3) and vaccination status (completed primary series vs not completed as defined by CDC). The randomization list for each stratum will be generated by block randomization with block size randomly selected from 6 and 9. Entire study is estimated to last up to 12 months from first participant enrolled until the end-of-study for the last participant enrolled. For each participant, the study lasts 15 weeks (3.5 months). Symptoms severity assessments, patient-reported outcomes (PROs), clinical assessments, and specimen collection will be performed at each time point.

2.1 Sample Size Estimation

We plan to enroll 200 participants total with 180 participants completing the 10-week follow-up, i.e., a 10% drop off rate. With a planned randomization ratio of 2:1, we expect approximately 120 participants receiving nirmatrelvir/ritonavir and 60 participants receiving placebo/ritonavir completing the 15 weeks follow-up at the end of the study. The primary outcome is symptom severity for fatigue, brain fog, dyspnea, body aches, CV symptoms, and GI symptoms in Likert-scale (0, 1, 2, and 3) at 10 weeks. For each aforementioned symptom, we plan to use the proportional odds regression model for ordinal outcomes to compare the Likert scale at week 10 (stratified by their baseline level) among participants experiencing the corresponding symptoms at the baseline. The overall comparison will be made based on averaging regression coefficients from comparisons of individual symptoms weighted by the sample size of the regression analysis, i.e. the number of participants experiencing the corresponding symptom at the baseline[1, 2]. The power estimation is made based on the following simplified assumptions:

(a) for each symptom, the Likert-scale at week 10 is uniformly distributed over 0, 1, 2, and 3 in the placebo/ritonavir group, i.e., the proportions of the participants with a Likert-scale being 0, 1, 2, and 3 are 25%, 25%, 25%, and 25%, respectively, among all participants in the study including participants without the symptom at the baseline.

(b) for each core symptom, the proportions of the participants in the nirmatrelvir/ritonavir group with a Likert-scale being 0, 1, 2, and 3 are 34.8%, 26.8%, 21.2%, and 17.2%, respectively, among all study participants.

(c) The z-scores of comparing 6 symptoms using all participants are positively correlated with a pair-wise correlation coefficient of 0.25.

Assumptions (a) and (b) satisfy the proportional odds model with a odds ratio of 1.6. Under this assumption, the z-score of individual comparison based on 60 participants in the placebo/ritonavir group and 120 participants in the nirmatrelvir/ritonavir group follows a normal distribution with mean of 1.66 and unit variance. The final test statistic is equivalent to the simple average of z-scores from analyses for individual symptoms, since the assumed alternatives are identical for all core symptoms. Under assumption (c) the Wald test statistic for the overall comparison follows a normal distribution with a mean of 2.71 and unit variance, providing a power of 77% at the two-sided significance level of 0.05. The proposed test is expected to have a higher power than that of the analysis discussed above for several reasons. First, the regression analysis excluding participants without the core symptom at the baseline is expected to generate a z-score with a greater mean value and a higher power, since the potential dilution effect from participants without the core symptom at the baseline are reduced. Second, the sample size weighting is expected to

generate a more efficient combination of test statistics from individual test than equal weighting. Lastly, the stratification by baseline Likert scale will also increase the power of comparing severity of individual symptom and also the power of overall comparison.

The interim analysis is mainly for assessing futility and safety and thus does not affect the sample size calculation. For estimating the timing of the interim and final analysis, we project to enroll one participant per day. Therefore, the interim analysis is expected to be conducted approximately 24 weeks after the first participant in, by which, the outcomes at 10-week follow up will be available for 90 participants. Complete data collection for the outcomes through the 15-week follow up is projected to occur approximately 46 weeks after the first participant in.

Since this is a phase II study, the proposed sample size does not necessarily guarantee sufficient power for all secondary endpoints.

2.2 Inclusion Criteria

In order to be eligible to participate in this study, an individual must meet all of the following criteria:

1. Adults 18 years and older
2. Weight > 40 Kg
3. Normal or near-normal renal function (eGFR ≥60 ml/min)
4. History of confirmed COVID-19 infection (SARS-CoV-2 positive PCR/NAAT, positive antigen, positive nucleocapsid antibodies, or positive spike antibodies before vaccination for SARS-CoV-2)
5. Post-COVID-19 symptoms persisting greater than three months (>90 days) since the initial COVID-19 infection that caused the Long COVID. Symptoms must be present for more days than not, were not present prior to COVID-19 infection, and are not explained by another cause
6. Participant's post-COVID symptom(s) must be at least 2 of the following 6 core symptoms or symptom clusters:
    a. Fatigue
    b. Brain fog (including difficulty with focus, memory, word-finding, processing, orientation, or multitasking)
    c. Shortness of breath
    d. Body aches (muscle or joint pain)
    e. Cardiovascular symptoms (including chest pain, tachycardia, palpitations, lightheadedness)
    f. Gastrointestinal symptoms (including nausea, vomiting, diarrhea, constipation, abdominal pain, decreased appetite)
7. Severity of at least two of the core PASC symptom(s) above must be moderate or severe, 2 or 3 on a Likert-scale of 0 to 3 (where 0 is absent, 1 is mild, 2 is moderate, and 3 is severe)
8. Willing to report all vaccinations received prior to and during the study, if any
9. Women of childbearing potential (not surgically sterile or 2 years postmenopausal) must use a medically accepted method of contraception during the treatment period and must agree to continue use of this method for 2 months after the last dose of the study intervention

10. Women of childbearing potential must be agreeable to a urine pregnancy screening test
11. Men whose partners may become pregnant must use adequate contraception during the treatment period and must agree to continue use of this method for 2 months after the last dose of the study intervention
12. Willing and able to adhere to study procedures and available for the duration of the study

2.3 Exclusion Criteria

An individual who meets any of the following criteria will be excluded from participation in this study:

1. Suspected or confirmed pregnancy or breastfeeding
2. Severe liver disease (Child-Pugh class C)
3. Use of the study drug within past 30 days of randomization or planned use of the study drug outside of FDA-authorized indication for the duration of the study
4. Receiving other COVID-19 specific treatments within 30 days of randomization
5. History of hypersensitivity or other contraindication to any components of the study drug
6. Current or expected use of any medications or substances including supplements and herbs that are highly dependent on CYP3A4 for clearance or are strong inducers of CYP3A4 or have known drug-drug interactions with study drug
7. Use or planned use of any supplements or herbs during study drug administration and potential additional time before and after this period as determined necessary by investigators, unless medically indicated (e.g., for nutrient deficiency) and determined to be safe by investigators
8. Known Human Immunodeficiency Virus (HIV) infection with viral load >50 copies/ml or taking prohibited medications for HIV treatment
9. Suspected or confirmed active SARS-CoV-2 infection within past 30 days
10. New or change in the dosing of immune-modulating or immunosuppressive medications within 30 days prior to enrollment until the primary endpoint (10 weeks) has been reached
11. Any other medical condition or concomitant medication/therapy that would compromise the patient's safety or compliance with the study protocol or significantly confound interpretation of the clinical and research tests, as determined by study investigators
12. History of COVID vaccine received within 28 days prior to enrollment or other vaccine (including influenza vaccine, shingles vaccine, etc.) within 14 days of enrollment or planned use of any investigational, authorized, or approved vaccine until the primary endpoint (10 weeks) has been reached
13. Current enrollment in, or discontinuation within the last 30 days from, a clinical trial involving any investigational drug or device, or concurrent enrollment in any other type of medical research judged not to be scientifically or medically compatible with this study
14. Inability to provide informed consent
15. Currently hospitalized

3. Endpoints

3.1 Primary Endpoints

The primary endpoints are core symptoms severity based on Likert-scale score at 10 weeks in participants. Core symptoms are defined as: fatigue, brain fog, dyspnea, body aches (muscle and/or joint pain), gastrointestinal symptoms (including nausea, vomiting, diarrhea, constipation, abdominal pain, or decreased appetite) and cardiovascular symptoms (including chest pain, fast heart rate, palpitations, or lightheadedness). The severity of the symptoms will be assessed on a 4-point Likert scale (0 = none, 1 = mild, 2 = moderate, and 3 = severe) with participants rating their symptom burden at its worst during the past seven days

3.2 Secondary Endpoints

Key secondary endpoints include

- Core symptom severity scores based on Likert-scale score at 15 days in participants treated with nirmatrelvir/ritonavir versus placebo/ritonavir, for each of the core symptoms
- Proportion of participants reporting relief at 10 weeks of at least one core symptom for 2 weeks. Relief defined as reduction of severity from moderate to none or severe to mild/none, ≥ 2-point Likert score change (see Section 4.1 for additional details)
- Proportion of participants with overall alleviation at 10 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir group for 2 weeks. Overall alleviation defined as both: (1) any core symptom(s) that are none/mild (Likert 0 or 1) at baseline are none/mild at 10 weeks *and* (2) any core symptom(s) that are moderate/severe (Likert 2 or 3) at baseline are none/mild at 10 weeks (see Section 4.1 for additional details)
- Severity of the most bothersome symptom at 5 weeks, 10 weeks, and 15 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir group
- Time to relief of each of the 6 core symptoms. Relief defined as above.
- Change in PROMIS Physical Function SF 4a v2.0 from baseline to 10 weeks
- Change in PROMIS Fatigue SF 7a v1.0 score between baseline and 10 weeks
- Change in PROMIS Dyspnea-Severity SF 5a v1.0 score between baseline and 10 weeks
- Change in PROMIS Cognitive Function Abilities score between baseline and 10 weeks
- Change in orthostatic vitals (difference in supine versus standing blood pressure and heart rate) from baseline and 10 weeks
- Change in 1-minute sit-to-stand test from baseline and 10 weeks
- Patient Global Impression of Severity scale (PGIS) at 15 days, 5 weeks, 10 weeks, and 15 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir group
- Patient Global Impression of Change scale (PGIC) at Day 15, 5 weeks, 10 weeks, and 15 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir group

Other secondary endpoints:

- Summative core symptoms severity measured by summation of Likert scale score of 6 core symptoms at 5 weeks, 10 weeks, and 15 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir group.
- Proportion of participants reporting relief of most bothersome symptom at 5 weeks, 10 weeks, and 15 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir group. Relief defined as above.
- Proportion of participants reporting relief for each of the 6 core symptoms at 5 weeks, 10 weeks, and 15 weeks. Relief defined as above.

- Time to relief of at least one core symptom without other core symptoms worsening at 5 weeks, 10 weeks and 15 weeks. Relief defined as above.
- Time to relief of the most bothersome symptom at 5 weeks, 10 weeks and 15 weeks. Relief defined as above.
- Expanded symptoms severity at 15 days, 5 weeks, 10 weeks, and 15 weeks (endpoint specified *post hoc*).
- Proportion of time (weeks) with mild or no symptoms after baseline among participants with the symptom at baseline, for each of the 6 core symptoms (endpoint specified *post hoc*).

3.3 Exploratory Endpoints

Exploratory endpoints include:

- Clinical laboratory test: d-dimer
- Stool RNA RT-qPCR to assess for presence of SARS-CoV-2 at baseline, 7 days, 15 days (immediately after treatment), 10 weeks (1)
- Difference in nirmatrelvir/ritonavir versus placebo/ritonavir and change from baseline to 7 days, 15 days, and/or 10 weeks of:
  - O-link plasma proteomic profiles, including specific cytokines and chemokines such as CCL11, others (2)
  - Autoantibodies – bead-based, multiplexed assays (3, 4)
  - Whole blood gene expression as measured by bulk RNA-Sequencing (5)
  - Anti-SARS-CoV-2, EBV, CMV and other IgG responses, measured in a combined panel of autoantibodies described above (3)
  - Proportions of blood cells measured by flow cytometry, including anti-SARS-CoV-2 T cell responses measured using spheromers (6)
- Other assays that may emerge regarding PASC pathophysiology
- Change in digital biometric wearables measures from baseline to 15 days, 5 weeks, 10 weeks, and 15 weeks in nirmatrelvir/ritonavir versus placebo/ritonavir groups for:
  - Physical activity (PA) measured by hand accelerometer
  - Ratio of daytime versus nighttime PA level
  - Time asleep and time in bed
  - Heart rate and resting heart rate
  - Heart rate variability
  - ECG rhythm abnormalities
  - O2 saturation
  - Blood pressure

4. Statistical Analysis Plan

4.1 General approach:

The statistical analysis will be conducted following the intention to treat principle. Randomization will be stratified by vaccination status (completed primary series vs not completed) and the number of symptoms at the baseline (2 or 3 symptoms vs >3 symptoms).

Core symptom surveys collected during the screening period will be treated as baseline. If a participant is missing a survey from the screening period, the week 0 survey will be treated as baseline. For participants who missed their 10 week survey, we will use the survey completed closest to 70 days and within the 10 week visit window (days 63-77 inclusive) for their week 10 symptom survey responses. All tests will be two-sided and performed at the alpha = 0.05 level of significance unless otherwise noted.

4.1.1 Definitions

- In the definitions below that require two consecutive surveys, two completed surveys can be considered consecutive if a single interval survey is missing between two surveys that were completed within a 3-week period. If two or more interval surveys are missing between two completed surveys, the two completed surveys will not be considered consecutive. The criteria must be met for the same symptom in consecutive weeks for the endpoint to be met.
- For a participant to meet the endpoint of relief at 10 weeks, they must 1) have a reduction of ≥ 2-points in at least one core symptom from the baseline Likert score at 10 weeks and 2) the reduction of ≥ 2-points for that core symptom must be present on at least 2 consecutive weekly symptom surveys including their 10 week survey.
- For a participant to meet the endpoint of alleviation at 10 weeks, they must meet all of the following criteria 1) any core symptom(s) that are none/mild (Likert 0 or 1) at baseline are none at 10 weeks, 2) any core symptom(s) that are moderate/severe (Likert 2 or 3) at baseline are none/mild at 10 weeks, and 3) the alleviation must last for at least 2 consecutive weekly symptom surveys including their 10 week survey.
- Time to relief is defined as time from randomization until the first survey where a participant meets the criteria for relief for at least two consecutive weekly symptom surveys. For example, if a participant meets the criteria for relief at weeks 6, 7, 8, 9, and 10, then their time to relief is 6 weeks. If a participant meets the criteria for relief at weeks 6, 8, and 10 but not at 7 or 9 weeks or any week after, then they should be right censored at the time of their last survey because they never met the relief endpoint. If a participant meets the criteria for relief at weeks 6, 8, and 10 and is missing their surveys for weeks 7 and 9, then they would meet the endpoint at week 6. Participants who ever meet this endpoint will be defined as “responders.”
- Severity of the most bothersome symptom at a given timepoint will be defined as the severity reported for the symptom that was selected as the most bothersome by the participant at baseline.
- The summative core symptoms severity at a given time point will be defined as the summation of the Likert scale scores for the six core symptoms at that time point. The scores assigned for each symptom will be None = 0, Mild = 1, Moderate = 2, and Severe = 3. The summative core symptoms severity can range from 0-18.

- Proportion of time (weeks) with mild or no symptoms after baseline will be calculated as (number of weeks with mild or no symptoms)/(number of weekly symptom surveys returned) for each symptom within each participant reporting the corresponding symptom at baseline.
- A participant is treatment compliant if they took at least 80% of the doses for both nirmatrelvir and ritonavir. The primary source of compliance will be the pill count; however, if the pill count is not available for a participant the number of doses will be determined by the participant's self-reported adherence survey responses.

4.2 Populations for Analyses

The intent-to-treat (ITT) population will include all randomized participants. Participants will be analyzed according to their assigned treatment arm. All efficacy analyses will be completed in the ITT population.

The modified ITT (ITT) population will include all randomized participants except participants who provide no follow-up data. Participants will be analyzed according to their assigned treatment arm.

The per-protocol (PP) population will include all randomized participants who have completed follow-up for the corresponding visit (e.g. participants who are missing 10-week data will be excluded from analyses looking at endpoints measured at 10 weeks) and were treatment compliant. All efficacy analyses will also be completed in the PP population as supportive evidence for the primary efficacy analysis.

The safety population will include all participants who received at least one dose of study treatment. Participants will be analyzed according to actual treatment received. All safety analyses will be completed in the safety population.

4.3 Demographic and Clinical Characteristics

Descriptive statistics (proportions for categorical variables, means, medians, standard deviations and/or interquartile ranges for continuous variables) will be reported for all key participant variables, including baseline and demographic characteristics, use of medications, compliance, and study completion status. The participants' characteristics at baseline will be summarized and compared between two treatment arms to examine the balance achieved via the randomization. The categorical variables will be summarized by frequency and proportion and the continuous variables will be summarized by mean and standard deviation or median and inter-quartile range as appropriate. Stratified descriptive analyses will be performed by the number of symptoms at the baseline (2 or 3 symptoms vs >3 symptoms).

4.4 Analysis Plan for Aim 1

We will evaluate the intention-to-treat (ITT) population for the primary efficacy analysis. All analyses will be stratified by the baseline Likert-scale and will be adjusted for vaccination status.

We will also analyze the mITT and the per-protocol (all randomized participants who completed follow-up, adhered to study procedures, and did not have diagnosed acute COVID-19 reinfection during the study) groups as supportive evidence for the primary efficacy analysis.

The primary endpoint is the 4-point Likert-scale of 6 core symptoms at week 10 follow-up. For participants who missed their 10 week visit, we will use the survey completed closest to 70 days and within the 10 week visit window (days 63-77 inclusive) for their symptom survey responses.

For the primary endpoint: 6 core symptoms severity at 10 weeks, we hypothesize

- H_0 = There is no difference in the symptom severity in *any* of the 6 core symptoms at 10 weeks between nirmatrelvir/ritonavir and placebo/ritonavir groups.
- H_1 = There is a difference in the symptom severity of at least one of the 6 core symptoms at 10 weeks between nirmatrelvir/ritonavir and placebo/ritonavir groups

In the subgroup of participants having a positive Likert-scale for each core symptom, we will compare the Likert-scale of the same symptom at week 10 between participants receiving nirmatrelvir/ritonavir for 15 days and participants receiving placebo/ritonavir using proportional odds regression for each core symptom(s). The regression model will be fit within each baseline symptom severity for each symptom (up to three models for each of the six symptoms, for a maximum total of 18 models). If a regression model in any of the strata has too few participants to be fit (e.g. the model did not converge), the stratum will be combined with the nearest symptom severity within the same core baseline symptom. If levels are collapsed, the number of model estimates to be combined will be less than 18. The overall comparison will be made based on a weighted average of the regression coefficient (the log odds) from each baseline severity and core symptom stratum weighted by the inverse variance of the coefficient. Multiple imputation will be used to impute missing survey responses. Rubin's rules[1] will be used to combine estimates across imputations and derive a standard error that accounts for uncertainty due to missingness. A permutation test will be used to generate the null distribution accounting for correlations among estimated regression coefficients for different symptoms[2, 3], and also the p-value for the overall comparison. This test aggregates the statistical evidence of treatment benefit for each core symptoms and is more powerful if the nirmatrelvir/ritonavir has a moderate beneficial effect in treating most or all six core symptoms. Note that since the comparison is stratified by the baseline Likert scale, comparing the symptom severity at the week 10 is equivalent to comparing the change in severity from baseline to week 10.

In a sensitivity analysis performed in the mITT population, we will repeat this analysis with weights equal to the sample size of the subgroup in which the regression analysis is conducted, i.e., the number of patients having the corresponding symptom and severity at baseline. Missing survey responses will not be imputed in this analysis.

4.5 Analysis Plan for Aim 2

The key secondary endpoints include core symptoms severity at 15 days, severity of most bothersome symptom at 5 weeks, 10 weeks and 15 weeks, proportion of subjects reporting relief for 2 weeks of at least one core symptom (reduction of severity from moderate to none or severe to mild/none, a 2-point or greater Likert score change) at 5 weeks, 10 weeks and 15 weeks, and time to relief of at least one core symptom without other symptom worsening, and time to relief of the most bothersome symptom. All models for secondary endpoints will adjust

for the randomization factor of 2-3 core symptoms vs >3 core symptoms. We will not adjust for the randomization factor of vaccinated vs not because only two randomized participants were not vaccinated.

We will fit proportional odds models to compare the severity of most bothersome symptom at each follow-up visit, severity of core symptoms at 15 days and 10 weeks, and severity of symptoms captured in the expanded symptom survey at 10 weeks. The odds ratio, i.e., the exponential of the regression coefficient of the treatment, for measuring the treatment benefit of nirmatrelvir/ritonavir for the corresponding symptom will be presented. The 95% confidence interval of the odds ratio will also be estimated. Subgroup analysis will be conducted by repeating the analyses for the core symptom endpoints according to the most bothersome symptom at the baseline.

Linear regression will be used to compare 1) the global impression of severity at follow-up visits, 2) the global impression of change scale as well, 3) PROMIS Fatigue-SF, 4) PROMIS Dyspnea Severity, 5) PROMIS Cognitive Function Abilities, and 5) the summative core symptoms severity. Logistic regression will be used to compare the probability of experiencing relief for at least one core symptoms. For secondary endpoints, we will also use a mixed effects model for repeated measurements (MMRM) to compare the longitudinal endpoints between arms across all weeks in the same model in an exploratory analysis to evaluate trajectory of symptoms. The MMRM will have a random effect to account for correlation in repeated measurements over time in the same participant. Day of survey completion will be used to model time non-linearly using smoothing splines. The MMRM analysis can borrow additional information from subjects not completing all follow-up visits in two group comparisons and also allow comparing the average endpoints across all longitudinal visits. Lastly, for each core symptom, we will use a Cox proportional hazards model using Efron's approximation for ties to compare the time to relief of the core symptoms. Participants who don't experience any symptom relief will be right censored at the time of their last observation.

There will be no formal adjustment for multiple testing in analyzing the secondary endpoints. The analysis results for all secondary endpoints will be reported regardless of their statistical significance level, allowing post-hoc adjustment of multiple testing and providing a global picture of the treatment benefit of nirmatrelvir/ritonavir for long COVID patients.

4.6 Exploratory Analysis

The exploratory endpoints include: d-dimer and other clotting assays, stool RNA RT-qPCR and metagenomic sequencing at baseline, 15 days (immediately after treatment), 10 weeks, and 15 weeks, change from baseline of O-link proteomic profiles, specific chemokine, Autoantibodies, Whole blood gene expression as measured by bulk RNA-sequencing, Anti-SARS-CoV-2, EBV, CMV and other IgG responses, and proportions of blood cells. Other assays may be included as they become available. nirmatrelvir/ritonavirLinear regression will be used to compare each of these endpoints between nirmatrelvir/ritonavir and placebo/ritonavir. These comparisons may include healthy controls obtained from other studies. Additional analyses will evaluate the treatment response separately by treatment responders vs not (as defined in Section 4.1.1). Additional definitions of "responder" may be considered in exploratory analyses.

Analyses of the wearable endpoints are described in a separate SAP.

4.7 Safety Analysis

The frequency of adverse events and serious adverse events will be tabulated by type and by treatment arm. AEs will be compared by arm using the Chi-squared test or Fisher's exact test, as appropriate, in the safety analysis set. We will additionally summarize the hospitalizations and emergency department visits by treatment arm.

4.8 Interim Analysis

We will include an interim analysis for safety and futility assessment, when 50% of the planned participants' outcomes at week 10 are available. To this end, for each core symptom at 10 weeks, same method outlined in 4.1 will be used to include all enrolled subjects who have completed at least one follow up visit at the interim analysis. A subject-specific random intercept will be included to account for within-subject correlations among symptoms severity at different visits. We will calculate the conditional power based on data available at the interim analysis, assuming the underlying treatment effect size is the same as that observed in the interim analysis. If the conditional power at the interim analysis is less than 10%, the trial will be stopped for futility. Furthermore, the reinfection rate, early drop off rate, and the proportion of participants with SARS-CoV-2 positive PCR/NAAT and/or positive antigen at the study baseline will be summarized and their impact on study power will be examined. The DSMB will evaluate overall safety and characteristics of the patient population including acute reinfections and potentially recommend enrollment adjustments.

For estimating the timing of the interim and final analysis, we project to enroll one participant per day. Therefore, the interim analysis is expected to be conducted approximately 24 weeks after the first participant in, by which, the outcomes at 10-week follow up will be available for 90 participants.

4.9 Sub-group Analysis

Subgroup analyses will be conducted to assess potential effect modification on the primary and secondary endpoints at 10 weeks by the following baseline participant characteristics:

- Race/ethnic groups
- Sex (male vs female)
- Age (< 50 vs ≥ 50)
- Number of known COVID-19 infections (1 vs > 1)
- Time since index COVID-19 infection (≤ 6 months vs > 6 months)
- Time since last COVID-19 vacccine (≤ 6 months vs > 6 months or not vaccinated)
- Number of core symptoms at baseline (2-3 vs > 3)
- Neurologic symptoms (reported at baseline vs not)
- Prior use of any anti-SARS-CoV-2 therapeutic (yes vs no)
- Did not have acute COVID-19 reinfection at baseline or during the study (either a positive antigen test or a positive PCR test)
- Positive treatment expectations (strongly or moderately disagree vs slightly disagree, neutral, or slightly agree vs strongly or moderately agree)
- Negative treatment expectations (strongly or moderately disagree vs slightly disagree, neutral, or slightly agree vs strongly or moderately agree)
- If a distinguishing biomarker is found, we may evaluate the baseline value as a potential effect modifier

The treatment effects across subgroups will be summarized using a forest plot. The appropriate treatment by covariate interaction will be tested for detecting treatment effect heterogeneity.

4.10 Additional Analyses

In a sensitivity analysis, we will adjust the primary and secondary efficacy endpoint analyses by the following covariates: race, ethnicity, sex, age (< 50 vs ≥ 50), number primary of known COVID-19 infections (1 vs > 1), time since index COVID-19 infection, number of core symptoms at baseline (2-3 vs > 3), and prior use of any anti-SARS-CoV-2 therapeutic.

We'll graphically compare the responses in our core symptom surveys for fatigue, shortness of breath, and brain fog to the validated PROMIS surveys (PROMIS Fatigue-SF, PROMIS Dyspnea Severity, and PROMIS Cognitive Function Abilities). For each symptom, we will create a boxplot showing the PROMIS scores within each severity level of our corresponding core symptom.

4.11 Handling of Missing Data

Quality control (QC) procedures will be implemented beginning with the data entry system and data QC checks that will be run on the database will be generated. Substantial missing data or data anomalies will be communicated to the study CRC for clarification/resolution.

Data that are missing on key participant characteristics and the outcome will be fully described, including any patterns of missingness (i.e., any relationships between missingness of a variable and participant characteristics). Graphical tools such as histograms, boxplots, and scatterplots will be created to assess quality of data and to display patterns over time.

As described above, multiple imputations will be used in the analysis of the primary endpoint. In the analyses of secondary endpoints, multiple imputation methods will be used to account for additional uncertainty when a participant fails to response a question or a variable is missing.

5. Potential Limitations

Power may be limited if many participants are censored due to active COVID.

This Phase 2 study was designed to identify useful endpoints in long COVID patients. If we find that a secondary endpoint is meaningful, we will investigate that endpoint more thoroughly.

6. Example Conclusions

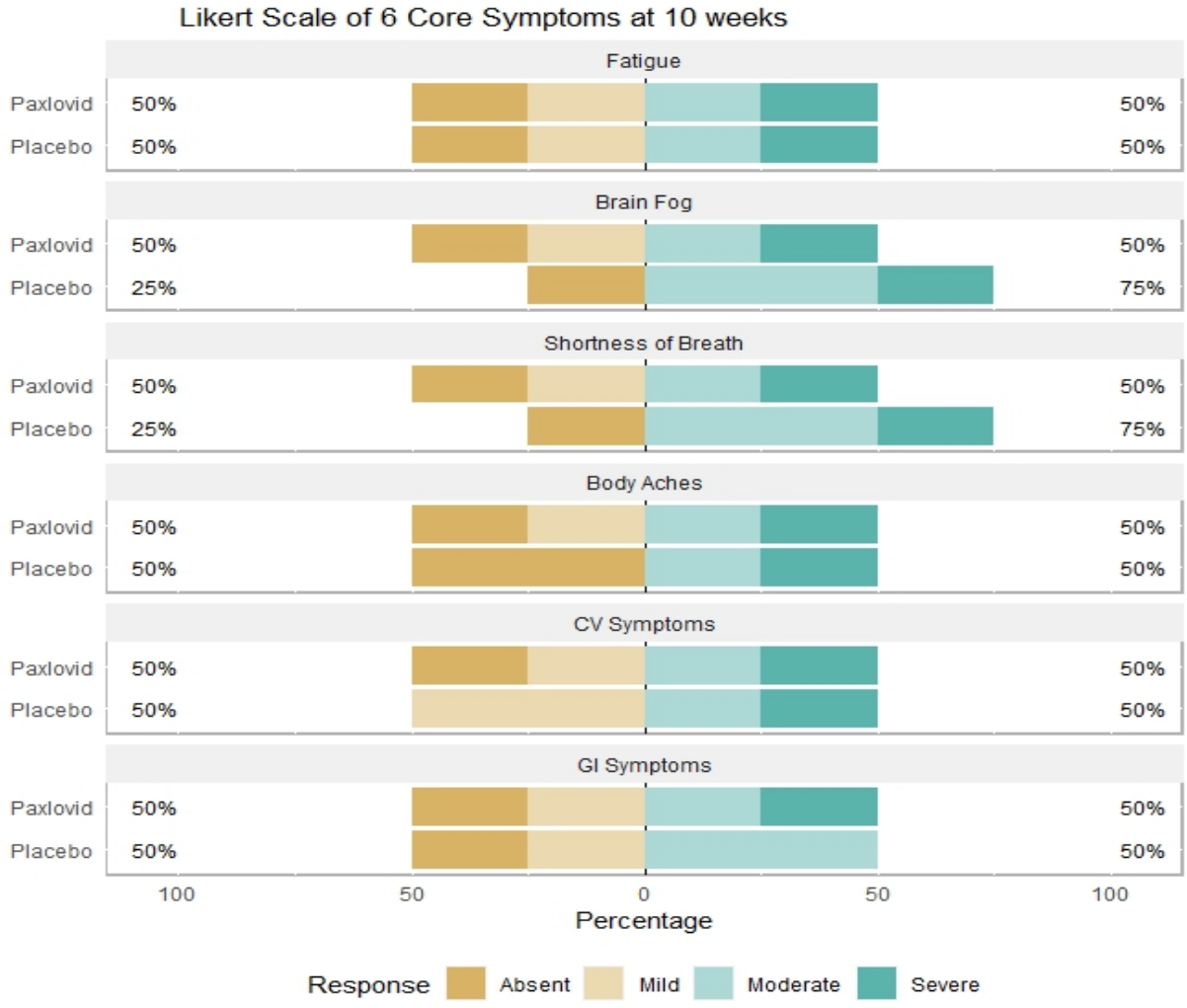
An example of the conclusions statement if the study is positive would read as follows:

"We found a difference in the symptom severity of at least one of the 6 core symptoms at 10 weeks between the nirmatrelvir/ritonavir and placebo/ritonavir groups."

An example of the conclusions statement if the study is negative would read as follows:

“We did not find a difference in the symptom severity of any of the 6 core symptoms at 10 weeks between the nirmatrelvir/ritonavir and placebo/ritonavir groups.”

In either case, we plan to show a figure demonstrating the distribution of responses in each group at 10 weeks. This figure can be created using the ‘likert’ package in R.



Additional figures:

- A heatmap of participant symptoms over time by treatment arm (one figure for each symptom) with patients sorted by symptom severity

- Trajectory over time for each of the 6 core symptoms in nirmatrelvir/ritonavir vs placebo/ritonavir

7. References

1. Marshall A, Altman D, Holder R, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Medical Research Methodology. 2009, 9 (1): 57-10.1186/1471-2288-9-57.
2. Wei LJ, Johnson WE. Combining dependent tests with incomplete repeated measurements. Biometrika. 1985;72(2):359-64.56.
3. Xu X, Tian L, Wei LJ. Combining dependent tests for linkage or association across multiple phenotypic traits. Biostatistics. 2003;4(2):223-9.

8. SAP Revision History

Revision	Date	Section/Page	Changes Made -- Reasons for the Change
1.0	10/7/22	N/A	Original
1.1	10/10/23	Section 4.1.1 on pg. 8, Section 4.2 on page 7-8, Section 4.4 on pg. 9-10Section 4.9 on pg. 10, Section 4.10 on pg. 11, throughout,	Definitions of endpoints are clarified; Analysis population section added and modified ITT population defined; Additional details provided for the primary analysis; Acute COVID-19 defined and specified as a sub-group analysis rather than part of the PP population definition; Removed duplicate language from the "Additional Analysis" section 4.10 and added an adjusted analysis as a sensitivity analysis; Additional clarifications added throughout;