Statistical Analysis Plan

Protocol Title: UTAH One (Understanding Treatment And Health in the Ongoing coroNa Epidemic): A Hydroxychloroquine Outpatient Study; formerly *Hydroxychloroquine for Outpatients with Confirmed COVID-19 (HCQ Trial)*

Protocol Version and Effective Date: v 1.6; July 6, 2021

Past Protocol Versions and Effective Dates: First (misnumbered as 0.1 but should have been 1.00, April 10, 2020); 1.01 (April 17, 2020); 1.2 (April 27, 2020); 1.3 (May 13, 2020); 1.4 (June 16, 2020); 1.5 (September 24, 2020)

SAP Authors: Bradley Barney, PhD; Tom Greene, PhD; Richard Holubkov, PhD (with contributions from Cody Olsen, MS, and reusing some DCC language)

SAP Version and Date: v1.4; February 23, 2022

SAP Version 1.3 Date: July 30, 2021 (created); Effective August 3, 2021
SAP Version 1.2 Date: May 26, 2021 (created); Effective May 27, 2021
SAP Version 1.1 Date: February 19, 2021 (created); Effective March 8, 2021
SAP Version 1.0 Date: November 18, 2020 (created); Effective December 1, 2020

CONFIDENTIAL

Approvals:

Approved By:

Adam M. Spivak MD		23 February 2022
Adam Spivak	Principal Investigator	Date
Rachel Hess, MD, MS		23 February 2022
Rachel Hess	Principal Investigator	Date
Richard Holubkov	Richard Holubkov, PhD	March 7 2022
Richard Holubkov	Trial Statistician	Date
Tom Greene, PhD		March 10, 2022
Tom Greene	Trial Statistician	Date
Digitally signed by Bradley J. Barney, PhD Date: 2022.03.10 16:03:46 -07'00' Bradley J. Barney, PhD		March 10, 2022
Bradley Barney	Trial Statistician	Date

Partial List of SAP Revisions:

Batch of changes for preliminary version of 24 May 2020

- Section 2.1.3: Changed a secondary outcome from duration of symptoms to average symptom level. Moved the duration of symptom outcomes to an exploratory outcome.
- Section 5.4: Added descriptions pertaining to the average of COVID-19 attributable symptoms. Added information to derive EQ-5D-5L VAS and index measure outcomes.
- Section 7.4.5: Revised analyses to incorporate baseline and Day 28 assessments, and detailed the statistical analyses of the two outcomes.
- Section 7.4.6: Revised analyses to incorporate baseline and Day 28 assessments.
- Section 7.4.7: Revised analyses to incorporate baseline and Day 28 assessments.
- Section 7.4.8: Revised analyses to incorporate baseline and Day 28 assessments

Batch of changes for preliminary version of 30 June 2020

- Section 4.2.6: Modified definition of per-protocol efficacy population (which doubles as the principal safety population) to require at least 80% of the assigned pills/tablets.
- Section 2.1.3: Included two additional core symptoms (chills and diarrhea).
- Section 5.4: Modified definition of symptom duration and start/end dates. Restricted to individuals symptomatic at baseline. Applied censoring at 15 days rather than using symptom assessments through 28 days because there is not daily symptom information after Day 15 (different from protocol). Added details for post-imputation processing of certain values for participants missing some assessments because they were deceased when the assessment was to have been administered.
- Section 7.4: Removed analysis of numerical viral shedding values.
- Section 7.4.5: Updated imputation approach and analytical methods of the EQ-5D-5L assessments (and subsequent assessments).

Batch of changes for preliminary version of 16 July 2020

- Signature Page: Restricted to PIs and trial statisticians, removing the clinical data manager line as this was deemed unnecessary.
- Section 2.1.3: Changed a secondary outcome from "Duration of viral shedding, measured on days 1 - 14 and 28" to "Persistence of viral shedding on day 28." The secondary outcome had essentially been the primary outcome with one later measurement and as such duration is not accurately captured beyond what is available in the primary outcome. The replacement secondary outcome maintains complementary status with the primary outcome and is straightforward.

- Section 4.2.4: Changed the statement that the ITT population "will be the primary population for most efficacy and safety analyses" to "will be the primary population for efficacy analyses."
- Section 4.2.5: Changed wording re sensitivity analyses of efficacy outcomes on the dITT population from stating they will be conducted to stating they may be conducted. Also clarified that the dITT population is restricted to subjects who are delivered the assigned study drug.
- Section 4.2.7: Defined the safety population to require only one pill/tablet of the delivered study drug to be administered, and that the arms for the SAFETY population are based on what was delivered rather than what was assigned, in case there are discrepancies. Previously the principal safety population for adverse events not already described as an outcome was the per-protocol population.
- Section 5.4: Added definition of assessment days and defined how to handle drug days and assessment days if the anchoring events (drug delivery and assessment surveys) are not available. Corrected censoring explanation for symptom duration to censor on last *symptomatic* day, as applicable. Added definition of comorbidity presence/absence.
- Section 7.1: Added the SAFETY population as a component of the CONSORT-like diagram. Clarified that footnotes may be used to describe some populations.
- Section 7.3.2: Modified to include all 15 daily assessments in defining the univariate average symptom level outcome. Further restricted imputation and analysis to individuals who are not known to have been hospitalized or died on or before $DayA15$, and who have at least one day with a non-missing average symptom level for $DayA1$–$DayA15$. Previous restriction had specified those not meeting the primary outcome or died within 14 days.
- Section 7.3.3: Moved what had been described in Section 7.4 as *Viral shedding at* $DayD28$ to this section. This reflects the substitution of a secondary outcome from *viral shedding duration through Day 28* to *persistence of viral shedding on day 28*.
- Section 7.4.1: Added a sensitivity analysis for symptom duration that assumes symptom resolution on the first asymptomatic day, regardless of subsequent symptom levels. Also emphasized the departure from the originally specified outcome in the protocol which also considered the Day 28 measurements.
- Section 7.4: Reinserted numerical viral shedding values as an exploratory outcome and described a tentative analysis plan.
- Section 7.5: Stated when safety (rather than per-protocol) and intent-to-treat populations would be used for safety analyses.

Batch of changes for preliminary version of 22 July 2020

- Section 2.1.3: Explained the change from a secondary outcome of *Duration of viral shedding, measured on days 1-14 and 28* to *Persistence of viral shedding on day 28*.
- Sections 5.4, 7.3.3: Described when and how missing Day 28 viral shedding values will be assumed positive or negative, with other values imputed unless the participant has died on or before $DayR + 30$. Explained there is a window for obtaining the Day 28 swabs.
- Section 5.4: Adjusted the definition of the adult household acquisition variable to replace "test" with "swab," assuming only the outcome of study-associated swabs will be systematically collected in the database. Also, changed from considering results "at any time" to looking at the results of the collected swabs because it is unclear that tests other than from non-study swabs will be available, and there is a bounded interval for when study swabs are to be collected.
- Section 7.3.2: Made the restriction less narrow to not exclude subjects who were hospitalized or who died on or before $DayA15$. Described how the last value carried forward approach will be used in conjunction with multiple imputation.
- Section 7.2.1: Clarified that the restricted mean duration and median duration analyses of the primary outcome will be stratified by *age group*. Explained how the confidence intervals will be created via nonparametric-bootstrap resampling and empirical percentiles.
- Section 7.4: Removed the viral shedding numerical outcome analysis paragraph (and from the list in Section 5.4).

Batch of changes for preliminary version of 27 July 2020

- Section 5.4: Clarified how assessment days after $DayA1$ are defined. Also noted that "swab days" are intended to align with drug-delivery days, and that like assessment days, a window is allowed for the $DayD28$ measurements.

Batch of changes for preliminary version of 28 July 2020

- Section 5.4: Clarified what to do if there are multiple daily averages on the same calendar day or if there are multiple swabs on the same calendar day. Affects the definition of the viral shedding duration (the primary outcome), the average symptom level (a secondary outcome), the persistence of viral shedding on day 28 (a secondary outcome), and the symptom duration (an exploratory outcome).
- Section 5.4: Amended the definition of the primary outcome to clarify that swabs after $DayD14$ are ignored when defining the primary outcome, even for purposes of confirmation of an earlier negative result.

- Section 5.4: Edited the first sentences describing the persistence of viral shedding on day 28 outcome to emphasize this is an outcome that replaced the originally specified outcome of viral shedding duration from days 1-14, 28.
- Section 7.4: Added a note about a possible exploratory analysis for numerical viral shedding measurements. The note was added to reconcile why the analysis of this exploratory outcome is not described in the current version of the SAP, despite the inclusion of such language in the original protocol.

Batch of changes for preliminary version of 29 July 2020

- Section 5.4: Defined several distinct outcomes for adult household contact viral acquisition. The main such outcome is now at the household level rather than at the person level. Additional outcomes are at the person level, as a dichotomous outcome and as a count outcome. These household contact acquisition outcomes, and the analysis methods applied to them, are still subject to change after consulting with an experienced infectious disease modeling person and better understanding incoming data.
- Section 7.3.2: Clarified Step 4 for post-processing of imputed values on hospitalized days.
- Section 7.4.1: Inserted a missing word, "a," in the first sentence.

Batch of changes for preliminary version of 03 August 2020

- Section 5.2: Added: "When multiple imputation is employed, it will be performed separately within each treatment arm if possible (by ITT arm for analyses of the ITT population or subpopulations thereof, or, if necessary, by SAFETY arm for analyses of the SAFETY population)."
- Section 7.3.2: Instead of including treatment as a variable in the imputation model, specified that imputation will be by treatment arm.
- Section 7.3.2: Added the last paragraph to clarify that the principal analysis for the average symptom level outcome is on the same set of subjects who are included in the imputation model for this outcome.
- Section 7.3.3: Added to the imputation model for the viral shedding at Day 28 outcome the following qualifier for the inclusion of a treatment arm indicator: "(if imputation is not already stratified by treatment arm)".

Batch of changes for preliminary version of 07 November 2020

- Section 5.4: Changed text to assume that collected surveys/swabs pertain to the intended day. As such removed text pertaining to assessments/swabs collected on the same calendar day. (Supersedes changes from 28 July 2020.)

- Sections 5.4 and 7.3.4: Clarified for the adult household viral acquisition outcome that it considers swabs from adult household contacts (rather than also considering swabs from the index subject).

Batch of changes for preliminary version of 11 November 2020

- Section 5.4: Changed text to assume that collected surveys/swabs pertain to the intended day for surveys, but to the participant-reported collection date for swabs. (Supersedes changes from 07 November 2020.) Added specific language about permissible windows for surveys and for "day 28" swabs.

Batch of changes for preliminary version of 13 November 2020

- Section 5.4: Removed redundant language about swab collection dates being per participant report, and corrected one reference to swabs that was supposed to refer to surveys.

Batch of changes for Version 1 of 18 November 2020

- Updated study name/acronym and latest protocol version number/date.
- Removed all references to the SAP version being preliminary.
- Section 7.5.2: Changed language from closed session of DSMB meetings being "tentatively planned for after approximately 100, 200, 300, and all subjects have been enrolled" to "tentatively planned for after approximately 100, 300, and all subjects have been enrolled." Within the last week, the DSMB meeting scheduled for November 30th was postponed. Nearly 200 were enrolled as of the accompanying October screen cutoff date that would have been used if the meeting were to take place November 30th. With a new tentative target of mid- to late-January for the next DSMB meeting, enrollment may be near 300.

Batch of changes for Version 1.1

- Section 5.4: Removed ", with the day 28 wording asking about the past two weeks" because the day 28 symptom assessment asks about one day, not for the past two weeks.
- Section 5.4: Corrected the erroneous statement about the symptom duration exploratory outcome: "Therefore, the $DayA15$ assessments asking about symptom levels over the previous two-week period will be ignored." It now states, "Therefore, the $DayA28$ symptom assessments will be ignored when determining symptom duration."
- Section 7.5.2: Removed the phrase "(tentatively planned for after approximately 100, 300, and all subjects have been enrolled)" in describing the closed session of DSMB meetings. The change of 18 November 2020 to remove "200" as a planned frequency

should not have been made as the meeting postponement was not intended to reduce the number of meetings, but rather to delay one meeting. The DSMB charter describes the ability of the DSMB to have flexibility in the frequency of meetings. The removed text is not necessary for the SAP, regardless.

- Corrected page headers to state "Statistical Analysis Plan" instead of "Provisional Draft of Statistical Analysis Plan for Internal Review."
- Cover page: Noted that September 24, 2020 is the effective date (in this case, also the IRB approval date) for version 1.5 of the protocol. Added prior protocol version numbers and effective dates. Added the effective date for version 1.0 of the SAP.

Batch of changes for Version 1.2

- Cover page: Added approval date for SAP Version 1.1 and creation date for SAP Version 1.2.
- Sections 5.4 and 7.3.4: Amended the adult household contact viral acquisition outcomes to be based on collected swabs *through Day*$D14$ for the main analysis. Previously these outcomes had been defined using all collected swabs. Noted additional exploratory analyses that modify the revised outcomes to also include the Day 28 swabs, as before.
- Section 5.4. Changed "will" to "may" in the following excerpt to make the outcome optional: "Another person-level outcome ~~will~~ may count the number of study-administered positive swabs from the person and the number of study-administered swabs with a result . . ."
- Section 5.4: Added a missing closing parenthesis to the definition of the EQ-5D-5L VAS.

Batch of changes for Version 1.3

- Cover page: Added approval date for SAP Version 1.2 and creation date for SAP Version 1.3. Updated the current protocol version to 1.6 and added its effective date. Noted use of some DCC language and contributions from Cody Olsen.
- Section 2.1.3: Instead of stating the protocol will be amended, noted that the protocol has been amended to include persistence of viral shedding at Day 28 as a secondary outcome.
- Edited the wording and definition of the primary outcome to better align with the protocol. Changes are noted with strikethrough (deleted text) and italics (new text) in the following:
  - Section 2.1.2: "The primary outcome is duration of viral shedding, as defined by days from randomization to the first of two consecutive negative swabs, measured ~~on days 1–14~~ *through 14 days after randomization* (formally defined in Section 5.4)."

  - Section 5.4: "Duration of viral shedding (~~until *DayD*14~~ *through 14 days after randomization, i.e., DayR* + 14): this is an integer outcome which is defined as (Date of first sustained negative swab *(ignoring any swab that was not in the interval DayR* + 1 *through DayR* + 14*, inclusive)* - *DayR*). The date of first sustained negative swab is the first day with a negative swab followed by the next available day's swab being negative. The requirement for confirmation will be waived for patients who are negative for shedding on ~~*DayD*14~~ *DayR* + 14. If no more positive swabs nor negative swabs are available ~~by *DayD*14~~ *by DayR* + 14 following a negative swab, the date of the negative swab will be considered confirmed for the primary analysis (using the last-observation-carried-forward approach). If the final swab from on or before ~~*DayD*14~~ *DayR*+14 which has a positive or negative result is positive and there were not two preceding consecutive negative swabs, the duration will be considered to be right censored on the day of the positive swab."
  - Section 7.2: "Due to the nature of the data collected over the *DayR*–(*DayR*+14) ~~*DayD*14~~ interval, and uncertainty about the distribution of shedding duration and the resulting hazard function, our primary analysis will apply the stratified log-rank test to compare the distribution of shedding duration between the hydroxychloroquine and control groups, with administrative right censoring at ~~*DayD*14~~ *DayR* + 14."
  - Section 7.2.1: "For example, in the unexpected scenario where the cumulative incidence curves for the ending of shedding actually cross during the ~~*DayR–DayD*14~~ *(DayR* + 1*) to (DayR* + 14*)* evaluation period".

- Section 5.4: Edited the wording and definition of a secondary outcome to better align with the new definition of the primary outcome: "The protocol originally specified a secondary outcome of duration of viral shedding for days 1–14 and 28. Because the primary outcome is viral shedding ~~through Day 14~~ *from DayR*+1 *through DayR*+14, the secondary outcome of persistence of viral shedding on Day 28 is substituted for duration through Day 28; there is limited precision between Days 15 and 28. The definition of the persistence of viral shedding on Day 28 outcome relies heavily on the Day 28 swab and is limited to those who have a test result for the Day 28 swab or who are not known to have died on or before *DayR* + 30 (there is a ±2 day window for collecting the Day 28 swab). If the Day 28 swab result is known, the result will be used to define the outcome ("yes" if positive, "no" if negative). Otherwise, subjects with a confirmed cessation of viral shedding ~~by Day 14~~ *when considering daily swab results from DayR*+1 *to DayR*+25*, inclusive (with the requirement for confirmation waived if the latest available daily result in this interval is negative),* and a missing Day 28 value will be assumed to be negative on Day 28; otherwise subjects hospitalized on any of the days *DayR*+26–*DayR*+30 with a missing Day 28 value will be assumed to

be positive on Day 28; otherwise, multiple imputation will be performed as described in Section 7.3.3."

- Edited the wording and definition of another secondary outcome to better align with the new definition of the primary outcome.
  - Section 5.4: "**Adult Household Contact Viral Acquisition** This outcome will be analyzed for households with at least two adults for which no other adult besides the index study subject is positive for COVID-19 at baseline and for which the index study subject is in the ITT population. This binary outcome will be at the household level and will be a "yes" if there is a positive swab by one or more adult household contacts for any of the study-administered swabs ~~through *DayD*14 (i.e., not including Day 28 swabs)~~ *from days* $DayR + 1$ *to* $DayR + 14$*, inclusive*. If there are no positive swabs but at least one negative swab, the outcome will be a "no;" otherwise it will be missing. For some additional analyses, an analogous outcome will be defined at the person level for each non-index adult from a household for which no other adult besides the index study subject is positive for COVID-19 at baseline and for which the index study subject is in the ITT population. Another person-level outcome may count the number of study-administered positive swabs from the person and the number of study-administered swabs with a result (e.g., as an offset in a Poisson regression model), again ~~through *DayD*14~~ *from days* $DayR + 1$ *to* $DayR + 14$*, inclusive*.
  - "As exploratory outcomes, the definitions in the previous paragraph will also include ~~the Day 28 swabs (i.e., all study-administered swabs).~~ *study-administered swabs after* $DayR + 14$ *through* $DayR + 30$*, if available*."
  - Section 7.3.4: "The analyses in this paragraph will be conducted twice, once when using the principal version of the outcomes (i.e., using swabs collected ~~through *DayD*14)~~ *from* $DayR + 1$ *to* $DayR + 14$*, inclusive)* and once when using the additional exploratory versions of the outcomes (i.e., using all collected swabs~~, and therefore also including Day 28 swabs~~ *from* $DayR + 1$ *to the Day 28 swabs, inclusive*)."

- Section 4.2.4: Added a restriction to the ITT population: "The Intention-to-Treat (ITT) population consists of all enrolled subjects *who were randomized before being administered study treatment*" and an accompanying footnote: "*After enrollment had ended, study statisticians became aware that due to a participant mix up, an unenrolled person was administered study treatment meant for somebody else, and that after this was initially discovered by study staff, the person was consented to the study (even though ineligible) and randomized to a treatment arm. However, because the treatment had already been administered, this person is not considered to be in the ITT population because there did not seem to have been an intention to follow-through on a post*

hoc randomization. This individual is included in the safety population for the arm corresponding to the administered treatment."

- Section 4.2.7: Added a clarification to the SAFETY population: "The safety population (SAFETY) consists of all ITT subjects who are administered at least one pill/tablet of the delivered study drug (HCQ or placebo). If the delivered study drug differs from the randomly assigned study drug, analysis of the safety population by arm will be according to drug delivered rather than drug assigned. *In addition, a post-hoc decision was made to include one person who was erroneously administered study treatment before consenting/randomization due to a participant mix up, so although this individual will be omitted from the study's ITT-based analysis for the primary manuscript, the person will still be considered to be in the safety population for the treatment arm that corresponds to the delivered study treatment.*"
- Section 5.4: Resolved an inconsistency in the truncation day for an exploratory outcome. "Post-randomization hospitalizations on or before day of study drug delivery~~, up to 14 days post-randomization~~ *(truncated at* $DayR + 13$*)*: this outcome will be assessed for the entire ITT population. This is 'yes' if a hospitalization (or death) is documented to have occurred after the time of randomization and *on or*before the earlier of the day of drug delivery ($DayD1$) or ~~*(*$DayR+14$*)*~~ *(*$DayR + 13$*)*. This is 'no' otherwise, assuming an absence of documentation of hospitalization/death in the medical chart is 'no.'"
- In keeping with the flexibility of DSMB meeting frequency, made the following edit to Section 6.1: "DSMB meetings to review safety data are ~~scheduled for after approximately 100, 200, 300, and all patients have enrolled~~ *as desired by the DSMB, as described in the DSMB charter*."
- Protocol version 1.6 ~~expanded~~ changed the windows for surveys at Day 28 and at 6 months. Induced changes to the SAP use the union of the windows and SAP edits are noted:
  - Section 5.4: "Note that while 28-day symptom and quality of life assessments will be considered to be on $DayA28$, there is an allowed ~~±2 day~~ -2 *day*/$+2$ *week* window for when they occur. Likewise, the 6-month *symptom and* quality of life assessments *for analysis purposes will use* ~~have~~ a ~~±2 week~~ -2 *week*/$+2$ *month* window... **Because swabs are intended to be delivered along with study drug, the drug delivery date will be assumed to correspond to swab delivery date, with the first swab to be collected on** $DayD1$**. There is likewise a window permitted for the Day 28 swab, *although it is narrower:*** $DayR + 26$ ***to*** $DayR + 30$***, inclusive.***
  - "Participants may complete surveys late (e.g., slow response by email) or dates may be incorrectly entered that suggest assessments were done early. Additionally,

surveys may be completed one day late by phone. It will be assumed that answers to any survey submitted 1 day before or after the intended day for the survey will be treated as the intended day. For example, if the day 4 and day 5 assessments are each completed by email four days after $DayD1$, it will be assumed that the day 4 responses pertain to $DayD4$ and the day 5 responses pertain to $DayD5$. However, surveys submitted more than one day before or after the intended day will not be analyzed (except for baseline surveys, which have a longer window because they may be done any day between randomization and the study treatment delivery, inclusive, ~~and~~ day 28 surveys, which have a ~~± 2 *day*~~ -2 *days*/$+2$ *weeks* window*, and 6 month surveys, with a window of* -2 *weeks*/$+2$ *months*). Swabs will be analyzed based on the collection date reported by the participant, which may not necessarily match the day on the swab label. If multiple swabs from a participant have the same participant-reported collection date but different results, a positive result will supersede a negative result, and a negative result will supersede an invalid or indeterminate result. If collected swab dates are clearly erroneous (e.g., before day of randomization), the associated swabs will be ignored. If ~~the "day 28"~~ *a* swab was collected within ± 2 days of ~~day 28~~ $DayR + 28$, ~~or on a day that coincides with any of days 1-14,~~ it will be ~~used~~ *regarded as a Day 28 swab*~~;~~ ~~otherwise, it will be treated as missing~~. ~~It will also be assumed swab specimens were collected on the intended day.~~" (The final sentence was deleted because it is confusing given the preceding language–it is assumed swabs were collected on the reported day, which is not necessarily the intended day.)

- Section 7.3.4: Deleted the redundancy in "~~rate of~~ household viral acquisition rates".
- Section 7.5.2: Corrected subject-verb disagreement to state "Formal statistical *analysis* ~~analyses~~ of adverse events is not planned".
- References: Corrected two instances of an author surname from "Lwe" to "Löwe".

Batch of changes for Version 1.4. Other than one inconsequential update on 23 Feb 2022 explicitly noted below, decisions below drafted and preliminarily approved by subset of statistical team 15–16 Feb 2022, approved by statistical team (BB, CO, RH, TG) on or before 23 Feb 2022, and then proposed to study PIs on 23 Feb 2022.

- Cover page: Added approval date for SAP Version 1.3 and creation date for SAP Version 1.4.
- Edited the handling of swabs on a hospitalized day, which affects the primary outcome. Background: The protocol had assumed swabs would not be collected during hospitalization and prescribed last value carried forward (LVCF), explicitly noting that positive swab results would be assumed during hospitalization if the last pre-hospitalization swab were positive. The SAP had noted swabs may not be collected during hospitalization and also specified LVCF imputation. Discussion with the statistical team

after the primary manuscript had been drafted revealed differences of interpretation among the statistical team as to the intent of the protocol (and SAP) language, but with the expectation that results would be qualitatively similar with various reasonable interpretations because of the small percentage of hospitalized participants and given the swab data available from these participants. The statistical team decided after extensive debate that all hospitalized days would be considered as having a positive swab, even if there were no swab or a negative swab on such a day, and considered this consistent with the original intent of the protocol and the convention already in the SAP that on days with mixed swab results, a positive swab would supersede a negative swab. Thus, a hospitalization's assumed positive swab results would supersede any obtained swabs those days. The edits to the SAP are tracked below.

  - Section 5.4: ~~Special handling will be required in the primary analysis of shedding duration for periods in which patients are hospitalized because swabs might not be obtained during hospitalizations. In the primary analysis, the last-value-carried-forward imputation approach will be applied during hospitalization, such that continued viral shedding during hospitalization will be assumed for patients whose last pre-hospitalization assessment is positive. A sensitivity analysis is described in Section 7.2.1.~~ *The primary outcome was viral-shedding duration over the first two weeks. Because the protocol stipulated that swabs would not be obtained during hospitalizations, and that positive results would be carried forward through hospitalization, swab status was considered positive for all hospital days. Hospitalizations or swabs documented after day 15 (i.e., DayR + 14) were not used for this outcome. Duration was considered to end on the first day of two consecutive negative swabs, allowing days without swab results between two negative swabs, or on the last available swab result if negative. Without negative confirmation, the outcome was considered right-censored on the day of their last positive result or day 15. In a sensitivity analysis, we treated hospitalized patients prior to day 15 without a prior confirmed negative as right-censored on the day of hospital admission or day 15.*
  - Section 7.2: The primary analysis of shedding duration will *consider every hospitalized day as having a positive swab result.* ~~apply a last-observation-carried-forward imputation approach for missing swabs that would have been obtained during hospitalization.~~ See section 5.4. [23 Feb 2022 note: The sentence "See section 5.4." had been preserved in this description of changes, but removed from Section 7.2 on 15 Feb 2022. This inconsequential discrepancy was unilaterally resolved by BB on 23 Feb 2022, after the statistical team had approved of the other decisions but before sending to the PIs, by adding this sentence back to Section 7.2.]

- Section 5.4: Added the following paragraph: "*Also of note is that we assumed all*

participants were positive on the day of randomization, even if a negative swab were available. It was uncommon for a swab to be available on the day of randomization." The handling of day-of-randomization swabs had not been particularly emphasized in the SAP, so this represents the decision that was implemented when obtaining the already-viewed results. Study swabs were generally unavailable from the day of randomization, and a study eligibility criterion required a sufficiently recent positive test for SARS-CoV-2. It seemed sensible to start all participants as being positive at the time of randomization.

Contents

Abbreviations

Abbreviation	Definition
CMH	Cochran-Mantel-Haenszel
CONSENTED	Consented Population
CRF	Case Report Form
$DayA1$	Study Assessment Day 1
$DayD1$	Day of study drug delivery
$DayR$	Day of randomization
DCC	Data Coordinating Center
dITT	Drug-delivered Intention-to-Treat
DSMB	Data and Safety Monitoring Board
ELIGIBLE	Eligible Population
EQ-5D-5L	5-level EuroQol- 5 Dimension
GAD-7	Generalized Anxiety Disorder 7-item
HCQ	hydroxychloroquine
HCQ Trial	Hydroxychloroquine for Outpatients with Confirmed COVID-19 Trial (original name of trial)
ITT	Intention-To-Treat
LVCF	Last Value Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
PCL-5	Post-traumatic Stress Disorder Checklist for DSM-5
PHQ-9	Patient Health Questionnaire 9
PP	Per-Protocol
(S)AE	(Serious) Adverse Event
SAFETY	Safety Population
SAP	Statistical Analysis Plan
SCREEN	Screening Population
UTAH One	Understanding Treatment And Health in the Ongoing coroNa Epidemic: A Hydroxychloroquine Outpatient Study
VAS	Visual Analog Scale (of EQ-5D-5L)

1 PREFACE

1.1 Purpose of SAP

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for the protocol: UTAH One (Understanding Treatment And Health in the Ongoing coroNa Epidemic): A Hydroxychloroquine Outpatient Study; formerly *Hydroxychloroquine for Outpatients with Confirmed COVID-19 (HCQ Trial)*.

The structure and content of this SAP provides sufficient detail to meet the requirements and standards set by the Data Coordinating Center (DCC).

1.2 Auxiliary/Other Documents

The following documents were reviewed in preparation of this SAP:

- Protocol: UTAH One (Understanding Treatment And Health in the Ongoing coroNa Epidemic): A Hydroxychloroquine Outpatient Study.
- Case Report Forms (CRFs) for the UTAH One protocol

The reader of this SAP is encouraged to read the protocol for details on the conduct of this study, and the operational aspects of clinical assessments.

The purpose of this SAP is to outline the planned analyses to be completed for the UTAH One trial. The planned analyses identified in this SAP will be included in future study abstracts and manuscripts. Also, exploratory analyses not necessarily identified in this SAP may be performed. Any post hoc, or unplanned, analyses not explicitly identified in this SAP will be clearly identified as such in any published reports from this study.

It is possible that, due to updates or identification of errors in specific statistical software discussed in this SAP, the exact technical specifications for carrying out a given analysis may be modified. This is considered acceptable as long as the original, prespecified statistical analytic approach is completely followed in the revised technical specifications.

2 STUDY OBJECTIVES AND OUTCOMES

2.1 Study Objectives

2.1.1 Primary Objective

The primary objective of the UTAH One is to assess the efficacy and safety of hydroxychloroquine (HCQ) for reduction of viral shedding and hospitalization in outpatients with confirmed COVID-19.

2.1.2 Primary Outcome(s)

The primary outcome is duration of viral shedding, as defined by days from randomization to the first of two consecutive negative swabs, measured through 14 days after randomization (formally defined in Section 5.4).

2.1.3 Secondary Outcome(s)

Secondary outcomes are:

- Average level of COVID-19-attributable symptoms
- Hospitalization within 14 days of enrollment
- Persistence of viral shedding on day 28
- Adult household contact viral acquisition

The secondary outcomes according to the original protocol included duration of COVID-19-attributable symptoms rather than average level of COVID-19-attributable symptoms. Reflecting the study PIs' desires, it was decided by May 21, 2020 to change this duration outcome to an exploratory outcome and include the average level as a new secondary outcome.

Another secondary outcome in the original protocol was duration of viral shedding, measured on days 1-14 and 28. Because of how similar this is to the primary outcome, this secondary outcome has been changed in the SAP(and since amended in the protocol) to persistence of viral shedding on day 28.

2.1.4 Exploratory Outcome(s)

Exploratory outcomes include, but are not limited to, the following:

- Duration of COVID-19 Symptoms
- Patient-reported instrument measures at baseline, 28 days, and 6 months

  - Patient Health Questionnaire 9 (PHQ-9) Total
  - Generalized Anxiety Disorder 7-item (GAD 7) Total
  - Post-traumatic Stress Disorder Checklist for DSM-5 (PCL-5) Total
  - 5-level EuroQol- 5 Dimension (EQ-5D-5L) Visual Analog Scale (VAS) and index measure

- Viral shedding numerical measurements
- Viral shedding at 28 days
- Hospital admissions within 14 days of randomization but prior to day of study drug delivery
- Cumulative mortality at 6 months
- Hospital admission by 6 months

2.1.5 Safety Outcome(s)

In addition to the following list, it should be noted that some of the secondary and exploratory outcomes also pertain to safety but are not necessarily restated here.

- Adverse event(s) by 14 days
- Mortality by 28 days
- Hospital admission by 28 days

3 STUDY DESIGN AND METHODS

3.1 Overall Study Design

The UTAH One Trial is a prospective, placebo-controlled, randomized controlled trial to evaluate the efficacy and safety of hydroxychloroquine for subjects with COVID-19. Individuals are randomized to either the hydroxychloroquine (HCQ) arm or placebo arm. Doses are scheduled for five days. Follow-up is most frequent through Day 15 but also is to occur at Day 28 and at 6 months.

3.2 Method of Treatment Assignment and Randomization

Randomization is stratified by age group and balanced between the two arms using random permuted blocks of size 2 or 4 (block size varies randomly after the initial block). A DCC faculty statistician has generated the sequence using R version 3.6.0 statistical software. The treatment assignments for study subjects are made available via a REDCap database.

3.2.1 Delivery of Randomization and Emergency Backup

Treatment assignment will be provided via a REDCap database. There are no emergency backup plans.

3.2.2 Handling of Incorrect Randomization in Study Analyses and Reports

It is possible that a patient may be randomized into the incorrect age stratum, due to incorrect birthdate information or improper specification of the age group. Assuming that any such occurrences are very occasional, legitimate errors, any such patients will be analyzed in their corrected stratum (i.e., the stratum defined by their true age) rather than in the incorrectly initially assigned stratum in all analyses in all reports. Restated, corrected stratum rather than initially incorrectly assigned randomization stratum is to be used for analyses in this document.

3.3 Treatment Masking (Blinding)

Because of the expedited timeline for implementing this study, it is infeasible to prepare a placebo in such a way that it is indistinguishable from the HCQ doses. However, study staff are not to inform the subject which treatment arm has been assigned to the subject to make this as nearly single-blinded as possible. For the same reason, it is not practical to implement blinding of research coordinators because they may be involved in preparing the kit (containing, among other study materials, study drug) that is delivered to study participants.

3.4 Study Intervention Compliance

Beginning the day after study medication is delivered, subjects are asked to indicate on five days whether or not study medication was administered the previous day and how many pills remain. The number of pills administered is thereby ascertained.

4 STUDY SUBJECTS AND ANALYSIS POPULATIONS

4.1 Eligibility

The inclusion/exclusion criteria for this study are described in the protocol.

4.2 Populations

4.2.1 Screening Population

The screening population (SCREEN) includes all subjects meeting all inclusion criteria who were entered on site screening logs provided to the DCC.

4.2.2 Eligible Population

The eligible population (ELIGIBLE) includes subjects meeting all inclusion criteria and none of the exclusion criteria.

4.2.3 Consented Population

The consented population (CONSENTED) includes all eligible subjects who consent to study participation.

4.2.4 Intention-to-Treat Population

The Intention-to-Treat (ITT) population consists of all enrolled subjects who were randomized before being administered study treatment[1]. Subjects are not considered enrolled until they are randomized to a treatment arm, so the day of randomization and day of enrollment are equivalent. Subjects are considered to be in the treatment arm to which they were randomized, regardless of whether or not they received any doses for the assigned arm. This will be the primary population for efficacy analyses.

4.2.5 Drug-delivered Intention-To-Treat

The drug-delivered Intention-To-Treat (dITT) population includes all ITT subjects who had the assigned study drug, HCQ or placebo, delivered to them. Efficacy analyses may be repeated on this subpopulation as sensitivity analyses.

[1] After enrollment had ended, study statisticians became aware that due to a participant mix up, an unenrolled person was administered study treatment meant for somebody else, and that after this was initially discovered by study staff, the person was consented to the study (even though ineligible) and randomized to a treatment arm. However, because the treatment had already been administered, this person is not considered to be in the ITT population because there did not seem to have been an intention to follow-through on a post hoc randomization. This individual is included in the safety population for the arm corresponding to the administered treatment.

4.2.6 Per-Protocol Efficacy Population

The Per-Protocol (PP) efficacy population includes all subjects in the dITT population who receive at least 80% of the assigned treatment's pills/tablets. Efficacy analyses will also be repeated on this subpopulation as sensitivity analyses.

4.2.7 Safety Populations

The safety population (SAFETY) consists of all ITT subjects who are administered at least one pill/tablet of the delivered study drug (HCQ or placebo). If the delivered study drug differs from the randomly assigned study drug, analysis of the safety population by arm will be according to drug delivered rather than drug assigned. In addition, a post-hoc decision was made to include one person who was erroneously administered study treatment before consenting/randomization due to a participant mix up, so although this individual will be omitted from the study's ITT-based analysis for the primary manuscript, the person will still be considered to be in the safety population for the treatment arm that corresponds to the delivered study treatment. The principal safety analyses for outcomes will be on the ITT population, while the principal analyses for adverse events will be on the SAFETY population. Adverse event summaries will also be prepared for the ITT population or as requested by the DSMB.

5 GENERAL ISSUES FOR STATISTICAL ANALYSES

5.1 Analysis Software

Analysis will be performed using SAS® Software version 9.4 or later whenever possible. Other software packages, including R and StatXact®, may be used for particular specialized procedures. Because of its specialized options for sequential regression imputation, IVEware is preferred for generating multiple imputations.

5.2 Methods for Withdrawals, Missing Data, and Outliers

Subjects with incomplete data for a variable needed in an analysis may, for example, be omitted from that analysis or have the last observation carried forward. Which technique is used depends on the analysis, as described later in the SAP.

Rates of missingness in the primary outcome will be compared between treatment arms for the ITT, dITT, and PP populations. If a substantial number of subjects have incomplete data for this outcome, analyses will be conducted to determine factors significantly related

to missingness. The factors assessed would include, at a minimum, assigned treatment arm and true age stratum.

When multiple imputation is employed, it will be performed separately within each treatment arm if possible (by ITT arm for analyses of the ITT population or subpopulations thereof, or, if necessary, by SAFETY arm for analyses of the SAFETY population).

5.3 Multiple Comparisons and Multiplicity

Explicit adjustments for multiple testing will not be applied. The secondary outcome of viral shedding from 1–14, 28 days will be regarded as exploratory unless the primary analysis of the primary outcome is significant.

5.4 Derived and Computed Variables

Definitions of Days For each subject, the day of randomization to treatment arm will be referred to as $DayR$. The day on which study drug is delivered is $DayD1$, with the next day being $DayD2$ and so forth. The day 1 daily study assessment is referred to as occurring on $DayA1$; the next calendar day is $DayA2$, and so forth through $DayA15 = DayA1+14$. Note that if the baseline study assessment were to occur on the same calendar day as $DayA1$, the form name in the database will be used to distinguish in such instances which values are used for baseline values and which are used for $DayA1$ values so that both sets of measurements can still be utilized. It is expected that for most subjects study drug will be delivered one day after randomization, i.e., that $DayD1 = DayR+1$. It is also expected that for most subjects $DayD1 = DayA1$. However, because these relationships are not guaranteed, separate nomenclature is sometimes used to define outcomes (typically based on intervals beginning at $DayR$) and study assessments (often based on days since drug delivery or assessment day). Note that while 28-day symptom and quality of life assessments will be considered to be on $DayA28$, there is an allowed -2 day/$+2$ week window for when they occur. Likewise, the 6-month symptom and quality of life assessments for analysis purposes will use a -2 week/$+2$ month window. If a randomized subject never provides a daily assessment, such that the anchoring event for defining $DayA1$ is unavailable, the convention will be taken to assume $DayA1 = DayR + 1$. Likewise, if study drug delivery date is not available, then it will be assumed that $DayD1 = DayR+1$. **Because swabs are intended to be delivered along with study drug, the drug delivery date will be assumed to correspond to swab delivery date, with the first swab to be collected on** $DayD1$**. There is likewise a window permitted for the Day 28 swab, although it is narrower:** $DayR+26$ **to** $DayR + 30$**, inclusive.**

Participants may complete surveys late (e.g., slow response by email) or dates may be incorrectly entered that suggest assessments were done early. Additionally, surveys may be completed one day late by phone. It will be assumed that answers to any survey submitted 1 day before or after the intended day for the survey will be treated as the intended day. For example, if the day 4 and day 5 assessments are each completed by email four days after $DayD1$, it will be assumed that the day 4 responses pertain to $DayD4$ and the day 5 responses pertain to $DayD5$. However, surveys submitted more than one day before or after the intended day will not be analyzed (except for baseline surveys, which have a longer window because they may be done any day between randomization and the study treatment delivery, inclusive, day 28 surveys, which have a -2 days/$+2$ weeks window, and 6 month surveys, with a window of -2 weeks/$+2$ months). Swabs will be analyzed based on the collection date reported by the participant, which may not necessarily match the day on the swab label. If multiple swabs from a participant have the same participant-reported collection date but different results, a positive result will supersede a negative result, and a negative result will supersede an invalid or indeterminate result. If collected swab dates are clearly erroneous (e.g., before day of randomization), the associated swabs will be ignored. If a swab was collected within ± 2 days of $DayR + 28$, it will be regarded as a Day 28 swab.

Primary Outcome Duration of viral shedding (through 14 days after randomization, i.e., $DayR + 14$): this is an integer outcome which is defined as (Date of first sustained negative swab (ignoring any swab that was not in the interval $DayR+1$ through $DayR+14$, inclusive) - $DayR$). The date of first sustained negative swab is the first day with a negative swab followed by the next available day's swab being negative. The requirement for confirmation will be waived for patients who are negative for shedding on $DayR + 14$. If no more positive swabs nor negative swabs are available *by* $DayR + 14$ following a negative swab, the date of the negative swab will be considered confirmed for the primary analysis (using the last-observation-carried-forward approach). If the final swab from on or before $DayR + 14$ which has a positive or negative result is positive and there were not two preceding consecutive negative swabs, the duration will be considered to be right censored on the day of the positive swab.

The primary outcome was viral-shedding duration over the first two weeks. Because the protocol stipulated that swabs would not be obtained during hospitalizations, and that positive results would be carried forward through hospitalization, swab status was considered positive for all hospital days. Hospitalizations or swabs documented after day 15 (i.e., $DayR + 14$) were not used for this outcome. Duration was considered to end on the first day of two consecutive negative swabs, allowing days without swab results between two negative swabs, or on the last available swab result if negative. Without negative confirmation, the outcome was considered right-censored on the day of their last positive result or day 15. In a sensitivity analysis, we treated hospitalized patients prior to day 15 without a prior confirmed

negative as right-censored on the day of hospital admission or day 15.

Also of note is that we assumed all participants were positive on the day of randomization, even if a negative swab were available. It was uncommon for a swab to be available on the day of randomization.

Hospitalization within 14 days of enrollment Hospitalization within 14 days: this is a binary outcome which is "yes" if (a) medical chart review (e.g., performed on $DayD28$) reveals a hospitalization from $DayR$ to $DayR$+13 (inclusive) or (b) a study assessment reports a hospitalization occurring in this span or (c) the subject is known to have died during this span. If a subjects meets none of these criteria, the outcome will assumed to be "no". If a subject meets none of these criteria but was lost to follow-up before $DayR$+13 and chart review is not permitted because of study withdrawal, the outcome will be considered missing.

Persistence of viral shedding on Day 28 The protocol originally specified a secondary outcome of duration of viral shedding for days 1–14 and 28. Because the primary outcome is viral shedding from $DayR + 1$ through $DayR + 14$, the secondary outcome of persistence of viral shedding on Day 28 is substituted for duration through Day 28; there is limited precision between Days 15 and 28. The definition of the persistence of viral shedding on Day 28 outcome relies heavily on the Day 28 swab and is limited to those who have a test result for the Day 28 swab or who are not known to have died on or before $DayR + 30$ (there is a ± 2 day window for collecting the Day 28 swab). If the Day 28 swab result is known, the result will be used to define the outcome ("yes" if positive, "no" if negative). Otherwise, subjects with a confirmed cessation of viral shedding when considering daily swab results from $DayR + 1$ to $DayR + 25$, inclusive (with the requirement for confirmation waived if the latest available daily result in this interval is negative), and a missing Day 28 value will be assumed to be negative on Day 28; otherwise subjects hospitalized on any of the days $DayR + 26$–$DayR + 30$ with a missing Day 28 value will be assumed to be positive on Day 28; otherwise, multiple imputation will be performed as described in Section 7.3.3.

Average of Core COVID-19 symptoms There are seven core symptoms: fever, chills, cough, tiredness, shortness of breath, diarrhea, and muscle aches. Symptoms are measured on a 6-point ordinal scale (not experiencing=0, extremely mild=1, mild=2, moderate=3, severe=4, extremely severe=5) at baseline, days 1-15, and day 28.If at most one of the core symptoms is missing for an assessment, the mean of the non-missing core symptoms' levels for that assessment will be calculated to obtain a symptom burden score. If two or more core symptoms are missing from the assessment, the average symptom level for that assessment will be considered missing.

Adult Household Contact Viral Acquisition This outcome will be analyzed for households with at least two adults for which no other adult besides the index study subject is positive for COVID-19 at baseline and for which the index study subject is in the ITT population. This binary outcome will be at the household level and will be a "yes" if there is a positive swab by one or more adult household contacts for any of the study-administered swabs from days $DayR + 1$ to $DayR + 14$, inclusive. If there are no positive swabs but at least one negative swab, the outcome will be a "no;" otherwise it will be missing. For some additional analyses, an analogous outcome will be defined at the person level for each non-index adult from a household for which no other adult besides the index study subject is positive for COVID-19 at baseline and for which the index study subject is in the ITT population. Another person-level outcome may count the number of study-administered positive swabs from the person and the number of study-administered swabs with a result (e.g., as an offset in a Poisson regression model), again from days $DayR$+1 to $DayR$+14, inclusive.

As exploratory outcomes, the definitions in the previous paragraph will also include study-administered swabs after $DayR + 14$ through $DayR + 30$, if available).

Exploratory/Safety Outcomes Duration of COVID-19 symptoms (through $DayA15$): this is an integer-valued outcome which is defined as (Date of first sustained asymptomatic day - Start date). To determine the end date, first each symptom assessment from baseline through $DayA15$, inclusive, will be classified as symptomatic, asymptomatic, or unknown. A symptomatic day is one in which at least one of the core symptoms (and the average of the non-baseline symptom levels for that symptom that day, if multiple non-baseline, non-missing symptom levels were provided for that symptom that day) is observed to exceed the permissible threshold: not experiencing for fever and chills; extremely mild for shortness of breath, diarrhea, and muscle aches; mild for cough and tiredness. If at most one symptom level is missing on a given day and all observed core symptoms' levels are at or below the permissible threshold, the day will be classified as asymptomatic. Otherwise (i.e., if at most five of the core symptoms have a reported symptom level and none of the reported symptom levels exceeded the threshold), the day will be considered unknown.

The analysis of this outcome is limited to participants whose status on the day of the baseline assessment is symptomatic, per the definition above. The start date is the date of the baseline assessment. The end date (i.e., date of the first sustained asymptomatic day) is the first asymptomatic day which is followed by the next known day being asymptomatic. The requirement for confirmation will be waived for subjects whose last available status as of $DayA15$ was asymptomatic, provided they were not subsequently hospitalized or deceased on or before $DayA15$; the date of the asymptomatic status will be considered confirmed. If the final day with known status from baseline through $DayA15$ is symptomatic and there were not two preceding consecutive asymptomatic days, the duration will be considered to

be right censored on the final symptomatic day (up through and including $DayA15$).

If a subject dies on or before $DayA15$ before a confirmed asymptomatic day, the end date will be censored on $DayA15$. If a subject is hospitalized on or before $DayA15$ before a confirmed asymptomatic day, the subject will be regarded as symptomatic through the date of discharge or $DayA15$, whichever comes first, which may push the date of censoring back.

Although the protocol mentioned analysis of the symptom duration through Day 28, daily symptom level assessment ceases on $DayA15$. Therefore, the $DayA28$ symptom assessments will be ignored when determining symptom duration.

- EQ-5D-5L measures (administered at baseline, 28 days, and 6 months)[1, 2]
  - Visual Analog Scale: 0 to 100 scale. If the value is missing because the individual is known to be deceased when the assessment was to have occurred, the VAS change since baseline will be post-processed to -101 (i.e., worse than possible for any survivor) after multiple imputation (see Section 7.4.5).
  - Index measure: The index measure will use coefficients from the cTTO Model (Model 1 in Table 2 of [2]) to provide a univariable summary of the five component dimensions. If the value is missing because the individual is known to be deceased when the assessment was to have occurred, the index measure will be set to 0. Otherwise, if any of the five dimensions are missing, the index measure will be regarded as missing.
- PCL-5 Total (baseline, 28 days, and 6 months): The 20 items are scored as 0–4 each, producing a total from 0–80.[3] The cited reference does not provide specific guidance on what percentage of missing values may be imputed with the mean of the corresponding non-missing values. In the absence of such a recommendation, if no more than 4 (i.e., 20%) of the relevant items are missing for an individual's total, then the average of the individual's observed item scores from the total will be substituted for the missing items to impute the total. If more than 4 items are missing, the total will be regarded as missing. In addition, subtotals of potential interest may be considered: cluster B (items 1–5), cluster C (6–7), cluster D (8–14), and cluster E (15–20); see [3]. Up to 20% of missing values within a cluster would be imputed with the average of the observed scores in the cluster for determining these subtotals. If the value is missing because the individual is known to be deceased when the assessment was to have occurred, the PCL-5 change since baseline will be post-processed as 81 (i.e., worse than possible for any survivor) after multiple imputation; similar adjustments will be made for cluster subtotals.
- GAD 7 Total (baseline, 28 days, and 6 months): The 7 items are scored as 0–3 each, producing a total from 0–21.[4] The cited reference does not provide specific guidance

on what percentage of missing values may be imputed with the mean of the corresponding non-missing values. In the absence of such a recommendation from this source, if no more than one of the relevant items are missing for an individual's total, then the average of the individual's observed item scores from the total will be substituted for the missing items to impute the total. If more than one item is missing, the total will be regarded as missing. If the value is missing because the individual is known to be deceased when the assessment was to have occurred, the GAD-7 change since baseline will be post-processed to 22 (i.e., worse than possible for any survivor) after multiple imputation.

- PHQ-9 Total (baseline, 28 days, and 6 months): The 9 items are scored as 0-3 each, producing a total from 0–27.[5] Several scoring approaches are noted, such as treating missing item scores as 0 to be "diagnostically conservative" or mean-imputing up to 20% of item scores.[5] Another validation article allowed 2 of the 9 items to be mean-imputed.[6] The latter approach will be used such that up to 2 items may be mean-imputed. If the value is missing because the individual is known to be deceased when the assessment was to have occurred, the PHQ-9 change in total since baseline will be post-processed to 28 (i.e., worse than possible for any survivor) after multiple imputation.

- Hospital admission by 6 months: this is a binary outcome which is "yes" if (a) the secondary 14-day hospitalization outcome is yes, or (b) any hospitalization, for any cause, is found in either the 28-day or 6-month chart review or from subject self-report, or (c) the subject is known to have died by the time of the 6-month chart review. This is "no" otherwise, assuming an absence of documentation of hospitalization in the medical chart is "no."

- Mortality by 14 days: all-cause mortality at $DayR$+13

- Mortality by 28 days: all-cause mortality at $DayR$+27

- Hospital admission by 28 days: this is a binary outcome which is "yes" if (a) the subject self-reported a hospitalization beginning between $DayR$ and $DayR$+27 inclusive, or (b) 28-day chart review discovered a hospitalization in this interval, or (c) the subject is known to have died during this interval.

- Post-randomization hospitalizations on or before day of study drug delivery (truncated at $DayR + 13$): this outcome will be assessed for the entire ITT population. This is "yes" if a hospitalization (or death) is documented to have occurred after the time of randomization and *on or* before the earlier of the day of drug delivery ($DayD1$) or ($DayR + 13$). This is "no" otherwise, assuming an absence of documentation of hospitalization/death in the medical chart is "no."

Subgroup Variables Age group will be determined based on the individual's age at randomization. The age groups will be 18–44, 45–59, 60–74, 75+.

Presence/absence of comorbidity The analysis and/or imputation models of some outcomes utilizes a dichotomous comorbidity variable. Presence/absence of comorbidities will be defined as "Yes" if any of the following are indicated on the medical history form (and "No" otherwise, including if status is not available):

- Diabetes mellitus
- Hypertension
- Chronic pulmonary disease
- Immunocompromised status
- Cardiovascular disease
- Congestive heart failure
- Chronic kidney disease
- Active malignancy
- Cerebrovascular disease
- Chronic liver disease
- Morbid obesity
- Chronic neurological disease

5.5 Independent Review

All statistical analyses for primary reporting of trial results will be independently verified through dual programming. Two statisticians will each program all datasets and analyses for the DSMB reports and the primary manuscript(s) and the results will be compared.

6 INTERIM ANALYSES

6.1 Frequency of and Timepoints for Interim Analysis

DSMB meetings to review safety data are as desired by the DSMB, as described in the DSMB charter. No interim efficacy analyses are planned for this trial.

In addition to the indicated interim safety analyses, the DSMB may request additional safety and efficacy analyses at their discretion.

6.2 Stopping Rules for Interim Efficacy Analysis

We do not anticipate applying formal stopping rules for either efficacy or futility in this trial. However, we will include feasibility evaluations to allow us to determine when the first wave of COVID-19 has resolved, such that further enrollment in the trial is unlikely. Given the totality of the evidence and enrollment in the trial to date, a decision may be made to suspend or terminate the trial if there is clear evidence that the first wave is over in Utah.

6.3 Blinding in the Interim Analysis

The closed-session DSMB reports will present generic treatment designations (e.g., "A" and "B") to promote objective decisions regarding study procedures and analyses. Nonetheless, the DSMB retains the prerogative to request precise treatment designations rather than generic labels.

7 PLANNED ANALYSES

7.1 Description of Subject Characteristics

The primary results will include key baseline characteristics for the ITT population, overall and by assigned arm. A partial list of these characteristics includes

- age
- sex
- race
- ethnicity.

The primary results will also include a CONSORT-like diagram describing patient flow. This diagram or its footnotes will depict, at minimum, counts in the SCREEN, ELIGIBLE, CONSENTED, ITT, dITT, SAFETY, and PP populations along with frequencies for categories describing why unretained patients were not part of subsequent subsets.

7.2 Primary Outcome Analysis

Due to the nature of the data collected over the $DayR$–($DayR$+14) interval, and uncertainty about the distribution of shedding duration and the resulting hazard function, our primary analysis will apply the stratified log-rank test to compare the distribution of shedding duration between the hydroxychloroquine and control groups, with administrative right censoring at $DayR+14$. The primary analysis will be conducted using a 2-sided α=0.05. The log-rank test will be stratified by the age group randomization stratum. The Efron approach will be

used to account for tied shedding duration times. Kaplan-Meier curves will summarize the shedding time distributions in the two randomized groups.

The primary analysis of shedding duration will consider every hospitalized day as having a positive swab result. See section 5.4.

7.2.1 Additional Analyses of Primary Outcome

A sensitivity analysis will right-censor follow-up on the day of hospital admission. These two analyses (primary and sensitivity) represent the range of plausible scenarios for shedding during hospitalization under the assumption that any association between hospitalization and shedding must be positive; i.e., conditional on past history, the risk of viral shedding on any given day is at least as large for hospitalized as for non-hospitalized subjects. It is expected that a relatively modest proportion of subjects will be hospitalized, limiting the impact of assumptions concerning shedding during hospitalizations on the primary analysis. Additional sensitivity analyses, including analyses in which the end date of shedding is modeled as a latent variable will be considered.

Quantification of the magnitude of "treatment effect," and indeed the interpretation of the trial results, will depend on the specific patterns of the shedding duration distribution. For example, in the unexpected scenario where the cumulative incidence curves for the ending of shedding actually cross during the $(DayR + 1)$ to $(DayR + 14)$ evaluation period, the treatment arm with lower 14-day shedding rate would likely be preferred regardless of any shorter-term benefit observed in the other arm. Assuming, however, that the assumption of proportional hazards is found to be consistent with the data, the effect of the treatment will be expressed as a hazard ratio with 95% confidence interval. We will also report the randomization age-group-stratified differences between the randomized groups in truncated mean shedding duration to 14 days and in median shedding duration with 95% confidence intervals based on estimates derived from Kaplan-Meier curves. The confidence intervals for each difference will be derived by first sampling with replacement the participant-level data from among the participants with at least some viral-shedding information, then fitting the stratified Kaplan-Meier curves, and finally estimating stratum-specific and overall median shedding durations and truncated mean durations. This will be repeated for a minimum of 500 bootstrap replicates. The 2.5th and 97.5th percentiles of each difference between treatments, overall and by age group randomization stratum, will be used as the confidence interval endpoints. The confidence interval construction is similar to the approach taken in Calkins et al.[7] for overall and stratified confidence intervals of the restricted mean duration, with a key difference being that no weights will be applied because treatment groups were randomly assigned.

As an additional exploratory analysis of the primary outcome, the primary analysis will be repeated when stratifying by duration of pre-randomization symptoms rather than by age group.

7.3 Secondary Outcome(s) Analyses

For each secondary outcome except the adult household contact viral acquisition, the population of interest is the ITT population, with assigned treatment arm used, even if it differs from administered treatment arm. The viral acquisition outcome is further restricted to households with at least two adults for which no other adult besides the index study subject is positive for COVID-19 at baseline.

7.3.1 14-day Hospitalization

The comparison of proportion of patients hospitalized by $DayR{+}13$ between the randomized hydroxychloroquine and placebo groups will be performed with a two-sided Cochran-Mantel-Haenszel (CMH) test stratifying by the age categories as used for randomization, with 2-sided α=0.05. As noted in Section 5.4, in the unlikely event that some patients die within 14 days after randomization without being hospitalized, these patients will be assigned to the hospitalization category in the analysis. Results will be presented as relative risks with 95% confidence intervals within each age stratum, and then pooled across strata. The pooled estimator is the estimator for the principal analysis.

7.3.2 Average of COVID-19 attributable symptoms

As noted in Section 2.1.3 of this SAP, the average symptom level of COVID-19-attributable symptoms was not originally an outcome in the protocol, but the PIs agreed to make this a secondary outcome and change the symptom duration outcome to be an exploratory outcome.

The average symptom level of the core symptoms will be analyzed as an average level for the 15 daily assessments, $DayA1$–$DayA15$. Missingness in outcomes is particularly acute because the univariate outcome is the average of 15 assessments. The following approach, which combines multiple imputation and the last value carried forward (LVCF), will therefore be taken.

1. Restrict the data to individuals who did not die on or before $DayA1$ and who have at least one nonmissing daily average from among the following collection of assessments: Baseline, $DayA1$, $DayA2$, …, $DayA15$.
2. For each participant, determine which of $DayA1$–$DayA15$ meet all of the following conditions: (a) the participant was hospitalized on at least some portion of the day, AND (b) the participant is not known to have died that day, AND (c) the participant's

daily symptom average is missing, as described in Section 5.4. If there are any such days, then

- Beginning at the earliest such day, determine if the immediately preceding period (i.e., baseline if for $DayA1$, $DayA1$ if for $DayA2$, etc.) had a nonmissing value. If so, assign the previous symptom average score to be the missing day's average and move to the next such day to determine if it will propagate a value forward or not. If not, leave the day as missing and proceed to the next such day to determine it if will propagate a value forward or not.

3. Impute any remaining missing values (whether hospitalized, dead, or not), separately for each ITT treatment arm, using chained equations with the following variables: daily core average from the baseline assessment and assessment days 1–15, comorbidity presence, age group, sex, ethnicity, and race. The imputation will be conducted in IVEware and will use an R-squared entry criterion of 0.01 for each variable's imputation; a minimum of 10 imputed data sets will be generated.
4. Post-process values that had been imputed during hospitalization by applying LVCF; the value to carry forward is from the most-recent period that (a) has an observed average score AND was earlier in the same hospitalization, if any, or otherwise (b) was the (possibly imputed) assessment period immediately before hospitalization. If there is no previous assessment because the hospitalization was before the baseline assessment, carry forward the imputed baseline value while the hospitalization is ongoing or through the period immediately before a nonmissing average was observed, whichever occurs first.
5. Calculate the average value for $DayA1$–$DayA15$. If the individual died on or before $DayA15$, average only the values from $DayA1$ to the day before death, inclusive.

The outcome will be modeled with linear regression while adjusting for the baseline average, age group, comorbidity presence, treatment, sex, race, ethnicity, and age group × comorbidity interaction effects. Rubin's rules will be used to make final inference, with interest on the two-sided p-value for the treatment effect, which will be declared statistically significant if < 0.05.

As with the imputation model, the main analysis model for the average symptom level will be restricted to ITT subjects who had at least one observed symptom average from among the baseline or $DayA1$–$DayA15$ measurements.

7.3.3 Persistence of Viral Shedding on Day 28

This analysis will omit subjects known to have died on or before $DayR + 30$ without having a Day 28 swab result. (The Day 28 swab may be in a ± 2 day window, so a participant may

have a "Day 28" swab even if the subject died on $DayD28$.) Some missing data will be assumed positive or negative, as described for this day 28 shedding outcome in Section 5.4. Then, multiple imputation will be employed for subjects with still-missing values for the Day 28 viral shedding outcome. There will be a minimum of $m = 10$ imputed datasets, and variables used to impute the outcome will include variables such as prior hospitalizations while on study, treatment arm (if imputation is not already stratified by treatment arm), age group, and viral shedding history.

Comparison of the proportion of index subjects with Day 28 viral shedding between the randomized hydroxychloroquine and placebo groups will be performed. This will be based on the Cochran-Mantel-Haenszel test statistic when stratifying by the age categories as used for randomization. It is desired to employ a test statistic that is approximately normally distributed under the null hypothesis of no treatment effect so that Rubin's rules may be used to make inference across imputations. Therefore, the signed square root of the CMH test statistic will be calculated for each imputation (negative if and only if the CMH odds ratio for the imputed data set is < 1). The two-sided test will use α=0.05.

7.3.4 Adult Household Contact Viral Acquisition

The principal analysis of adult household contact viral acquisition will be restricted to households with at least two adults for which no other adult besides the index study subject is positive for COVID-19 at baseline. We will evaluate the effect of the hydroxychloroquine intervention on the risk of adult household viral acquisition by applying a Cochran-Mantel-Haenszel test stratifying by the randomization age categories and total number of adults within the household (with this total number being regardless of whether or not the adults contributed to the study's swab collection). Results will be presented as relative risks with 95% confidence intervals within each stratum, and then pooled across strata. A secondary analysis will assess the effect of treatment arm on household viral acquisition rates averaged across all non-index adults within the household by using a modified Poisson regression model incorporating correlation between participants in the same household, and controlling for age category and number of household members. The analyses in this paragraph will be conducted twice, once when using the principal version of the outcomes (i.e., using swabs collected from $DayR + 1$ to $DayR + 14$, inclusive) and once when using the additional exploratory versions of the outcomes (i.e., using all collected swabsfrom $DayR + 1$ to the Day 28 swabs, inclusive).

7.4 Exploratory Outcomes Analyses

The original protocol mentioned an analysis of viral shedding numerical measurements. Such an analysis may or may not be conducted; if conducted, the analysis will be described

elsewhere and regarded as an exploratory analysis.

7.4.1 Duration of COVID-19 attributable symptoms

As noted in Section 5.4 of this SAP, we will analyze symptom duration through $DayA15$, rather than Day 28, which is a departure from what was described in the protocol.

The duration of COVID-19-attributable symptoms will be summarized graphically using Kaplan-Meier curves, and compared between randomized groups by applying the Gehan test with stratification by the age group.

As a sensitivity analysis, the duration of symptoms will be recalculated without any restrictions for having confirmation. That is, the first asymptomatic day in the window will be regarded as the end date, even if the next available symptom assessment indicates a symptomatic day. The symptom duration analysis will be repeated on this modified version of the outcome.

7.4.2 Six-month mortality

Subjects for whom this outcome cannot be ascertained will be excluded from analysis of this exploratory outcome.

The frequency and percentage of ITT subjects known to have died by six months will be calculated, overall and by treatment arm. In addition, the estimated relative risk within each stratum will be reported along with a 95% confidence interval, and the Cochran-Mantel-Haenszel estimate of a common relative risk will be reported along with a 95% confidence interval.

7.4.3 Six-month hospitalization

Subjects for whom this outcome cannot be ascertained will be excluded from analysis of this exploratory outcome.

Analysis will be conducted as for the six-month mortality outcome.

7.4.4 Hospitalizations before $DayD1$

For this outcome, the frequency and percentage of ITT subjects who are hospitalized or die on or before the earlier of the day before study drug delivery and $DayR + 13$ will be calculated, overall and by treatment arm.

7.4.5 Quality of life (EQ-5D-5L Assessment)

The EQ-5D-5L quality of life assessment is administered at baseline, 28 days, and 6 months and we will analyze the VAS and an index measure, as described in Section 5.4. Multiple imputation of missing values will be implemented prior to analysis to improve inference given the expectation that missingness may depend on other characteristics. Imputation will include a death indicator for subjects who were known to have died before the assessment was to have occurred. After imputation, differences between day 28 (6 months) and baseline values of the VAS will be calculated. Any subjects whose measurement at day 28 (6 months) is missing because the individual died will then be post-processed to have a value worse than the worst possible difference among survivors (i.e., -101). Mean differences between day 28 (month 6) values and baseline will be compared between arms using a suitable approach (e.g., Van Elteren test), while adjusting for age stratum.

A similar analytical approach will be taken for the EQ-5D-5L index measure, except that the index measure will be set to 0 for individuals who were missing the assessment because they were deceased. No other missing index measure values will be imputed. Note that survivors may have a lower value for this index measure (i.e., <0) than those who are deceased.

7.4.6 Six-month depression (PHQ-9 Total)

Analysis of the PHQ-9 Total will be conducted in the manner analogous to that of the 28-day and six-month EQ-5D-5L VAS.

7.4.7 Six-month anxiety (GAD-7 Total)

Analysis of the GAD-7 Total will be conducted in the manner analogous to that of the 28-day and six-month EQ-5D-5L VAS.

7.4.8 Post-traumatic stress disorder (PCL-5 Total)

Analysis of the PCL-5 Total will be conducted in the manner analogous to that of the 28-day and six-month EQ-5D-5L VAS.

7.4.9 Additional Analyses for Time-to-event Outcomes

Kaplan-Meier incidence curves will also be created by treatment group for the following:

- Time to death
- Time to first hospitalization and death out of the hospital (as competing risks)
- Time to no symptoms and death (as competing risks)

For all but the first, the cumulative incidence curves for the composite outcome including death would be depicted along with cumulative incidence curves for death and for the other outcome overlaid on the plot.

7.5 Safety Analyses

7.5.1 Formal Safety Outcome(s)

In addition to outcomes described earlier in this section, additional exploratory safety outcomes for this trial include:

- Mortality by 14 days
- Mortality by 28 days
- Hospital admission by 28 days

The analysis of each of these additional safety outcomes will be conducted in analogous fashion to the six-month mortality and six-month hospitalization endpoints (see Sections 7.4.2 and 7.4.3). For these exploratory outcomes, the analysis will be conducted on the ITT and SAFETY populations, with emphasis on the former.

7.5.2 Adverse Events

Per the protocol, AEs will be recorded through the earlier of hospitalization and Day 14.

Collected adverse events The protocol identifies adverse events (AEs) to be collected in AE report forms of the database. These include

- Serious adverse events (SAEs)
- Non-serious AEs judged by investigator as related to study drug or study procedures or of uncertain relationship
- AEs resulting in permanent discontinuation of study drug

Summaries Formal statistical analysis of adverse events is not planned, other than what has been noted previously in this SAP for adverse events that are used to define trial outcomes. However, descriptive summaries of adverse events will be created to assist the DSMB in monitoring safety.

For the closed session of DSMB meetings the overall number and percentage of patients with AEs and SAEs will be summarized, overall and by Medical Dictionary for Regulatory

Activities (MedDRA) coding. AEs will be classified by relatedness, expectedness, seriousness, intensity, action taken, outcome, and modification to study treatment. AEs will be summarized by seriousness, relatedness, and expectedness, and SAEs will be summarized by relatedness and expectedness. This will be done by treatment arm for the SAFETY population (principal focus) and for the ITT population. Generic treatment designations are preferred for interim DSMB meetings unless the DSMB requests otherwise.

8 SAMPLE SIZE DETERMINATION

The protocol contains a detailed sample size justification.

References

[1] EuroQol Research Foundation. EQ-5D-5L User Guide, 2019. Available from: https://euroqol.org/publications/user-guides.

[2] A. Simon Pickard, Ernest H. Law, Ruixuan Jiang, Eleanor Pullenayegum, James W. Shaw, Feng Xie, Mark Oppe, Kristina S. Boye, Richard H. Chapman, Cynthia L. Gong, Alan Balch, Jan J.V. Busschbach. United States Valuation of EQ-5D-5L Health States Using an International Protocol, *Value in Health*, 22(8), 2019, 931–941. https://doi.org/10.1016/j.jval.2019.02.009.

[3] U.S. Department of Veterans Affairs. PTSD Checklist for DSM-5 (PCL-5). Available from: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp.

[4] Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), May 2006, 1092–1097.

[5] Kroenke, K., Spitzer, R.L., Williams, J.B.W., Löwe, B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General Hospital Psychiatry*, 32(4), 2010, 345–359. doi: https://doi.org/10.1016/j.genhosppsych.2010.03.006.

[6] Arrieta, J., Aguerrebere, M., Raviola, G., Flores, H., Elliott, P., Espinosa, A., Reyes, A., Ortiz-Panozo, E., Rodriguez-Gutierrez, E.G., Mukherjee, J., Palazuelos, D., Franke, M.F. Validity and Utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for Screening and Diagnosis of Depression in Rural Chiapas, Mexico: A Cross-Sectional Study. *J. Clin. Psychol.*, 73, 2017, 1076–1090.

[7] Calkins, K.L., Canan, C.E., Moore, R.D., Lesko, C.R., Lau, B. An application of restricted mean survival time in a competing risks setting: comparing time to ART initiation by injection drug use. *BMC Med Res Methodol* 18, Article 27, 2018. https://doi.org/10.1186/s12874-018-0484-z.