Official Title: Hydroxychloroquine vs. Azithromycin for Hospitalized Patients With Suspected or Confirmed COVID-19 (HAHPS): A Prospective Pragmatic Trial

NCT #: NCT04329832

Document Date: Jun 29 2020

HAHPS Study
Statistical Analysis Plan version 1.0, 29JUN2020

Responsible analytic personnel

Tom Greene, PhD
Professor of Medicine and Epidemiology
University of Utah

Allison Butler, MStat
Biostatistician, Intermountain Medical Center

Danielle Groat, PhD
Biostatistician, Intermountain Medical Center

Responsible investigators

Samuel Brown, MD
Associate Professor of Medicine
Intermountain Medical Center/University of Utah

Lindsay Leither, DO
Physician investigator
Intermountain Medical Center

Robert Paine, MD
Professor of Medicine
University of Utah

Ithan Peltan, MD
Assistant Professor of Medicine
Intermountain Medical Center

Table of Contents

Background

The HAHPS trial is a randomized, open label, active comparator trial of hydroxychloroquine vs. azithromycin for hospitalized patients with COVID-19. The HAHPS trial enrolled patients admitted to the hospital with suspected or confirmed COVID-19. As outlined in the protocol and a prior publication,[1] the eligibility criteria excluded patients with markers of increased risk associated with administration of hydroxychloroquine or azithromycin. The trial is being conducted at Intermountain and University of Utah hospitals. While the trial is open-label, outcomes assessors and investigators are blinded to outcomes by treatment group. This SAP was written by blinded personnel, without knowledge of study outcomes.

Summary of analytic approach

General. Descriptive summaries will be produced for subject disposition, demographic and clinical characteristics of the randomized hydroxychloroquine and azithromycin groups. The primary analysis and other secondary analyses of efficacy outcomes will be performed in accordance with the intention-to-treat (ITT) principle, with all patients analyzed according to their randomized treatment assignment, irrespective of compliance. In addition to the formal statistical analyses described below, the proportions of subjects with each level of the 8-point COVID scale will be displayed in frequency tables and in graphical summaries for each day, from day 0 (baseline) through day 14 of follow-up. Other ordinal secondary outcomes will generally be summarized using medians, 10^{th}, 25^{th}, 75^{th}, and 90^{th} percentiles and plots of cumulative distribution functions. In cases where a substantial proportion of subjects fall into an extreme category, several of the indicated quantiles may take on the value of this category and a quantile-based summary may not be fully informative. In these cases, we will tabulate the proportions of subjects which fall into designated categories which are defined based on clinical relevance.

Adherence to the study medications will be summarized by providing the distributions of the proportion of doses of study medication which were administered to each patient overall and by randomized group. We will compute two proportions: 1) The proportion of protocol-intended dose, in which doses for which the patient had a protocol-specified reason not to receive study medication (e.g., prolonged QTc, died, or discharged home) will not be counted in the denominator for proportion of scheduled doses received. 2) The proportion of the total ideal dose, in which doses for which the patient had a protocol-specified reason not to receive study medication will be counted. We will also describe the median (IQR) for total doses received for both protocol-intended dose and ideal dose.

Analysis Sets. The primary analysis and analyses of secondary efficacy outcomes will be performed in the ITT population, which consists of all randomized patients. Although the trial was launched with the flexibility to enroll patients with suspected rather than confirmed COVID-19, no patients were

actually enrolled under that criterion. The ITT population thus simplifies to patients with laboratory-confirmed COVID-19. Summaries of safety outcomes will be performed in the safety population consisting of all patients who receive at least one dose of study medication. In a subgroup of the ITT population, we will compare study endpoints among patients who had not received azithromycin before randomization.

Primary Analysis. The prespecified primary analysis will compare the day 14 assessment of the 8-level COVID Ordinal Outcomes Scale between the randomized hydroxychloroquine and azithromycin groups. We will perform this analysis using a proportional odds logistic regression model (also called an ordinal regression model),[2] with randomized treatment group as the independent variable. Patient age, comorbidities (dichotomized as either no comorbidities or any comorbidities, with the comorbidities drawn from hypertension, COPD, and the Charlson constituent comorbidities[3] (Table1), and the baseline level of the COVID Ordinal Outcomes Scale will be included in the analysis as covariates. Due to restriction of the cohort to hospitalized patients, the baseline COVID Ordinal Outcomes Scale distribution will include only patients with levels 3 through 7. In a sensitivity analysis, we will include the total amount of pre-randomization azithromycin received as another covariate in the primary model.

Table 1. Charlson Comorbidities
Moderate/Severe Liver Disease
Cerebrovascular Disease (CVD)
Peripheral Vascular Disease (surgical)†
Renal disease
Peripheral Vascular Disease (diagnosis) †
Paralysis (Hemiplegia or Paraplegia)
Dementia

Mild liver disease
Congestive heart failure (CHF)
Chronic Obstructive Pulmonary Disease (COPD)
Diabetes with complications
Diabetes

Under the ordinal regression model, the treatment effect is expressed as the odds ratio comparing the odds of patients having a score that exceeds a given threshold vs. a score that is less than or equal to that threshold, and those odds are compared between the hydroxychloroquine and azithromycin groups. An odds ratio less than 1 thus represents treatment benefit for hydroxychloroquine over azithromycin. The model assumes that this odds ratio is the same across all possible thresholds across the 8-level ordinal scale. However, the primary analysis using the proportional odds model is closely linked to the Wilcoxon-rank-sums test and is thus expected to provide approximately valid inference even if the proportional odds assumption is violated.[4] Nonetheless, we will graphically assess the compatibility of the data with the proportional odds assumption by running a series of binary logistic regressions with varying cutpoints on the dependent variable and visually checking the deviation of coefficients across cutpoints. If the proportional odds assumption is clearly violated, our primary analysis under the proportional odds model will be supplemented with separate odds ratios comparing different groups of categories for which the proportional odds assumption is violated.

Following a structure proposed by Harrell and collaborators (http://hbiostat.org/proj/covid19/bayesplan.html), the primary analysis will be performed using a Bayesian framework, with a conservative normal prior distribution assumed for the log transformed odds ratio. The normal prior will be assigned a mean of 0 (thus taking an agnostic position as to the direction of the treatment effect), and a standard deviation of 0.354. This prior distribution assigns a probability of 0.95 to an odds ratio between 0.5 and 2.0, with a probability of 0.025 that the odds ratio

is below 0.50 and also a probability of 0.025 that the odds ratio is greater than 2.0. The use of this moderately conservative Bayesian prior will tend to shrink the observed proportional odds ratio towards 0, and thus moderately discount extreme observed odds ratios which may occur particularly for early interim analyses with relatively small sample size.

The results of the Bayesian analysis will be expressed as (1) a graphical display of the posterior distribution of the odds ratio, (2) the posterior mean, median and mode with 95% credible intervals, and (3) the following probabilities of clinical relevance based on the posterior distribution of the odds ratio (OR) given the data:

P1 = Pr(OR < 1), indicating the evidence for any benefit of hydroxychloroquine over azithromycin;

P2 = Pr(OR < 1/1.25), indicating the evidence for a moderate benefit or greater of hydroxychloroquine over azithromycin;

P3 = Pr(OR > 1), indicating the evidence for any harm of hydroxychloroquine over azithromycin;

P4 = Pr(OR > 1.25), indicating the evidence for a moderate harm or greater of hydroxychloroquine over azithromycin;

P5 = Pr(1/1.2 < OR < 1.2), indicating the evidence for similarity between the two treatments.

For the purposes of primary inference, we will employ the 95% posterior probability of P1 or P3 as evidence for either benefit or harm of hydroxychloroquine over azithromycin. Evidence for moderate harm of hydroxychloroquine (P4) would be considered to be demonstrated at a lower 75% posterior probability threshold. In any case, the distributions of P1–P5 will be displayed for reference in the statistical report of the primary analysis. For clarity and simplicity, posterior Bayesian probabilities will be estimated by applying a normal approximation to the maximum likelihood estimate of the proportional odds ratio obtained by fitting the likelihood function to the data. This simplification will

avoid the complication of specifying prior distributions for the intercept coefficients in the proportional odds model.

Secondary Analyses. We will apply similar Bayesian proportional odds models to evaluate the effectiveness of the hydroxychloroquine vs. azithromycin interventions for efficacy analyses of all ordinal outcomes. We will also present numbers (proportions) of patients in each ordinal status for all endpoints for descriptive purposes and to assist with future meta analyses. If the graphical diagnostics (described above) indicate that the proportional odds model is untenable for certain outcomes, we will report that fact, and supplement the primary analysis using the proportional odds model with separate logistic regression for different cutpoints or quantile regression in order to fully characterize the relationship of each outcome with treatment assignment.

Ordinal secondary outcomes include ventilator-, ICU-, and hospital-free days through day 28 (with death assigned as a worst-rank ordinal level[5,6] to avoid survivorship bias for each of these outcomes). We will use the last-off method—only time after the last liberation from, e.g., ventilation counts toward the total number of -free days. Similar to the primary outcome measured at 14 days, we will apply a primary proportional odds model similar to that used for the primary analysis at day 14 to evaluate the effect of the randomized treatments on the COVID Ordinal Outcomes Scale at day 7. We will present the results of each of these analyses as the posterior median odds ratio with 95% credible interval under the moderately skeptical Bayesian prior described above. For ease of use for investigators performing meta-analyses, we will include detailed descriptions including number of patients for each value of the endpoints by treatment group as tables in the online supplement.

We include two time-to-event endpoints, censored at 28 days: time to a 1-point decrease in the COVID Ordinal scale and time to recovery (defined as in the ACTT-1 trial). The time to recovery endpoint was added after publication of the methods paper but before unblinding of investigators. It was added to assist with future meta analyses given increasing interest in it as an endpoint. Both time-to-event endpoints will be displayed by treatment group (for descriptive purposes) as Kaplan-

Meier curves. Patients who die prior to day 28 are assigned an artificial follow-up time of 28 days censored at that point. This will assign to these patients a worse outcome than any surviving patients, and avoid conflating prolonged hospitalization with death. Descriptively, these time-to-event endpoints will be displayed as median (IQR). If fewer than half of patients meet the event before censoring, we will choose an appropriate quantile (e.g., first quartile) for description and display the proportions with events during the respective weeks of follow-up. For the sake of consistency, we will use the same ordinal regression model with the same prior distribution for the odds ratio as was used for the primary analysis to generate median posterior odds ratios and credible intervals for these endpoints.

Sub-groups. On a purely exploratory level, we will summarize effect variation in sub-groups defined by (a) age ≥ 65 years, (b) symptom duration ≥ 10 days, (c) in ICU at enrollment, (d) presence of at least one comorbidity, (e) race/ethnicity, and (f) receipt of any clinical azithromycin before enrollment. At early interim analyses, the sample size will be too small to support meaningful statistical inference, and we will limit presentation of results in sub-groups to descriptive summaries using the summary measures described for each outcome above.

Missing Data. We expect fewer than 5% of subjects will have missing data on either the primary or key secondary outcomes listed above. Hence, our primary analyses will be restricted to non-missing observations, without imputation. For the primary outcome, a sensitivity analysis will be performed by applying single imputation. Where daily COVID Ordinal Outcomes Scale scores are missing after hospital discharge, we will use last carried forward with the following modifications—patients discharged to a facility other than home will be classified as “home with limitations” (score 2), and patients found subsequently to be home with limitations will be rated as “home with limitations” from hospital discharge through that follow-up point. Patients who are discharged to home and found to have no limitations at day 14 will be rated as “home without” limitations starting at the time of hospital

discharge. This approach will be used because persisting limitations are unlikely to have resolved and then recurred by 14 days. Similar single imputation rules will be devised for sensitivity analyses of secondary efficacy endpoints consistent with prior practice in the field and will be determined before the given analysis is performed to avoid bias from knowledge of outcomes.

Safety reporting

The HAHPS trial uses the adverse event reporting developed for the PROWESS-SHOCK trial.[7,8] This approach is based on the high incidence of conditions resulting from the underlying disease that would be classified as adverse events if they occurred in healthy outpatients but are unlikely to be related to the study drug. Instead of being classified as adverse events, such events are monitored as study clinical or safety outcomes (Table 2). The exception to this rule are events which are either serious and study-related or which lead to permanent discontinuation of study drug. Those latter are reported as adverse events and are classified within the MedDRA taxonomy.

For reporting purposes, for each safety outcome or adverse event each patient will be classified as having experienced the outcome at least once (vs. never having experienced the outcome). We will separately provide counts (proportions) of safety outcomes and individual adverse events. In early interim analyses, comparisons of safety outcomes between treatment groups will be viewed as primarily descriptive. However, we will provide nominal p-values to compare adverse events and safety outcomes between groups without multiple comparison adjustment using the Chi square or Fisher exact test as appropriate to provide a crude assessment of the compatibility of observed differences between treatment groups with chance. Isolated p-values which are < 0.05 will be interpreted recognizing that 5% of such p-values will reach this threshold by chance in the absence of an actual effect of the treatment.

QTc was monitored at 24–48h after randomization as a study procedure. QTc monitoring at baseline was not mandatory but was described where it was available. We also described the change between baseline and 24–48h after randomization in QTc. These results will be summarized

descriptively by treatment group. If the baseline QTc is available in at least 80% of participants, an analysis of covariance will be performed to compare follow-up QTc between treatment groups with adjustment for baseline QTc. Otherwise, a 2-sample t-test will be used to compare the follow-up QTc levels.

Table 2 Safety outcomes (measured daily for first five days)
Death unrelated to study drug or procedures
Seizure
Receipt of vasopressors
Atrial or ventricular arrhythmia
Cardiomyopathy
cardiac arrest
Hypoxemia requiring supplemental oxygen
ARDS
Mechanical ventilation
ECMO
Elevated LFTs
Pancreatitis
Acute kidney injury
Renal replacement therapy
Symptomatic hypoglycemia
Hematologic or coagulation events
Cytopenias
Severe dermatologic reaction

Nausea/vomiting
Visual changes
C. diff infection

Concomitant medications

The classes of concomitant medications listed in Table 3 were measured at baseline (before randomization) and through day 5 after randomization. Concomitant medications will be descriptively summarized overall and by treatment group. In addition, some patients in the hydroxychloroquine arm received azithromycin either before or in some cases after randomization. The proportion receiving azithromycin was recorded, and in exploratory summaries, safety outcomes will be displayed in that group as well as in the groups who received only azithromycin or only hydroxychloroquine.

Table 3. Concomitant medications
ACE inhibitors
ARBs
NSAIDs
Glucocorticoids
Immune-suppression
Chloroquine
Hydroxychloroquine
Azithromycin
Zinc
Vitamin C

Interim Monitoring.

Under the Bayesian design, the posterior distribution describing the accumulating evidence provided by the data for treatment benefit or harm was updated at successive interim analyses during the trial.[9] After consultation with the DSMB and consistent with the DSMB charter written before the first interim analysis, we performed interim analyses after enrollment of every 60 patients. This cadence for interim analyses was confirmed through simulation to yield performance characteristics consistent with a trial-wise alpha of approximately 0.05 if a frequentist approach had been taken.[10] We adopted a Bayesian approach to allow the DSMB to recommend continuation or closure of the trial based on the totality of evidence, including data from the present trial and from other relevant experiments or trials. While the trial is not designed to be stopped for futility based on the trial data alone, the DSMB is authorized to incorporate the totality of evidence in making decisions about recommending closure of the trial to further enrollment.

Sample size and power. We generated an approximate sample size and power calculation by providing the odds ratios closest to 1 (corresponding to the minimum detectable treatment effects, which are indicated by odds ratios less than 1, reflecting treatment benefit for hydroxychloroquine over azithromycin) which can be detected in a frequentist analysis of the primary outcome under a proportional model without covariate adjustment. The target sample size of this trial is 300 randomized subjects; Table 4 also considers smaller sample sizes to indicate the power the trial will reach under various scenarios for enrollment depending on state-wide disease dynamics. The power calculations were performed using statistical simulation (N=10,000 simulations per combination of parameters), and provide results very close to those provided by an approximation provided by Whitehead.[11] For the sample size calculation, we considered three scenarios for the distribution in the azithromycin group for the 8 categories of the COVID Ordinal Outcomes Scale: Scenario 1: Uniform distribution with probabilities of 0.125 for each score level, Scenario 2: Overall probabilities at 14 days

averaging across both groups of the recent COVID-19 trial by Cao et al.[12] (1= 0, 2= 0.36, 3=0.16, 4=0.23, 5=0.05, 6=0.02, 7=0.02, 8=0.16), and Scenario 3, with probabilities averaged between Scenarios 1 and 2.

Table 4. Maximum Detectable Odds Ratios (below 1) with 80% Power, 2-sided α= 0.05, Assuming 95% Follow-up			
Sample size	Scenario 1 (uniform)	Scenario 2	Scenario 3
100	0.36	0.35	0.36
150	0.44	0.42	0.44
200	0.49	0.47	0.49
250	0.52	0.52	0.52
300	0.56	0.55	0.55

We note that when expressed as risk ratios the minimum detectable effects are closer to 1 than the odds ratios in Table 4. For example, in Scenario 3 with N = 300, the detectable odds ratio of 0.55 for a COVID Ordinal Outcomes Scale score > 3 translates to a detectable risk ratio of 0.702 = 0.337/0.48. Thus the target sample size of 300 provides sufficient power to detect a substantial signal for superiority of hydroxychloroquine over azithromycin.

References

1. Brown SM, Peltan ID, Webb B, et al. Hydroxychloroquine vs. Azithromycin for Hospitalized Patients with Suspected or Confirmed COVID-19 (HAHPS): Protocol for a Pragmatic, Open Label, Active Comparator Trial. Ann Am Thorac Soc 2020 (In eng). DOI: 10.1513/AnnalsATS.202004-309SD.
2. McCullagh P. Regression models for ordinal data. Journal of the Royal Statistical Society: Series B (Methodological) 1980;42(2):109-127.
3. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases 1987;40(5):373-83. (http://www.ncbi.nlm.nih.gov/pubmed/3558716).
4. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer, 2015.
5. Colantuoni E, Scharfstein DO, Wang C, et al. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. BMJ (Clinical research ed) 2018;360:j5748. (In eng). DOI: 10.1136/bmj.j5748.
6. Novack V, Beitler JR, Yitshak-Sade M, et al. Alive and Ventilator Free: A Hierarchical, Composite Outcome for Clinical Trials in the Acute Respiratory Distress Syndrome. Critical care medicine 2020;48(2):158-166. (In eng). DOI: 10.1097/ccm.0000000000004104.
7. Ranieri VM, Thompson BT, Barie PS, et al. Drotrecogin alfa (activated) in adults with septic shock. New England Journal of Medicine 2012;366(22):2055-2064.
8. Finfer S, Ranieri VM, Thompson BT, et al. Design, conduct, analysis and reporting of a multi-national placebo-controlled trial of activated protein C for persistent septic shock. Intensive care medicine 2008;34(11):1935-1947.
9. FAYERS PM, ASHBY D, PARMAR MK. Tutorial in biostatistics: Bayesian data monitoring in clinical trials. Statistics in medicine 1997;16(12):1413-1430.
10. Hobbs BP, Carlin BP. Practical Bayesian design and analysis for drug and device clinical trials. Journal of biopharmaceutical statistics 2007;18(1):54-80.
11. Whitehead J. Sample size calculations for ordered categorical data. Statistics in medicine 1993;12(24):2257-2271.
12. Cao B, Wang Y, Wen D, et al. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. New England Journal of Medicine 2020.