Supplemental Online Content

Di Stefano L, et al. Hydroxychloroquine/Chloroquine for the Treatment of Hospitalized Patients with COVID-19: An Individual Participant Data Meta-Analysis.

eMethods. Description of the Primary Outcome Model and Estimands

eTable 1. Data Dictionary from Data Harmonization Spreadsheet

eTable 2. Risk of Bias Assessment

eTable 3. Changes to the Prespecified Statistical Analysis Plan

eTable 4. Primary, Secondary, and Safety Outcomes

eTable 5. Trial Characteristics: Treatment Groups, Participant Assessment, and Inclusion/Exclusion **Criteria**

eTable 6. Merging Trial Arms

eFigure 1. Trial Selection/RCT Selection Process in Detail

eFigure 2. Estimated Study Coefficients

eFigure 3. Conditional Covariate Effects

eFigure 4. Estimated Mortality Rate in Subgroups Under Both Control and HCQ/CQ

eFigure 5. Posterior Predictive Check of Primary Outcomes by Site

eReferences

eAppendix. Collaborators from the 8 Studies Analyzed

eMethods. Description of the primary outcome model and estimands

Model Description

We fit a Bayesian proportional odds ordinal regression model for ordinal score measured between day 28 and day 35 post enrollment. For individuals with multiple measurements, the outcome was taken as the earliest measurement within the time window. Deaths recorded before the time window were carried forward. The following baseline covariates were included in the model: sex, age, number of comorbidities, body mass index, COVID ordinal scale). The coding and reference levels for the baseline variables are indicated in the table below.

Missing baseline covariate data was imputed using multiple imputation via the R package mice (version 3.12).¹ Treatment assignment and outcome were not used the imputation process. All posterior computations described below were pooled across the imputations. Individuals with missing outcome data were excluded from the model fitting; individuals with missing baseline data on a given covariate were excluded from the corresponding subgroup effect estimates.

Let i index an individual patient. Each patient has a vector of baseline covariates X_i and a treatment assignment T_i (1 for hydroxychloroquine or chloroquine or 0 for control). Let the primary outcome for individual \vec{l} be denoted by Y_i with levels $l = 1, ...,$ 7*.* The proportional odds model takes the form:

$$
logit P[Y_i \le l | X_i, T_i] = \theta_l - \eta_i; l = 1, ..., 6
$$

where $\theta_1, \ldots, \theta_6$ are cutpoints that are common to all individuals (prior - ordered Student t prior with 3 degrees of freedom and scale parameter 2.5);

$$
\eta_i = Z_i^T \beta_0 + \alpha_{0, study} + \delta_{0, NCOSS} + T_i (\tau + Z_i^T \beta_1 + \alpha_{1, study} + \delta_{1, NCOSS})
$$

is the linear predictor; Z_i —a function of X_i —is a vector including sex, natural cubic splines (with 3 degrees of freedom) for age, BMI and number of comorbidities, and 5 minus ordinal score; β_0 and β_1 are fixed effects (prior - uniform); τ is a fixed effect (prior – uniform); δ_0 and δ_1 are independent normally distributed mean zero random effects (prior for standard deviation parameters — a half Student-t distribution with 3 degrees of freedom and scale parameter 10).

The model was fit using R, and the library "brms" (version 2.15).^{2,3}

Effect Estimates of Interest

We produced two kinds of effect estimates:

1. Standardized effect estimates*.* These represent the effect of the treatment, averaged over the empirical distribution of individual-level covariates.

We estimated two effects: a *proportional odds ratio* and *risk difference for mortality*. We produced these estimates as follows.

Let $\pi_i^l(t)$ denote the predicted probability that an individual with covariates X_i has outcome level *l* under treatment *t*. Let $\pi^l(t)$ = 1 $\frac{1}{n}\sum_{i=1}^{n} \pi_i^l(t)$ be the predicted probability of outcome level *l* under treatment *t* in a population with the same covariate distribution as in our study.

Proportional Odds: For each iteration of the MCMC algorithm, our estimation of the standardized effect is tantamount to estimation by simulation with the following three repetitive steps: (1) draw a vector of covariates from the empirical distribution; (2) compute the predicted outcome probabilities under treatment and under control; and (3) use these predicted probabilities to draw outcomes under treatment and control (independently). Repeat steps (1) to (3) to generate a very large dataset and fit a proportional odds model with treatment indicator as the sole covariate. The resulting regression coefficient (i.e., log proportional odds ratio) is our standardized treatment effect on this iteration of the MCMC. In practice, for each iteration of the MCMC we fit a weighted proportional odds model with treatment as the sole covariates to a dataset with seven outcome levels crossed with two treatment levels; the weights for outcome level l and treatment level t are proportional to $\pi^l(t)$.

The associated estimand is the odds ratio from the closest fitting proportional odds model (with treatment as the sole covariate) to the true outcome probabilities under treatment and under control for a population with same distribution of covariates as in our pooled dataset.

Risk Difference: We utilize the procedure above with the exception that there is no need to fit the proportional odds model; we simply utilize the predicted probabilities of death under treatment and control and compute the difference. The estimand is the true risk difference under treatment versus control for a population with same distribution of covariates as in our pooled dataset.

Subgroup effects: We apply the above estimation procedure where we restrict the covariate distribution to the specific subgroup of interest. The estimands are subgroup specific.

2. Conditional effect estimates. For each level of a given covariate of interest and within iteration of the MCMC, we compute (1) predicted probabilities of mechanical ventilation or death under treatment and under control as well as the associated relative risk and (2) difference in the values of the linear predictor under treatment and under control, with all other covariates set to their reference values. In these computations, the quantities are marginalized over the study random effects. Posterior summaries of these quantities are plotted.

Model Diagnostics and Sensitivity Analysis

To assess the within-sample fit of our model, we compared the observed outcome data with draws from the posterior predictive distribution and examined Dunn-Smyth randomized quantile residuals.⁴

To assess the sensitivity of our conclusions to modeling choices, we:

- repeated the analysis with weakly informative $N(0, 5^2)$ priors on the fixed effect coefficients, and more conservative half Student-*t* (df = 3, scale = 5) priors on the group-level standard deviations,
- explored the impact of adding to the model the additional variables (1) randomization to treatment with azithromycin and (2) time between symptom onset and enrollment, and
- (post-hoc) fit a version of the model without individual-level treatment interaction terms.

To compare the model fits, we recomputed the primary outcome and mortality estimates as well as estimated leave-one-out predictive (log) densities (LOO-ELPD) using the R package "loo."5,6 The results are as follows:

These results indicate that the model without interaction provides the lowest cross-validated error. See main manuscript for our interpretation of these findings.

eTable 1. Data Dictionary from Data Harmonization Spreadsheet

In addition to the variables below, we requested inclusion and exclusion criteria from each trial.

eTable 2. Risk of Bias Assessment

We utilized the Cochrane Risk of Bias 2 (RoB 2) tool to rate specific risk of bias domains and "overall risk of bias." We considered the following domains of bias, using trial protocols, IPD, and other information provided by investigators: (1) bias arising from the randomization process (methods used to generate and conceal the allocation sequence), (2) bias due to deviations from intended interventions (whether participants and health professionals were masked to assigned intervention and methods used to ensure that participants received allocated intervention), (3) bias due to missing outcome data, and (4) bias in measurement of the outcome. Since we analyzed IPD, we excluded the fifth domain "risk of bias in selection of the reported result." We followed the recommended algorithms to reach an overall "risk of bias" assessment for each trial.

We assessed 6 studies at low risk of bias in the first domain of "Bias arising from the randomization process" (ORCHID, WU352, NCT04335552, TEACH, COVID MED, OAHU-COVID19); COVID MED was assessed as low risk because despite allowance for study arm shifting due to drug supply interruption, this never took place. Two studies (HAHPS, NCT04344444) did not have information on allocation concealment, and NCT04344444 had noticeably different sample sizes for each treatment group. All but two studies had low risk of bias for the second domain of "Bias due to deviations from the intended intervention"; NCT04344444 and OAHU-COVID19 elicited some concerns due to lack of information in the former and no masking in the latter. TEACH was the only study to elicit some concerns for "Bias due to missing outcome data" due to data missing for >20% of the study population. Finally, 5 studies scored "some concerns" for the domain of "Bias in the measurement of the outcome" (WU352, NCT04335552, HAHPS, NCT04344444, OAHU-COVID19). The primary reason for this rating was due to the outcome assessors being aware of the treatment each participant received; the rating was not assessed as "high risk of bias" because our primary outcome measurement is a hard endpoint for which most decision making is protocolized by objective oxygenation and ventilation respiratory status numbers. Our "Overall risk of bias" assessment was "low" for 2 and "some concerns" for 6 of the 8 studies.

eTable 3. Changes to the Prespecified Statistical Analysis Plan

eTable 4. Primary, Secondary, and Safety Outcomes

eTable 5. Trial Characteristics: Treatment Groups, Participant Assessment, and Inclusion/Exclusion Criteria

eTable 6. Merging Trial Arms

eFigure 1. Trial Selection/RCT Selection Process in Detail *(see legend, next page)*

eFigure 1. Two of the trials included in our analysis did not have study acronyms (only trial registration numbers). COVID MED indicates Comparison Of Therapeutics for Hospitalized Patients Infected With SARS-CoV-2 In a Pragmatic aDaptive randoMizED Clinical Trial During the COVID-19 Pandemic; CTSA, Clinical and Translational Science Awards; HAHPS, Hydroxychloroquine vs. Azithromycin for Hospitalized Patients With Suspected or Confirmed COVID-19; HCQ/CQ, hydroxychloroquine or chloroquine; OAHU-COVID19, A Randomized, Controlled Clinical Trial of the Safety and Efficacy of Hydroxychloroquine for the Treatment of COVID-19 in Hospitalized Patients; ORCHID, Outcomes Related to COVID-19 Treated With Hydroxychloroquine Among In-patients With Symptomatic Disease; TEACH, Treating COVID-19 With Hydroxychloroquine; WU352, Washington University 352: Open-label, Randomized Controlled Trial of Hydroxychloroquine Alone or Hydroxychloroquine Plus Azithromycin or Chloroquine Alone or Chloroquine Plus Azithromycin in the Treatment of SARS CoV-2 Infection.

Starting on April 30, 2020, a Trial Innovation Network team aiming to pool RCT data completed multiple systematic outreaches to the Clinical and Translational Science Awards (CTSA) Program community. We connected with interested study teams via the COVID-19 Collaboration Platform website and additional directed outreach. Trialists were asked to upload protocols to the COVID-19 Collaboration Platform repository in an effort to promote collaboration and data aggregation and reduce duplication. We utilized this multi-pathway approach to assure the widest possible knowledge of our effort across established groups of likely investigators in the United States.

One of the sites contacted by the COVID-19 Collaboration Platform, Bassett, had initiated a collaboration registry effort and performed systematic searches of ClinicalTrials.gov using search words "COVID-19" and "hydroxychloroquine" or "chloroquine," and study status of "recruiting" on May 9, 2020, yielding 9 COVID-19 HCQ/CQ RCTs, and again on May 21, 2020; the latter search and subsequent recruiting elicited a list of 19 RCTs (18 from ClinicalTrials.gov, 1 from personal communication). PIs from 18 registered RCTs in the Bassett/COVID-19 Collaboration Platform list were invited to participate in a pooled analysis project focusing on HCQ/CQ; outreach was initiated by Bassett and subsequently completed by the Trial Innovation Network team. Eight RCTs were selected after excluding 2 that declined participation or did not respond; 3 that had differing trial designs including outpatient and prophylaxis studies; 2 that had no enrollment; 1 that was not registered on ClinicalTrials.gov; and 3 that had sites located outside the US (e.g., RECOVERY and DisCoVeRy, part of SOLIDARITY). (We sought to avoid cumbersome international data sharing regulatory delays such as the General Data Protection Regulation [GDPR].) We were able to include 93.3% of patients from the targeted list of US studies. Excluding data from the international trials in our original list, on the other hand, led to the exclusion of a large number of patients, 6,569 in total.

Simultaneously, the Trial Innovation Network team completed a systematic search of ClinicalTrials.gov using search terms "COVID-19" or "SARS-CoV-2," and USbased, on June 2, 2020, yielding 207 studies. The list was filtered to CTSA or CTSA lead sites in part because of consistent data collection at these sites, yielding 103 studies. The team surveyed this CTSA consortium; 26 expressed interest in collaboration. These searches and outreach activities, in addition to review of the studies' key variables, resulted in identifying 13 HCQ/CQ studies. A consensus investigator meeting made the decision to focus on inpatient trials. The rationale was that despite a substantial event rate occurring in inpatient trials, many were experiencing recruitment difficulty. Conversely, the outpatient trials, with lower event rates, had robust recruitment and desire to complete recruitment prior to pooling. Of the 13 trials, 10 were excluded: 4 with prophylactic trial designs, 2 outpatient trials, and 4 that had no enrollment (some were excluded for more than one reason). The outpatient studies were redirected to an outpatient pooling effort (Gates Foundation). Five of the Trial Innovation Network's list of 13 trials overlapped with the Bassett/COVID-19 Collaboration Platform list, and 3 were included in our analysis (ORCHID, WU352, COVID MED).

The list developed by Bassett and the COVID-19 Collaboration Platform was the primary driver for study inclusion/exclusion in our pooled analysis, with augmentation and refinement by the Trial Innovation Network's outreach and search efforts. None of the selected studies were published prior to selection.

eFigure 2. Estimated Study Coefficients

Estimated study coefficients from the primary outcome model, with 66% and 95% credible intervals indicated. "Control" terms represent differences in predicted outcomes between sites under control, for individuals with similar baseline covariates, on the proportional log-odds scale. Treatment effect terms represent differences in predicted treatment effect between sites, again for individuals with similar baseline covariates. Positive coefficients represent, respectively, better ordinal scale outcomes under control and benefit of HCQ/CQ at days 28-35 post-enrollment.

HCQ/CQ indicates hydroxychloroquine or chloroquine.

eFigure 3. Conditional Covariate Effects

eFigure 3 (continued)

Conditional effects from the Bayesian proportional odds model. Shown are (1) the relative risk of mechanical ventilation or death at day 28-35; (2) the estimated probabilities of mechanical ventilation or death at day 28-35 under control and HCQ/CQ; and (3) the log proportional odds ratio comparing HCQ/CQ and control. Each of these effects are shown for reference individuals with the following covariate values: age 60, BMI 25, no baseline comorbidities, baseline ordinal score of 5, and sex coefficient set between male and female values. Curves for continuous covariates are accompanied by 50% and 95% credible bands; intervals for discrete covariates are accompanied by 66% and 95% credible intervals.

BMI indicates body mass index; HCQ/CQ, hydroxychloroquine or chloroquine.

eFigure 4. Estimated Mortality Rate in Subgroups Under Both Control and HCQ/CQ

Estimated mortality rates (measured at day 28-35) by subgroup under control and HCQ/CQ. Shown are both plug-in estimates (based on the proportion of deaths in each subgroup) along with 95% confidence intervals, and model-adjusted estimates with 95% credible intervals. The model used is the same as for the primary outcome analysis.

HCQ/CQ indicates hydroxychloroquine or chloroquine.

eFigure 5. Posterior Predictive Check of Primary Outcomes by Site

Ordinal scale at day 28-35

Posterior predictive check of the main analysis model. Shown are observed outcome data (first row) and draws for the posterior predictive distribution (subsequent rows) of the ordinal outcome scale at day 28-35, plotted against the expected linear predictor for each individual. Each column corresponds to one study in our analysis. Data points have been jittered for clarity.

eReferences

- 1. Buuren S van, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011;45:1-67. doi:10.18637/jss.v045.i03
- 2. Bürkner PC. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*. 2018;10(1):395-411. https://journal.r-project.org/archive/2018/RJ-2018-017/index.html
- 3. Bürkner PC. brms: an R package for Bayesian multilevel models using stan. *Journal of Statistical Software*. 2017;80:1-28. doi:10.18637/jss.v080.i01
- 4. Dunn PK, Smyth GK. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*. 1996;5(3):236-244. doi:10.2307/1390802
- 5. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. 2017;27(5):1413-1432. doi:10.1007/s11222-016-9696-4
- 6. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2019;182(2):389-402. doi:10.1111/rssa.12378
- 7. Didelez V, Stensrud MJ. On the logic of collapsibility for causal effect measures. *Biometrical Journal Biometrische Zeitschrift*. February 2021. doi:10.1002/bimj.202000305
- 8. Greenland S. Noncollapsibility, confounding, and sparse-data bias. Part 1: the oddities of odds. *Journal of Clinical Epidemiology*. 2021;138:178-181. doi:10.1016/j.jclinepi.2021.06.007
- 9. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. *Annals of Internal Medicine*. 2020;172(1):35-45. doi:10.7326/M18-3667
- 10. Self WH, Semler MW, Leither LM, et al. Effect of hydroxychloroquine on clinical status at 14 days in hospitalized patients with COVID-19: a randomized clinical trial. *JAMA*. 2020;324(21):2165-2176. doi:10.1001/jama.2020.22240
- 11. Ulrich RJ, Troxel AB, Carmody E, et al. Treating COVID-19 with hydroxychloroquine (TEACH): a multicenter, double-blind randomized controlled trial in hospitalized patients. *Open Forum Infectious Diseases*. 2020;7(10):ofaa446. doi:10.1093/ofid/ofaa446
- 12. Brown SM, Peltan ID, Webb B, et al. Hydroxychloroquine versus azithromycin for hospitalized patients with suspected or confirmed COVID-19 (HAHPS). Protocol for a pragmatic, open-label, active comparator trial. *Annals of the American Thoracic Society*. 2020;17(8):1008-1015. doi:10.1513/AnnalsATS.202004-309SD
- 13. Brown SM, Peltan I, Kumar N, et al. Hydroxychloroquine vs. azithromycin for hospitalized patients with COVID-19 (HAHPS): results of a randomized, active comparator trial. *Annals of the American Thoracic Society*. 2020;18(4):590-597. doi:10.1513/AnnalsATS.202008-940OC

eAppendix. Collaborators from the 8 Studies Analyzed

