Estimating the effect of preventative use of hydroxychloroquine and interaction between age in four severity biomarkers after COVID infection

Chunhui Gu
Department of Biostatistics, School of Public Health
University of Texas Health Science Center

Abstract

The effect of preventative use of hydroxychloroquine (HCQ) after COVID infection in four commonly used biomarker of severity: D-dimer, partial thromboplastin time (PTT), C-Reactive protein (CRP), and procalcitonin were assessed by Multivariate Analysis of Variance (MANOVA) and Analysis of Variance (ANOVA), as well as the interaction between age and HCQ use. We had 24 patients who had HCQ use in the case group and 24 patients who didn't have HCQ use in the control group. Patients were divided into four groups based on their age group and HCQ use. PTT, CRP, and procalcitonin were considered as multivariate normally distributed after removing two potential outliers by Royston test for three of the four groups. The variance-covariance matrices were considered the same by the Box-M test ($p = 0.001$). No interaction between age and preventative use of HCQ was found in PTT, CRP, and procalcitonin by MANOVA ($p = 0.3$) and in D-dimer ($p = 0.3$). A marginally significant effect of preventative HCQ use was found in PTT, CRP, and procalcitonin by MANOVA ($p = 0.1$) but a significant effect was found in D-dimer ($p < 0.001$). The study suggested a potentially preventative effect of HCQ use but further study with a larger sample size should be done for a more accurate result.

Introduction

The pandemics of COVID-19 in the United States had led to about 30 million infections and 540 thousand deaths (Centers for Disease Control and Prevention, 2021) and still hasn't shown a sign of slowing down. It imposed an urgent need for an effective treatment for COVID-19 than any time. Since the emergence of COVID-19, a variety of treatments have been proposed to use off-label to treat severe COVID cases in hope of having some benefits. Among those treatments, hydroxychloroquine (HCQ) was considered the most prospective one because its underlying pharmacological mechanism could block the pathway that coronavirus used to invade cells (Schrezenmeier & Dörner, 2020).

Many studies had been done in exploring the potential benefits of HCQ in treating COVID showed that there was no or even opposite effect in the probability of encountering an adverse event or whether the time-to-event of adverse events changed (Jorge, 2020; Kashour et al., 2021). Other studies that tried to ascertain whether HCQ could help to prevent COVID infection also failed to see a significant reduction of infection rate in people using HCQ for preventative purposes. However, very few studies had been done about how HCQ could prevent the development of severity after a person get COVID.

D-dimer test and partial thromboplastin time (PTT) test are two commonly used tests to evaluate the coagulation function. C-Reactive Protein (CRP) test and procalcitonin test are two widely used tests for inflammation activity in the body. Those four tests alone or along with some other tests are used together to evaluate the severity of a patient from a biochemical scope.

In this study, we were interested in how preventative use of HCQ could affect those four lab tests when after diagnosed as COVID and whether there was an interaction between the effect of HCQ and age.

Method

Data extraction

The subjects we included in this study were retrieved from OptumCOVID, which was a massive electronic health record (EHR) database containing necessary information from COVID patients across the nation. The version which we used in this study stored records up to 2021 Jan 22. All queries were done MongoDB, a non-relational database management software through its Python interface.

Records of different kinds were recorded chronologically and stored in different collections of the database. A variety of constraints were enforced to retrieve variables needed as accurately as possible. (For example, the COVID diagnosis date was defined as the date of the earliest diagnosis after 2020/01/01. The completed list of definitions of variables in this study can be found in Supplementary document 1.)

Since reinfection rarely happened in a patient already infected by COVID-19 at the time that this study was done (Centers for Disease Control and Prevention, 2020). In this study, we assumed that each patient only got COVID infection one time and the earliest recorded date of COVID positive diagnosis records was treated as the COVID diagnosis time of that patient.

For the values of the outcome variables: D-dimer, PTT, CRP, and procalcitonin, the first available lab test result of each of those four tests for each patient within 3 days after he or she was diagnosed as COVID was used in this study.

Eligibility of patients

Include Criterion: Patients who were diagnosed as COVID either by tests or professionals' judgment between 2020/01/01 and 2021/01/22.

Exclude Criterion: Patients who have missing information in gender, ethnicity, division, age, smoking status, and BMI.

Case and control groups

A patient was assigned to the HCQ case group if it had a record of HCQ administration at least 2 weeks before being diagnosed no early than 10 weeks. The indications that HCQ was used were unclear for patients in the database and regular contraindications of HCQ were assumed. Patients in the case group and control group were called cases and controls respectively for short.

Commented [CG1]: The restriction is proposed to make sure the HCQ usage is used for treating COVID for that patient.

Matching of control patients to treatment patients

Matching was used to adjust other possible confounders to the treatment effect of the HCQ. Matching was performed by R 4.0.2 and package MatchIt 4.0.1. Cases and controls were exactly matched on gender, ethnicity, region division on a fixed ratio of 1:1 without replacement. The matched data are similar for the covariates used in matching except for the age group and whether they use HCQ or not. Base on their HCQ use case and age group matched patients were classified into four groups: no HCQ use and low age group, HCQ use and low age group, no HCQ use and high age group, and HCQ use and high age group. Patients old than 18 but younger than 65 years old were classified as low age group and patients equal to or old than 65 years old were classified into the high age group.

Commented [CG2]: Race and ethnicity, and region and division are moderately to highly correlated. Should we just use one of each pair?

Commented [CG3R2]: Use just ethnicity and division instead of race and region

Methods

The log-transformation was used for D-dimer, partial thromboplastin time (PTT), and procalcitonin test results. For CRP, square root transformation was used. When mentioned in this study, those four outcome variables were used in their transformed form except explained explicitly. The Box-plot was used to check potential univariate outliers for each of those four outcome variables respectively.

Multivariate analysis of variance (MANOVA) using Pillar's trace as test statistic was used to test whether the means of the outcomes were the same in different groups. MANOVA was used to adjust the correlation between the four outcome variables for better power in detecting effects. Analysis of variance (ANOVA) will be used for outcome variables that didn't satisfy the MANOVA assumption. MANOVA and ANOVA tests were performed by the "Manova" function and "Anova" function in package "car" 3.0-10 respectively in R (Fox et al., 2019).

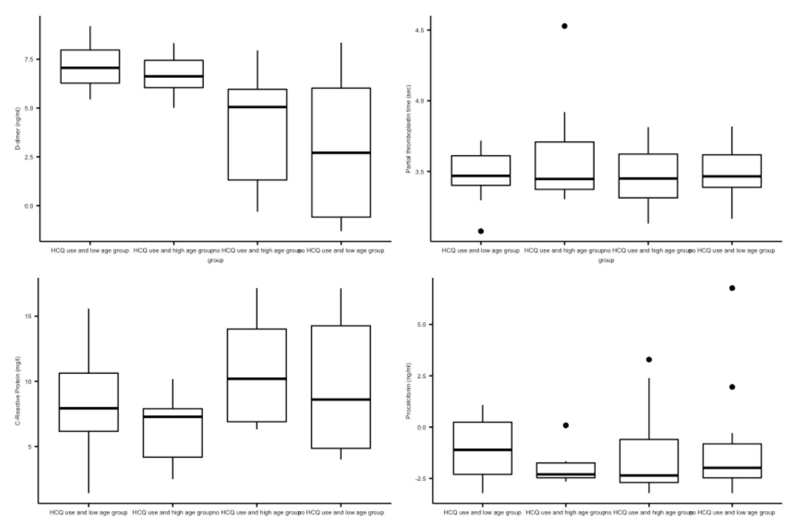
The following assumptions were made for MANOVA: patients were independent samplings from a large population (independence assumption), the population of each group was multivariate normally distributed (multivariate normality assumption), the populations of those four groups had a common variance-covariance of the outcome variables (homogeneity of variance-covariance assumption), and the outcome variables were not extremely highly correlated with each other (absence of multicollinearity assumption). Similarly, univariate assumptions were made for ANOVA.

Shapiro-Wilk test was used to test the univariate normality of the four transformed outcome variables. The Royston test was used to test the multivariate normality assumption for the outcome variables and a Q-Q plot was drawn for a visual check of the multivariate normality assumption. Shapiro-Wilk test and Royston test were performed by the "mvn" function in "MVN" package 5.8 (Korkmaz et al., 2014). Box-m test was used to test the homogeneity of variance-covariance assumption and was performed by "boxM" function in "biotools" package 4.1 (da Silva et al., 2021)

. A correlation matrix was made for the four outcome variables to check the potential multicollinearity problem.

Result:

Two patients were removed as potential outliers due to extremely high procalcitonin values. The homogeneity variance assumption could be violated for the D-dimer variable based on the boxplot after removing the two potential outliers (**Figure 1**). No severity multicollinearity problem was found in the data and the largest correlation was 0.235, which was between CRP and procalcitonin (**Table 2**).



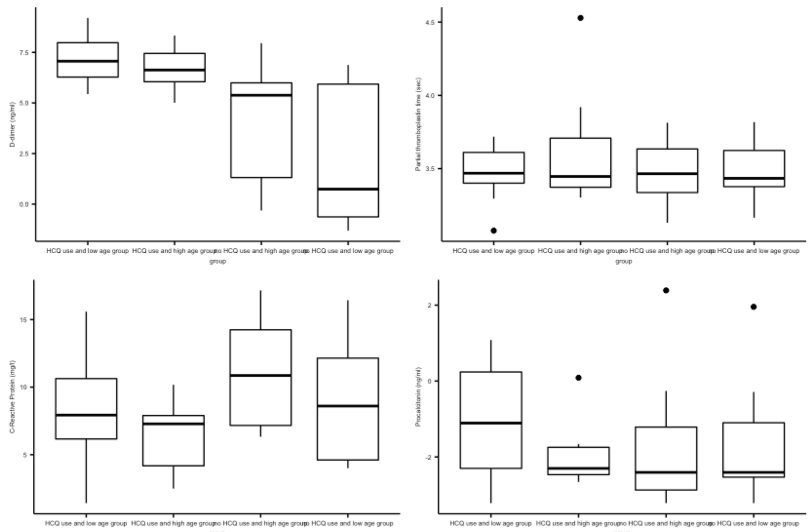


Figure1 Box-plot of the four outcome variables for the case and control group before removing outliers (upper) and after removing outliers (lower).

After removing the two potential outliers, the covariates used in matching were still considered as balanced between the case and control group. Besides, the marginal D-dimer, CRP, and procalcitonin were quite different between the case and control groups (**Table 1**).

Table 1: Descriptive table

	HCQ use (N = 24)	no HCQ use (N = 22)
age		
[18,65)	17 (70.83%)	11 (50.00%)
>= 65	7 (29.17%)	11 (50.00%)
gender		
Female	17 (70.83%)	16 (72.73%)
Male	7 (29.17%)	6 (27.27%)
race		
Caucasian	13 (54.17%)	11 (50.00%)
African American	8 (33.33%)	8 (36.36%)
log D-dimer (ng/ml ddu)		
mean (sd)	7.05 ± 1.15	3.28 ± 3.20
log PTT (sec)		
mean (sd)	3.52 ± 0.27	3.49 ± 0.20
Sqrt CRP (mg/L)		
mean (sd)	7.45 ± 3.67	9.99 ± 4.26
log procalcitonin (ng/ml)		
mean (sd)	-1.35 ± 1.36	-1.73 ± 1.57

Table2: Correlation between outcome variables

	D-dimer	PTT	CRP	procalcitonin
D-dimer	1	-0.159	-0.15	0.235
PTT	-0.16	1.000	0.14	0.059
CRP	-0.15	0.136	1	0.230
procalcitonin	0.24	0.059	0.23	1

The Royston test rejected the multivariate normality of four outcome variables for most of the groups (p-values not shown here) and the D-dimer was considered as the one that could cause the violation for its univariate heterogeneity of variance in Figure 1. After the D-dimer was removed from the dependent variables in the multivariate model, three of the four groups passed the Royston multivariate normality test (**Table 3**). The only group that could depart from multivariate normality was the HCQ use and low age group ($p < 0.019$). This could be because the number of patients in this group was too small. From the Q-Q plot, there was no obvious systematic departure pattern from a straight line. As a result, we still assumed multivariate normality in this group.

Using the criterion of $P < 0.001$, we considered the variance-covariance matrices of those four groups were the same from the Box-m test ($P = 0.001$).

Table 3: Univariate normality and multivariate normality tests

	Shapiro-Wilk test			Royston Test	
Group	Variable	Statistic	P-value	H	P-value
No HCQ use and low age group	PTT	0.98	0.973	5.7	0.13
	CRP	0.89	0.142		
	Procalcitonin	0.87	0.069		
HCQ use and low age group	PTT	0.95	0.42	2.4	0.5
	CRP	0.96	0.56		
	Procalcitonin	0.94	0.30		
No HCQ use and high age group	PTT	0.98	0.964	7.5	0.057
	CRP	0.91	0.257		
	Procalcitonin	0.81	**0.011**		
HCQ use and low age group	PTT	0.77	**0.019**	9.9	**0.019**
	CRP	0.94	0.673		
	Procalcitonin	0.78	**0.027**		

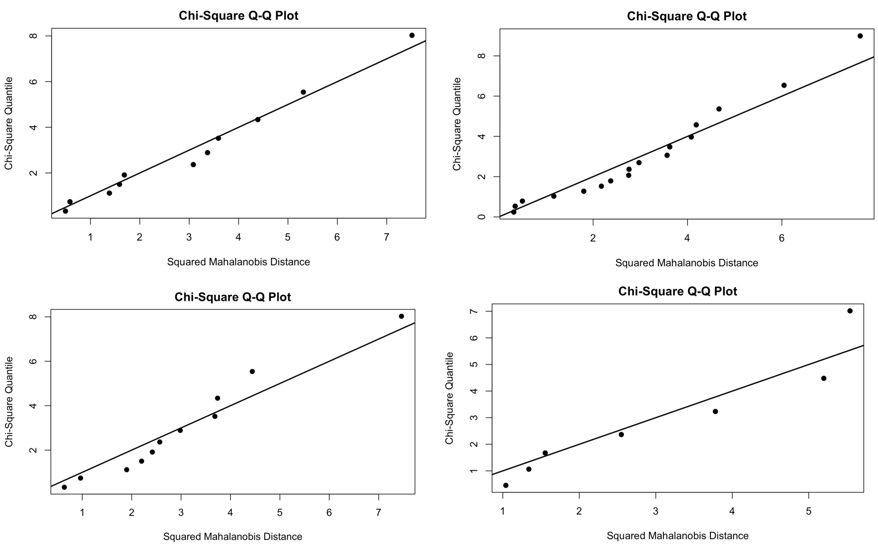


Figure 2 Multivariate Q-Q plots of the three outcome variables four the four groups: no HCQ use and low age group (top left), HCQ use and low age group (top right), no HCQ use and high age group (bottom left), and HCQ use and high age group (bottom right).

The MANOVA analysis showed that there was no significant interaction between age group and HCQ use (P-value = 0.3, **Table 4**). The interaction term was removed from the model as result to test whether there was a main effect of HCQ use or age group. No significant effect of HCQ usage (p-value = 0.1) or age group (p-value = 0.5) were found from the model (**Table 5**).

Table 4: Type III MANOVA test for interaction between age group and HCQ use

Effect	Df	Pillai's trace	Approximate F	Numerator Df	Denominator Df	Pr (>F)
HCQ use	1	0.04	0.59	3	40	0.6
Age group	1	0.10	1.5	3	40	0.2
HCQ use X Age group	1	0.09	1.4	3	40	0.3

Table 5: Type III MANOVA test for the main effect of age group and HCQ use

Effect	Df	Pillai's trace	Approximate F	Numerator Df	Denominator Df	Pr (>F)
HCQ use	1	0.14	2.2	3	41	0.1
Age group	1	0.05	0.74	3	41	0.5

Since ANOVA test was robust to normality assumption and univariate homogeneity of variance assumption. ANOVA test was used for D-dimer. HCQ use had significant effect in D-dimer ($p < 0.001$, **Table 6**).

Table 6: Type III ANOVA test for the main effect of age group and HCQ use for D-dimer

Effect	Sum square	Df	F value	Pr (>F)
HCQ use	168	1	30.07	<0.001
Age group	5	1	0.86	0.36
Residuals	240	43		

Analysis using data without removing potential outliers

When no potential outlier was removed, the Royston test rejected the multivariate normality assumption for most of the groups (**Table 7**).

Table 7: Univariate normality and multivariate normality tests

	Shapiro-Wilk test			Royston Test	
Group	Variable	Statistic	P-value	H	P-value
No HCQ use and low age group	PTT	0.99	> 0.999	13	**0.006**
	CRP	0.89	0.118		
	Procalcitonin	0.74	**0.002**		
HCQ use and low age group	PTT	0.95	0.42	2.4	0.5
	CRP	0.96	0.56		
	Procalcitonin	0.94	0.30		
No HCQ use and high age group	PTT	0.97	0.877	8.9	**0.031**
	CRP	0.89	0.116		
	Procalcitonin	0.80	**0.010**		
HCQ use and low age group	PTT	0.77	**0.019**	9.9	**0.019**
	CRP	0.94	0.673		
	Procalcitonin	0.78	**0.027**		

Four individual ANOVA tests were performed on each of those outcome variables. Besides D-dimer, the use of HCQ was significant in CRP (**Table 8**).

Table 8: Type III ANOVA test for the main effect of age group and HCQ use for CRP with outliers

Effect	Sum square	Df	F value	Pr (>F)
HCQ use	85	1	5.05	**0.03**
Age group	0	1	0.02	0.90
Residuals	755	45		

Comparison with the results from the original literature

The data in the study was extracted from the Optum COVID database by the author and not from a previous study.

Conclusion

The preventative use of HCQ could reduce the level of D-dimer in blood both statistically and clinically significant after COVID infection. The effect of preventative use of HCQ was the same between different age groups. However, whether the reduced level of D-dimer was a good indicator of COVID severity needs to be checked.

Limitations and future work

We only included patients that had complete information in four outcome variables, this could introduce selection bias to the data. The multiple imputation method that fills missing values based on available outcome variables may be useful to solve this problem.

The mean of CRP, PTT, procalcitonin was highly different between the HCQ use and no HCQ use group. However, the MANOVA p-value was close to 0.1 but not significant could be because the sample size was not large enough. Further study that has a larger sample size should be done to check whether preventative HCQ use affects or not in PTT, CRP, and procalcitonin.

We detected an effect of preventative use of HCQ in D-dimer. However, the result should be interpreted with caution despite the univariate ANOVA was robust to the homogeneity of variance assumption.

Reference

Centers for Disease Control and Prevention. (2020). *Reinfection with COVID-19*.

https://www.cdc.gov/coronavirus/2019-ncov/your-health/reinfection.html
Centers for Disease Control and Prevention. (2021). *COVID Data Tracker*. United States COVID-19 Cases and Deaths by State. https://covid.cdc.gov/covid-data-tracker/#datatracker-home
Jorge, A. (2020). Hydroxychloroquine in the prevention of COVID-19 mortality. *The Lancet Rheumatology*. https://doi.org/10.1016/s2665-9913(20)30390-8
Kashour, Z., Riaz, M., Garbati, M. A., Aldosary, O., Tlayjeh, H., Gerberi, D., Murad, M. H., Sohail, M. R., Kashour, T., & Tleyjeh, I. M. (2021). Efficacy of chloroquine or hydroxychloroquine in COVID-19 patients: a systematic review and meta-analysis. *Journal of Antimicrobial Chemotherapy*, *76*(1), 30–42. https://doi.org/10.1093/jac/dkaa403
Schrezenmeier, E., & Dörner, T. (2020). Mechanisms of action of hydroxychloroquine and chloroquine: implications for rheumatology. *Nature Reviews Rheumatology*, *16*(3), 155–166. https://doi.org/10.1038/s41584-020-0372-x

Fox J, Weisberg S (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/
da Silva AR (2021). *biotools: Tools for Biometry and Applied Statistics in Agricultural Science*. R package version 4.1, https://cran.r-project.org/package=biotools.
Korkmaz S, Goksuluk D, Zararsiz G (2014). “MVN: An R Package for Assessing Multivariate Normality.” *The R Journal*, 6(2), 151–162. https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf.