Authors: Patrick Poleshuk, Derek Moore, Jared Rebuyon

The majority of the analysis was re-created by me, Patrick Poleshuk, and reinterpreted from our original project. However, the simple linear regression findings below, using "D6" as the response variable, as well as the write-ups for the potential variable relationships and portions of the analysis outline were generous contributions from my groupmates. All code and comments in the corresponding R-Markdown file were written solely by me.

The original clinical experiment can be found here:
"https://reader.elsevier.com/reader/sd/pii/S0924857920300996?token=DA9F3ACA910281A2DA3D37D275997DFF7BA1013EF4550EE270991C32CC1CB5AB95DB0F57AB13A121257B5544CFFE3C55&originRegion=us-east-1&originCreation=20210821224531"

Problem

Covid-19, later called SARS-CoV-2, is a respiratory virus that began in Wuhan, China, and has spread worldwide infecting millions. While typically only one in three people develop symptoms, the infection can be extremely harmful to older people and those at a higher health risk. If someone at a higher risk gets infected, it could result in severe illness and other symptoms, and in extreme cases, it can even be fatal. Because of the potential danger of the virus, the world has shifted into a pandemic with our reliance on social distancing and quarantining to disrupt the spread of the virus.

In order to combat the recent Covid-19 pandemic, scientists and health officials have worked in conjunction to discover an effective cure. One French medical team ran a case study in which they analyzed the efficacy of two different drugs as treatment for Covid-19: Hydroxychloroquine and azithromycin. These two treatments demonstrated promising results in isolated experiments by the medical team. Because of this, they believed these treatments may be effective outside of such isolated settings. To determine whether these treatments were effective, the team treated patients who were Covid-19 positive with a 600mg dose of hydroxychloroquine. Depending on whether or not they volunteered, some were dosed with azithromycin as well. The trial was orchestrated by The Méditerranée Infection University Hospital Institute in Marseille, but other patients from hospitals around South France who did not receive hydroxychloroquine treatment were used as part of the control group.

Procedure

The experiment location varied depending on whether patients were assigned to the treatment or control group, but all were treated at hospitals within the same geographic area of South France. All patients prior to the study were exposed to a sufficiently high SARS-CoV-2 RNA concentration, with hospitalized patients being eligible if they were aged above 12 years and had a documented SARS-CoV-2 carriage, but participation was excluded for breastfeeding and pregnant women and those with known allergic reactions to the treatment drug. Patients, prior to participating, had all given informed consent and safe and ethical practices were all insured by local and global regulatory bodies.

At the beginning of the clinical experiment, data was collected for all patients, and daily thereafter up until day 14. The primary endpoint, however, occurred on day 6 and consisted of testing for virological clearance of SARS-CoV-2 RNA concentration through real-time reverse transcription-PCR. All those who complied with the hydroxychloroquine treatment were located at Marseille center, with the 600 mg dosage being spread out in thirds throughout the day. All patients in the study were categorized as either being asymptomatic, suffering from upper respiratory tract infection (URTI), or lower respiratory tract infection (LRTI), with symptoms, evaluated through standardized questionnaires and nasopharyngeal samples, that would gradually disappear as the study progressed to its final days.

Data Variables

Each of the dataset value names are described below. Several nominal and numerical variables are included and reflect the extent of the original researchers' study.

age: The age (in years) of the subject.
sex: Coded either "M" for Male or "F" for Female.
status: Every patient is either classified as asymptomatic or suffering symptoms of upper respiratory tract infection (UTRI) or lower respiratory tract infection (LRTI).

inclusion time: Days between the onset of symptoms and day of inclusion.
chloroquine: Coded either "No" for the control group or "Yes" for the treatment group.
conc: Hydroxychloroquine dose concentration per patient.
azithromycin: Coded either "No" for no exposure to the Azithromycin dosage or "Yes" for exposure.
D0: The SARS-CoV-2 RNA concentration from nasopharyngeal samples, prior to the clinical experiment
D1-D6: The SARS-CoV-2 RNA concentration from nasopharyngeal samples, observed daily through the experiment's primary endpoint.
atrtCode: A binary factor variable corresponding to the data of "chloroquine", where "No" = 0 & "Yes" = 1.
ctrtCode: A binary factor variable corresponding to the data of "azithromycin", where "No" = 0 & "Yes" = 1.
sexCode: A binary factor variable corresponding to the data of "sex", where "M" = 0 & "F" = 1.

t0: A binary factor variable, corresponding to the "D0", where 0 represents a negative PCR test result for SARS-CoV-2 viral capacity. A "1" tells us that the patient has tested positive for COVID-19.
t1-t6: A binary factor variable, corresponding to the "D1-D6", where 0 represents a negative PCR test result for SARS-CoV-2 viral capacity. A "1" tells us that the patient has tested positive for COVID-19.

Results

In total, the sample size consisted of 26 patients in the hydroxychloroquine treatment group and 16 in the control group. It is important to note that some attrition was observed when a total of 6 patients had dropped out, from either adverse effects from the drug, being rushed into an intensive care unit, or in one extreme case from death. All of this attrition occurred within the treatment group, as all 16 observations in the control were preserved.

The data that is presented only contains follow-ups up to the primary endpoint, and among the SARS-CoV-2 carrying patients, 16.7% were asymptomatic 61.1% suffered from URTI symptoms, while 22.2% suffered from LRTI symptoms. With the exception of age, the average being 51.2 and 37.3 for the treated and control group respectively, no systematic differences were observed between the two groups. Out of the subjects exposed to the hydroxychloroquine dosage, 6 of them had supplemented with azithromycin.

The hydroxychloroquine treatment showed a significant lagged effect throughout the 3rd to 6th day. After that last day, it was shown that 70% of the treated patients were virologically cured with only 12.5% of those in the control group experiencing that same benefit. These findings are proven to be statistically significant at the 1% significance level. Moreover, when we separate the treatment group into those who were treated with the azithromycin supplement, we find that 100% of these subjects were virologically cured after the 6th day. In contrast, only 57.1% of those who took the base treatment were virologically cured after that same period. These findings are also significant at the 1% significance level.

Interpretations

Conclusively, these findings are shown to be very significant. A notable difference between the treated and control group, with respect to their viral nasopharyngeal carriage, was even observed after only 3 days. This is even more shocking when we consider existing Chinese research, pointing towards uninterrupted viral shedding with a mean duration of up to 20 days.

If we wanted to expand upon these findings, future research should be devoted to analyzing the cost-benefit relationship between the synergistic effect of hydroxychloroquine and azithromycin, specifically if azithromycin acts as a significant counter to bacterial super-infections. Additionally, more time could be allocated towards understanding why hydroxychloroquine had failed to work for some subjects, through a genome analysis of these patients. Nevertheless, while the need for future research remains paramount, findings such as these can be said to have great international influence both in eliminating the spread of COVID-19 and restoring global stability.

Regression Analysis

The depth of the publication's observed data invites multiple motivations for constructing appropriate inferences, which include the extrapolation of the 36-sample size to a national or global population and the determination among several variable relationships as causal. To contextualize the given information, a number of values are illustrated below. In the interest of our research, patients treated with both hydroxychloroquine and azithromycin

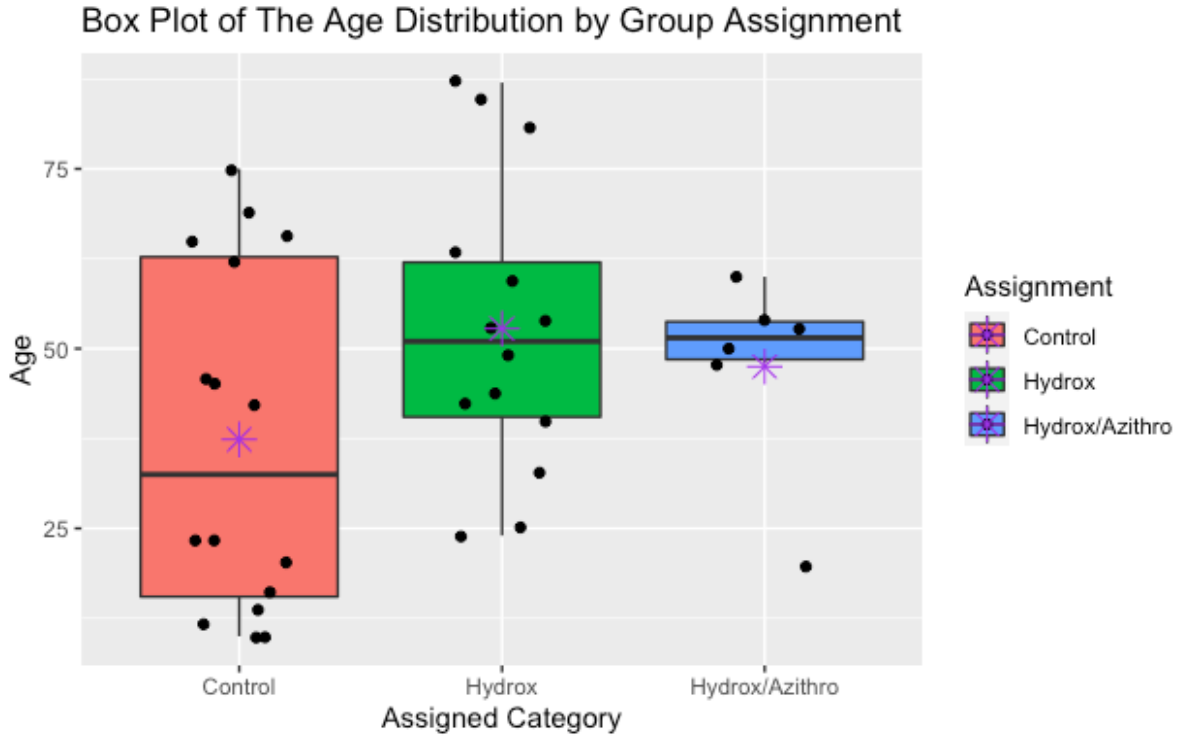
will be distinguished from those who only received hydroxychloroquine in the treatment group, as the negative tests of the former intimate more distinct observations.

Since the main goal of the published study was to observe virological clearance at day-6 post-inclusion, many of the variables were chosen as potential predictors of a received negative PCR test at day 6 (D6). Accordingly, we want to build a set of regression models to determine which, if any, variables possess a causal relationship with a negative test result at this period of time. As the data set provides binary interpretations for all of our D0-D6 data, we will be using logistic regressions to approximate the log-odds interpretation that one receives a positive or negative test result, given what parameter estimates we've observed. For our classification-type regressions, we can make inferences using the "t6" column, derived from the data of "D6", while simple linear regression can be used with just "D6" as our response variable. Our approach consequently follows making Bayesian inferences over the classical alternatives because of the relatively small sample size for the dataset, as well as its importance for conducting future experiments on such a severe situation. All of the results from our findings will be presented here, but it is in the R markdown file where one can see the code necessary to make these calculations, as well as recreate the original clinical findings, along with comments giving more detailed probabilistic interpretations of our log-odds effects.

t6 ~ Age

A majority of older patients among both treatment and control groups were acknowledged, but the significance between this trend and its correlation with PCR testing was left undetermined by the original paper. An investigation as to whether such a relationship existed was examined.

Figure 1: Interquartile Ranges for Age among Three Groups



Age Values for Sample Set

	Mean ± SDs	Median	IQR
Control (N=16)	37.375 ± 23.97186	32.5	47.25
Hydrox. Only (N=14)	52.78571 ± 20.59193	51	21.5
Hydrox. and Azith. (N=6)	47.5 ± 14.0819	51.5	5.25

While disparate amounts among the three groups are notable, an older demographic, by about a decade, is observed, with a wider interquartile range of ages and lesser median value for the control group.

Day 6 Test results versus standardized age for the entire dataset was modeled. Parameter estimates with 95% Bayesian Credible Intervals are provided below. Significant figures were given up to four decimal places.

	Median	MAD_SD
(Intercept)	0.3574	0.3485
Standardized Age	0.0486	0.3468

	5%	95%
(Intercept)	-0.2166	0.9552
Standardized Age	-0.5116	0.6569

From the data, the intercept median value of 0.3574 and 95% CI range of ~ -0.2166 to ~ 0.9552 indicates that someone of average age has a 0.3574 log-odds effect of testing positive for COVID-19. Additionally, our credible interval tells us that there is a 95% probability that 0 would be a plausible value for this effect. The standardized age variable alters the chance of receiving a non-negative test by about 0.0486, an otherwise negligible effect. We can further see two visuals for the insignificance of the effect that age has on the probabilistic outcome of COVID test results.

Figure 2:

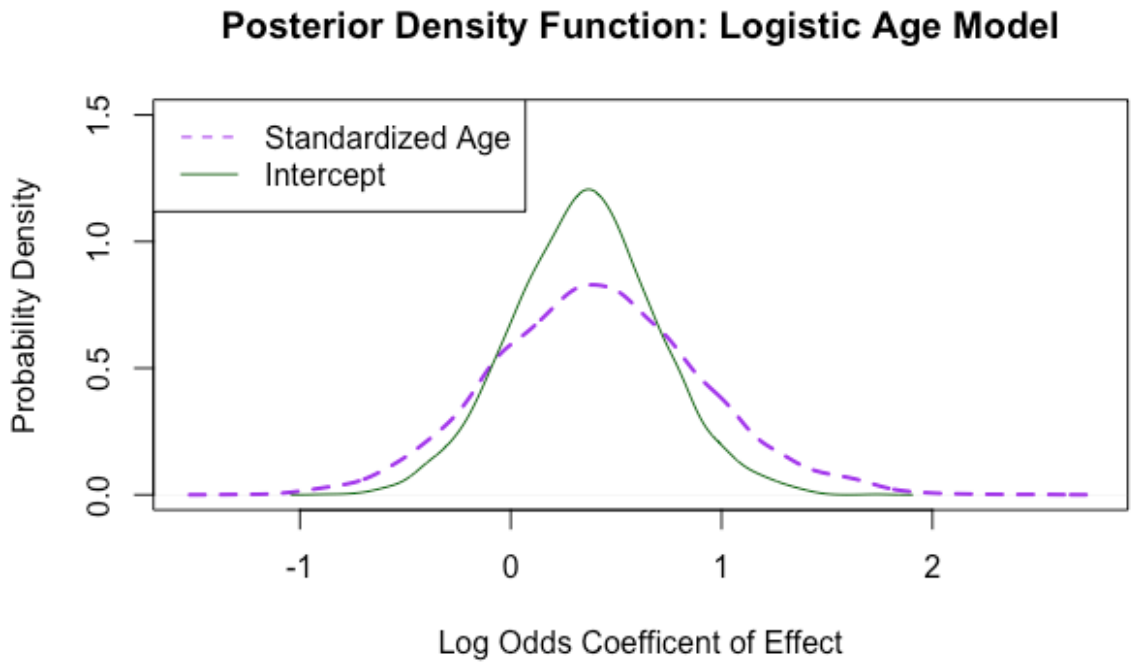
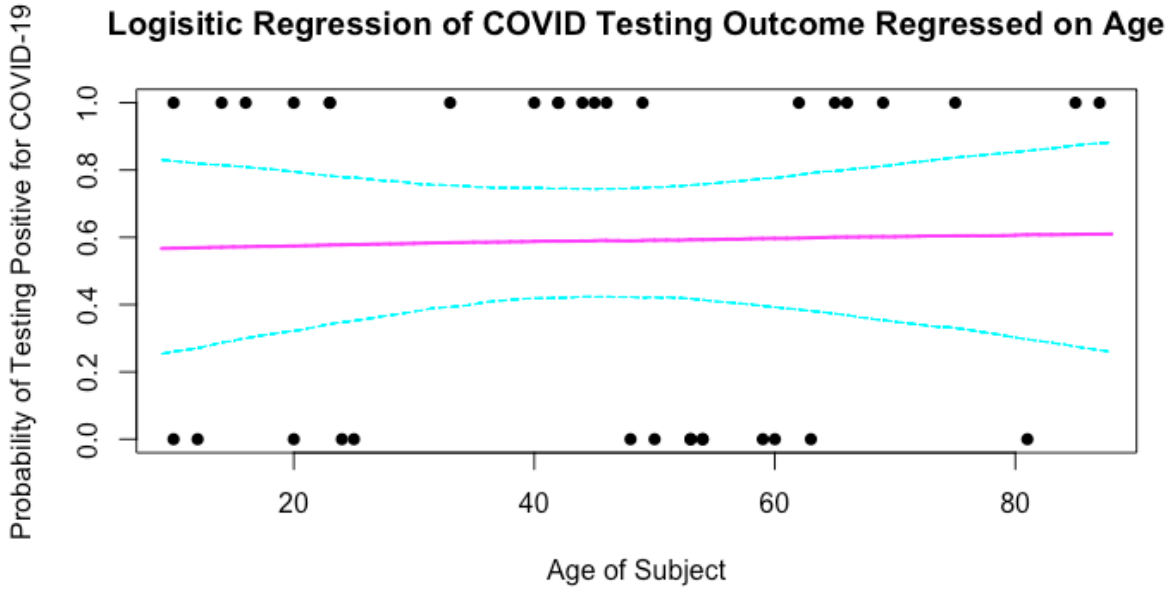


Figure 3:



We can consequently determine that, for age across both groups, a minimal correlation exists between age and day 6 positive test results, such that one is more likely to not test negative after six days if the individual is near middle-aged.

t6 ~ Sex

Per the original publication, no significant difference was observed between the two groups and the biological gender of each individual. However, a greater frequency of female patients for both treatment and control sets can still be observed. Likewise, we can examine

this variable against day 6 test results by modeling the data with parameter estimates and 95% Bayesian CIs. Beforehand, we can also visualize these two observations.

Figure 4:

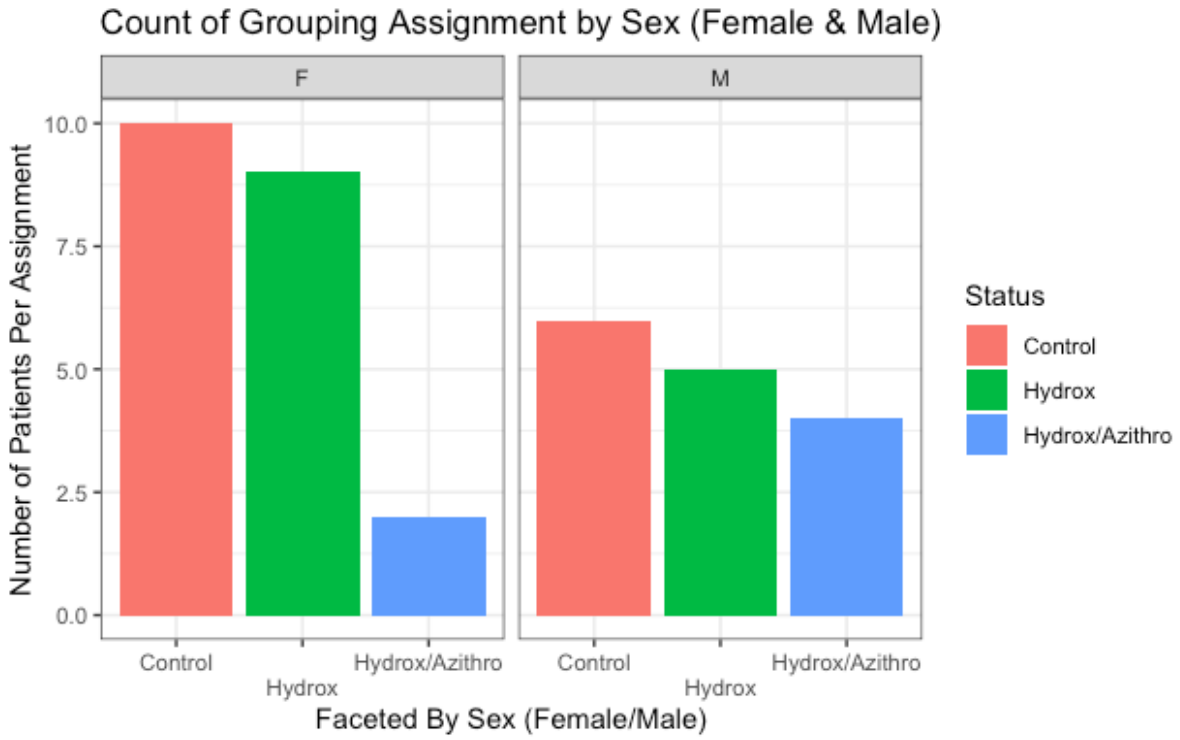
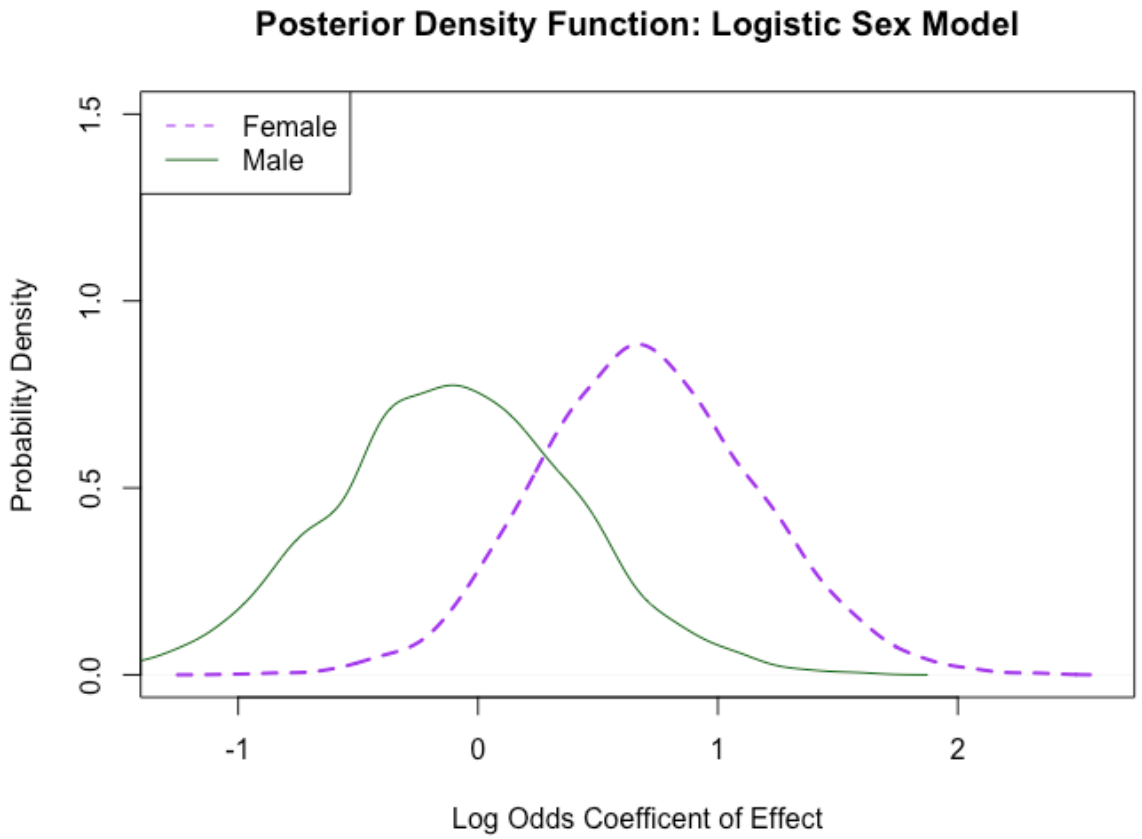


Figure 5:



	Median	MAD_SD
Male	-0.1067	0.5036
Female	0.8016	0.6790

	5%	95%
Male	-0.9458	0.7318
Female	-0.2892	1.9661

The model suggests that, for the entire dataset, women are more likely to test positive after six days than men within the same period. This trend is observed, in the first table, detailing the log-odds effect of being either Male or Female. However, and as we can see imagine from the large Mean Absolute Deviation and above visual, these credible intervals for the effect are extremely large. We, therefore, can't put too much weight in the log-odds effect.

T6 ~ Clinical Status

A majority of the patients from both the treatment and control groups consisted of individuals diagnosed with upper respiratory tract infection (URTI), but patients with lower respiratory tract infection (LRTI) outnumbered those who were asymptomatic only for the treatment group.[1] Those in the control set had more members showing no symptoms than those with LRTI. From the publication, clinical status held no significance towards the group, yet the disparate membership is worth another look. Two visuals for how the makeup of these patients are distributed among the control and treatment group can be seen below, as well as the predicted log-odds values for testing positive.

Figure 6:

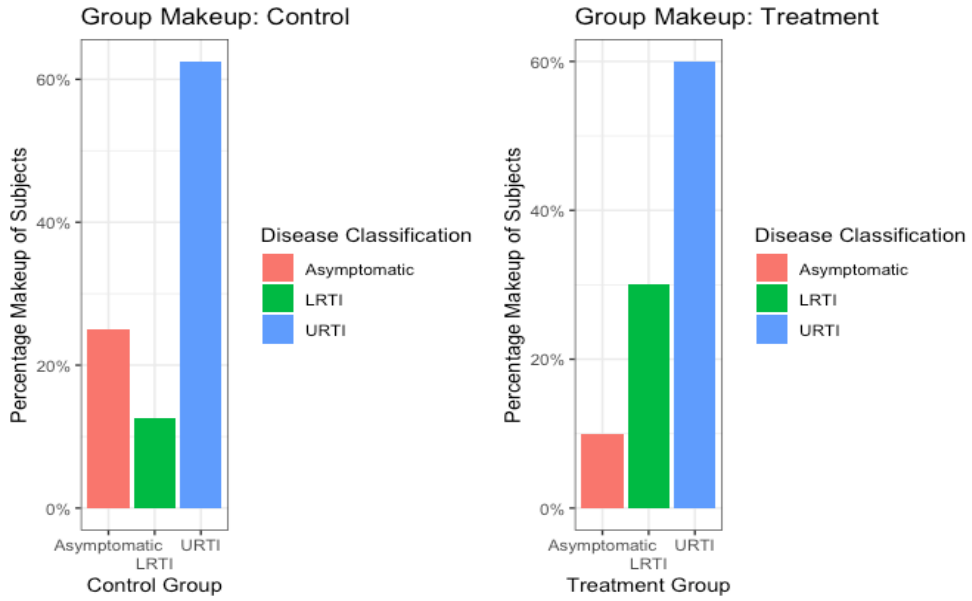


Figure 7:

[1] A data value under "status" featured a typo in one of the rows. "LRTI" was spelled "LTRI" and provided another bar in the corresponding chart, as a result, and the mistake was corrected in the original document.

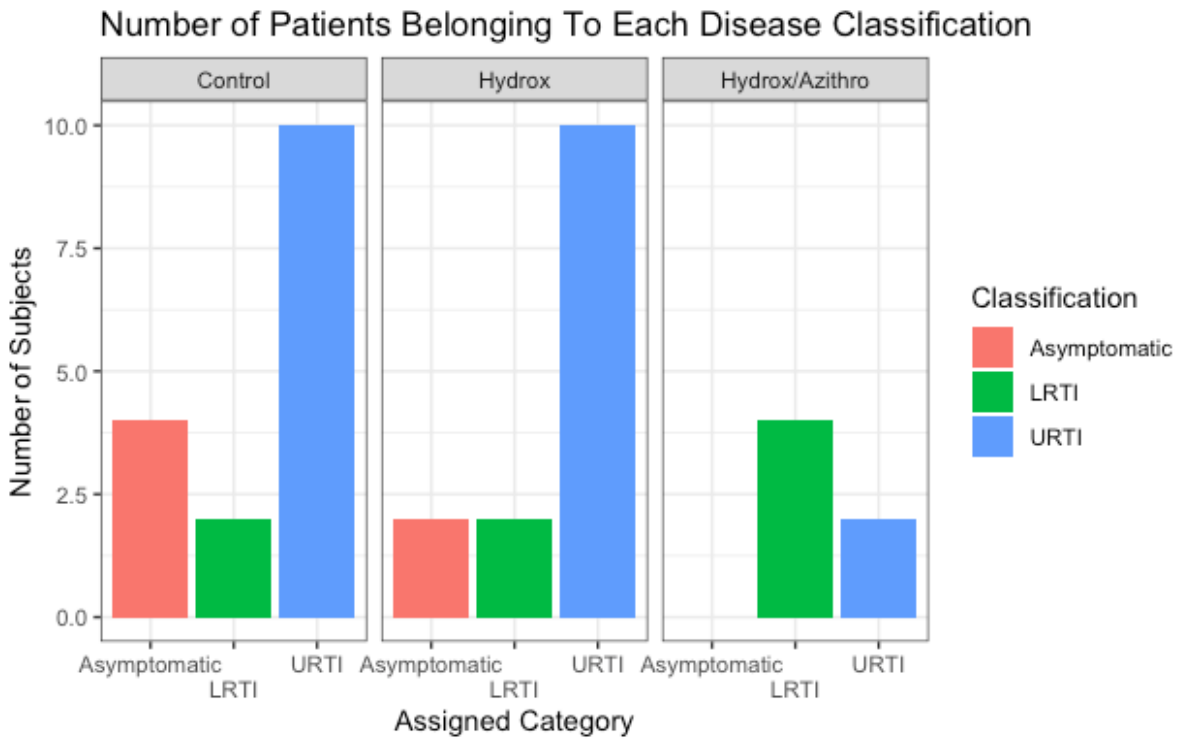
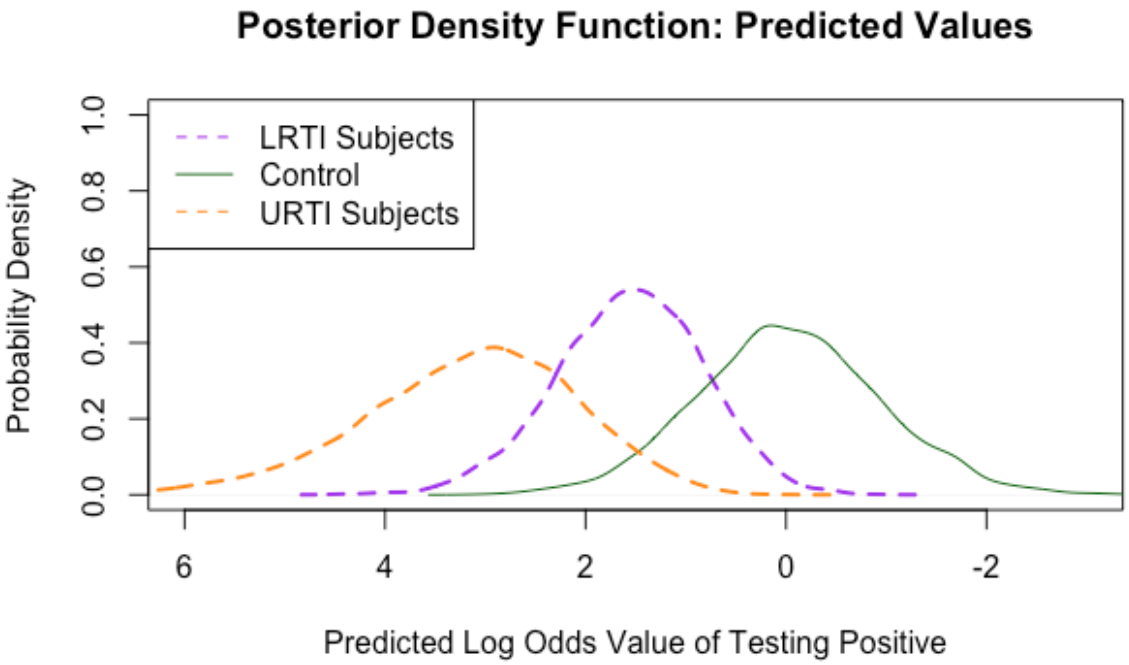


Figure 8:



Because clinical status was evaluated as a nominal variable in the original data set, a numerical equivalent denoted as "statCode" was created for the purpose of creating the corresponding model, where 0 = asymptomatic, 1 = LRTI, and 2 = URTI. Like before, parameter estimates of the regression model are given, along with 95% Bayesian CIs.

	Median	MAD_SD

Asymptomatic	-1.0880	0.7625
LRTI, HRTI = 2*LRTI	1.0018	0.5078

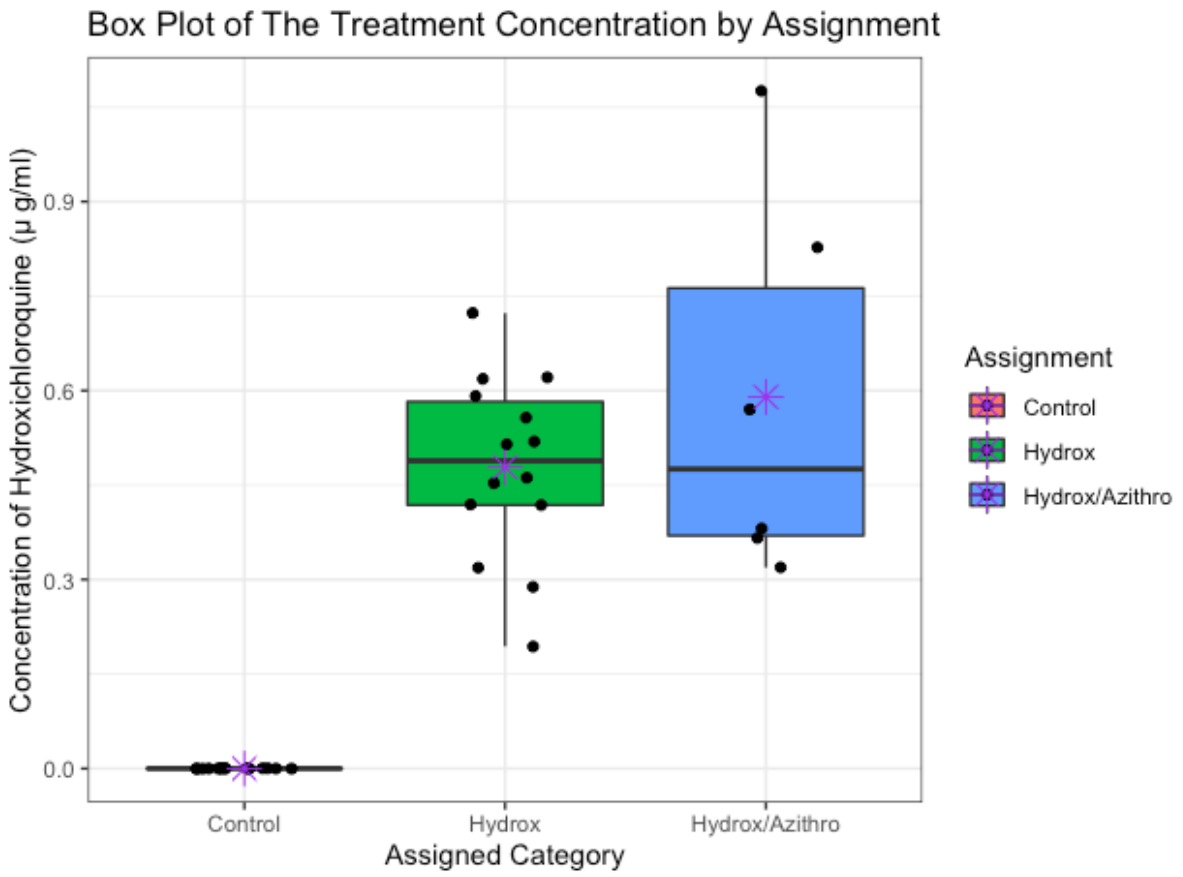
	Median	MAD_SD
Asymptomatic	-2.4867	0.1377
LRTI, HRTI = 2*LRTI	0.2254	1.8779

The main observation to take away from this should be that there is a 95% probability that our log-odds effect for our beta1 coefficient is positive. In the R markdown file, where these results came from, I explore this relationship further, by running the same type of regression while controlling for HCQ dose concentration. Still, we find that these results are significant.

t6 ~ Hydroxychloroquine Dosage

Patients that were part of the treatment group received azithromycin, or not, in spite of each of the aforementioned variables. Consequently, over-representation of certain factors can be observed from those who received varying concentrations of hydroxychloroquine, such as how patients given azithromycin had been given generally larger amounts of hydroxychloroquine in comparison to those that did not.

Figure 9: Interquartile Ranges for Dose Concentration of All Groups



Dose Concentration of Sample Set

	Mean ± SDs	Median	IQR
Hydrox. Only (N=14)	0.4784286 ± 0.1447575	0.4885	0.16425
Hydrox. and Azith. (N=6)	0.5898333 ± 0.3033357	0.4755	0.393

We can model the concentration dosage with day 6 test results to further observe the likelihood that the 100% virological removal of patients given both hydroxychloroquine and azithromycin may have been the result of other contributors.

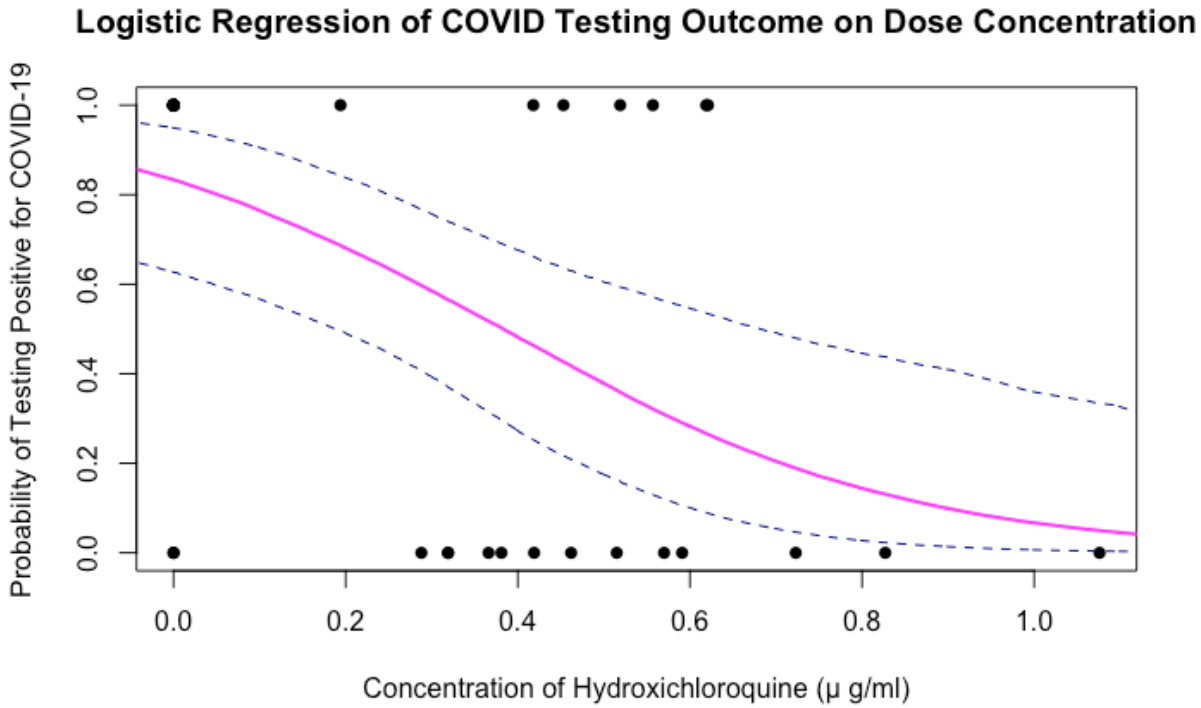
	Median	MAD_SD
(Intercept)	1.6095	0.5953
Concentration	-4.2069	1.4791

	5%	95%

(Intercept)	0.6699	2.6672
Concentration	-6.9077	-1.9394

The probabilistic interpretation of these coefficients is, once again, expanded on in the R markdown file. It is important to note, though, that these effects are not only significant but remain fairly large. Visually we can also see that Hydroxichloroquine dose concentration serves as a decently strong probabilistic indicator for COVID-19 test results, as it resembles a sigmoid curve.

Figure 10:



D6 ~ Variables and Group Designation

For more information on the effect of our variables on COVID-19 test results, we can use simple linear regression and treat our response variable as the SARS-CoV-2 RNA concentration, taken from nasopharyngeal samples. Having observed each of the previous variables as predictors for both treatment and control groups, the effect of whether a patient received hydroxychloroquine (HCQ) or hydroxychloroquine and azithromycin was modeled for the previous four models. All of the consequent results provided proof that using either treatment improved an individual's chance of testing negative after six days of inclusion, in which any given median value decreases between -.43 to -.38 for HCQ treatment and -.5 to -.45 for HCQ and azithromycin. For credible intervals, lower bounds for HCQ treatment decrease between -.85 to -.65 and upper bounds decrease between -.21 and .08. The lower and upper bounds of HCQ and azithromycin treatment featured greater decreases as well,

with values ranging from -.89 to -.78 and -.22 to -.14, respectively. Values corresponding to patients given both drugs were always less than those corresponding to HCQ treatment, azithromycin included or not, and we can accordingly attribute this inequality to either the proposed synergistic effect of HCQ and azithromycin or the greater HCQ dose concentration prescribed to patients that were also given azithromycin.

D6 ~ Age

	Median	MAD SD
(Intercept)	0.7149	0.1593
Age	0.0042	0.0033
ctrtCode	-0.4394	0.1555
atrtCode	-0.4773	0.1978

	5%	95%
(Intercept)	0.451500997	0.986734532
Age	-0.001285103	0.009804379
ctrtCode	-0.702270086	-0.177334154
atrtCode	-0.806010500	-0.144670144

D6 ~ Biological Gender

	Median	MAD SD
Male	0.8186	0.1308
Female	0.9125	0.1113
ctrtCode	-0.3817	0.1431
atrtCode	-0.4643	0.2022

	5%	95%
(Intercept)	0.6075671	1.0497516
Age	0.7259785	1.1043453
ctrtCode	-0.6362546	-0.1364318

atrtCode	-0.7860098	-0.1242547

D6 ~ Clinical Status

	Median	MAD SD
Asymptomatic	0.4695	0.1518
LRTI	0.9716	0.1710
URTI	1.0137	0.1049
ctrtCode	-0.4299	0.1307
atrtCode	-0.5544	0.1973

	5%	95%
Asymptomatic	0.2226787	0.7350301
LRTI	0.6918225	1.2712229
URTI	0.8396167	1.1874672
ctrtCode	-0.6480096	-0.2104125
atrtCode	-0.8887301	-0.2275514

D6 ~ HCQ Concentration

	Median	MAD SD
(Intercept)	0.8757	0.1048
Concentration	0.0143	0.4859
ctrtCode	-0.3837	0.2750
atrtCode	-0.4953	0.2116

	5%	95%
(Intercept)	0.7027349	1.04745377
Concentration	-0.7646952	0.79543712
ctrtCode	-0.8293836	0.07808454

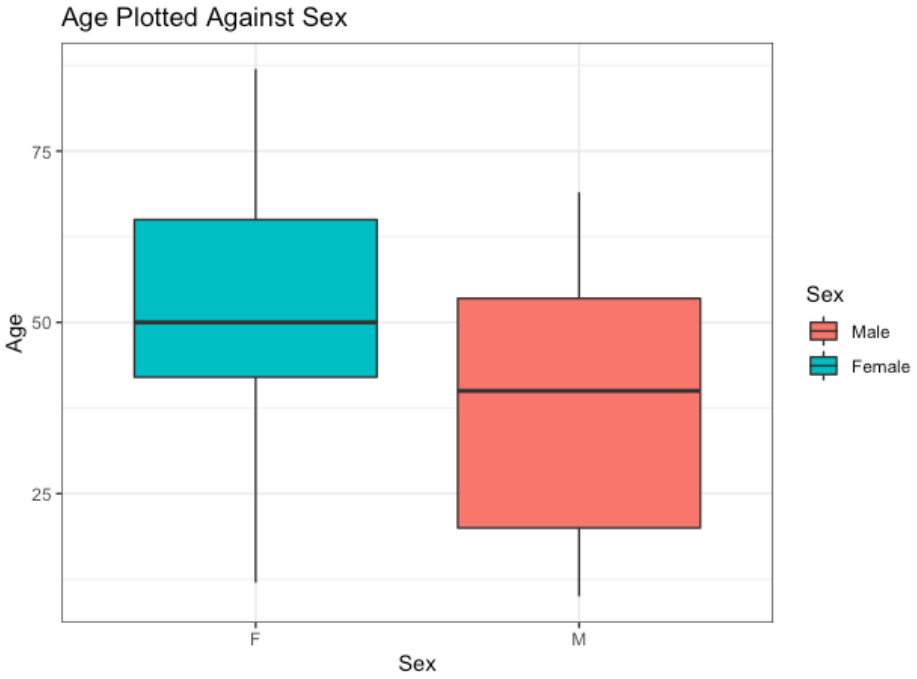
atrtCode	-0.8430710	-0.15324362

Potential Variable Relationships

To conclude our regression analysis of the data, we chose to examine relationships between two chosen variables to see the extent of the aforementioned over-representation of a given nominal or numerical value

Age versus Gender

Plotting age versus biological gender suggests that female patients were generally older than their male peers.

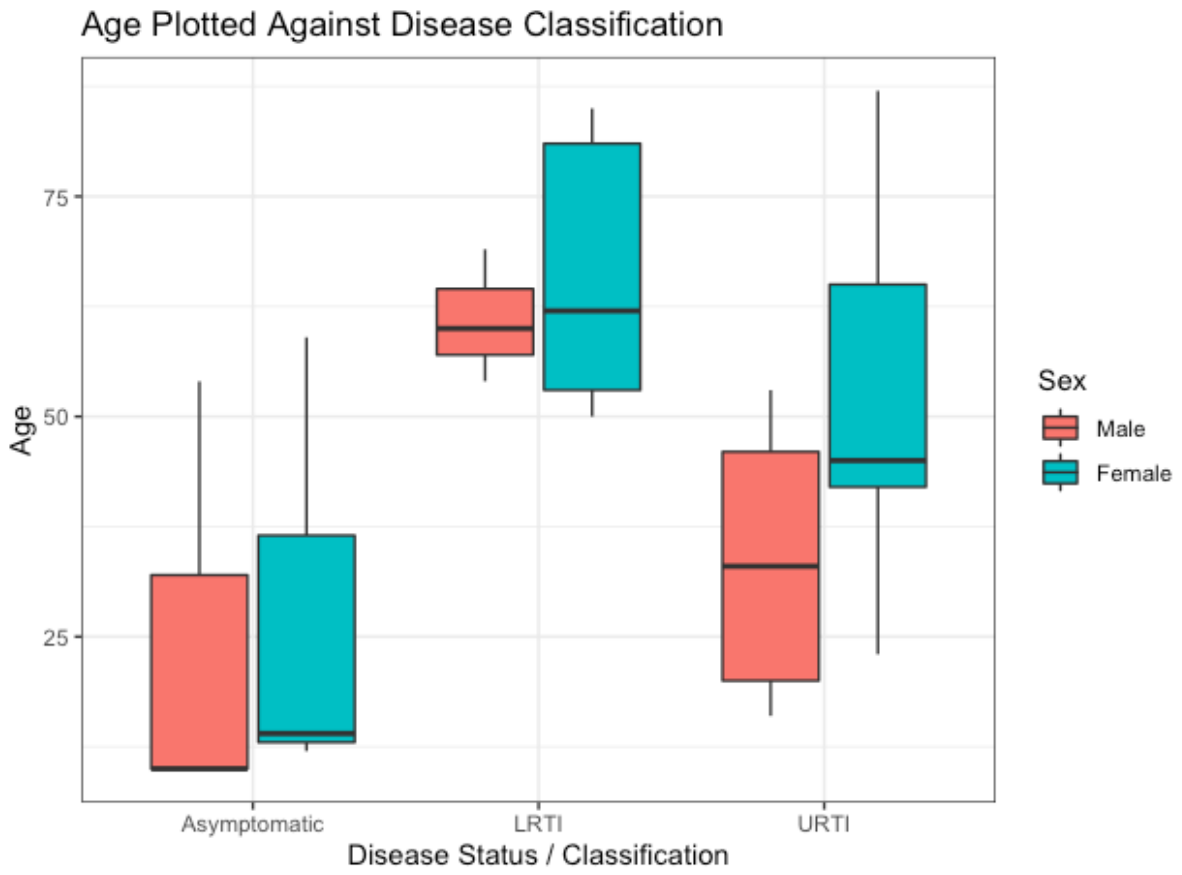


Age versus Status

The oldest patients were diagnosed with LRTI, followed by URTI and no symptoms thereafter.

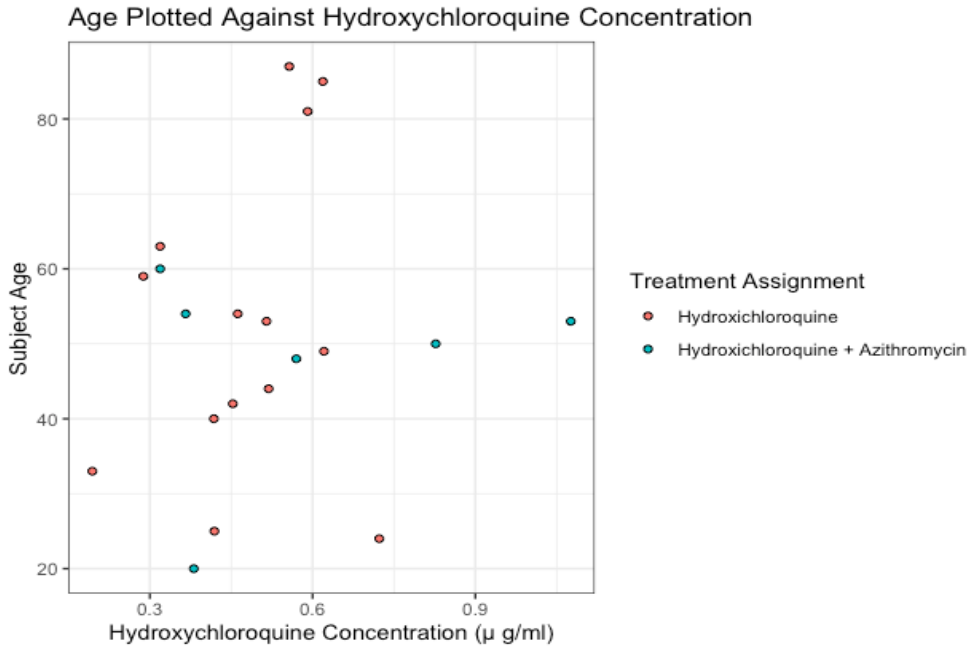
Gender versus Status

Male patients appear to represent the majority of patients with either URTI or LRTI, while an even distribution of both genders occurs for asymptomatic individuals.



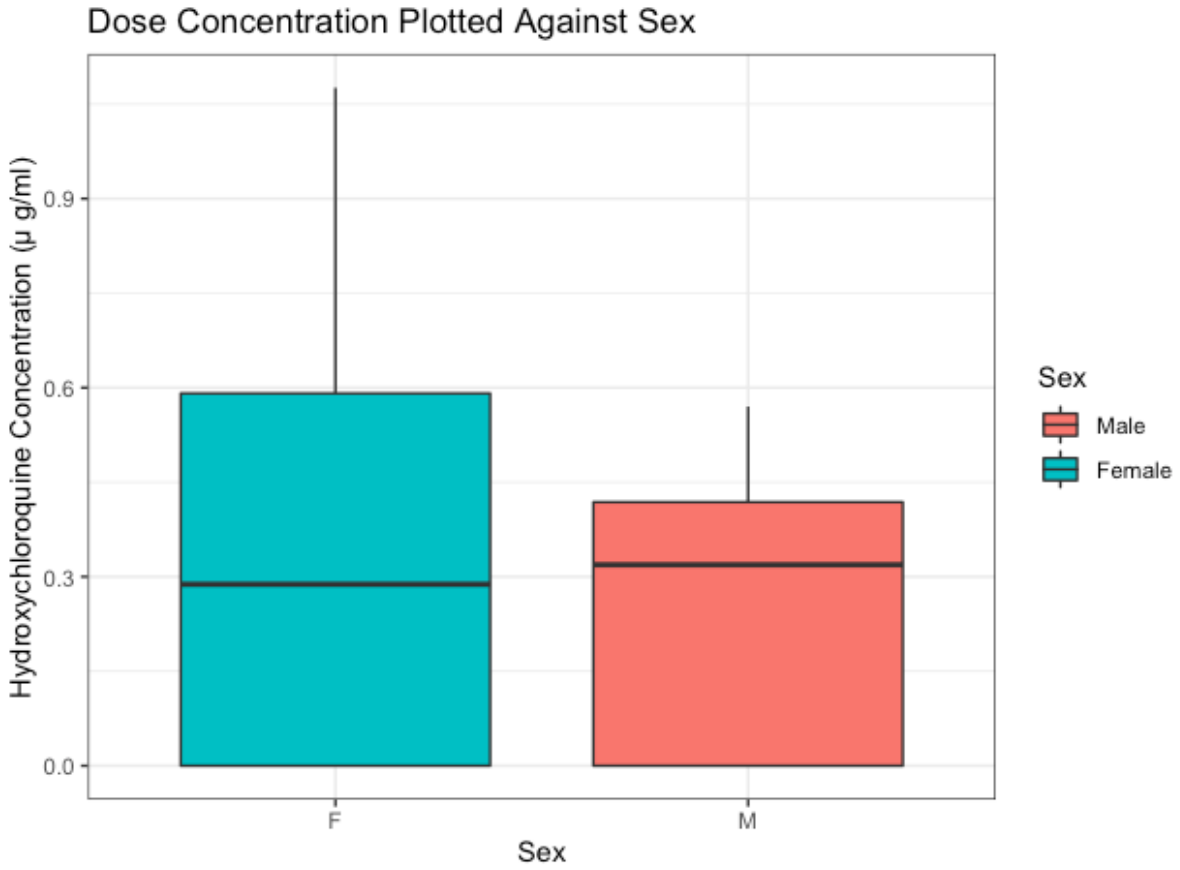
Age versus Concentration

A nebulous pattern, or lack of any trends, appears in the scatterplot of the sample set's age and HCQ concentration values. These two variables appear to have the best distribution for research.



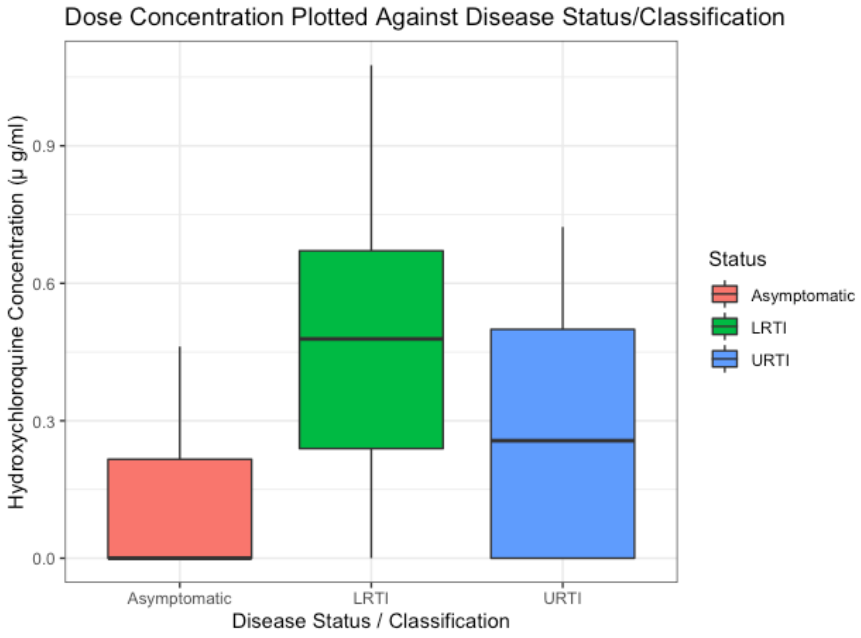
Gender versus Concentration

Female patients received greater concentrations than their male counterparts.



Status versus Concentration

Patients with LRTI were given the greatest concentration dosages of HCQ, followed by patients with URTI and patients that were asymptomatic.



Observations/Interpretations

Overall, the paper does a decent job of illustrating the effects of hydroxychloroquine and azithromycin on reducing PCR results. While the study takes into account some simple variables for each person, it may be beneficial to expand the study to include more variables and examine the effects that those variables have on the treatment. To do this, the team might try to account for other factors by increasing the size of the sample group. Since the group that was treated with both drugs only included six people, a larger sample group will illustrate a more accurate representation of the efficacy of the drug. We should also be skeptical of the relatively large amount of attrition bias in the study.

Another factor the team might want to consider is continuing the trial past day six. The paper discludes data for days after six. While this may be the primary endpoint, it does not show the true effect of the drug over time since some might retest positive after the primary endpoint. This occurred once for the six people in the group treated by both drugs. The final results, however, seem to be significant and warrant further research into the efficacy of these drugs.

Recreation of the results to the best of our abilities (in LaTeX):

		M $\pm SD[AGE]$	t	p	Male-count	p	Asymp	URTI	LRTI	p	M$\pm SD[INCLUS]$	t	p
1	Control(N=16)	37 $\pm$24			6		4	10	2		3.9 $\pm$2.81		
2	Hydroxychloroquine(N=20)	51 $\pm$19	-1.9	.06	9	.66	2	12	6	.31	4.06 $\pm$2.62	-.14	.88
3	All(36)	45.06 $\pm$22.0			15		6	22	8		4 $\pm$2.6		

		Neg/Pos3	%	p3	Neg/Pos4	%	p4	Neg/Pos5	%	p5	Neg/Pos6	%	p6
1	Hydroxychloroquine(N=20)	10/20	50%		12/20	60%		13/20	56%		14/20	70%	
2	Control(N=16)	1/16	6.3%	.003	4/16	25%	0.04	3/16	18.8%	.004	2/16	12.5%	.0001

References

"Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial", Philippe Gautret

Meyerowitz, Eric A, et al. "Rethinking the Role of Hydroxychloroquine in the Treatment of COVID-19." FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology, John Wiley and Sons Inc., May 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7267640/.

Is Chloroquine or Hydroxychloroquine Useful in Treating People with COVID-19, or in Preventing Infection in People Who Have Been Exposed to the Virus?, 25 Mar. 2021, www.cochrane.org/news/chloroquine-or-hydroxychloroquine-useful-treating-people-covid-19-or-preventing-infection.

"WHO 'Strongly' against Hydroxychloroquine Use for COVID-19 Prevention." Medical News Today, MediLexicon International, www.medicalnewstoday.com/articles/who-strongly-against-hydroxychloroquine-use-for-covid-19-prevention.