Stat 139 Project Final Paper

Group Members: Alex Cheng, Eton Shon, and Erin Yuan

Project Title: Risk Factors and Severity Outcomes for Post-Vaccination COVID-19 Infections: A Case Study of Yucatan, Mexico

I. Introduction

The first cases of the coronavirus disease 2019 (COVID-19), the infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), were reported in late 2019. Since then, the disease has spread worldwide and was declared to be a pandemic by the World Health Organization (WHO) in March 2020.[1] Research on vaccine development against the disease has been ongoing by governments, nonprofit institutions, private companies, and universities and over thirty vaccines have been approved for use worldwide with over thirteen billion vaccine doses given globally.[2]

The toll of the pandemic felt by countries around the world has greatly varied based on income status and vaccine access.[3] Since the WHO declared Latin America to be the new COVID-19 epicenter in May 2020,[4] Mexico is one country that has been greatly affected.[5] Mexico is an important area for analysis especially because its government health authority, the Federal Commission for Protection against Health Risks (COFEPRIS), offers one of the most diverse selections of vaccines in the world[6], yet there remains a lack of literature on vaccination and the impact of COVID-19 on its population. In particular, the scope of this paper centers on Mexico's

---

[1] WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2022. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[2] Mathieu E, Ritchie H, Rodés-Guirao L, et al. Coronavirus Pandemic (COVID-19). *Our World in Data*. Published online March 5, 2020. https://ourworldindata.org/covid-vaccinations

[3] Ning C, Wang H, Wu J, Chen Q, Pei H, Gao H. The COVID-19 Vaccination and Vaccine Inequity Worldwide: An Empirical Study Based on Global Data. *Int J Environ Res Public Health*. 2022;19(9):5267. doi:10.3390/ijerph19095267

[4] WHO says the Americas are new COVID-19 epicenter as deaths surge in Latin America. *Reuters*. https://www.reuters.com/article/us-health-coronavirus-latam-idUSKBN2322G6. Published May 26, 2020. Accessed December 10, 2022.

[5] Ibarra-Nava I, Cardenas-de la Garza JA, Ruiz-Lozano RE, Salazar-Montalvo RG. Mexico and the COVID-19 Response. *Disaster Med Public Health Prep*.:1-2. doi:10.1017/dmp.2020.260

[6] Exteriores S de R. Mexico has one of the broadest vaccine portfolios in the world. gob.mx. http://www.gob.mx/sre/prensa/mexico-has-one-of-the-broadest-vaccine-portfolios-in-the-world

southeastern Yucatan peninsula because it continued to see a large tourism presence[7] and remained highly connected to the U.S. and worldwide. Thus, Yucatan serves as a complex setting for its viral dynamics in terms of importing new variants and contributing to subsequent spread.

Three main types of vaccines have been in use in Mexico and are the focus here: mRNA vaccines, viral vector vaccines, and protein-based vaccines.[8] While all vaccines work by triggering the immune system to recognize a part of the viral protein to produce antibodies against the invader, specific vaccines differ in their methods of delivering this effect. mRNA vaccines, which work by introducing parts of the viral genome that encodes the viral protein into cells, include the Pfizer-BioNTech Comirnaty (BNT162b2) and the Moderna Spikvax (mRNA-1279), with among the highest respective efficacies of 91%[9] and 93-98%.[10] Next, viral vector vaccines, which use a harmless virus as a vector to enter cells, include the Oxford/AstraZeneca Vaxzevria (ChAdOx1-S or AZD1222) with 75% efficacy,[11] the CanSino Convidecia (Ad5-nCoV) with 58-92% efficacy,[12] the Gamaleya Sputnik V vaccine with 75-92% efficacy,[13] and the single-dose Janssen/Johnson & Johnson (Ad26.COV2.S) vaccine with 72% efficacy.[14] The last broad category features protein-based and inactivated vaccines that supply parts of the virus for the immune response to recognize, including the Novavax vaccine

---

[7] Tourism statistics (January - August 2021). https://embamex.sre.gob.mx/eua/index.php/en/2016-04-09-20-40-51/tourism/1762-tourism-statistics-2

[8] Zimmer C, Corum J, Wee SL, Kristoffersen M. Coronavirus Vaccine Tracker. *The New York Times*. https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html. Published June 10, 2020. Accessed December 10, 2022.

[9] Six Month Safety and Efficacy of the BNT162b2 mRNA COVID-19 Vaccine | medRxiv. https://www.medrxiv.org/content/10.1101/2021.07.28.21261159v1

[10] El Sahly HM, Baden LR, Essink B, et al. Efficacy of the mRNA-1273 SARS-CoV-2 Vaccine at Completion of Blinded Phase. *New England Journal of Medicine*. 2021;385(19):1774-1785. doi:10.1056/NEJMoa2113017

[11] The Oxford/AstraZeneca (ChAdOx1-S [recombinant] vaccine) COVID-19 vaccine: what you need to know. https://www.who.int/news-room/feature-stories/detail/the-oxford-astrazeneca-covid-19-vaccine-what-you-need-to-know

[12] Halperin SA, Ye L, MacKinnon-Cameron D, et al. Final efficacy analysis, interim safety analysis, and immunogenicity of a single dose of recombinant novel coronavirus vaccine (adenovirus type 5 vector) in adults 18 years and older: an international, multicentre, randomised, double-blinded, placebo-controlled phase 3 trial. *The Lancet*. 2022;399(10321):237-248. doi:10.1016/S0140-6736(21)02753-7

[13] Russia's Vaccine Is Safe and Effective, Published Study Shows - The New York Times. https://www.nytimes.com/2021/02/02/world/europe/russia-vaccine-safe-effective.html

[14] The Janssen Ad26.COV2.S COVID-19 vaccine: What you need to know. https://www.who.int/news-room/feature-stories/detail/the-j-j-covid-19-vaccine-what-you-need-to-know

(NVX-CoV2373) with 90% efficacy,[15] the Chinese Sinopharm (BBIBP-CorV) vaccine at 78% efficacy,[16] and the SinoVac CoronaVac vaccine with 84% efficacy.[17]

Despite the protection afforded by vaccines against infection, severe cases requiring hospitalization, and death, some individuals may still become infected with SARS-CoV-2 even after vaccination.[18] Known as breakthrough infections, these cases remain an active area of study in order to further understand vaccine efficacy, identify populations most at risk for infection, and inform public health policy regarding additional doses and the need for updated vaccine formulations. As a result, the purpose of this paper is to 1) describe risk factors associated with post-vaccination SARS-CoV-2 infection and compare these to risk factors in unvaccinated individuals, 2) assess illness severity and symptom profiles in vaccinated vs. unvaccinated individuals with SARS-CoV-2 infection, and 3) compare the effectiveness of different vaccine types against hospitalization given infection.

II. Data and EDA

The data comes from a study conducted by the U.S. National Institutes of Health (NIH) in collaboration with the Autonomous University of Yucatan (UADY). The study population is defined as patients with COVID-19-like symptoms residing in the Yucatan peninsula who were reported to have visited a healthcare institution in the public or private sector of the Mexican healthcare system between January 1st, 2021 and February 1st, 2022. The study population had no travel history in the past 14 days, no prior SARS-CoV-2 infection, and were not pregnant. The variables in the data set include (for $n = 134,679$ patients):

- **`gender`:** Gender of the patient: `female` = female, `male` = male
- **`age`:** Age of patient
- **`location`:** Residency in Yucatan divided into 3 jurisdictions (“counties”), 1 = around Merida (capital), 2 = around Tizimin, 3 = around Ticul
- **`indigenous`:** Contact with indigenous populations, 0 = no contact, 1 = contact
- **`job`:** Occupation: `HCW` = healthcare worker, `home` = at home (homemaker, retired, or unemployed), `office` = professionals in office settings, `school` = education-related settings (teachers and students), `tradesmen` = trade-based occupations (farmers,

---

[15] Efficacy and Safety of NVX-CoV2373 in Adults in the United States and Mexico | NEJM. https://www.nejm.org/doi/full/10.1056/NEJMoa2116185

[16] Wee SL. China’s Vaccine Diplomacy Just Got a Big Win. But Can the Country Deliver? *The New York Times*. https://www.nytimes.com/2021/05/07/business/economy/china-sinopharm-vaccine-who.html. Published May 7, 2021.

[17] Tanriover MD, Doğanay HL, Akova M, et al. Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): interim results of a double-blind, randomised, placebo-controlled, phase 3 trial in Turkey. *The Lancet*. 2021;398(10296):213-222. doi:10.1016/S0140-6736(21)01429-X

[18] Birhane M, Bressler S, Chang G, et al. COVID-19 Vaccine Breakthrough Infections Reported to CDC — United States, January 1–April 30, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(21):792-793. doi:10.15585/mmwr.mm7021e3

drivers, food market workers, factory workers, construction workers), `other` = other occupations

- **`medication_hiv`:** Patient on antiretroviral treatment for human immunodeficiency virus (HIV): 0 = no, 1 = yes
- **`medication_flu`:** Patient taking ibuprofen for influenza-like illness: 0 = no, 1 = yes
- **`symptoms_fever`:** Fever or chills, binary outcome: 1 = yes, 0 = no
- **`symptoms_cough`:** Cough, binary outcome: 1 = yes, 0 = no
- **`symptoms_throat`:** Sore throat, binary outcome: 1 = yes, 0 = no
- **`symptoms_respiratory`:** Difficulty breathing, chest pain, rapid respiration, or cyanosis (skin discoloration), binary outcome: 1 = yes, 0 = no
- **`symptoms_anxiety`:** Irritabilities, binary outcome: 1 = yes, 0 = no
- **`symptoms_aches`:** Muscle aches, joint stiffness, malaise, or headache, binary outcome: 1 = yes, 0 = no
- **`symptoms_eyenose`:** Runny nose or pink eye, binary outcome: 1 = yes, 0 = no
- **`symptoms_GI`:** Abdominal pain, vomiting, or diarrhea, binary outcome: 1 = yes, 0 = no
- **`symptoms_senses`:** Partial or full loss of smell or taste, binary outcome: 1 = yes, 0 = no
- **`comorb_heart`:** Cardiovascular disease or hypertension, binary outcome: 1 = yes, 0 = no
- **`comorb_lung:** Chronic obstructive pulmonary disease (COPD), smoking, or asthma, binary outcome: 1 = yes, 0 = no
- **`comorb_immuno`:** Immunocompromised, HIV/AIDS, or chronic renal insufficiency or chronic kidney disease (CKD), binary outcome: 1 = yes, 0 = no
- **`comorb_endocrine`:** Diabetes, binary outcome: 1 = yes, 0 = no
- **`comorb_other`:** Obesity or other conditions, binary outcome: 1 = yes, 0 = no
- **`status`:** Patient hospital status: `outpatient` = no overnight stay, `inpatient` = overnight stay
- **`outcome`:** Patient outcome status: `recovered` = discharged after hospitalization, `death` = death
- **`diagnosis`:** Diagnosis by medical professional: `ILI` = influenza-like illness, `SARI` = severe acute respiratory infection
- **`flu`:** Influenza test status: `positive` = confirmed positive by laboratory test, `suspected` = suspected of having influenza without confirmation from laboratory test, `negative` = confirmed negative by laboratory test
- **`covid`:** COVID-19 test status: `positive` = first positive confirmed by RT-PCR or rapid antigen test, `negative` = negative confirmed by RT-PCR or rapid antigen test
- **`vaccine`:** Vaccine dosage: `complete` = complete series (2 doses for AstraZeneca, Gamaleya, Moderna, Novavax, Pfizer, Sinopharm, SinoVac; 1 dose for CanSino, Johnson & Johnson), `incomplete` = incomplete series (only 1 dose of 2 doses for AstraZeneca, Gamaleya, Moderna, Novavax, Pfizer, Sinopharm, SinoVac), `none` = no vaccine received
- **`vaccine_name`:** Vaccine type received: `mRNA` = Pfizer or Moderna; `Non-replicating viral vector` = AstraZeneca, CanSino, Gamaleya, or Johnson & Johnson; `Whole or protein` = NovaVax, Sinopharm, or SinoVac

- **`vaccine_time`:** Difference between date of last dose of vaccination to date of testing positive or negative for COVID-19

Study Design

To narrow the scope of the analysis to the research question on the impact of vaccination formulation in breakthrough infections, we split the data into cases and controls to implement a case-control analysis. Breakthrough cases were defined as patients who tested positive after receiving a complete or incomplete vaccine dosage (case 1, $n = 10,355$). Case 2 consisted of unvaccinated patients who tested positive ($n = 38,456$), control 1 was vaccinated patients who tested negative after receiving a complete or incomplete vaccine dosage ($n = 19,317$), and control 2 was unvaccinated patients who tested negative ($n = 66,549$). We compared case 1 to control 1 and case 2 to control 2 to identify risk factors associated with post-vaccination SARS-CoV-2 infection compared to unvaccinated SARS-CoV-2 infection. To understand disease outcomes in unvaccinated and vaccinated COVID-19 positive individuals, we compared case 1 to case 2. We removed observations with invalid or unknown COVID-19 test results and unknown vaccine status and grouped symptoms and comorbidities together for more predictive power using cluster analysis.

EDA

To inspect trends between variables, we began with an exploratory data analysis (EDA) to compare vaccines to COVID-19 outcomes. Most patients in Yucatan who visited the healthcare system with COVID-like symptoms did not have any vaccination history, while the next most popular vaccines were viral vector, mRNA, and lastly whole or protein (Figure 1a). 30.3% of patients who received an mRNA vaccine tested positive for COVID-19, which was lower than the positivity rate of 36.3% for patients with a viral vector vaccine, 36.6% for no vaccine, and 40% for whole or protein vaccines (Figure 1b).

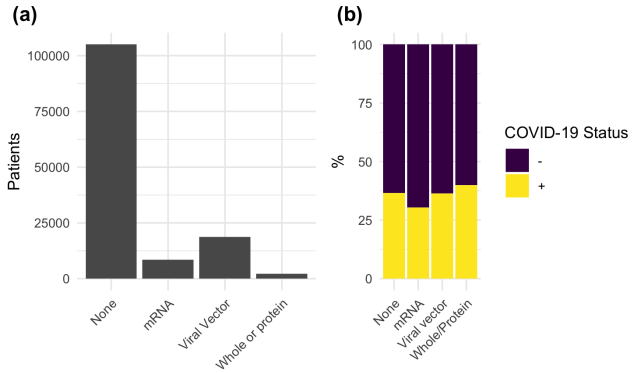

Figure 1: (a) Distribution of vaccines received by the study population for $n = 134,409$ patients with reported vaccine data. (b) Proportions of COVID-19 positive and negative outcomes for patients who received no vaccine ($n = 105,039$), an mRNA vaccine ($n = 8,437$), a viral vector vaccine ($n = 18,723$), and whole or protein vaccines ($n = 2210$).

Demographics and outcomes varied between patients in both cases and controls (Table 1). Patients with breakthrough infections were represented by both sexes, had a median age of 42, lived mostly near the capital of Merida, had the most common comorbidity of heart conditions, did not have contact with indigenous populations, and were not taking flu medication (Figure 2). In terms of outcomes for vaccinated patients, still some had to be hospitalized, passed away, and were diagnosed with severe respiratory disease, but seemed to experience less symptoms (Figure 3).

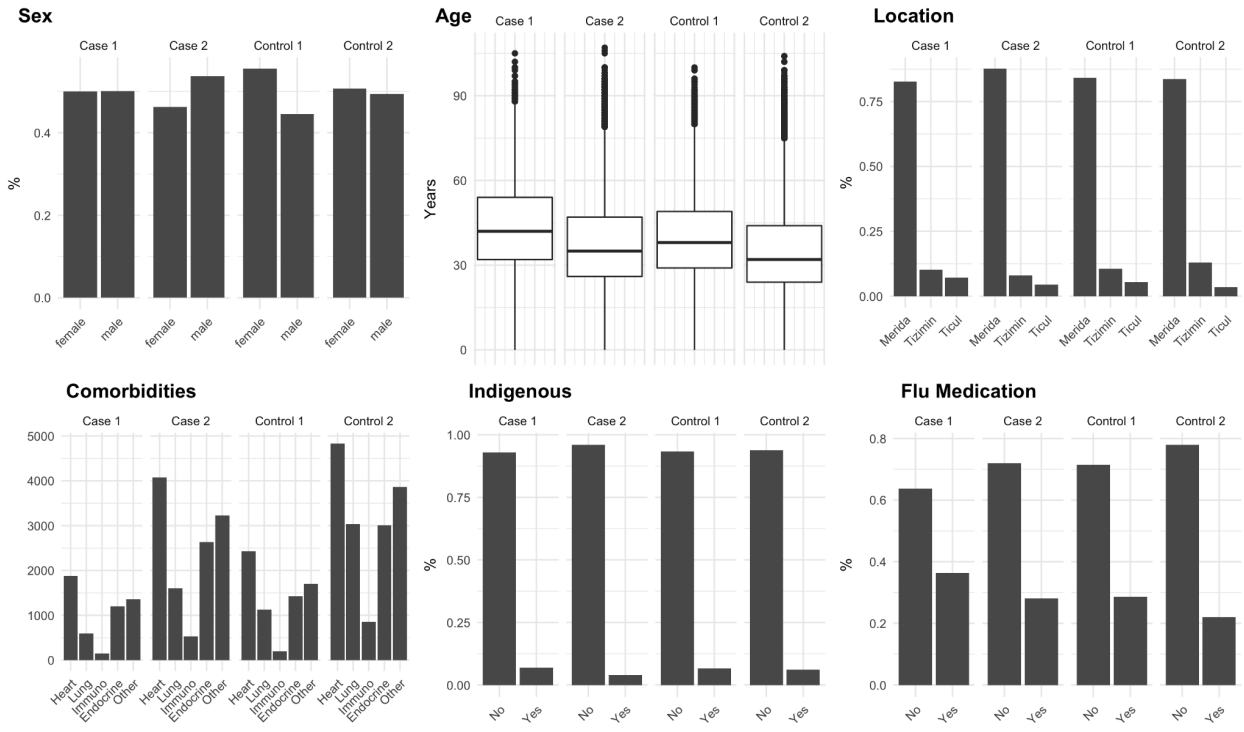

Figure 2: Collected demographic information on patients who visited a healthcare institution in Yucatan, Mexico.

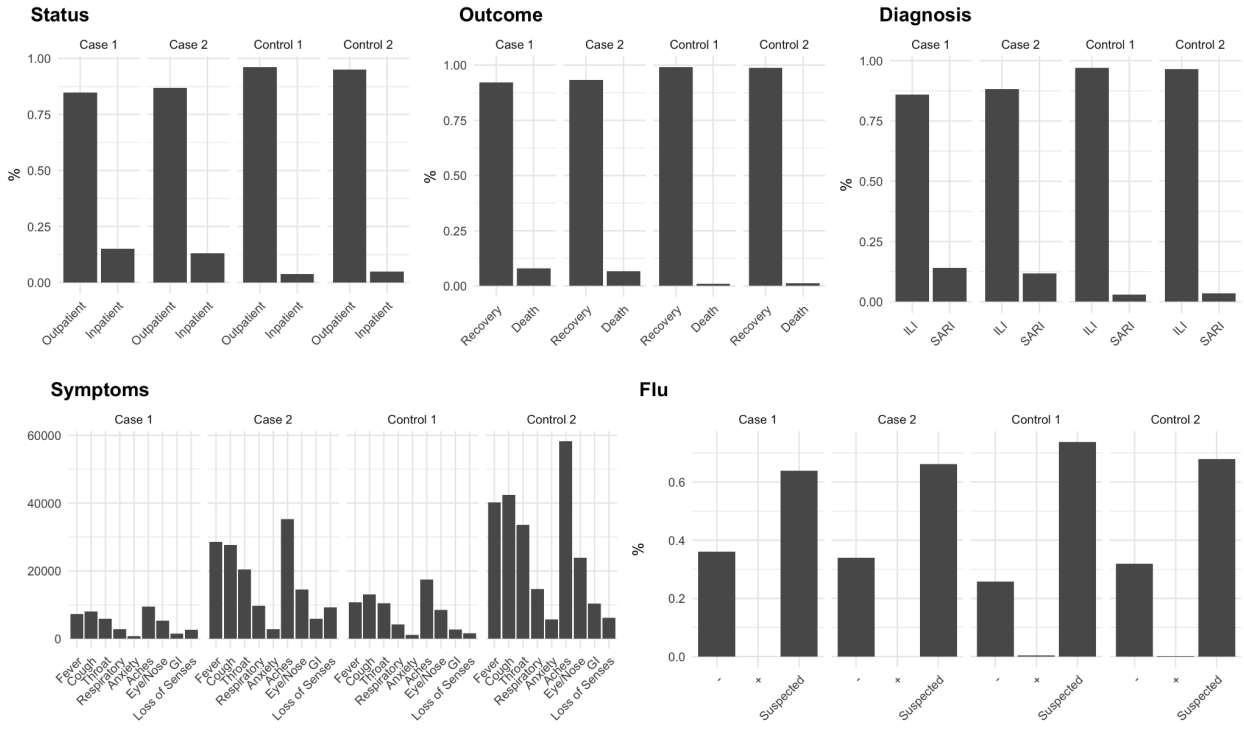

Figure 3: Measured outcomes for patients who visited a healthcare institution with COVID-like symptoms in Yucatan, Mexico.

```{r, echo = FALSE, eval = TRUE}
# split into case 1, case 2, control 1, and control 2
clean_comps <- clean %>%
  select(-system, -id, -seq_id, -sequenced) %>%
  mutate(comps = ifelse((vaccine == "complete" | vaccine == "incomplete") & covid == "positive",
"case1",
                        ifelse(vaccine == "none" & covid == "positive", "case2",
                        ifelse((vaccine == "complete" | vaccine == "incomplete") & covid == "negative",
"control1",
                        "control2")))) %>%
  mutate(comps = factor(comps))
```

```{r, eval = TRUE, echo = FALSE, message = FALSE, warning = FALSE}
table1 = clean_comps %>%
  mutate(num_comorbs = as.numeric(num_comorbs)) %>%
  tbl_summary(
        by = comps, # split table by group
        missing = "no" # don't list missing data separately
  ) %>%
  add_n() %>% # add column with total number of non-missing observations
```

```
  add_p() %>% # test for a difference between groups
  modify_header(label = "**Variable**") %>% # update the column header
  bold_labels()
table1
```

Table 1: Characteristics of study participants in case 1, case 2, control 1, and control 2.

III. Methods

Logistic Regression Models

We used multivariable logistic regression models to identify the risk factors for post-vaccination COVID-19 infection and understand the disease severity based on vaccination status. We split the data into three case-control studies conditioning on COVID-19 positive status, vaccination, and no vaccination, respectively. Several methods to develop an inferential model were tested, and the best performing model was selected for interpretation based on Akaike information criterion (AIC). An initial baseline logistic model was generated from all predictors in the data, along with a backward selection model, a tuned least absolute shrinkage and selection operator (LASSO) model using cross validation, and a tuned ridge model using cross validation. After selecting the model with the lowest AIC in each case-controlled study, a new OLS logistic regression model was generated using the coefficients of that model for interpretability (Supplementary Information A).

Predicting Hospitalization from Vaccine Type

One useful question is to predict the chances of a patient having a severe case, as indicated by an overnight hospital stay, based on different vaccine types. We conducted a secondary analysis using Bayes' Rule to compare the probability of hospitalization from three different vaccine types—mRNA, viral vector, and whole or protein—given a COVID-19 infection. Bayesian inference is necessary because the target probability is unobtainable from the study data since it is conditioned on patients who visited the hospital and does not accurately generalize to all Yucatan residents who have received a given vaccine, so previous methods of logistic regression also do not apply to answer this question. Bayes' rule follows:

(1) $$P(\text{Hosp. with COVID-19|No Vaccination}) = \frac{P(\text{No Vaccination|Hosp. with COVID-19})P(\text{Hosp. with COVID-19})}{P(\text{No Vaccination})}$$

(2) $$P(\text{Hosp. with COVID-19|mRNA}) = \frac{P(\text{mRNA|Hosp. with COVID-19})P(\text{Hosp. with COVID-19})}{P(\text{mRNA})}$$

(3) $$P(\text{Hosp. with Covid|Viral Vector}) = \frac{P(\text{Viral Vector|Hosp. with COVID-19})P(\text{Hosp. with COVID-19})}{P(\text{Viral Vector})}$$

(4) $$P(\text{Hosp. with Covid|Whole or Protein}) = \frac{P(\text{Whole or Protein|Hosp. with COVID-19})P(\text{Hosp. with COVID-19})}{P(\text{Whole or Protein})}$$

In the formulas above, $P(\text{Hosp. with COVID-19|Vaccine type})$ is defined as the probability of hospitalization given the patient has taken a certain vaccine, $P(\text{Vaccine type|Hosp. with COVID-19})$ is defined as the probability of the patient having taken a certain vaccine given they have been hospitalized with COVID-19, $P(\text{Vaccine Type})$ is defined as the probability of a random person in Yucatan having taken a certain vaccine type, and $P(\text{Hosp. with COVID-19})$ is defined as the probability of a person in Yucatan being hospitalized with COVID-19. Note that our goal is to identify significant differences between the three vaccine types groups in terms of relative performance in protecting against a hospital stay, so $P\text{(Hosp. with COVID-19)}$ is neglected from further calculations and treated as a constant. While this quantity could be estimated using daily cases in Yucatan, given that there were only 140,000 cases of COVID-19 and the population of Yucatan is close to 2.3 million, the rate of hospitalization with COVID-19 would be so small that the relative difference between the probabilities for different vaccines would be uninformative. Keeping the probability of hospitalization as a constant also prevents us from further bias given that this probability is impossible to estimate from the dataset and is not available from Yucatan-wide data collection efforts. We estimate the probabilities of receiving each vaccine type conditioned on being hospitalized directly from our dataset (Table 2).

Table 2: Distribution of vaccine types conditioned on hospitalized with COVID-19. Of the patients, the majority have no vaccination (76.9%), followed by non-replicating viral vector vaccines (15.0%), followed by mRNA (5.19%) and lastly whole or protein vaccines (2.91%).

```{r}
test = subset(clean,covid == "positive" )
hospitalized = subset(test, status == "inpatient")
vaccine_table_hosp = table(hospitalized$vaccine_name)
prop_hosp_vac = vaccine_table_hosp/sum(vaccine_table_hosp)

rownames = c("Probability")
colnames = c("None", "mRNA", "Non-replicating Viral Vector", "Whole or Protein")
table_1 = matrix(prop_hosp_vac, 1, 4, dimnames = list(rownames, colnames))
table_1

```

To find the overall probability of receiving each vaccine type, we approximate this value from the proportions of vaccine types among patients who tested negative for COVID-19 (Table 3). This method reduces bias related to patients because a certain vaccine type may be overrepresented in the overall distribution if it leads to more hospitalizations (for example, if the viral vector vaccine was significantly less effective, there could be more patients in the hospital

with that particular vaccine, but it may not represent the general population). We therefore achieve a less skewed distribution for the proportion of vaccine types received by patients by filtering for patients that visited in the hospital for non-COVID-19 reasons.

Table 3: Distribution of vaccine types conditioned on no COVID-19. Of all people living in Yucatan, we estimate that the majority have taken no vaccine (77.6%), followed by non-replicating viral vector vaccine (13.9%), followed by mRNA (6.87%) and lastly whole or protein vaccines (1.55%).

```{r}
noncovid = subset(clean, covid == "negative")
vaccine_table = table(noncovid$vaccine_name)
prop_vac = vaccine_table/sum(vaccine_table)
prop_vac
```

We then divided these probabilities according to formulas (1), (2), and (3). Finally, we performed bootstrapping to obtain confidence intervals for each of the vaccine types to test if there was a statistically significant difference between vaccine types and the outcome of hospitalization status. We resampled the data with replacement with permutation invariance and 1000 simulations on both the hospitalized ($n = 1,515$) and not hospitalized ($n = 19,155$) datasets using the quantile function.

IV. Results

For our first case-control study, comparing COVID positive vaccinated individuals against COVID positive unvaccinated individuals, the best performing model was a ridge regression model. Generating an ordinary least squares (OLS) model from the coefficient selected by the ridge model for interpretability (Table 4), we focus on the most severe outcomes: status, outcome, and diagnosis. Based off of the regression model, holding all other variables constant, a patient who was COVID-positive and vaccinated is 34.6% less likely to be an inpatient than an outpatient compared to a patient who was COVID-positive and unvaccinated ($p < 0.05$). Similarly, a patient who was COVID-positive and vaccinated is 41.4% less likely to die from COVID than a patient who was COVID-positive and unvaccinated ($p < 0.05$).

Table 4: Beta coefficients from the logistic regression model for comparing COVID-positive vaccinated individuals against COVID-positive unvaccinated individuals.

For our second case-control study comparing patients who were vaccinated and tested positive for COVID-19 against those who were vaccinated and tested negative for COVID-19, the best performing model was a ridge regression model. We again generate an OLS model from the predictors selected by the ridge model for interpretability and examine the beta coefficients for vaccine dosage and vaccine name to find the impact of vaccines on testing positive for

COVID-19 (Table 5). For a patient who received a full vaccine series, holding all other variables constant, a patient who does not have a complete vaccine series is 41.2\% less likely to test positive for COVID than a patient who has not any COVID vaccine series, significant at the $p = 0.05$ level. For the three different vaccine types, holding all other variables constant, a patient who received a viral vector vaccine was 18.5\% more likely than a patient who received an mRNA vaccine to test positive for COVID-19, significant at the $p = 0.05$ level, and a patient who received a whole or protein vaccine was 26.4\% more likely than a patient who received an mRNA vaccine to test positive for COVID-19, significant at the $p = 0.05$ level.

Table 5: Beta coefficients of the logistic regression model for comparing patients who were vaccinated and tested positive for COVID-19 against those who were vaccinated tested and tested negative for COVID-19.

For our third case-control study, comparing patients who were unvaccinated and tested positive for COVID against those who were unvaccinated tested and tested negative for COVID, the best performing model was a ridge regression model. After generating an OLS model from the coefficient selected by the ridge model for interpretability, we compare the predictors to our second case-control study in order to determine if there are any important risk factors .

Table 6: Beta coefficients of the logistic regression model for comparing patients who were unvaccinated and tested positive for COVID-19 against those who were unvaccinated tested and tested negative for COVID-19

From the regression output, there are four clear predictors that appear in the model for the third case-control study but do not appear in the model for the second case-control study. The beta coefficient for taking HIV medication is 0.2550841, indicating that a patient taking HIV medication is much more susceptible to testing positive for COVID-19 if they are vaccinated compared to if they are not vaccinated ($p < 0.05$). Similarly, the beta coefficients for having an office job and having a school job, which are the only significant coefficients for jobs at the $p = 0.05$ level, are both positive at 0.1328969 and 0.1245775 respectively. This indicates that having one of these types of jobs makes one more susceptible to testing positive for COVID-19 if they are not vaccinated. Finally, the beta coefficient for being immunocompromised was -0.3096261, meaning that individuals who were immunocompromised and unvaccinated were less likely to test positive for COVID-19 compared to individuals who were immunocompromised and vaccinated. This may be due to the fact that these individuals were extra cautious due to their own health conditions, and avoided travel and other circumstances that could lead to contracting COVID-19.

Our investigation of vaccine types continues with the Bayesian inference analysis. From inspecting the raw results obtained by division via the procedure above, we conclude that mRNA vaccines are most effective, viral vector vaccines are next, and the whole or protein vaccines rank last (Table 7).

Table 7: Values of $P(\text{Hosp. with COVID-19|Vaccine type}) / P(\text{Hosp. with COVID-19})$ or $P(\text{Vaccine Type|Hosp. with COVID-19}) / {P(\text{Vaccine type})$. Note that these values are not probabilities, which are restricted to the range of 0 to 1. However, the relative order is still informative.

```{r}
prop = prop_hosp_vac/prop_vac
prop
```

Upon visual inspection of the initial point estimates, it would seem that whole or protein vaccines are considerably less effective than every other vaccine type. More surprisingly, the results would suggest that no vaccine is more effective than both the whole or protein vaccines and the viral vector vaccines. If viral vector and whole or protein vaccines added minimal protection, it is possible that vaccinated people would take on more risk resulting in an inflated percentage of hospitalizations. The bootstrap suggests that mRNA vaccines are statistically significantly lower in their probability of hospitalizations. Both the viral vector and whole or protein vaccines performed worse than no vaccine, as whole or protein vaccines are ranked last in effectiveness against hospitalizations.

Table 8: Bootstrap Confidence Intervals comparing the probability of the three different vaccination types conditioning on hospitalization divided by the distribution of the three vaccination types.

```{r}
set.seed(139)
nsims = 5
nsize1 = sum(vaccine_table_hosp)
nsize2 = sum(vaccine_table)
hosp_data = hospitalized$vaccine_name
nonhosp_data = noncovid$vaccine_name
prop.means = matrix(NA, nsims, 4)

for (i in (1:nsims)) {
  prophosp.boot = table(sample(hosp_data, nsize1, replace=TRUE))/nsize1
  nonhosp.boot = table(sample(nonhosp_data, nsize2, replace=TRUE))/nsize2
  prop.means[i,1] = prophosp.boot[1]/nonhosp.boot[1] #None Probability
  prop.means[i,2] = prophosp.boot[2]/nonhosp.boot[2] #mRNA Probability
  prop.means[i,3] = prophosp.boot[3]/nonhosp.boot[3] #Viral Vector Probability
  prop.means[i,4] = prophosp.boot[4]/nonhosp.boot[4] #Whole or Protein Probability
}

rownames = c("None", "mRNA", "Non-replicating Viral Vector", "Whole or Protein")
colnames = c("Mean", "CI Lower Bound", "CI Upper Bound")
```

```
table_data = c(mean(prop.means[,1]), mean(prop.means[,2]), mean(prop.means[,3]),
mean(prop.means[,4]), quantile(prop.means[,1], 0.025), quantile(prop.means[,2], 0.025),
quantile(prop.means[,3], 0.025), quantile(prop.means[,4], 0.025), quantile(prop.means[,1],
0.975),
            quantile(prop.means[,2], 0.975), quantile(prop.means[,3], 0.975),
quantile(prop.means[,4], 0.975))

table_comp = matrix(table_data, 4, 3, dimnames = list(rownames, colnames))

#hist(prop.means[,1])
#hist(prop.means[,2])
#hist(prop.means[,3])
#hist(prop.means[,4])

table_comp
```

V. Conclusion and Discussion

The risks for contracting a SARS-CoV-2 infection and the severity of the resulting disease is highly dependent on vaccination status and vaccine type. From the analyses using logistic regression and Bayesian inference, evidence supports clinical efficacy studies that mRNA is the most effective vaccine type in resisting both COVID-19 exposure and hospitalization. Additionally, both models suggest that whole or protein vaccines are the least effective with higher associated rates of COVID-19 and hospitalization.

The primary risk factors for COVID-19, as evidenced by the logistic regression are working a certain occupation, experiencing sore throat symptoms, immunocompromised status and receiving antiretroviral treatment for HIV/AIDS. There are two main considerations with the given risk factors. Occupation can greatly determine one's in person interactions, which then increases the possibility of exposure. For example, nurses and doctors are much more likely to be exposed to COVID-19 relative to someone that can work from home. Additionally, comorbidities can affect a person's behavior. If someone is immunocompromised, they may be more likely to stay inside because they are aware of the increased risk given their condition.

However, these conclusions are not without limitations. Notably, the study population is restricted to patients who voluntarily visited a healthcare institution for COVID-like symptoms. We attempted to draw inferences for certain COVID-19 outcomes, such as hospitalization status, using Bayes' rule, but this constraint still affected our probability estimates. Additionally, because not all residents of Yucatan have the same accessibility to healthcare services at a hospital, which means the data may be skewed toward patients with a higher socioeconomic status or may not represent indigenous populations. Finally, this paper is situated within the

geographical and social context of Yucatan, Mexico, and results may not be generalizable to other locations with different styles of living, political institutions, and cultural backgrounds.

References

Supplementary Information

Acknowledgements

The authors would like to thank Professor Kevin Rader, teaching fellow Will Nickols, Dr. Nídia Trovão, Chelsea Hansen, and Professor Guadalupe Talavera for their contributions to this project.

The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of HHS, or the U.S. government.