Social Determinants Relating to Covid-19 Infection Risk, Vaccination Rates, and Telehealth Utilization

By
Jordy Bernard

November 2021

Background

The COVID pandemic has highlighted the real impacts of social determinants of health (SDoH) on health risks and outcomes. Studies have found that African Americans, Latino individuals, and Native Americans have experienced a disproportionate burden of COVID-19-related infections and deaths (Price-Haygood Burton 2020, Andrasfay Goldman 2021). Further, numerous reports have shown that Black, Latinos, and Native Americans are least likely to receive the COVID-19 vaccination (Cuellar NG). Since, Alaska is a unique state culturally, demographically, and environmentally, and social factors can influence population health at various spatial scales, an investigation into Covid-19 risk and vaccination at the Alaskan Community Health Center (CHC) level is merited.

By coupling risk stratification (i.e., the separation of patients into high risk and low risk populations) with interventions that target high-risk groups, health care systems can be improved. Auxiliary data regarding a patient's environment and socioeconomics can be used to inform risk stratification algorithms, which can be used to improve the accuracy of diagnosis, which can in-turn lower health care costs and improve patient outcomes.

In this project, we used a statistical technique called survival analysis to synthesize up to date Covid-19 related information and identify factors that place CHC patients at an increased risk of contracting Covid-19. A proportional hazards model was used to measure patient risk. Logistic regression was also used to identify potential factors influencing telehealth utilization and access.

The results of this study will be used to evaluate the effectiveness of telehealth services in treating patients with increased Covid-19 related risk. This information will help policy makers as they evaluate the utility of telehealth services and image the future of telehealth in the State of Alaska. The results of this study will also be used to make recommendation for how the risk stratification algorithms can be modified to reflect Covid-19 related risk more accurately.

Objectives

The objectives of this study are to:

1: Estimate the probability that a patient from a given patient population has not contracted the Covid-19 infection at any given point in time and identify factors associated with increased Covid-19 infection risk.

2: Estimate the probability that a patient from a given population has not received a Covid-19 vaccination at any point in time and identify factors associated unvaccinated populations.

3: Estimate the probability that a patient from a given patient population has utilized telehealth services in the past month and identify factors affecting telehealth utilization.

Data Reduction

A person's health is affected by a number of social determinants. This includes individual factors such as who they are (age, sex, and genetic factors). This also includes behavioral factors such as what they do (diet, physical activity, and drug use). Health is also affected by conditions in which people are born, grow, live, work, and age. These include social and community networks; socio-economic, cultural, and environmental conditions; and health systems. In this analysis we considered, 10 different predictors of health (i.e., explanatory variables). The measures considered are detailed in the Table 1.

Patient Level Data Analyzed	
Variable Description	Variable Type
Was the patient diagnosed with a new case of Covid-19 during the encounter (Yes or No)	Response
Was the patient vaccinated during the encounter (Yes or No)	Response
Was this a telehealth appointment (Yes or No)	Response
Health center where the encounter occurred	Explanatory
Patient's age (1, 2, 3, …)	Explanatory
Patient's sex (Male or Female)	Explanatory
Patient's race (White, Native American/Alaskan, Black, Asian, or Other)	Explanatory
Patient's ethnicity (Hispanic or Non-Hispanic)	Explanatory
Patient's language (English or Other)	Explanatory
Patient's insurance category (Uninsured, Medicaid, Medicare, or Private)	Explanatory
Does the patient have an active depression diagnosis during the study period (Yes or No)	Explanatory
Did the patient identify as a tobacco user in the past 24 months (Yes or No)	Explanatory
Was the patient identified as being an unhealthy alcohol user during the past 24 months (Yes or No)	Explanatory

Table 1: De-identified patient level data that was analyzed

In this analysis, 21 months of data ranging from 01/12/2020 to 10/31/2021 was considered. Patient data coming from 23 clinics spread across the state of Alaska was considered.

Data Analysis

An extension of the usual proportional hazards model was used to identify patient populations that are at increased risk of being infected with the Covid-19 virus (Rodríguez 2007) and identify factors associated with increased vaccination rates. In developing the statistical framework of the model, we will start by defining the hazard and survival functions in discrete time.

Let T be a discrete random variable representing the month in which patient receives a positive Covid-19 diagnosis. Here, T takes the values $t_1 < t_2 < \cdots$ with probabilities $f(t_j) = f_j = \Pr(T = t_j)$. The "survival function" is defined at time t_j as the probability that the time of infection T is at least t_j:

$$S(t_j) = S_j = \Pr(T \geq t_j) = \sum_{k=j}^{\infty} f_j.$$

Next, the hazard function λ_j is defined at time t_j as the probability of being diagnosed with a Covid-19 infection at that time given that the individual has not been infected up until that time.

We now let $i \in \{1,2,\cdots,n\}$ be an index representing the patient. In this study, we will consider the patient population to be the set of all patients with any meaningful contact between the patient and a health center in the time duration considered in this study. Additionally, we let x_i be a vector representing the health determinants (i.e., explanatory variables) detailed in Table 1. The proportional hazard model is now defined as

$$\mathrm{logit}\lambda(t_j|x_i) = \alpha_j + x_i^T \beta.$$

In this equation, $\alpha_j = \mathrm{logit}\lambda_0(t_j)$ where $\lambda_0(t_j)$ is the baseline hazard. The baseline hazard function is analogous to the intercept term in a multiple regression or logistic regression. β reflects the effect of each determinant of health on the logit of the baseline hazard. The hazard ratio, $\lambda(t_j|x_i)/\lambda_0(t_j)$ is regarded as the relative risk that patient i is infected with the Covid-19 virus at time j.

Logistic regression was used to assess which patient populations have been utilizing telehealth services. Letting $Y_{ij} = 0$ be the event that the i^{th} appointment was an in-person encounter, and letting $Y_i = 1$ be the event that the i^{th} appointment was a telehealth encounter, the probability that the i^{th} patient utilized telehealth services was modeled as

$$\mathrm{logit}p_i = \alpha + x_i^T \beta.$$

β reflects the effect of each variable on probability of telehealth utilization. A separate model was fit for each and every subset of explanatory variables listed in Table 1, and AIC was used as the model selection criterion. The final models that we considered contained all the explanatory variables detailed in Table 1.

Summary of Results

Covid-19 Infection Risk

Covid-19 infection rates peaked in September 2021. More Covid-19 new cases were documented in this month than in any other month on record. During this month, 2.6% of patients that entered a health center were diagnosed with a new case of Covid-19. As of November 2021, approximately 95% of the patients that visit a community health center have not received a positive Covid-19 diagnosis by the health center in which the encounter occurred. Of the predictor variables considered in this study, the health center (a proxy for a patient's geographic location) was found to have the greatest effect upon a patient's Covid-19 risk of infection. Native American/Alaskans were identified as the racial group with the greatest risk of infection. We also found differences in Covid-19 infection rates between ethnicities. Hispanic patients were found to have lower infection rates than non-Hispanics. Younger patients (particularly those between the ages of 20 and 60) were found to have increased Covid-19 infection rates. The remaining variables that we considered did not have a meaningful impact upon Covid-19 risk of infection.

Covid-19 Vaccination

Vaccine administration rates peaked in February 2021. While there was a significant amount of variation between health centers, approximately 15 percent of patients encountered by a health center in February 2021 were given a Covid-19 vaccination by that health center. Rates of vaccine administration have decreased since February 2021. As of September 2021, approximately 80 percent of patients that enter a health center have been immunized by the health center. Older patients and those on Medicare were found to have the highest immunization rates. Though the effect wasn't as strong as age and insurance, language, ethnicity, and race were related immunization rates.

Telehealth Utilization

The utilization of telehealth services peaked in April 2020. Since this time, the utilization of these services has declined steadily. Approximately 10 percent of al patient encounters occurred by telehealth in November 2021. A patient's geographic location had the largest effect upon telehealth utilization. Patients from the Anchorage metropolitan area utilized telehealth services at the highest rates. Of the patient level variables we considered, depression status was found to have the greatest effect on telehealth utilization. Patients with active depression diagnosis were more likely to utilize telehealth services than patients without depression diagnosis. A patient's insurance class was the next most influential patient level variable on predicting telehealth utilization. We found that telehealth utilization rates were the highest for uninsured patients and those on Medicaid. Additionally, patients between the ages of 20 and 60 were found to utilize telehealth services at the highest rates.

References

Cuellar NG. Vaccination Hesitancy. *Journal of Transcultural Nursing*. 2021;32(3):197-197. doi:10.1177/1043659621999703

Collett, D. (2015). The Cox Regression Model. In Survival Data in Medical Research (Third, pp. 57–130). essay, CRC Press.

Rodríguez, G. (2007). *Lecture Notes on Generalized Linear* Models. https://data.princeton.edu/wws509/notes/c7s6

Rodríguez, G. (2007). *Lecture Notes on Generalized Linear* Models. https://data.princeton.edu/wws509/notes/c3.pdf

Penn State Eberly College of Science. Epidemiological Research Methods. URL: Lesson 13: Proportional Hazards Regression | STAT 507 (psu.edu)

Price-Haygood EG, Burton J, et al. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. N Engl J Med. 2020. DOI: https://doi.org/10.1056/NEJMsa2011686external icon.

Andrasfay, T., & Goldman, N. (2021). Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. Proceedings of the National Academy of Sciences, 118(5).