Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	April 20, 2022
Alisha Kalangara	Date

COVID-19 Variant Surveillance Data as an Early Indicator of COVID-19 Case Surges

By

Alisha Kalangara
Master of Public Health

Epidemiology

Kristin Bratton Nelson, PhD, MPH
Committee Chair

CDR Benjamin J. Silk, PhD, MPH
Committee Member

COVID-19 Variant Surveillance Data as an Early Indicator of COVID-19 Case Surges

By

Alisha Kalangara

B.S., Biology
The University of Texas at Austin
2019

Thesis Committee Chair: Kristin Bratton Nelson, PhD, MPH

An abstract of
A thesis submitted to the Faculty of the Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of Master of Public Health
in Epidemiology
2022

Abstract

COVID-19 Variant Surveillance Data as an Early Indicator of COVID-19 Case Surges
By Alisha Kalangara

Introduction: The COVID-19 pandemic is the first pandemic to occur in the modern genomic sequencing era. Genomic surveillance has been utilized during the COVID-19 pandemic to study the pandemic origins, aid in outbreak investigations, study potential immune escape, and to actively monitor the emergence and prevalence of new SARS-CoV-2 variants. Given the public health significance of SARS-CoV-2 variants, we investigate whether genomic surveillance data could help anticipate domestic surges in case incidence at the state level.
Methods: The objective of this study was to assess whether modeling of Delta variant proportion data could act as an early indicator for reported COVID-19 case surges. Publicly available datasets were utilized to capture longitudinal data during the Delta variant's circulation in the US (April 19 – October 24 of 2021). Case incidence for each state was calculated using July 2021 US Census Bureau population estimates. Using generalized estimating equations with an autoregressive correlation matrix, the relationship between changes in Delta variant proportion and changes in case incidence was estimated using non-lagged, 2-week lagged, and 4-week lagged data, while adjusting for vaccination rates, infection induced seroprevalence, and case age distribution.
Results: At the state level, the 2-week lagged model had the strongest association between the Delta variant proportion data and a surge in COVID-19 case incidence (OR: 14.30, 95% CI: 7.12 - 29.08). The model with four-week lag suggested a weaker association between change in Delta variant proportion and a COVID-19 surge would occur 3 or 4 weeks later (OR: 2.12, 95% CI: 0.88 - 5.10). The non-lagged model showed a strong positive association, demonstrating simultaneous rises of Delta variant proportions and overall COVID-19 case incidence (OR: 8.85, 95% CI: 4.18 - 18.77).
Conclusion: Our results suggest that monitoring changes in COVID-19 variant proportion data can act as a leading indicator of COVID-19 case incidence surges. This genomic surveillance strategy is important for anticipating a surge, which allows for appropriate public health and healthcare capacity measures to be prepared to lessen or avoid the consequences of a surge.

COVID-19 Variant Surveillance Data as an Early Indicator of COVID-19 Case Surges

By

Alisha Kalangara

B.S., Biology
The University of Texas at Austin
2019

Thesis Committee Chair: Kristin Bratton Nelson, PhD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of Master of Public Health
in Epidemiology
2022

ACKNOWLEDGMENTS

I thank Dr. Kristin Nelson and Dr. Benjamin J. Silk for their mentorship, feedback and time on this project. I would also like to thank my family and friends for checking in on me, and supporting me in this journey.

TABLE OF CONTENTS

Introduction……..……..1

Methods……..……4

Time Period……..……..4

Exposure Variable……..……....4

Outcome Variable……5

Confounders……..……6

Figure 1. Directed Acyclic Graph Diagram for Analysis ……...……..6

Modelling Strategy…….....8

Results……8

COVID-19 case incidence……8

Figure 2. COVID-19 Case Incidence per 100k……9

Prevalence of SARS-CoV-2 Delta Variant……9

Figure 3. Percent of Cases Attributable to SARS-CoV-2 Delta Variant Among US States Through Time……10

Figure 4. 2-week Change in Delta Variant Proportion and Percent Change in Incidence Across Time in Select US States……11

Generalized Estimating Equation Models……...12

Table 1. Lagged and Nonlagged Model Results……12

Discussion……12

Model Results……12

Limitations……..…13

Context and Public Health Significance……14

References……16

Appendix……..20

Emory IRB Exemption……20

Introduction

The ongoing coronavirus disease 2019 (COVID-19) pandemic is suspected to have begun in late 2019 in Wuhan, China.[1] On January 30, 2020 WHO declared a Public Health Emergency of International Concern and human to human transmission was confirmed.[1] On February 11 the novel coronavirus was named SARS-CoV-2 by the International Committee on Taxonomy of Viruses (ICTV) and WHO named the disease as COVID-19.[1] With the first case of COVID-19 reported outside of China in Thailand in mid-January of 2020, and outbreaks being detected in South Korea, Japan, Italy, and Iran by mid-February, WHO declared COVID-19 as a pandemic on March 11, 2020.[1]

As of April 14, 2022, there have been over 500 million confirmed cases of COVID-19 worldwide.[2] The most common symptoms of SARS-CoV-2 infection are fever, congestion, and cough; but fatigue, diarrhea, chest tightness, confusion, vomiting, nausea, sore throat, sneezing, sputum production, anosmia and dyspepsia, rash, discoloration of fingers or toes, and viral conjunctivitis have also been reported.[3,4] Though most cases are mild (80.9%), there are severe (13.8%) and critical (4.7%) cases, along with those that result in death (2.3%).[5] Of note, asymptomatic cases play a significant role in transmission as it is estimated that > 50% of transmission occurs from asymptomatic people.[6.] It appears that increasing age (especially > 60 years of age), male sex, or having an underlying condition is associated with greater risk of severe COVID-19 illness.[3,5,7] Milder cases resolve in 1-2 weeks but more severe cases could progress toward pneumonia, acute respiratory distress syndrome, septic shock, cardiac dysfunction, an exaggerated inflammatory response, exacerbation of underlying comorbidities or hepatic, renal, central nervous system, or thrombotic disease.[5,8] It has also been noted that 35% of patients treated for COVID-19 on outpatient basis, and around 87% of hospitalized patients experience residual symptoms greater than 3 weeks past the onset of symptoms in what is termed “long COVID”.[9]

In the United States, the first COVID-19 case was identified on January 19 of 2020 in Washington in a patient who had traveled to Wuhan, China.[10] By February, community transmission had been observed, and by mid-March SARS-CoV-2 infections were detected in all 50 states.[10] The virus is

thought to spread both directly through human to human transmission (through droplets produced by coughing, sneezing, or talking), and indirectly (exposure to contaminated objects and airborne virus).[4] Consequently, the CDC recommended the use of facial coverings (masks), social distancing, isolation and quarantine, and stay-at-home orders to help mitigate the spread of the virus.[11] Though there is evidence to suggest that use of masks and social distancing measures helped to reduce transmission and deaths in the United States, these measures were not universally adopted and widespread adherence is not sustainable.[12, 13, 14] Among states, wide variations in case incidence, testing frequency, and mortality were observed possibly due to staggered introduction of the virus, population density, timing of community mitigation strategies, availability of and approaches to SARS-CoV-2 testing, the occurrence of mass gatherings and differences in prevalence of underlying conditions.[15, 16]

The United States has utilized SARS-CoV-2 genomic sequencing to monitor the changing landscape of variants that have emerged during the pandemic. As COVID-19 is arguably the first pandemic in the modern genomic era, genomic surveillance has burgeoned under the pandemic as a result of quicker and cheaper next generation sequencing capabilities.[17] The sequence for SARS-CoV-2 was available relatively early in the pandemic, and since then, an unprecedented number of sequences have been uploaded to GISAID.[17] These data have wide applications for medicine and public health, including studying the origins of the pandemic, aiding in outbreak investigations in a variety of settings, studying potential immune escape , and analyzing drivers of epidemiological trends as in this analysis.[18] Genomic surveillance is especially important for active monitoring of viral evolution at the population-level, including the emergence, spread, and real time prevalence of new SARS-CoV-2 variants. In the United States, CDC has established a national surveillance system for SARS-CoV-2 variants using genomic sequencing and phylogenetic analysis. Initially, the system utilized only specimens submitted from public health laboratories to the National SARS-CoV-2 Strain Surveillance (NS3) program.[19] These specimens are submitted to CDC for assessment, sequencing, and genomic analysis. Phylogenetic lineages are assigned and sequences are made available in public repositories.[20] As of December 2020, however, the system also incorporates specimens from contracted commercial laboratories, which increases the

geographic coverage of surveillance.[20] Both NS3 and commercial labs provide a weekly target number of specimens and sequences that are geographically representative.[20] NS3 also supplies demographic data though demographic data from commercial labs are limited.[20] CDC has also developed statistical weighting and variance estimation methods to generate population-based estimates of the proportions of SARS-CoV-2 variants circulating nationwide and in each of the 10 U.S. Department of Health and Human Services (HHS) geographic regions.[20] In effect, this information can be used to monitor the emergence of new variants and their predominance.[20]

Genetic differences between SARS-CoV-2 variants have the potential to affect transmissability if they are associated with differences in disease severity, the ability to detect and treat infection or the duration and strength of immunity.[20] During the first months of the pandemic, relatively few SARS-CoV-2 variants had been identified. However, four variants of concern were detected in the fall of 2020.[21] Among these was Delta which, as per the definition of a variant of concern, showed evidence of an increase in transmissibility, more severe disease, significant reduction in neutralization by antibodies generated during previous infection or vaccination, reduced effectiveness of treatments or vaccines, or diagnostic detection failures .[22] Notably, spike mutations of the Delta variant allowed for an estimated 40–60% increase in transmissibility compared to the Alpha variant, which itself was twice as transmissible as the original strain from Wuhan.[23] Additionally, according to a Canadian study utilizing a retrospective cohort of people who tested positive for SARS-CoV-2 in Ontario, the Delta variant was associated with an 108% increase in hospitalization risk, 235% increase in risk of ICU admission, and 133% higher risk of death than the original variant.[24] In the United States, the Delta variant was first detected in March of 2021 and became especially prevalent in under-vaccinated populations where social distancing and mask usage were less common.[23, 25] Consequently, the Southeastern United States experienced a higher number of cases and hospitalizations due to the Delta variant, as compared to the rest of the United States.[25]

Given the public health significance of SARS-CoV-2 variants and the proven value of genomic surveillance for monitoring their prevalence, it is important to consider how to best utilize genomic surveillance information to manage the pandemic and anticipate surges in case incidence. This analysis

utilizes publicly available datasets to evaluate whether changes in SARS-CoV-2 variant proportions estimated from genomic surveillance data are associated with state-level case surges during a period when the Delta variant became predominant, while adjusting for age distributions among cases, population-level seroprevalence from infection, and vaccination coverage rates.

Methods

Time Period

The time period for analysis was April 19, 2021 to October 24, 2021. The time interval of interest for the analysis was 2-weeks, as the Delta variant proportions data were available in 2-week intervals. The start date was selected to correspond to when Delta variant roughly began circulating in the United States. The end date was selected to include the peak of the Delta surge while also predating the subsequent rise in the Omicron variant.

Exposure variable

The primary exposure of interest was a significant increase in the Delta variant proportion from the previous time interval. Delta variant proportion refers to the estimated proportion of genomic sequences in a state that is attributable to the Delta variant. After testing various cut off points, "a significant increase" was defined as the first increase in Delta variant proportion by 30 percentage points or more compared to the previous interval. The Delta variant proportion data for each 2-week interval for each state were taken from state specific genomic surveillance data displayed on the Regeneron dashboard, which utilizes publicly available SARS-CoV-2 sequence data from the Global Initiative on Sharing All Influenza Data (GISAID) repository.[26, 27] These data were available by 2-week interval and so the difference between consecutive 2-week intervals was taken to be the change in Delta variant proportion. So as to focus on earliest detection, the first time interval when the change was greater than or equal to 30% was used, rather than all weeks that had greater than or equal to 30%.

Outcome variable

The outcome of interest is the occurrence of a COVID-19 surge in case incidence. This was defined as the 2-week period when cases in a state began to rise in the period between mid-May and early June. Analytically this was measured by when the percent change in 2-week incidence of COVID-19 cases in a state first switched from negative to positive. COVID-19 is nationally notifiable and state- level case surveillance data from CDC were utilized to derive this metric.[28, 29] These data were publicly available in line list format starting in January of 2021 and are updated twice a day. Data were subset to the time period of interest, and only new confirmed cases were retained. Of note, cases from New York City were reported separately from the rest of New York State, but combined for this analysis. Probable cases were excluded.

The cumulative incidence for each two-week time interval was calculated by summing the number of new cases reported in each interval by each state and then dividing by each state's population. State population estimates were taken from the U.S. Census's Population Estimates Program, which provided state-specific population estimates for July 2021 based on the 2020 U.S. Census.[30] These state population estimates are estimated by taking the base population estimate, adding births, subtracting deaths, and adding net migration.[31] Of note, for July of 2021 estimates, the base population was the 2020 decennial Census supplemented with the 2020 Census PL 94-171 Redistricting File, and 2020 Demographic Analysis (DA) estimates.[31] These supplements were added in part due to COVID-19 disruption of the 2020 Census.[31] Births and deaths data came from vital statistics from the National Center for Health Statistics (NCHS) net migration calculations utilize data sources such as Internal Revenue Service (IRS) tax return data, Medicare enrollment data, Social Security Administration's (SSA) Numerical Identification Files, Demographic Characteristics File (DCF), and American Community Survey (ACS) among other sources.[31] We assumed that state populations did not vary significantly throughout the time period of interest. The percentage change in incidence was calculated by comparing the incidence between consecutive two-week intervals; percentage change in incidence was evaluated to identify the first week when this change switched from negative (indicating declining case incidence) to

positive (indicating increasing case incidence). Of note, this evaluation could not be performed for initial two-week intervals, so these baseline periods were excluded from the analysis.

Confounders

In examining an exposure-outcome relationship, it was hypothesized that several additional factors could influence the primary relationship (Figure 1).

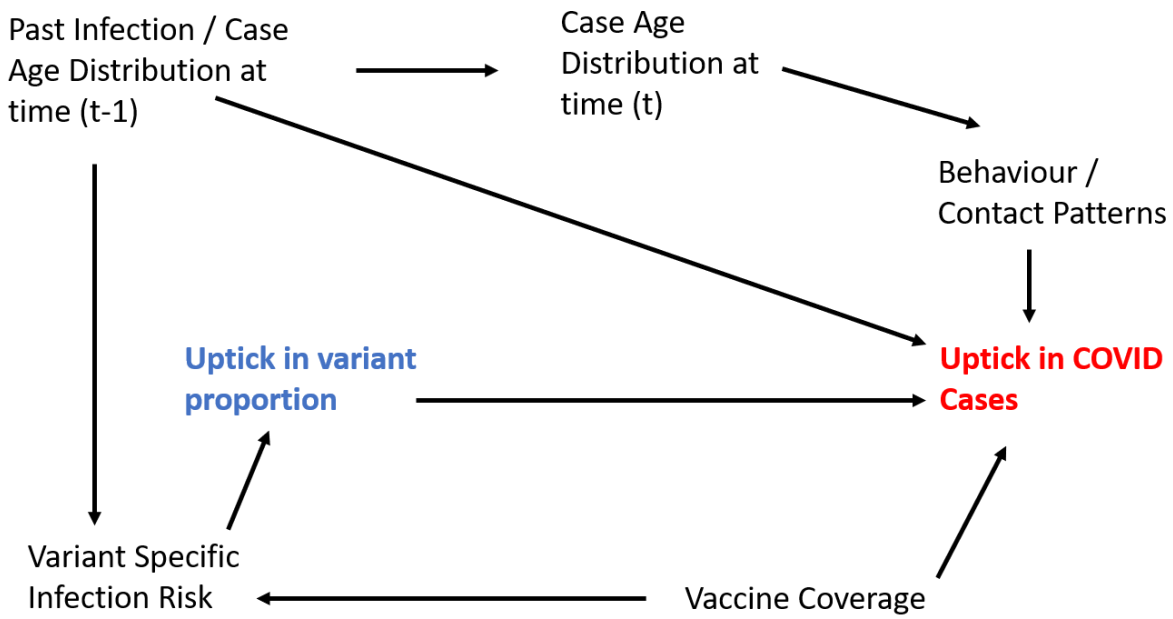


Figure 1. Directed Acyclic Graph Diagram for Analysis. Exposure of interest is shown in blue and the outcome in red.

Population vaccination coverage could be a confounding variable, as a decrease in cases could be related to increased population-level immunity because of high vaccination rates, and vaccine-induced immunity may protect differentially against different SARS-CoV-2 variants. To adjust for this, we used publicly available data from CDC on vaccination rates by state.[32] The dataset draws from jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities; these data are used to obtain the daily state percentage of fully vaccinated (i.e., one dose for a one-dose vaccine or two doses of a two-dose series) individuals among those eligible from December 13,

2020. For this analysis, the average percentage was taken across the days in each 2-week interval for each state.

SARS-CoV-2 seroprevalence data were also included as a potential confounder, which reflect population-level immunity acquired from previous infection. In this case, if a state's population has a relatively higher estimated seroprevalence, there may be a decrease in cases regardless of variant proportion, as infection-induced immunity from previous variants might influence the probability of infection with new variants. Publicly available, nationwide seroprevalence data were used.[33] These data originated from 17 blood donation locations that represent the 50 states and have a catchment area of 74% of the U.S. population.[34] Blood samples from these locations were screened for the SARS-CoV-2 nucleocapsid protein, which is indicative of a past infection (i.e., not COVID-19 vaccination). As blood donation is only allowed for those 16 and older, the estimates in the dataset represent the percentage of the population 16 and older presumed to have had a previous SARS-CoV-2 infection. These data are updated monthly, and though most states had a single value for infection-induced seroprevalence for a given period, data were given at the regional level; 9 of the 51 states were subdivided across several regions. For these subdivided states, the maximum among regional estimates was used. Additionally, because data were collected monthly, it was assumed that for 2-week intervals that comprise a month, seroprevalence measures were similar enough to use a single monthly estimate across both time intervals.

The case age distribution in each state was also considered a confounder. Upticks in cases may be caused by a shifting age distribution of infections, rather than due to increased transmission of a novel variant. To account for this, the proportions of confirmed cases in a given month among 0 – 17 year olds, 18 – 49 year olds, 50 – 64 year olds, and 65 + year olds were included based on CDC's publicly available data at the county level, which are updated monthly.[35] This county level line listing was aggregated to the state level as case counts for each age group during each of the 2-week time intervals. Like the seroprevalence data, it was assumed that for the 2-week intervals that comprise a month, case age proportions calculated were similar enough to use a single monthly estimate across both time

intervals. Additionally, some states did not have information for all months in the period of interest. For these state and month combinations, the distribution of the closest month was used assuming that across the time period, cases' age proportions are similar to one another.

Modelling Strategy

To model the relationship between a significant increase in Delta variant proportions and whether or not there was a COVID-19 surge, generalized estimating equations (GEE) with an autoregressive correlation matrix was used. GEE was used because data are longitudinal and each unit of analysis (states) has repeated measures (2-week time intervals that span selected time period) that are not expected to be independent of on another, but rather related to one another. The correlation matrix defines how the intervals are expected to be related to one another in a state. The autoregressive correlation matrix used assumes that within each state or district, intervals that are closer together in time are more related to one another than intervals that are further apart in time.

Additionally, in assessing whether significant increases in Delta variant proportions can serve as an early surveillance indicator of rises in case incidence, the influence of a lag was also tested. A non lagged model evalates the relationship between significant Delta variant proportion in a given 2-week interval and whether or not there is a COVID-19 surge in that same interval. A 2-week lagged model and 4-week lagged model were also used to evualte the relationship between the exposure in a given 2-week period and the outcome in the upcoming two weeks and three to four weeks afterwards respectively.

Results

COVID-19 case incidence

Increases in COVID-19 incidence begin around mid June of 2021. (Figure 2) Afterwards, states generally experienced an increase in case incidence before peaking, and experiencing a decline in case incidence. States' case incidences appear to peak at different times with some states peaking as early as

August 2021, and others as late as early October 2021.

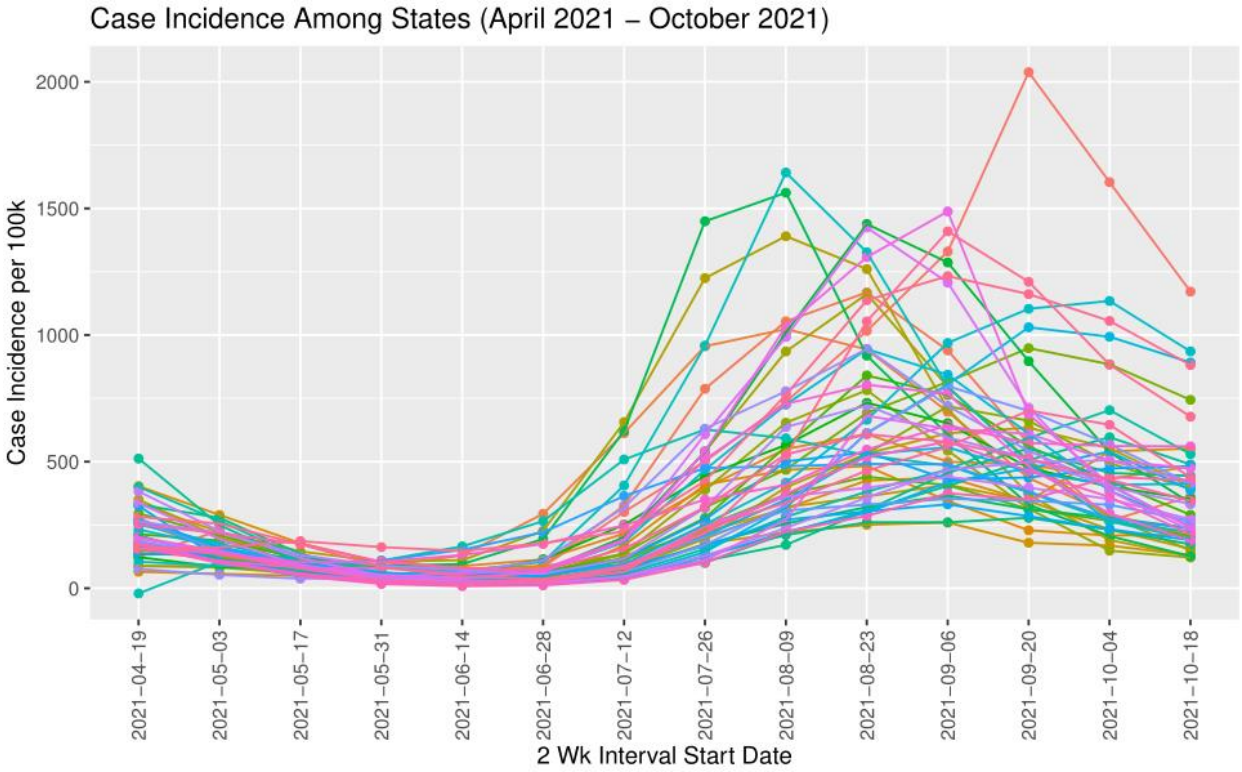


Figure 2. COVID-19 Case Incidence per 100k. Each line represents case incidence in a U.S. state, using U.S. Census state population estimates for July 2021. The legend has been omitted to display the general trend over time.

Prevalence of SARS-CoV-2 Delta variant

In examining the SARS-CoV-2 Delta variant proportions over time, a similar trend emerges. In general, Delta variant prevalence was low during late April and early May 2021 for most states (Figure 3). Afterwards, states experienced an increase in Delta variant proportion until approximately 100% of cases were attributable to Delta variant, which occurred for most states in late June 2021. Afterwards, Delta variant proportions appear to remain close to 100% for the remainder of the selected time interval, until mid-October of 2021.

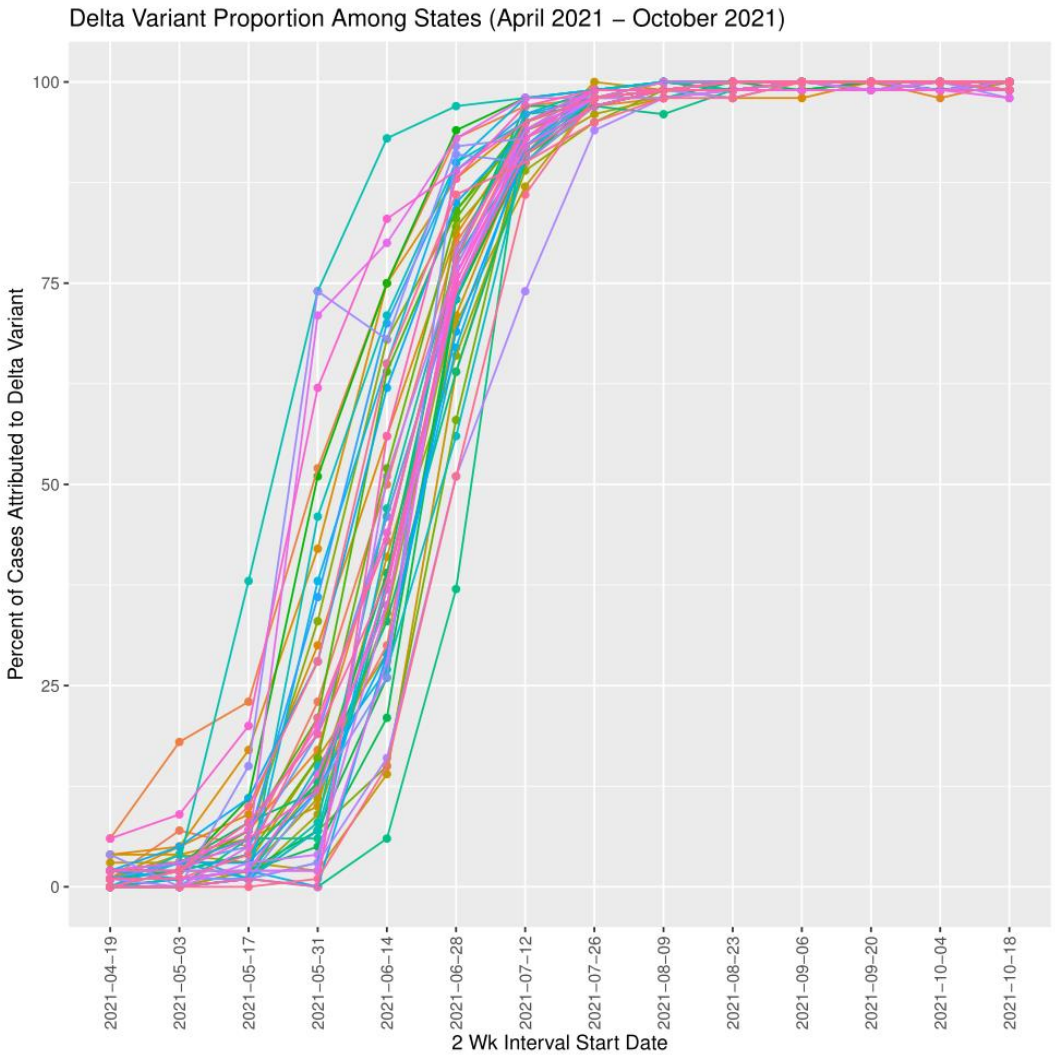


Figure 3. Percent of Cases Attributable to SARS-CoV-2 Delta Variant Among U.S. States Through Time. Each line represents the SARS-CoV-2 Delta variant proportion in a US state. The legend has been omitted to display the general trend over time.

When examining the change in both proportion of cases caused by the Delta variant and COVID-19 case incidence over time, it is evident that there are separate rises and falls for both changes in Delta variant proportion and 2-week case incidence. Looking at the combination of these patterns, it is observed that the rise in Delta variant proportion precedes the peak in percent change of incidence (Figure 4). This appears to support the hypothesis that Delta variant proportion can be used as a leading indicator for case incidence.

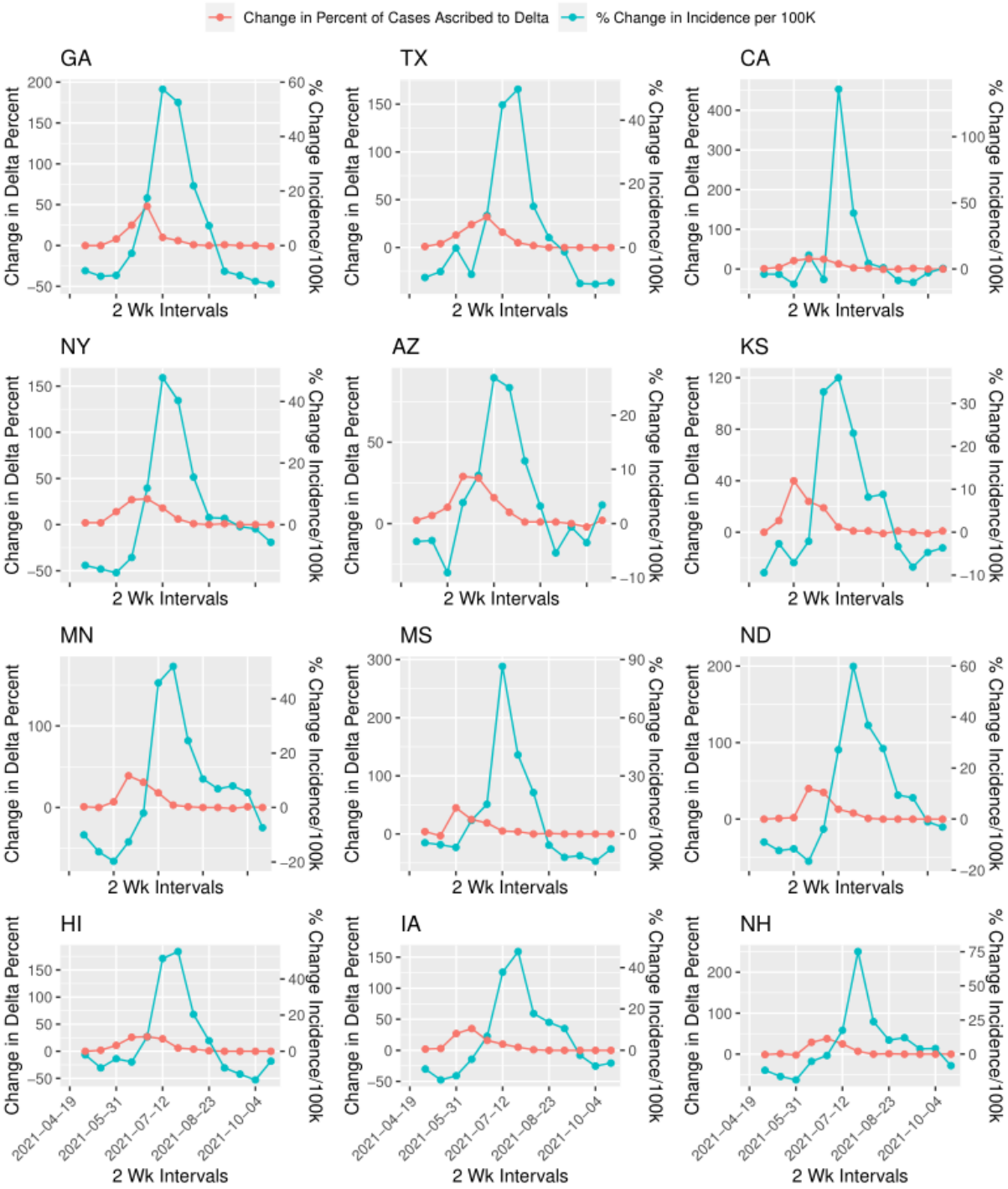


Figure 4. 2-week Change in Delta Variant Proportion and Percent Change in Incidence Across Time in Select U.S. States

Generalized Estimating Equation Models

The results of the generalized estimating equations are given in Table 1. In a model with no lag, during a 2-week period in states when there was an initial 30% or greater increase in Delta variant proportion, the odds of experiencing a surge in COVID-19 cases at that time was 8.85 (95% CI: 4.18 – 18.77) times more likely compared to during a 2-week period in states when there was not an initial increase in Delta variant proportion, when adjusting for vaccination rate, case age distribution, and infection-induced seroprevalence.

This association is stronger when incorporating a 2-week lag (OR:14.30, 95% CI: 7.12 – 29.08). In this case, during a 2-week period in states when there was an initial 30% or higher increase in Delta variant proportion, the odds of experiencing a surge in COVID-19 cases in the upcoming two weeks is 14.30 times more likely compared to a 2-week period in states when there was not an increase in Delta variant proportion, when adjusting for vaccination rate, case age distribution, and infection-induced seroprevalence.

When introducing 4-week lag, this relation is weaker. The odds of experiencing a surge in COVID-19 cases in the third and fourth weeks were 2.12 (95% CI: 0.88 –5.10) times as likely for a 2-week period in states when there was an initial 30% or greater increase in Delta variant proportion, compared to those that did not, when adjusting for vaccination rate, case age distribution, and infection-induced seroprevalence. This was not found to be statistically significant (alpha level = 0.05), however.

Table 1. Lagged and Nonlagged Model Results

Exposure Indicator	**Lag Type**	**OR**	**95% CI**	**P value**
First Delta Change >= 30 Percentage Points	No Lag	8.85	(4.18, 18.77)	<0.001
	2 Week Lag	14.30	(7.12, 29.08)	<0.001
	4 Week Lag	2.12	(0.88, 5.10)	0.095

Discussion

Model Results

By utilizing publicly available data on over a period of high Delta variant circulation in 2021 (April 19 – October 24), we assessed the relationship between a significant change in Delta variant proportion and surges in COVID-19 case rates while adjusting for vaccination rates, infection induced seroprevalence, and case age distribution. Overall, Delta variant proportion and case incidences were strongly related. The model with a 2-week lag showed the strongest association between rising Delta variant proportion and overall case incidence, suggesting that the change in Delta variant proportion data can be used to determine whether a COVID-19 surge is expected in the upcoming two weeks (OR: 14.30 , 95% CI: 7.12 – 29.08). The model with a 4-week lag suggested a relatively weaker relationship between change in Delta variant proportion and if a COVID-19 surge would occur 3 or 4 weeks later (OR: 2.12 , 95% CI: 0.88 – 5.10). We also found a strong positive association in the no-lag model, demonstrating simultaneous rises of Delta variant proportions and overall COVID-19 case incidence (OR: 8.85 , 95% CI: 4.18 – 18.77).

Limitations

This analysis has several limitations. Case surveillance data provided by CDC are based on reports by U.S. states and autonomous reporting entities, including New York City and the District of Columbia, as well as U.S. territories and affiliates. Case reporting completeness may be affected as COVID-19 usually results in mild illness which may be difficult to detect on time or altogether Additionally, symptoms might not appear immediately resulting in delays in testing or reporting, which may influence when cases are attributed to each time interval. This was demonstrated early in the pandemic through a serosurvey conducted in two Georgia counties. Though an estimated one half of seropositive persons recalled having had COVID-19 symptoms, only one third of seropositive persons sought medical care, and fewer had a test for SARS-CoV-2 infection.[36] A more recent 2021 estimate suggests that 60% of all cases have been unreported in the United States as a whole.[37] However, we expect that incomplete surveillance data and any reporting lags would not be differential by the exposure or outcome, and therefore would not influence our estimates of epidemiologic association.

Similarly, the data used for infection-induced seroprevalence represents selection bias as these data are based on blood donors who are representative of 74% of the United States geographically, and include only those 16 years and up. If the blood donors for a particular region are not reflective of the region's age make up, the seroprevalence rates may be an under or overestimate. It is unclear how this would influence our estimates of epidemiologic association.

Lastly, Delta variant proportion data are based on sequences available in GISAID. Early in the study period, states had relatively less robust systems for sequencing and depositing sequence data meaning that the precision around the variant proportion estimates might have varied through time. Additionally, genome sequencing capabilities continue to vary by state, as some states sequence far more specimens compared to other states. In this case, states with greater capacity may be drawing from sequences that are more representative of variant circulation.

Aside from dataset limitations, this model does not address variant specific infection risk or behavior and contact patterns, which could also play a role in the relationship between changes in variant proportion and surges in case rates (Figure 1). Surges in cases may be related to masking and social distancing policies in effect, and adherence to these policies. For variant-specific infection risk, this could mean that some variants are more transmissible than others. Increased transmissibility would result in higher case incidence and also enable higher variant proportions. This metric, however, is hard to determine.

Context and Public Health Importance

Despite these limitations, the analysis conducted is useful in the context of acting as an early surveillance indicator. Other proposed systems for COVID-19 surge detection that have emerged through the course of the pandemic include wastewater surveillance and using PCR cycle threshold (CT) as a metric for viral load and proxy for transmissability.[38, 39] Wastewater surveillance is used as a community level surveillance strategy that relies viral shedding in feces from those infected by COVID-19.[38, 40] Through monitoring sewage for variants through genomic methods, outbreaks of COVID-19 can be anticipated.[36, 38] While this approach is promising, drawbacks are that many U.S. areas do not have

established wastewater surveillance capacity and globally there is inadequate sewage infrastructure in developing countries that may seek to use this approach.[40] In contrast, the conducted analysis makes use of diagnostic testing an genomic surveillance data that are likely already being collected in most countries.

The PCR cycle threshold (CT) value approach also evades an infrastructure issue. This approach suggests that a higher proportion of COVID-19 positive samples with low CT values corresponding to high viral load, is indicative of an upcoming surge.[37] This approach can be easy to implement using diagnostic testing results, but a drawback of this method is that it depends heavily on PCR test conditions and kits used.[37] As a result, it may not be reliably used across different locations and testing platforms. The conducted analysis does not have this limitation, as variants only need to be identified via molecular methods.

Rather than focusing on detection, forecasting is an alternate method for determining surges in COVID-19. A number of mathematical models have emerged to predict future case rates, hospitalizations, and deaths.[41] One such endeavor is the COVID-19 Forecast Hub consisting of many international research groups that each submit a weekly model to be included in an ensemble modelling approach.[42] The ensemble model works well, but is limited in that models are based on the natural history parameters of the Alpha variant, including its transmissibility.[42] As a result, the predictions may not be as accurate for other variants, such as Delta, which have different transmissibility characteristics. Additionally, more reported cases than expected fell outside the forecast prediction intervals for extended periods of time indicating low reliability.[43] Using a statistical approach as done in this analysis, could complement the forecasting models and strengthen reliability.

Ultimately, the strategies presented can be used independently or in tandem as an early surveillance indicator to notify of a potential COVID-19 surge. By anticipating surges, appropriate public health measures, such as increased used of nonpharmaceutical interventions (masking, social distancing) or hospital capacity preparations, can be implemented to lessen or avert the consequences of a surge.

Given that this analysis was conducted with publicly available surveillance data, it is especially important to upkeep COVID-19 surveillance in order to be able to detect and anticipate surges.

References

[1] K. K. W. To *et al.*, "Lessons learned 1 year after SARS-CoV-2 emergence leading to COVID-19 pandemic," *Emerg. Microbes Infect.*, vol. 10, no. 1, pp. 507–535, 2021.

[2] World Health Organization. (2022). WHO Coronavirus (COVID-19) Dashboard [Dashboard]. Retrieved from https://covid19.who.int/

[3] X. Y. Huang, F. X. Wei, L. Hu, L. J. Wen, and K. Chen, "Epidemiology and Clinical Characteristics of COVID-19," *Arch. Iran. Med.*, vol. 23, no. 4, pp. 268–271, 2020.

[4] M. Lotfi, M. R. Hamblin, and N. Rezaei, "COVID-19: Transmission, prevention, and potential therapeutic opportunities," *Clin. Chim. ACTA*, vol. 508, pp. 254–266, 2020.

[5] Y. F. Jin *et al.*, "Virology, Epidemiology, Pathogenesis, and Control of COVID-19," *VIRUSES-BASEL*, vol. 12, no. 4, 2020.

[6] Johansson, M. A., Quandelacy, T. M., Kada, S., Prasad, P. V., Steele, M., Brooks, J. T., … Butler, J. C. (2021). SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Network Open*, *4*(1). https://doi.org/10.1001/JAMANETWORKOPEN.2020.35057

[7] Shoaib, N., Noureen, N., Munir, R., Shah, F. A., Ishtiaq, N., Jamil, N., … Zaidi, N. (2021). COVID-19 severity: Studying the clinical and demographic risk factors for adverse outcomes. *PLOS ONE*, *16*(8), e0255999. https://doi.org/10.1371/JOURNAL.PONE.0255999

[8] Clinical Spectrum | COVID-19 Treatment Guidelines. (n.d.). Retrieved April 18, 2022, from https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/

[9] Raveendran, A. V., Jayadevan, R., & Sashidharan, S. (2021). Long COVID: An overview. *Diabetes & Metabolic Syndrome*, *15*(3), 869. https://doi.org/10.1016/J.DSX.2021.04.007

[10] J. R. Fauver *et al.*, "Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States," *Cell*, vol. 181, no. 5, pp. 990-+, 2020.

[11] V. Chernozhukov, H. Kasahara, and P. Schrimpf, "Y Causal impact of masks, policies, behavior on early covid-19 pandemic in the US," *J. Econom.*, vol. 220, no. 1, pp. 23–62, 2021.

[12] C. Courtemanche, J. Garuccio, A. Le, J. Pinkston, and A. Yelowitz, "Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate," *Health Aff.*, vol. 39, no. 7, pp. 1237–1246, 2020.

[13] W. Lyu and G. L. Wehby, "Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US," *Health Aff.*, vol. 39, no. 8, pp. 1419–1425, 2020.

[14] Andrejko, K. L., Pry, J. M., Myers, J. F., Fukui, N., DeGuzman, J. L., Openshaw, J., … Jain, S. (2022). Effectiveness of Face Mask or Respirator Use in Indoor Public Settings for Prevention of SARS-CoV-2 Infection — California, February–December 2021. *MMWR. Morbidity and Mortality Weekly Report*, *71*(6), 212–216. https://doi.org/10.15585/MMWR.MM7106E1

[15] R. S. Loomba *et al.*, "Disparities in case frequency and mortality of coronavirus disease 2019 (COVID-19) among various states in the United States," *Ann. Med.*, vol. 53, no. 1, pp. 151–159, 2021.

[16] Bialek, S., Bowen, V., Chow, N., Curns, A., Gierke, R., Hall, A., … Wen, J. (2020). Geographic Differences in COVID-19 Cases, Deaths, and Incidence — United States, February 12–April 7, 2020. *MMWR. Morbidity and Mortality Weekly Report, 69*(15), 465–471. https://doi.org/10.15585/MMWR.MM6915E4

[17] L. van Dorp, C. J. Houldcroft, D. Richard, and F. Balloux, "COVID-19, the first pandemic in the post-genomic era," *Curr. Opin. Virol.*, vol. 50, pp. 40–48, 2021.

[18] B. B. Oude Munnink *et al.*, "The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology," *Nat. Med.*, vol. 27, no. 9, pp. 1518–1524, 2021.

[19] NATIONAL SARS-COV-2 strain surveillance (NS3). (n.d.). Retrieved April 18, 2022, from https://www.aphl.org/programs/preparedness/Crisis-Management/COVID-19-Response/Pages/Sequence-Based-Surveillance-Submission.aspx

[20] P. Paul *et al.*, "Genomic Surveillance for SARS-CoV-2 Variants Circulating in the United States, December 2020-May 2021," *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 70, no. 23, pp. 846–850, Jun. 2021.

[21] M. Schmidt *et al.*, "The evolving faces of the sars-cov-2 genome," *Viruses*, vol. 13, no. 9, 2021.

[22] SARS-COV-2 variant classifications and definitions. (2021, December 1). Retrieved April 18, 2022, from https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html

[23] S. Shiehzadegan, N. Alaghemand, M. Fox, and V. Venketaraman, "Analysis of the Delta Variant B.1.617.2 COVID-19," *Clinics and Practice* , vol. 11, no. 4. 2021.

[24] Fisman, D. N., & Tuite, A. R. (2021). Evaluation of the relative virulence of novel SARS-CoV-2 variants: a retrospective cohort study in Ontario, Canada. *CMAJ*, *193*(42), E1619–E1625. https://doi.org/10.1503/CMAJ.211248

[25] Del Rio, C., Malani, P. N., & Omer, S. B. (2021). Confronting the Delta Variant of SARS-CoV-2, Summer 2021. *JAMA*, *326*(11), 1001–1002.

[26] Chen, Calvin, Kaitlyn Gayvert, Richard Copin, Sheldon McKay, John Sikorski, and Regis James. "Top Variants Over Time." Regeneron covid-19 dashboard. Regeneron Pharmaceuticals Inc. Accessed January 2022. https://covid19dashboard.regeneron.com/?tab=Variant_Graphs&subTab=Top_Variants_Over_Time_%28Select_Location%29&Country=any&State=any.

[27] Elbe, Stefan, and Gemma Buckland-Merrett. "Data, Disease and Diplomacy: Gisaid's Innovative Contribution to Global Health." *Global Challenges* 1, no. 1 (January 10, 2017): 33–46. https://doi.org/10.1002/gch2.1018.

[28] NNDSS supports the COVID-19 response. (2021, September 24). Retrieved April 18, 2022, from https://www.cdc.gov/nndss/action/covid-19-response.html

[29] Centers For Disease Control and Prevention. "United States COVID-19 Cases and Deaths by State over Time," June 11, 2020.

[30] United States Census Bureau. Accessed March 16, 2022. https://www.census.gov/quickfacts/fact/table/US/PST045221.

[31] *Methodology For The United States Population Estimates: Vintage 2021* [PDF]. (2021, December). U.S. Department of Commerce.

[32] Centers for Disease Control and Prevention. "COVID-19 Vaccinations in the United States, Jurisdiction," May 24, 2021.

[33] Centers For Disease Control and Prevention. "Nationwide Blood Donor Seroprevalence Survey Infection-Induced Seroprevalence Estimates," September 27, 2021.

[34] Jones, Jefferson M., Mars Stone, Hasan Sulaeman, Rebecca V. Fink, Honey Dave, Matthew E. Levy, Clara Di Germanio, et al. "Estimated US Infection- and Vaccine-Induced SARS-COV-2 Seroprevalence Based on Blood Donations, July 2020-May 2021." *JAMA* 326, no. 14 (2021): 1400. https://doi.org/10.1001/jama.2021.15161.

[35] Centers For Disease Control and Prevention. "COVID-19 Case Surveillance Public Use Data with Geography," February 3, 2021.

[36] Biggs, H. M., Harris, J. B., Breakwell, L., Dahlgren, F. S., Abedi, G. R., Szablewski, C. M., … Zacks, R. (2020). Estimated Community Seroprevalence of SARS-CoV-2 Antibodies — Two Georgia Counties, April 28–May 3, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(29), 965–970. https://doi.org/10.15585/MMWR.MM6929E2

[37]Irons, N. J., & Raftery, A. E. (2021). Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(31). https://doi.org/10.1073/PNAS.2103272118/SUPPL_FILE/PNAS.2103272118.SAPP.PDF

[38] J. Martin *et al.*, "Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during COVID-19 Pandemic," *VIRUSES-BASEL*, vol. 12, no. 10, 2020.

[39] B. Mishra *et al.*, "High proportion of low cycle threshold value as an early indicator of COVID-19 surge," *J. Med. Virol.*

[40] S. Shah, S. X. W. Gwee, J. Q. X. Ng, N. Lau, J. Koh, and J. Pang, "Wastewater surveillance to infer COVID-19 transmission: A systematic review," *Sci. Total Environ.*, vol. 804, p. 150060, Jan. 2022.

[41] Borchering, R. K., Viboud, C., Howerton, E., Smith, C. P., Truelove, S., Runge, M. C., … Lessler, J. (2021). Modeling of Future COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Rates and Nonpharmaceutical Intervention Scenarios — United States, April–September 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*(19), 719–724. https://doi.org/10.15585/MMWR.MM7019E3

[42] Cramer, E. Y., Ray, E. L., Lopez, V. K., Bracher, J., Brennen, A., Castro Rivadeneira, A. J., … Reich, N. G. (2022). Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences*, *119*(15). https://doi.org/10.1073/PNAS.2113561119

[43] Covid-19 forecasts: Cases. (2022, March 23). Retrieved April 20, 2022, from https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/forecasts-cases.html

Appendix

IRB Exemption Form

EMORY UNIVERSITY
Institutional Review Board

NON-HUMAN SUBJECTS RESEARCH DETERMINATION FORM

Emory does not require IRB review of studies that do not meet the definitions of "human subjects research" (DHHS) or "clinical investigation" (FDA). This tool is to help you define your project and to ensure proper review and regulatory requirements are met.

If the tool results in an outcome of "no IRB review required," this form will serve as your documentation of that determination. Please keep the completed copy in your records.

AUDIT: The IRB will periodically audit completed forms and your written proposal to ensure that the tool is providing accurate results.

NOTE: this tool should only be used for projects completed by Emory/EHC affiliates doing work for Emory purposes. When answering the questions in this determination tool, consider only the project activities performed by Emory/EHC affiliates in the current proposed project (e.g. if your study is a secondary data analysis, do not include the primary data collection activities when considering your responses.) Emory/EHC affiliates who are completing a project for academic credit at a different institution should seek a determination from that institution's IRB.

...

Hi, Alisha. When you submit this form, the owner will see your name and email address.

* Required

1

Project Title *

Utilizing state-level surveillance data to detect changes in COVID-19 case incidence rates in SARS

2

PROJECT LEADER (not necessarily the person filling in this form) *

Alisha Kalangara

3

FUNDING *

Will these activities be supported by a DHHS award (e.g., NIH, NSF, DoE, DoD) through a grant, contract, subaward/subcontract, or cooperative agreement?

NOTE: If Emory is the prime recipient of a DHHS award and the funding application indicates that human subjects will be involved, IRB submission is required.

Also, if Emory is the prime recipient of a DHHS award, but contracting with another site to carry out all non-exempt human subjects research activities for that award, please contact the Emory IRB for guidance instead of using this form.
If Emory is the subrecipient, only the activities done by Emory should be considered for this form, even if other sites are performing human subjects research.

- ○ Yes
- ○ No

4

SHARING DATA OUTSIDE OF EMORY *

Will you be sharing data (identified or de-identified) outside of Emory? If yes, you need to contact OTT (ott@emory.edu) to determine if a Data Use Agreement is needed.

- ○ Yes

- ○ No

5

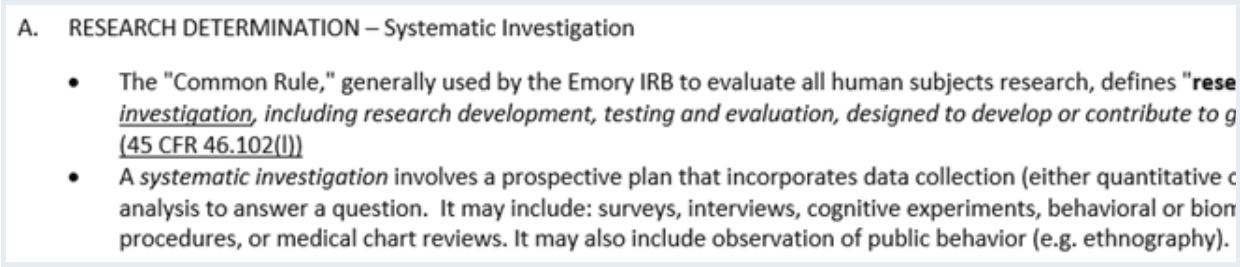
Does the project involve Veterans Affairs?
(e.g. study site, data source, researcher's affiliation) *

- ○ Yes
- ○ No

6

RESEARCH DETERMINATION- Systematic Investigation *

Is the proposed project a "systematic investigation?" For example: are you conducting online or in-person surveys, focus group discussions, or data analysis?



A. RESEARCH DETERMINATION – Systematic Investigation

- The "Common Rule," generally used by the Emory IRB to evaluate all human subjects research, defines "**rese** *investigation, including research development, testing and evaluation, designed to develop or contribute to g* (45 CFR 46.102(l))
- A *systematic investigation* involves a prospective plan that incorporates data collection (either quantitative c analysis to answer a question. It may include: surveys, interviews, cognitive experiments, behavioral or biom procedures, or medical chart reviews. It may also include observation of public behavior (e.g. ethnography).

- ○ No
- ◉ Yes

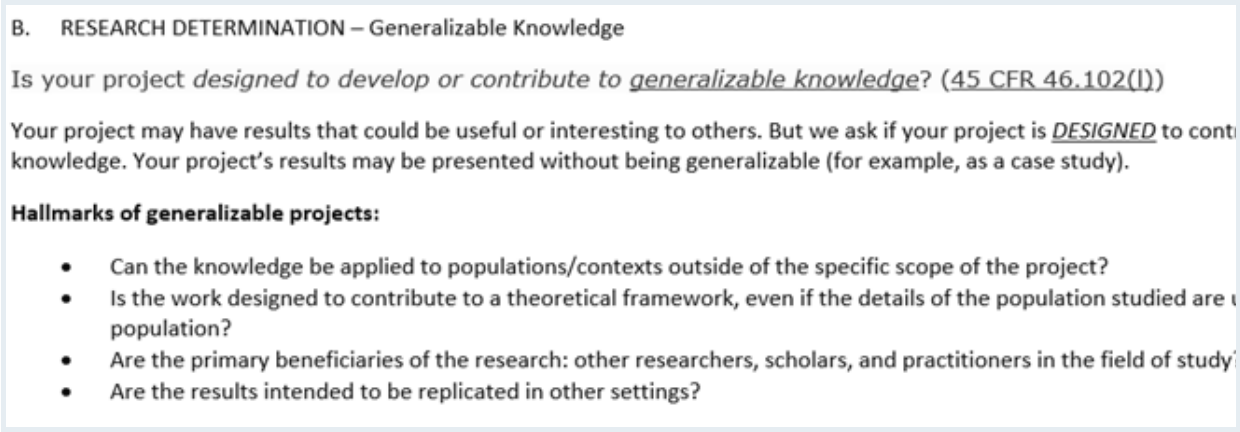
7

RESEARCH DETERMINATION- Generalizable Knowledge
Is the proposed project "designed to develop or contribute to generalizable knowledge?" *

Review these links if your project falls into one of the following categories:
Case Studies/Series (http://irb.emory.edu/forms/review/casestudy.html)
Classroom Activities (http://irb.emory.edu/forms/review/classroom.html)
Public Health Practice (http://irb.emory.edu/forms/review/PH.html)

Program Evaluations (http://irb.emory.edu/forms/review/programeval.html)
Quality Improvement (http://irb.emory.edu/forms/review/QI.html)
Sociobehavioral research: Oral History/Journalism and Ethnography/Anthropology (http://irb.emory.edu/forms/socio.html)

If you still have questions, you can call our office for clarification at (404) 712-0720.



B. RESEARCH DETERMINATION – Generalizable Knowledge

Is your project *designed to develop or contribute to generalizable knowledge*? (45 CFR 46.102(l))

Your project may have results that could be useful or interesting to others. But we ask if your project is *DESIGNED* to cont knowledge. Your project's results may be presented without being generalizable (for example, as a case study).

Hallmarks of generalizable projects:

- Can the knowledge be applied to populations/contexts outside of the specific scope of the project?
- Is the work designed to contribute to a theoretical framework, even if the details of the population studied are population?
- Are the primary beneficiaries of the research: other researchers, scholars, and practitioners in the field of study
- Are the results intended to be replicated in other settings?

- (●) Yes
- ( ) No

8

HUMAN SUBJECTS DETERMINATION *

Does this study involve obtaining information about living individuals? Answer "yes" if you're obtaining de-identified data or anonymous survey results if the results contain information about living people.

- (●) Yes
- ( ) No

9

If yes, does the study involve intervention or interaction with the individuals (e.g., online or in-person surveys [even if generating anonymous results], prospective collection of specimens, scans, etc.)?

- ○ Yes
- ○ No

10

Do the activities involve accessing or generating individually identifiable and private information about living individuals?

Please review the list of identifiers for more information (http://www.irb.emory.edu/documents/phi_identifiers.pdf)

- ○ Yes
- ◉ No

11

Does the study involve analysis of existing data/specimens, where ALL data and/or specimens already exist prior to the start of the study? (Important: all parts of this question must apply if answering Yes.)

- ◉ Yes
- ○ No

12

If yes, would ANY member of the research team be able to reidentify the data/specimens, either directly, or via a code and key?
* If anyone on the newly-proposed study team took part in the original collection of the existing specimens or data, your should answer Yes.
* If there are codes on the data, but no one on the study team has access to a link: you may answer "No" to this question only if you have a documented agreement with the data/specimen providers that prohibits your team from having access to the link.

- ○ Yes
- ○ No

13

HUMAN SUBJECTS RESEARCH DETERMINATION - FDA

Will any individual be a recipient of any test article (i.e., drug, medical device) or be used as a control?

FDA 21 CFR 56.102 (23c&e)

Human Subject- an individual who is or becomes a participant in research, either as a recipient of the test articl
Clinical Investigation- any experiment that involves a test article and one or more human subjects.

- ○ Yes
- ○ No

14

Will any device be tested (including software, apps, in-vitro assays) using any individual's specimens or data, even if completely deidentified?

- ○ Yes
- ○ No

15

This project does not require IRB review because it is not research with "human subjects", nor is it a "clinical investigation" as defined in the federal regulations. Please use the Microsoft Print to PDF or Microsoft XPS Document Writer option to save a copy of your responses to this form. *

- (●) There is no eIRB submission necessary. I will protect the confidentiality of information accessed or obtained in this project. I will keep a copy of my responses to this form for my records.

☑ Send me an email receipt of my responses

Submit