

# Emergence to dominance: Estimating time to dominance of SARS-CoV-2 variants using nonlinear statistical models

Srishti Awasthi<sup>1,✉</sup>, Maryam Zolfaghari Dehkharghani<sup>1,✉</sup>, Miguel Fudolig<sup>2\*</sup>,

**1** Department of Healthcare Administration and Policy, School of Public Health, University of Nevada Las Vegas, Las Vegas, NV, USA, 89119

**2** Department of Epidemiology and Biostatistics, School of Public Health, University of Nevada Las Vegas, Las Vegas, NV, USA, 89119

✉These authors contributed equally to this work.

\* miguel.fudolig@unlv.edu

## Abstract

**Background/Objective:** Relative proportion of cases in a multi-strain pandemic like the COVID-19 pandemic provides insight on how fast a newly emergent variant dominates the infected population. However, the behavior of relative proportion of emerging variants is an understudied field. We investigated the emerging behavior of dominant COVID-19 variants using nonlinear statistical methods and calculated the time to dominance of each variant.

**Method:** We used a phenomenological approach to model national- and regional-level variant share data from the national genomic surveillance system provided by the Centers for Disease Control and Prevention to determine the best model to describe the emergence of two recent dominant variants of the SARS-CoV-2 virus: XBB.1.5 and JN.1. The proportions were modeled using logistic, Weibull, and generalized additive models. Model performance was evaluated using the Akaike Information Criteria (AIC) and the root mean square error (RMSE).

**Findings:** The Weibull model performed the worst out of all three approaches. The generalized additive model approach slightly outperformed the logistic model based on fit statistics, but lacked in interpretability compared to the logistic model. These models were then used to estimate the time elapsed from emergence to dominance in the infected population, denoted by the time to dominance (TTD). All three models yielded similar TTD estimates. The XBB.1.5 variant was found to dominate the population faster compared to the JN.1 variant, especially in HHS Region 2 (New York) where the XBB.1.5 was believed to emerge. This research expounds on how emerging viral strains transition to dominance, informing public health interventions against future emergent COVID-19 variants and other infectious diseases.

## Introduction

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, a member of the coronavirus family, has had a significant impact on global health, economy, and societies. Coronaviruses, including COVID-19, Middle East Respiratory Syndrome (MERS), and Severe Acute Respiratory Syndrome (SARS), are a group of respiratory viruses that can cause diseases in animals and humans. They are named after the crown-like spikes on their surface. Human coronaviruses were first discovered in the

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

mid-1960s, and public health officials closely monitor them. The SARS-CoV-2 virus, prone to genetic alterations over time, experiences mutations during replication. Since the emergence of the SARS-COV-2 virus, it has undergone dynamic evolution characterized by various changes resulting in the emergence of different strains. The ongoing mutation of SARS-CoV-2, a characteristic of RNA viruses, occurs due to errors in viral replication, as RNA polymerases lack proofreading mechanisms. These genetic changes lead to the emergence of new variants, some of which have significant public health implications due to increased transmissibility or severity of illness [1]. These mutations can cause an emergence of new variants that have different characteristics such as vaccine resistance, immunity resistance and reinfection. Each strain has distinct traits and health concerns of its own. Since the start of the pandemic, SARS-CoV-2 has shown itself to be capable of genetic evolution. The ongoing mutation of SARS-CoV-2, a characteristic of RNA viruses, happens because mistakes in viral replication are not fixed as RNA polymerases do not have inspection mechanisms. The genetic heterogeneity of the virus enables it to adapt and perhaps improve its ability to survive and transmit. In order to address this issue, it is imperative that we closely monitor these mutations and adjust our interventions accordingly.

## Recently Emerged COVID-19 Variants

Since the onset of the pandemic, several variants of SARS-CoV-2 have been identified and classified by health authorities as Variants of Interest (VOIs) or Variants of Concern (VOCs). Notable variants include Alpha (B.1.1.7), first discovered in the United Kingdom, Beta (B.1.351) from South Africa, Delta (B.1.617.2) from India, and Omicron (B.1.1.529) in Botswana and South Africa each showing unique traits and health impacts [1,2].

The Omicron variant (B.1.1.529) emerged in late 2021 in Botswana and South Africa and quickly spread all over the world. The omicron variant of the SARS-COV-2 virus was notable for its numerous spike protein mutations, raising concerns about immune evasion and vaccine effectiveness [3,4]. Two of the most recent variants that dominated the infected population in the United States (US) are XBB.1.5, also known as the Kraken variant, and JN.1. The XBB.1.5 variant is a descendent lineage of XBB, which is a recombinant of Omicron BA.2.10 and BA.2.75 descendant lineages [5–7]. The XBB.1.5 emerged in New York City and rapidly spread in the region in November-December 2022 and was identified to be responsible for most of the cases in the national level in early 2023. On the other hand, the JN.1 variant is the offspring of the Omicron BA.2.86 variant was first identified in the US in September 2023 and has already become the dominant variant of COVID-19 infections in the US. The mathematical modeling of the emergence and rapid spread of the XBB.1.5 and JN.1 variants remains a significant gap in the current research landscape. As of press time, there have not been any published research on the mathematical modeling of the JN.1 strain in the United States. Cheng and colleagues [8] used a multi-strain SIR model and variant proportion data to estimate the transmission rates and reproduction number of the XBB strain using surveillance data and variant proportions data. One of their recommendations was to include XBB.1.5 as a separate compartment as it dominated other XBB subvariants. However, there was no discussion on how quickly the XBB.1.5 dominated the XBB infections compared to other strains. Although there is a plethora of studies on the mathematical modeling of multi-strain epidemics, most of these studies are concerned with calculating transmissibility coefficients [9,10], estimating reproduction numbers [9,11,12], and performing stability analyses [13,14]. Current epidemic models often focus on well-established parameters, neglecting a crucial element: the emergence and dominance of highly transmissible COVID-19 variants. Understanding how these new variants arise and outcompete existing variants is essential for accurate modeling in

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

multi-strain epidemics. Furthermore, it is vital to analyze the behavior of these dominant variants within the US population to effectively predict future spread and inform public health interventions. The relative rate of emergence of each variant can be monitored using variant proportion data from genomic surveillance, which is provided by the Centers for Disease Control and Prevention (CDC) [15] and the Global Initiative on Sharing Avian Influenza Data [16], and wastewater surveillance [17, 18]. An important parameter in describing the rate of dominance of a newly emerged variant is the time required for an emergent variant to reach the majority status in the infected population, also known as the time to dominance (TTD).

## Time to Dominance

With the rapid mutation of SARS-CoV-2 in the United States general population, there is a critical need to determine how fast a newly emerged variant dominates the infected population. Newly emerged variants might exhibit vaccine resistance, increased transmissibility, and higher mortality rates. Once a newly emerged variant with these characteristics make up the majority of the infections, stricter public health policies might need to be reinstated to curb the spread of these variants. For instance, mask mandates were reinstated in some parts of the United States in early 2022 because of the emergence of the Omicron variant [19, 20]. Public health officials have limited time to decide on the best strategy to characterize and devise strategies against these new emergent variants. One benchmark we can to measure the speed of emergence of a viral strain is the time to dominance (TTD). We define the TTD as the time it takes for an emergent viral strain to make up the overall majority, i.e. at least 50%, of the infected cases in a multi-strain epidemic. Fudolig [21] performed a simulation-based experiment to test the effect of vaccination and transmission on the TTD on a two-strain epidemic model. A three-parameter logistic growth model was used to describe the increase of variant share in the simulated epidemics and estimate TTD. Simulations showed that a more transmissible emergent strain relative to the existing strains was found to dominate the infected population faster. In addition, higher vaccination rates and coverages could lead to lower TTD. However, this method was not applied to data from real-world multi-strain epidemics such as COVID-19. While the use of the logistic model to estimate the TTD values for each simulation was sufficient for simulated data, other growth models such as a generalized logistic or Weibull growth models might yield a better fit and explain the emergence behavior better than logistic models. Semi-parametric approaches such as generalized additive models would also be a great option in modeling the non-linear growth of rapidly emerging SARS-CoV-2 variants. To the best of our knowledge, there has not been any prior research that compared the model performance of different models of COVID-19 variant proportion shares.

## Objectives and Significance of the Study

The JN.1 and XBB.1.5 variants were two of the most recently emerged variants to record variant shares of over 50% based on genomic surveillance data. This study estimated the duration required for the JN.1 and XBB.1.5 variants to dominate the COVID-19 infected population in the United States. Specifically, we focused on modeling the emergence of these variants using the following modeling approaches: the logistic growth model, the Weibull model, and the generalized additive model (GAM). These models were then used to estimate the TTD based on variant proportion data reported by the CDC from 2021-2024 [15]. While the CDC provides model-based projected estimates of the variant share for the weeks when samples are being processed, the underlying data-driven model is not provided by the CDC. The results of this study presents preliminary data that would inform future studies on the emergence of new

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

variants in multi-strain epidemic models. Only the weighted estimates of the variant share during reported dates were modeled in this study. The confidence limits provided by the CDC were not analyzed. By investigating the behavior of variant proportion shares of COVID-19, we addressed the literature gap on the behavior leading to the dominance of emergent variants in multi-variant epidemics.

## Materials and methods

### Dataset

COVID-19 variant proportion data was downloaded from the CDC website [15] on May 7, 2024. The variant proportions were calculated based on sequenced samples collected from different regions of the US. The regional division was defined by the United States Department of Health and Human Services (HHS). The member states of each region can be accessed through the following link:

<https://www.hhs.gov/about/agencies/iea/regional-offices/index.html>. The dataset included variant proportions every two weeks from January 2022 to April 2024. We investigated the most recent SARS-CoV-2 variants, JN.1 and XBB.1.5, that were reported to have a majority of COVID-19 cases nationwide.

## Theory and Calculation

### Variant Proportions

**Fig 1.** National- and regional-level variant proportion data of the JN.1 subvariant.

Historically, there are multiple SARS-CoV-2 variants that coexist in different regions of the US. Fig 1 shows the estimated proportion share of the JN.1 variant as reported by the CDC. These estimates were recorded every two weeks based on empirical genomic sequencing data. The share of SARS-CoV-2 variant proportions appear to follow a non-linear trend in time. We modeled this non-linear trend using three different approaches: generalized additive models (GAM), logistic function curve fit, and Weibull function curve fit.

### Modeling Non-linear Emergence

#### Generalized Additive Models

Generalized additive models (GAM) are generalized linear models that utilize smoothing splines to accurately model non-linear trends [22]. GAMs are made up of a linear combination of a linear predictor and smoothing functions to describe any smooth monotonic curves, making it ideal to model variant proportion data. An intercept-only GAM was used in this study which had the following form:

$$f(t) = \beta_0 + s(t), \tag{1}$$

where  $\beta_0$  is the model intercept and  $s(t)$  is the smooth function of the time  $t$ . The semi-parametric nature of GAMs make it difficult to provide confidence intervals of inverse estimates such as the case of estimating TTD values, which will be explained in the section . The `mgcv` package in R was used to fit a GAM on the CDC variant proportion share data.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

## Logistic and Weibull Models

We fit non-linear functions such as the logistic and Weibull functions to model the non-linear increase of JN.1 and XBB.1.5 variant proportions upon emergence. Fudolig [21] previously used the logistic function to model the increase in emerging variant proportion in a multi-strain epidemic in calculating TTD values. The five-parameter generalized logistic function can be expressed as

$$f(t) = c + \frac{d - c}{[1 + \exp(b(t - e))]^f}, \quad (2)$$

where the time  $t$  will be measured in days from the first emergence.  $b$  is the scale factor,  $c$  and  $d$  are the respective upper and lower asymptotes of the logistic curve,  $e$  is the inflection point, and  $f$  is the asymmetry factor. The logistic function is also known as the Boltzmann sigmoidal function [23]. A closely related model to the logistic function is the Weibull function. The Weibull function is another function typically used to model growth curves that provides more flexibility in modeling non-monotonic functions. The Weibull function can be expressed as

$$f(t) = c + (d - c)(\exp[-\exp(b(\log(t) - e))]), \quad (3)$$

where  $b$  is a scale factor,  $c$  is the lower asymptote,  $d$  is the upper asymptote, and  $e$  is the point of inflection. Both models have been used to describe growth of the infected population in epidemics [24–28], but there is a literature gap in using these models to describe variant proportion shares. Curve fitting for both functions can be implemented using the R package `drc`, which is commonly used to model dose response curves [23]. Even though both logistic and Weibull models are lacking in flexibility compared to the GAM, both logistic and Weibull models offer easily interpretable results that could translate to action items for policy makers and public health officials.

## Estimating the Time to Dominance (TTD, $t_D$ )

The time  $t = 0$  was set two weeks before a share greater than 0.01% was reported for the variant. Based on the criterion for  $t = 0$  being the reporting date before the variant proportion share was reported to be above 0.01%, the start of the emergence occurred on October 15, 2022 for XBB.1.5 and Sept 2, 2023 for JN.1. All models were implemented from  $t = 0$  to  $t_{MAX}$ , the time at which the variant proportion share is maximum for each region and variant. The fit of each model was assessed using the Akaike information criteria (AIC) and the root mean square error (RMSE). Lower AIC and RMSE values imply a better model fit on the data.

The values of  $t_D$  were estimated for regional and national level data. The TTD value, denoted by  $t_D$ , were estimated for each model by numerically solving for  $t_D$  such that  $f(t = t_D) = 0.5$ . The R packages `drc` and `mgcv` were used to implement the logistic, Weibull, and generalized additive models on the regional and national variant proportion estimates from the CDC. We wish to clarify that the  $t_D$  does not equate to ED50 as neither JN.1 and XBB.1.5 reached a maximum share equal to 1. The root finding function `uniroot` in R was used to estimate  $t_D$  for both JN.1 and XBB.1.5. The TTD estimates obtained from each model were compared and analyzed for each region.

## Results and Discussion

### Model Performance

Figs 2 and 3 show the graph of the absolute values of the AIC and RMSE values for each model as applied to the national and regional COVID-19 variant proportion data.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Lower AIC and RMSE values suggest a better model fit. The Weibull model appears to perform the worst out of the three models, while the logistic model and GAM performed comparably well in estimating the variant proportion curve during the emergence of the two variants. 187  
188  
189  
190

**Fig 2.** The magnitude of the AIC values (-AIC) for each fit function plotted for the national and regional variant proportions data set. All calculated AIC values were negative. A higher magnitude of the AIC corresponds to a better fit.

**Fig 3.** The root mean square error (RMSE) values for each fit function plotted for the national and regional variant proportions data set. A lower magnitude of the RMSE corresponds to a better fit.

The GAM outperformed both logistic and Weibull curve fits in modeling the emergence of both variants at the national level. Figs 4 and 5 show the three different models superimposed with the actual variant share data for XBB.1.5 and JN.1, respectively. The GAM and logistic model cannot be visually perceived implying very similar fit. Upon visual inspection, the Weibull model seems to underestimate the "knee" of the sigmoid curve of the XBB.1.5 variant share. The Weibull model fails to correctly account for the gradual increase of the variant share at  $t = 70$  and  $t = 84$ , but appears to intersect the other two models close to the line  $y = 0.5$ . This behavior is reflected in Fig 3 which displays how the RMSE of the Weibull model is significantly higher compared to the other two models. 191  
192  
193  
194  
195  
196  
197  
198  
199  
200

**Fig 4.** The actual variant proportion data for XBB.1.5 in the US, shown as black dots, plotted with the GAM (black line), logistic model (red), and Weibull model (green) fitted values. The blue dashed line marks the 50% share level used for calculating the TTD values.

**Fig 5.** The actual variant proportion data for JN.1 in the US, shown as black dots, plotted with the GAM (black line), logistic model (red), and Weibull model (green) fitted values. The blue dashed line marks the 50% share level used for calculating the TTD values.

This large discrepancy in model performance is also observed in the regional level. The Weibull model performed relatively poorly in modeling the emergence of JN.1 and XBB.1.5 in all regions. Fig 6 displays the performance of the three models against the actual XBB.1.5 variant proportion data in HHS Region 5 (Chicago), which also shows the Weibull underestimation at  $t = 70$  and  $t = 84$  and the similarity between GAM and logistic models observed in the national data. The same trend was observed in the JN.1 variant proportion as shown in Fig 7, which includes the three models and the actual JN.1 variant share data in HHS Region 1 (Boston). 201  
202  
203  
204  
205  
206  
207  
208

**Fig 6.** The actual variant proportion data for XBB.1.5 in HHS Region 5, shown as black dots, plotted with the GAM (black line), logistic model (red), and Weibull model (green) fitted values. The blue dashed line marks the 50% share level used for calculating the TTD values.

It is noteworthy that the Weibull model underestimation at the "knee", specifically  $t = 70$  and  $t = 84$ , occur for both strains at the national level. For XBB.1.5, the dates 209  
210

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

**Fig 7.** The actual variant proportion data for JN.1 in HHS Region 1, shown as black dots, plotted with the GAM (black line), logistic model (red), and Weibull model (green) fitted values. The blue dashed line marks the 50% share level used for calculating the TTD values.

corresponding to  $t = 70$  and  $t = 84$  are December 24, 2022 and January 7, 2023, respectively. During these weeks, festive holidays such as Christmas, Hanukkah, and New Year's Eve often entail numerous social gatherings and meetups that could have potentially lead to super-spreader events. After these dates, the variant proportion increased before plateauing at the maximum which was recorded in mid-March (March 18, 2023). As for JN.1,  $t = 70$  and  $t = 84$  occurred on November 11, 2023 and November 25, 2023. These weeks include Thanksgiving 2023, which was celebrated in the US on November 17, 2023. Like the aforementioned holidays, Thanksgiving is a holiday in the US that involves social and family gatherings all over the country. It is possible that the Weibull model could not account for the sudden increase in variant share leading up to reaching the simple majority of the variant proportion cases for both cases.

The poor performance of the Weibull model in this study contrasts with previous studies that assess the performance of the Weibull function in modeling COVID-19 incidence, prevalence, and mortality data. Attanayake and colleagues [29] determined phenomenologically that the Weibull growth curve performed the best in modeling the cumulative number of COVID-19 infections in the US from the case of first appearance to July 2, 2020. Al-Ani and colleagues [24] also found that the Weibull model performed the best in modeling the cumulative number of COVID-19 cases in Saudi Arabia. There is a plethora of research that use modified Weibull distributions to model daily and cumulative number of cases and deaths of COVID-19 patients from all over the world [28, 30, 31]. Despite its wide usage in modeling other aspects of the COVID-19 pandemic, the Weibull function appears to perform underwhelmingly in modeling COVID-19 variant proportion share data in the United States. Based on our findings, the logistic or generalized additive models are more recommended to use in modeling the share of emerging COVID-19 variants that achieve majority dominance in the United States. While both logistic model and GAM perform similarly, the two models provide different insights about the curve. The logistic model provides more information on the shape of the curve, i.e. asymptotes, slopes, and inflection point. Moreover, it can provide an estimate of the share at any given time. Its limitations lie when we want to model the behavior of the variant proportion share past the emergence phase. After COVID-19 variants reach their maximum share, the variant share would naturally start decreasing as the number of cases decrease. This decrease can be naturally observed or could have been caused by the emergence of a more transmissible variant. The GAM approach would be more useful if a researcher intends to investigate the entire variant share curve. GAMs provides precise estimates for variant proportion shares during and after the emergence phase without constraint. While the GAM is generally the better performing model, the model does not have an interpretable closed form because of the general nature of the smoothing function  $s(t)$ . It is essential to establish specific research questions to determine which of the two models should be used in analyzing variable proportion shares. In the next section, we illustrate how these models can be used to measure the time to dominance of the XBB.1.5 and JN.1 strains, which is an important aspect in characterizing a variant's emergent pattern.

## Model Application: Estimating TTD ( $t_D$ )

Fig 8 shows the estimates of the  $t_D$  values for the two variants for all models as applied to national and regional level data. While there are discrepancies between the models

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

**Fig 8.** National- and regional-level estimates of the  $t_D$  values for JN.1 and XBB.1.5.

with respect to their AIC and RMSE values, the differences between  $t_D$  estimates of the three models are quite low. The highest difference between the  $t_D$  estimates from each model was 1.70 days for JN.1 (HHS Region 10, GAM vs. Weibull) and 2.04 days for XBB.1.5 (HHS Region 4, GAM vs. Weibull). In most cases, the GAM estimates are slightly higher compared to both logistic and Weibull estimates. Even though it did not perfectly capture the entire growth curve, the Weibull model's performance in estimating  $t_D$  was promising for JN.1 and XBB.1.5. This observation was not surprising as Figs 4 to 7 showed the three curves converging close to  $y = 0.5$  as mentioned in Model Performance section in the Results. For ease of reporting from this point onwards, we averaged the  $t_D$  estimates from all three models.

**Fig 9.** Side-by-side comparison of  $t_D$  values for JN.1 and XBB.1.5 evaluated for each model.

Fig 9 shows the side-by-side comparison of the  $t_D$  values for both JN.1 and XBB.1.5 variants in each region. The XBB.1.5 variant yielded lower time to dominance estimates across all regions except HHS Region 9 (San Francisco), in which the JN.1 variant had a slightly lower estimate. Even though the CDC insinuated that the JN.1 is more transmissible compared to other variants present in December 2023, the CDC did not find an increased risk to public health. Individuals who have updated vaccinations were also reported to be protected from the JN and XBB variants [32]. HHS Region 2 (New York) yielded the lowest  $t_D$  value for both XBB.1.5 and JN.1 with respective  $t_D$  estimates of 79.28 and 107.82 days. On the other hand, the highest  $t_D$  GAM estimates for XBB.1.5 was recorded to be in HHS Region 10 (Seattle) at 124.73 days. It is important to highlight that the time to dominance for HHS Region 2 was typically the time where the "knee" of the logistic curve occurred. We can observe in Fig 10 that the "knee" appeared earlier compared to other regions shown in Figs 4 to 7. According to Luoma [5], New York could be the epicenter of the XBB.1.5 rapid spread which can explain the lower time to dominance in HHS Region 2 and surrounding HHS regions. HHS Regions 1 (Boston,  $t_D = 88.58$  days) and 3 (Philadelphia,  $t_D = 95.97$ ) closely follow as second and third lowest TTD values. On the other hand, HHS Region 10 (Seattle) is located in the opposite coast of the country and has a lower population density than HHS Regions 1,2, and 3, which could have contributed to a slower spread of the variant. Meanwhile, the highest  $t_D$  estimates for JN.1 was observed in HHS Region 5 at 125.63 days. The rapid dominance of JN.1 and XBB.1.5 variants in eastern HHS Regions (1, 2, and 3) likely stems from the extensive interconnectivity of urban centers in these regions. Air, land, and water transport between these urban centers are possible, which could not be said for other HHS regions.

**Fig 10.** The actual variant proportion data for XBB.1.5 in HHS Region 2, shown as black dots, plotted with the GAM (black line), logistic model (red), and Weibull model (green) fitted values. The blue dashed line marks the 50% share level used for calculating the TTD values.

## Limitations of the Study

The phenomenological approach of the study provides insight on the temporal behavior of variant proportions for XBB.1.5 and JN.1. However, we must remain cautious of generalizing this approach to previous dominant variants like the Alpha, Delta, and



It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

early Omicron variants because of the different public health interventions in place at the time of emergence. XBB.1.5 and JN.1 emerged after vaccines were provided to the public, mask mandates were dropped, and social distancing protocols were not enforced. We directed our study to the two most variants as they are the more relevant to current public health officials dealing with relaxed interventions. A separate study must be done to investigate the time to dominance for earlier variants.

We would also emphasize that these trends were analyzed from the genomic surveillance data of COVID-19 in the United States provided by the CDC. While this study focused on genomic sequencing data, future research will incorporate estimated variant proportions from wastewater surveillance. This rich data source holds promise for revealing potentially divergent trends in newly emerged COVID-19 variants.

## Conclusion

We have modeled the variant share data during the emergence of two variants that dominated the COVID-19 infected population in the US most recently: XBB.1.5 and JN.1. Logistic, Weibull, and generalized additive models (GAMs) were considered for both national and regional-level data for the proportion of confirmed XBB.1.5 and JN.1 cases provided by the CDC. After evaluating model performance based on AIC and RMSE values, the Weibull model was determined to perform the worst among the three models. The logistic and GAM approaches yielded similar results, with GAM providing slightly lower RMSE values. The advantage in model performance and versatility provided by GAM could be compensated with the interpretability of the logistic model in modeling variant emergence in a multi-variant epidemic such as COVID-19.

We were also able to calculate the time to dominance (TTD)  $t_D$ , which measures the time required for the variants to reach majority status in the infected population. Despite its subpar performance in modeling variant emergence, the  $t_D$  estimates from the Weibull model did not deviate largely from the  $t_D$  estimates from the logistic model and GAM. We also determined that the TTD was the lowest in HHS Region 2 (New York), which indicates a faster spread of XBB.1.5 and JN.1 during emergence compared to other regions. The dominance of JN.1 and XBB.1.5 variants concentrated in eastern HHS Regions (1, 2, and 3) likely stems from these regions' unique characteristics. These regions boast densely populated urban centers extensively interconnected by air, land, and water transportation networks, facilitating rapid viral spread compared to other regions, where public transport linking urban centers is limited. The TTD estimates for both variants ranged from 2.6 to 4 months depending on the region, which could be used as an early temporal checkpoint to assess whether current strategies against COVID-19 infections should change to fight against these new variants.

We also observed how the timing of the "knee" of the variant share curve could affect whether new variants could dominate the population. The rapid increase of both XBB.1.5 and JN.1 variant shares occurred during holiday seasons in which people in the US typically gather in large groups, which might have made it easier for these variants to dominate. This temporal association suggests that holiday gatherings might act as a catalyst in decreasing the TTD of new variants in the US. We could use the findings of this study to analyze the similarities in the emergence of the FLiRT variants (KP.2, KP.3, KP.1.1) compared to the previously dominant variants such as JN.1 and XBB.1.5 in this period where COVID-19 policies is less stringent compared to the policies during the pandemic. While none of these FLiRT variants have a 50% share of the cases as of press time, the results of the study could inform which one of these variants could dominate the infected population in the US in the following weeks, especially after the independence day weekend celebrations.

In addition to researching future emerging COVID-19 variants, further research is

needed to determine relative model performance between logistic, Weibull, and GAM approaches hold for other variants of interest (VOIs) such as BA.2.86 that achieved majority status before the relaxation of COVID-19 public health and social measures. We would also like to include VOIs that did not achieve majority status such as EG.5 and HV.1 in future work. Expanding the study on data from outside the United States is also recommended to enable us to compare different optimal control strategies in containing the spread of newly emergent COVID-19 variants all around the world.

## Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Acknowledgments

The authors would like to acknowledge the University of Nevada, Las Vegas School of Public Health for their provided support in this study.

## References

1. Vasireddy D, Vanaparthi R, Mohan G, Malayala SV, Atluri P. Review of COVID-19 Variants and COVID-19 Vaccine Efficacy: What the Clinician Should Know? *Journal of Clinical Medicine Research*. 2021;13(6).
2. Musa SS, Gyeltshen D, Manirambona E, Ayuba D, Lucero-Prisno DE. The new COVID-19 omicron variant: Africa must watch its spread! *Clinical Epidemiology and Global Health*. 2022;13:100961. doi:<https://doi.org/10.1016/j.cegh.2022.100961>.
3. Ao D, Lan T, He X, Liu J, Chen L, Baptista-Hon DT, et al. SARS-CoV-2 Omicron variant: Immune escape and vaccine development. *MedComm*. 2022;3(1):e126. doi:10.1002/mco2.126.
4. Shrestha LB, Foster C, Rawlinson W, Tedla N, Bull RA. Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. *Reviews in Medical Virology*. 2022;32(5):e2381. doi:10.1002/rmv.2381.
5. Luoma E. Notes from the field: epidemiologic characteristics of SARS-CoV-2 recombinant variant XBB. 1.5—New York City, November 1, 2022–January 4, 2023. *MMWR Morbidity and mortality weekly report*. 2023;72.
6. Gámbaro F, Duerr R, Dimartino D, Marier C, Iturrate E, Mulligan MJ, et al. Emergence and dissemination of SARS-CoV-2 XBB.1.5 in New York. *Virus Evolution*. 2024;10(1):veae035. doi:10.1093/ve/veae035.
7. Graham F. Daily briefing: Is subvariant XBB.1.5 a global threat? *Nature*. 2023;doi:10.1038/d41586-023-00052-x.
8. Cheng Z, Lai Y, Jin K, Zhang M, Wang J. Modeling the XBB strain of SARS-CoV-2: Competition between variants and impact of reinfection. *Journal of Theoretical Biology*. 2023;574:111611. doi:10.1016/j.jtbi.2023.111611.

9. Lazebnik T, Bunimovich-Mendrazitsky S. Generic approach for mathematical model of multi-strain pandemics. *PLOS ONE*. 2022;17(4):e0260683. doi:10.1371/journal.pone.0260683. 383  
384  
385
10. Yagan O, Sridhar A, Eletreby R, Levin S, Plotkin JB, Poor HV. Modeling and analysis of the spread of COVID-19 under a multiple-strain model with mutations. *Harv Data Sci Rev*. 2021;4. 386  
387  
388
11. de León UAP, Avila-Vales E, Huang Kl. Modeling COVID-19 dynamic using a two-strain model with vaccination. *Chaos, Solitons & Fractals*. 2022;157:111927. doi:10.1016/j.chaos.2022.111927. 389  
390  
391
12. Tchoumi SY, Rwezaura H, Tchuenche JM. Dynamic of a two-strain COVID-19 model with vaccination. *Results in Physics*. 2022;39:105777. doi:10.1016/j.rinp.2022.105777. 392  
393  
394
13. Fudolig M, Howard R. The local stability of a modified multi-strain SIR model for emerging viral strains. *PloS one*. 2020;15(12):e0243408. 395  
396
14. Sharma N, Rana PS, Negi SS. Mathematical modeling and case study analysis for COVID-19 pandemic in India. In: *AIP Conference Proceedings*. vol. 2435. AIP Publishing; 2022. 397  
398  
399
15. Centers for Disease Control and Prevention. COVID Data Tracker; 2020. Available from: <https://covid.cdc.gov/covid-data-tracker>. 400  
401
16. Global Initiative on Sharing Avian Influenza Data. GISAID - hCov19 Variants; 2024. Available from: <https://gisaid.org/hcov19-variants/>. 402  
403
17. Shah S, Gwee SXW, Ng JQX, Lau N, Koh J, Pang J. Wastewater surveillance to infer COVID-19 transmission: A systematic review. *Science of The Total Environment*. 2022;804:150060. doi:10.1016/j.scitotenv.2021.150060. 404  
405  
406
18. Farkas K, Hillary LS, Malham SK, McDonald JE, Jones DL. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Current Opinion in Environmental Science & Health*. 2020;17:14–20. doi:10.1016/j.coesh.2020.06.001. 407  
408  
409  
410
19. Meng F. This Week in COVID-19: Cases rise following spring break, Tufts reinstates mask mandate-The Tufts Daily. *UWIRE Text*. 2022; p. 1–1. 411  
412
20. Wang N, Xue J, Xu T, Li H, Liu B. A weapon to fight against pervasive Omicron: systematic actions transiting to pre-COVID normal. *Frontiers in Public Health*. 2023;11:1204275. 413  
414  
415
21. Fudolig M. Effect of Transmission and Vaccination on Time to Dominance of Emerging Viral Strains: A Simulation-Based Study. *Microorganisms*. 2023;11(4):860. doi:10.3390/microorganisms11040860. 416  
417  
418
22. Wood SN. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC; 2017. 419  
420
23. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. *PLOS ONE*. 2015;10(12):e0146021. doi:10.1371/journal.pone.0146021. 421  
422
24. Al-Ani BG. Statistical modeling of the novel COVID-19 epidemic in Iraq. *Epidemiologic Methods*. 2021;10(s1). doi:10.1515/em-2020-0025. 423  
424

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

25. Allaman IB, Jelihovschi EG. Estimation of the Critical Points of an Epidemic by Means of a Logistic Growth Model. *Brazilian Journal of Biometrics*. 2022;40(2). doi:10.28951/bjb.v40i2.576. 425  
426  
427
26. Kaniadakis G, Baldi MM, Deisboeck TS, Grisolia G, Hristopulos DT, Scarfone AM, et al. The  $\kappa$ -statistics approach to epidemiology. *Scientific Reports*. 2020;10(1):19949. doi:10.1038/s41598-020-76673-3. 428  
429  
430
27. Moreau VH. Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths. *Brazilian Journal of Microbiology*. 2020;51(3):1109–1115. doi:10.1007/s42770-020-00331-z. 431  
432  
433  
434
28. Moreau VH. Using the Weibull distribution to model COVID-19 epidemic data. *Model Assisted Statistics and Applications*. 2021;16(1):5–14. doi:10.3233/MAS-210510. 435  
436  
437
29. Attanayake A, Perera S, Jayasinghe S. Phenomenological Modelling of COVID-19 Epidemics in Sri Lanka, Italy, the United States, and Hebei Province of China. *Computational and mathematical methods in medicine*. 2020;2020(1):6397063. 438  
439  
440
30. Ahmad Z, Almaspoor Z, Khan F, El-Morshedy M. On predictive modeling using a new flexible Weibull distribution and machine learning approach: Analyzing the COVID-19 data. *Mathematics*. 2022;10(11):1792. 441  
442  
443
31. Bantan R, Ahmad Z, Khan F, Elgarhy M, Almaspoor Z, Hamedani G, et al. Predictive modeling of the COVID-19 data using a new version of the flexible Weibull model and machine learning techniques. *Math Biosci Eng*. 2023;20(2):2847–2873. 444  
445  
446  
447
32. Link-Gelles R. Early estimates of updated 2023–2024 (monovalent XBB. 1.5) COVID-19 vaccine effectiveness against symptomatic SARS-CoV-2 infection attributable to co-circulating Omicron variants among immunocompetent adults—Increasing Community Access to Testing Program, United States, September 2023–January 2024. *MMWR Morbidity and Mortality Weekly Report*. 2024;73. 448  
449  
450  
451  
452  
453





















