2023

Modeling Behavior and Vaccine Hesitancy for Predicting Daily Vaccination Inoculations Using Trends, Case, Death, and Twitter Sentiment Data

Talal Daghriri
University of Central Florida

MODELING BEHAVIOR AND VACCINE HESITANCY FOR PREDICTING DAILY VACCINATION INOCULATIONS USING TRENDS, CASE, DEATH, AND TWITTER SENTIMENT DATA

by

TALAL MOHAMMED DAGHRIRI
B.S. Industrial Engineering, Jazan University, 2011
M.S. Industrial Engineering, University of New Haven, 2018

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2023

Major Professor: Michael D. Proctor

ABSTRACT

Over the past 100 years, epidemiological models have evolved to incorporate greater facets of the process. With the advent of social networking, massive computational power, population sentiment analysis can now be added to the epidemiological modeling process. Sentiment analysis is greater understanding of the fears, uncertainties, motivation, and trends of the public with respect to vaccination and associated events. Lack of public confidence in the efficacy of models, safety of vaccines, and appropriateness of policies confounds vaccine inoculation prediction. Sentiment analysis of social media is a seminal technique that accesses shared users' contents and tweets on the Twitter platform for daily fast and accurate modeling of public sentiment. As an applied contribution to this science, we present sentiment-based models for predicting United States daily COVID-19 vaccine inoculations. The research methodology encompasses predictive regression models spanning three phases of the U.S. pandemic including a baseline COVID-19 phase, a Delta variant phase, and Omicron variant phase that when combined span the period June 1, 2021, to March 31, 2022. Additionally, the models incorporate U.S. population behavior responses during the CDC recommended first dose interval, second dose interval, and booster intervals. Investigation of variables influencing daily inoculations identified CDC VOC phase, daily cases, daily deaths, and positive and negative Twitter Tweets as statistically significant for first dose and booster dose intervals exceeding a predictive R square of 77% and 84% respectively. The best regression model for the second dose interval proved to be a three variable- phases, cases, and negative tweets - inoculation model that exceeded a predictive R square of 53%. Limiting tweets collection to geolocated tweets does not encompass the entire U.S. Twitter population. However, Kaiser Family Foundation (KFF) surveys results appear to generally support the regression factors common to the First Dose and Booster Dose regression models and their results.

ACKNOWLEDGMENTS

First and foremost, I want to thank God for his blessings throughout my research work, which led to the successful completion of my Ph.D. dissertation.

I would like to express my deep and sincere gratitude to my advisor, Professor Michael Proctor and Dr. Sarah Matthews for their continuous support and guidance throughout this research. It was a great privilege and honor to work and study under their guidance.

I also would like to thank the members of my committee: Dr. Waldemar Karwowski, and Dr. Ivan Garibay for their time and continued support.

I would like to thank my parents Mohammed Daghriri and Fatimah Sahly. They guided me and supported me in whatever I chose throughout my life.

I would like to appreciate the encouragement and optimistic support of all my brothers and sisters during my Ph.D. journey

Special thanks to my friends Ahmad Kabli and Nami Al Solami for always listening, supporting, and motivating me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

STATEMENT OF CONTRIBUTION

Investigating public opinions on Twitter platform toward taking COVID-19 vaccination can be used to predict daily vaccine inoculation in the USA. Including other factors into the research was helpful step to generate a great modeling and analysis that show how the integrating different fears directions, different factors and different levels impacted intentions toward vaccination. Therefore, Twitter users' tweets were classified sentimentally, then they were compared to CDC vaccines data in order to explore the nature and strength of association between opinions on Twitter and daily vaccination. Additional factors included adding phases factor to the Pandemic based on CDC Variant of Concern (VOC) announcements investigated the effects of each phase on the vaccine inoculation, where each phase has its own characteristics and different public response. Moreover, regression modeling has been built using virus cases, deaths, social media sentiment (vaccination positive and negative tweets), and the first, second, booster dose, which built a developed vaccine inoculation predictive model. Vaccination approvals events haven added as they on order to quantify their effects on the vaccines inoculation. Furthermore, lag times between VOCs, FDA vaccine approvals, and a considerable increase in vaccination inoculations in the USA have been measured to explores the times between events and the public response. As different events and factors generated different public response levels which explained the impact of public opinions and response during selected phases for unvaccinated and partially vaccinated population. In addition, predictive power of the regression models for predicting first, second, and booster dose daily vaccination inoculation in the USA have been assessed and analyzed. Results have been validated with KFF vaccination opinions surveys, then, limitations and future research goals were identified.

<u>CHAPTER ONE: INRODUCTION</u>

Historically global pandemics strain public health services and leave millions if not tens of millions dead in their wake (He et al., 2020). Pandemics negatively impact the social behavior of individuals, national economies, and the vitality of the global economy (Squazzoni et al., 2020). Pandemic spread predictions, public health protocols, and vaccine adoption models are critical for advising governments, non-government organizations, heads of households, and individuals on behaviors and policies that may reduce their risk and slow the spread of the disease (Correia, Luck, & Verner, 2020). Since model error can lead to even more disaster, the re-occurring question is, what can be done to improve the efficacy of modeling?

In response to COVID19 virus predictive modeling and public health protocol recommendations, governments imposed widespread lockdowns in Europe. After appearing to have initially stopped the spread of the pandemic, the lockdowns proved to be ineffective in stopping emergence of the feared “second wave” (Cacciapaglia, Cot, & Sannino, 2020). When new lockdown measures were proposed, social media channels suddenly came alive eventually resulting in acts of civil disobedience across the continent (Hernandez, 2020).

In the United States, model guidance for “fifteen days to slow the spread” locked the nation down (Fauci, 2020). Health protocols such as social distancing, mask wearing, and hand sanitizing led to an initial drop in cases (Wang et al., 2020) (Cascini et al., 2020). Inconsistency in lockdown guidance and protocol implementation resulted in significant revisions to model predictions and a drop in public confidence in governing authorities and the protocols promoted. More recently, with as many as one in five Americans expressing vaccine hesitancy, vaccine adoption rates in the U.S. peaked at around 50% of the population fully vaccinated (Monmouth, 2021; Kukulka, 2021).

With the onset of the Delta variant and corresponding rise in infections, hospitalizations, and deaths, a new surge in vaccinations occurred though far less than hoped (Neel, 2021). As a result of unfilled expectation, the Biden administration resorted to mandates which have been met with widespread resistance as to their scope and appropriateness (Rogers & Stolberg, 2021; Wadman, 2021: Erman & Manojna, 2021) While vaccine hesitancy isn't a new phenomenon, vaccine hesitancy has the potential of prolonging the longevity of the pandemic and may even enable its re-emergence (Edwards, et al, 2016).

Clearly government-imposed protocols and policies, no matter how well intended, degrade over time due to mounting social pressures and inevitable civil disobedience ("Protests, policing, and COVID-19," 2021). Given significant errors in original model predictions and policies (Magness, 2020; Zanin & Papo, 2020), the question arises, what is the state of incorporating social media and sentiment analysis into modeling analysis? And finally, do we yet have evidence that incorporation of social media factors into models improve the quality of analysis and success of public health protocols and policies?

Disease Spread and Health Behavior Modeling

We a lay a foundation for discussion of these questions with a literature review of models listed and characterized in Table 23 For review purposes below, those models are classified into the following categories: *Epidemiological State models, Epidemiological Statistical Prediction models, Traditional Systems Dynamics models, Agent-based models and Multiagent systems models, and Machine Learning and Hybrid models.*

Epidemiological State Models

One of the simplest approaches to epidemiological modeling is based on a chain of disease states as observed within a population. Epidemiological State models are among the earliest and longest lasting modeling technique. At a minimum, State models report a snapshot of selective states of a disease within a population. Epidemiological state data observed over time may be used to generate mathematical relationships to produce patterns in disease behavior within the population (Brown & Ozanne, 2019). Epidemiological State models considered in Table 1 may forecast future trend of disease spread among various populations but typically do so based solely on observations (Brauer, Driessche, & Wu, 2008).

Table 1: Epidemiological State Models

Susceptible, Infected, Recovered (SIR)
Susceptible, Infected, Susceptible (SIS)
Susceptible, Infected, Recovered, Deceased (SIRD)
Maternally derived immunity, Susceptible, Infected, Recovered (MSIR)
Susceptible, Exposed, Infected (SEI)
Susceptible, Exposed, Infected, Recovered (SEIR)
Susceptible, Exposed, Infected, Susceptible (SEIS)
Maternally derived immunity, Susceptible, Exposed, Infected, Recovered (MSEIR)
Maternally derived immunity, Susceptible, Exposed, Infected, Recovered, Susceptible (MSEIRS)
Susceptible-Latent-Infected-Recovered-Dead-Susceptible (SLIRDS)
Exposed, Infected, Hospitalized (EIH)
Susceptible, Infected, Hospitalized, Recovered (SIHR)

Susceptible, Infected, Recovered (SIR) Model:

Developed in the 1920's, Kermack and McKendrick (1991) SIR model (Figure 1) represents the effect of the flow of the disease through the population from susceptible (S), infected (I), or recovered (R) states (Cooper, Mondal, & Antonopoulos, 2020). Observed stochastic probabilities bIS and gI may be derived from the observed likelihood of disease movement from one state to the next state (Jiang, Yu, Ji, & Shi, 2011). A primary outcome of the model is numerical (Figure 1) or graphical representation over time (Figure 2) of the level of the disease for each state within a population.

Figure 1: SIR Model

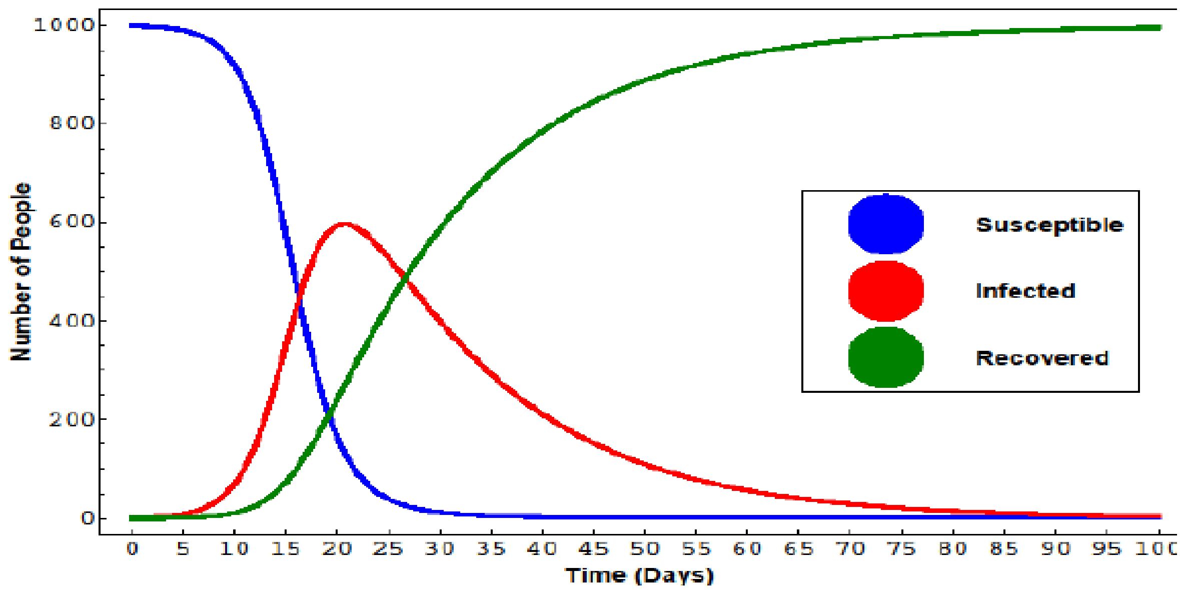


Figure 2: SIR Model Curves (Macal, 2010)

Susceptible-Infected-Susceptible (SIS) Model:

SIS model (Figure 3) represents different disease states or flows (Vargas-De-León, 2011) (Fernández-Villaverde & Jones, 2020) (Bin, Sun, & Chen, C. C. 2019). In contrast with the SIR model, the SIS model describes diseases for which there is no immunity. After a person has been infected and has healed, he or she is again susceptible to the disease (Vargas-De-León, 2011) (Qi, Liu, & Meng, 2017).

Figure 3: SIS Model (adapted from Crossin, 2020)

Susceptible-Infected-Recovered-Deceased (SIRD) Model:

SIRD model (Figure 4) distinguishes between Recovered and Deceased. (Fernández-Villaverde & Jones, 2020) (Caccavo, 2020).

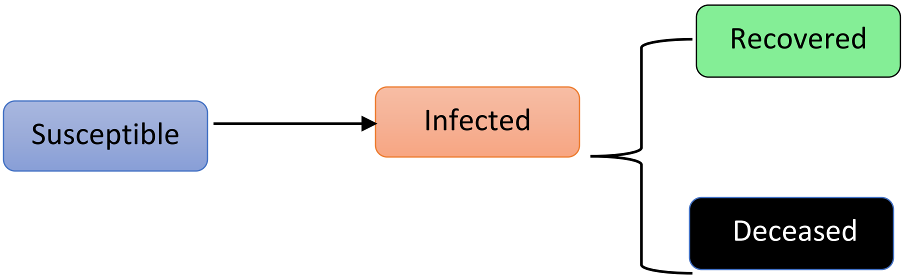


Figure 4: SIRD Model (adapted from Jamakayala, 2020)

Maternally Derived Immunity, Susceptible, Infected, Recovered (MSIR) Model:

MSIR model (Figure 5) adds to the SIR model a state M for derived immunity that includes babies with passive immunity (Seyoum Desta, 2019) (Fajar, 2019).

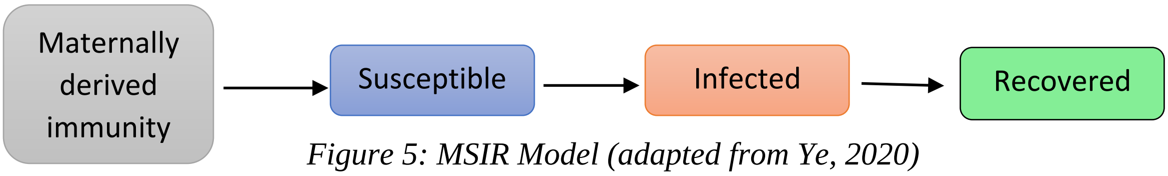


Figure 5: MSIR Model (adapted from Ye, 2020)

Susceptible, Exposed, Infected (SEI) Model

SEI model (Figure 6) adds the state E, where this model assumes that susceptible person goes through a exposed period before becoming infected (Kim & Lin, 2008).

Figure 6: SEI Model (adapted from Kim & Lin, 2008)

Susceptible, Exposed, Infected, Recovered (SEIR) Model:

SEIR model (Figure 7) adds an exposure state to the SIR model where recovered includes a level of immunity (Lekone & Finkenstädt, 2006).

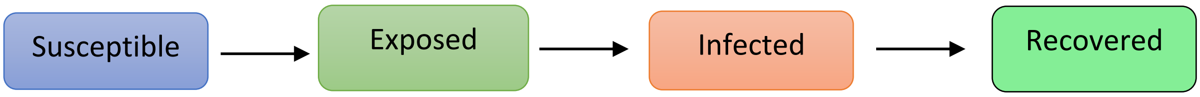


Figure 7: SEIR Model (adapted from Gupta et al., 2020)

Susceptible, Exposed, Infected, Susceptible (SEIS) Model:

SEIS model (Figure 8) is identical to the SEIR model (above) except that there is no immunity obtained at the end. (Fan, Li, & Wang, 2001).

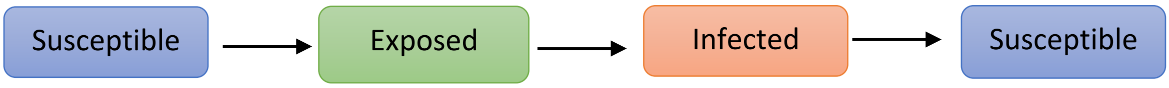


Figure 8: SEIS Model (adapted from Fan, Li, & Wang, 2001)

Maternally Derived Immunity, Susceptible, Exposed, Infected, Recovered (MSEIR) Model:

MSEIR model (Figure 9) is like SEIR model, but it considers a passive immunity as an additional factor (Li & Zhang, 2012). Passive immunity is obtained when an individual is given antibodies to a disease rather than created by his immune system (Qureshi & Yusuf, 2019).

Figure 9: MSEIR Model (adapted from Hethcote, 2000)

Maternally Derived Immunity, Susceptible, Exposed, Infected, Recovered, Susceptible (MSEIRS) Model:

MSEIRS model (Figure 10) is similar to the MSEIR, except the R-class immunity would be temporary, meaning people would recover their susceptibility after their immunity expired (Mosavi, 2020).

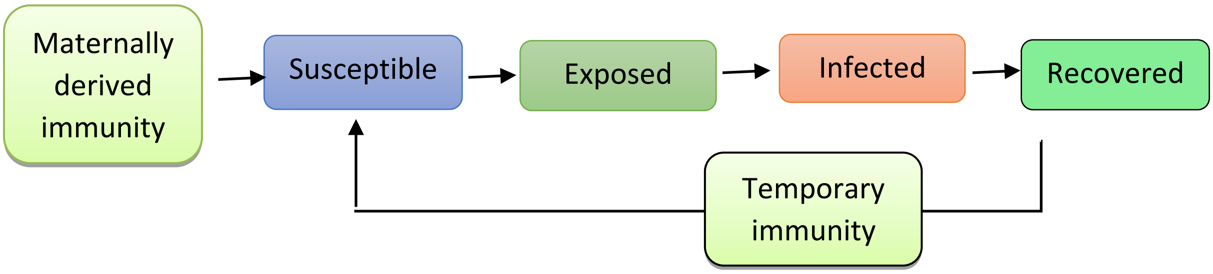


Figure 10: MSEIRS Model (adapted from Ajisafe, 2018)

Susceptible-Latent-Infected-Recovered-Dead-Susceptible (SLIRDS) Model:

SLIRDS model (Figure 11) (Bin, Sun, & Chen, C. C. 2019) incorporates population growth, sex ratio, and age structure to determine the model's evolutionary rules. The latent state in the model means that virus is found in the body but remains in a resting (latent) state without creating more viruses. Latent viral infection usually causes no visible symptoms and can last a long time before it becomes active and causes symptoms (Bin, Sun, & Chen, C. C. 2019) ("Viral latency | ClinicalInfo," 2020).

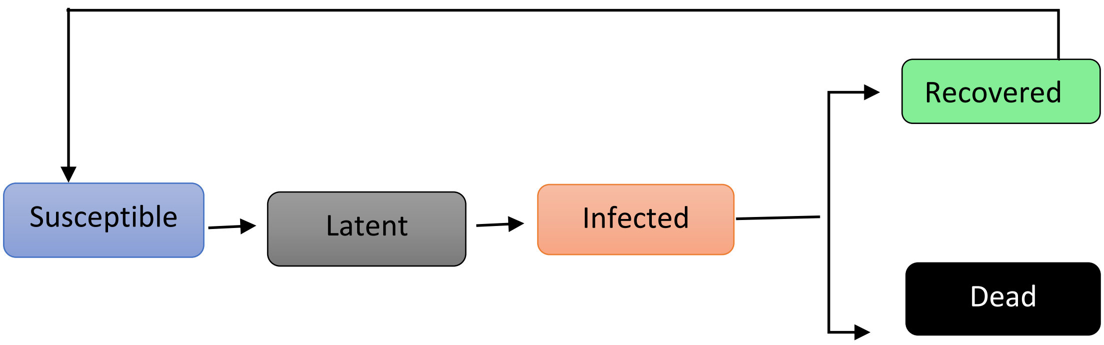


Figure 11: SLIRDS Model (adapted from Bin, Sun, & Chen, C. C. 2019)

Exposed, Infected, Hospitalized (EIH) Model:

EIH model (Figure 12) adds the state H, where this model assumes that infected person goes through a hospitalized period after becoming infected **(**Sooknanan & Comissiong, 2020**).** As discussed below, during the COVID19 pandemic, hospitalization state models proved their importance when linked to hospital capacity.

Figure 12: EIH Model (adapted from Sooknanan & Comissiong, 2020)

Susceptible, Infected, Hospitalized, Recovered (SIHR) Model:

SIHR model (Figure 13) assumes that susceptible person goes through an infected, hospitalized, then to recovered status **(**Sooknanan & Comissiong, 2020**).** So, it is like SIR model except it involves hospitalized state.

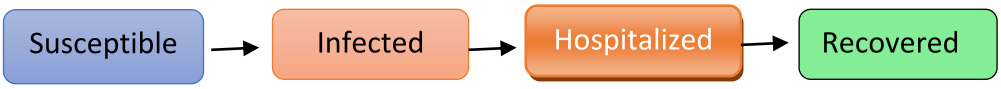


Figure 13: SIHR Model (adapted from Jiao, & Huang, 2020)

<u>Epidemiology Statistical Forecast Models</u>

Epidemiological Statistical Forecast models build on the fore mentioned State models to project trends into the future. In the current COVID19 pandemic, the John Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/data) is well known for tracking Epidemiological State data and reports various statistical trends of interests to decision makers. Table 2 identifies the Statistical Forecast models discussed herein.

Table 2: Statistical Forecast Models

Differential Equations Leads to Predictions of Hospitalizations and Infections (DELPHI)
Auto regressive integrated moving average (ARIMA)
Los Alamos National Laboratory COVID-19 forecasting using Fast Evaluations and Estimation (LANL COFFEF)
John Hopkins model COVID-19 prediction (JHU COVID-19 prediction)

Differential Equations Leads to Predictions of Hospitalizations and Infections (DELPHI) Model

DELPHI Model (Figure 14) is known for its statistical forecast of future COVID-19 infected cases, hospitalizations, and deaths rates during the pandemic. The model builds on the SEIR state model to project trends while incorporating government interventions ("Coronavirus disease 2019 (COVID-19)," 2020), where there are four phases for that represent the governmental interventions.

Phase I: This phase represents the immediate solution as the government is just beginning to implement measures to combat the outbreak. Some people will have changed their behavior in response to reports of an epidemic, but a large portion of the population will continue to live their lives normally.

Phase II: This process is distinguished by a sharp reduction in infection rates as measures to regulate the spread are fully implemented (e.g., the closure of a portion of the economy) and the population as a whole suffers a shock case.

Phase III: This process simulates the response's eventual flattening out as the measurements hit saturation. This is expressed by the decreasing marginal returns (i.e. convexity) in the infection rate decrease.

Phase IV: This phase represents the revival of cases triggered by the premature removal of social distancing interventions and people returning to their normal behaviors.

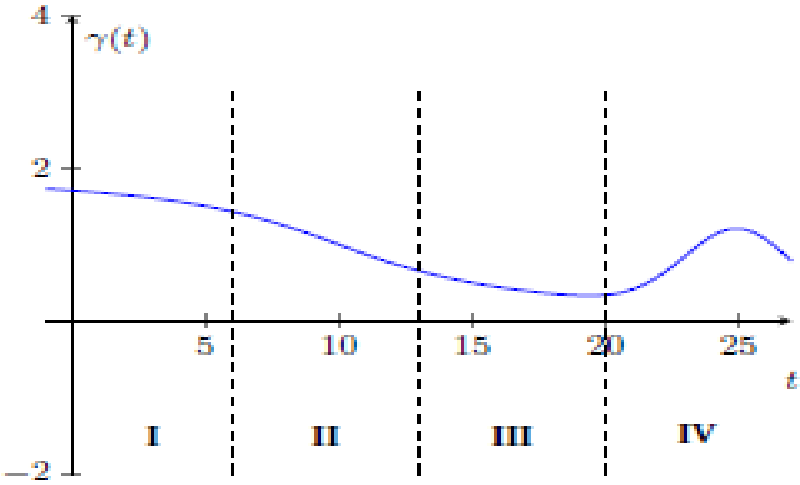


Figure 14: DELPHI Model (adapted from "COVIDAnalytics,"2020)

Auto Regressive Integrated Moving Average (ARIMA) Model

The ARIMA model builds on the SIRD state model using time series regression model to provide a COVID-19 case forecast with an 95% expected range of the forecast that expands with time. Figure 15 illustrates one projection of cases and its associated 95% expected range of the cases in the model forecast. Additional forecasts with associated ranges may be derived from the “daily total confirmed cases, total confirmed new cases, total deaths, total new deaths, growth rate in confirmed cases, and growth rate in deaths ” (Tran, Pham, and Ngo, 2020).

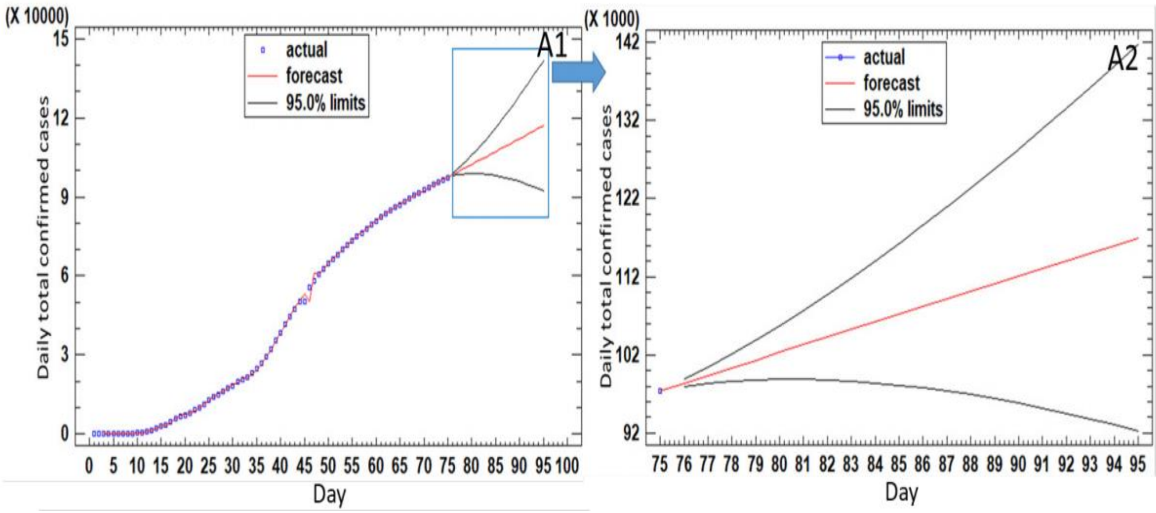


Figure 15: Daily Total Confirmed Cases (adapted from Tran, Pham, and Ngo, 2020). Los Alamos National Laboratory COVID-19 Forecasting Using Fast Evaluations and Estimation (LANL COFFEF) Model

LANL COFFEF model also produces a case/death forecast with more detailed prediction intervals for any state with at least one confirmed case/death of COVID-19 and every country with at least 100 confirmed cases of COVID-19 and 20 fatalities. Building on the SIRD state model, COFFEF model takes geographical structure, growth rate, reporting process, and case fatality rate into consideration while predicting infects and deaths (figure 16). The model can generate short- and long-term forecasts to help decision-makers obtain useful insights about possible pandemic outcomes in the near future. The figures below show that “Recent observations (grey points) and forecasts for reported confirmed cases and deaths. Colored points are forecast medians and represent the model's best guess. The darker colored bar is the 50% prediction interval, while the lighter color bar is the 90% prediction interval”.

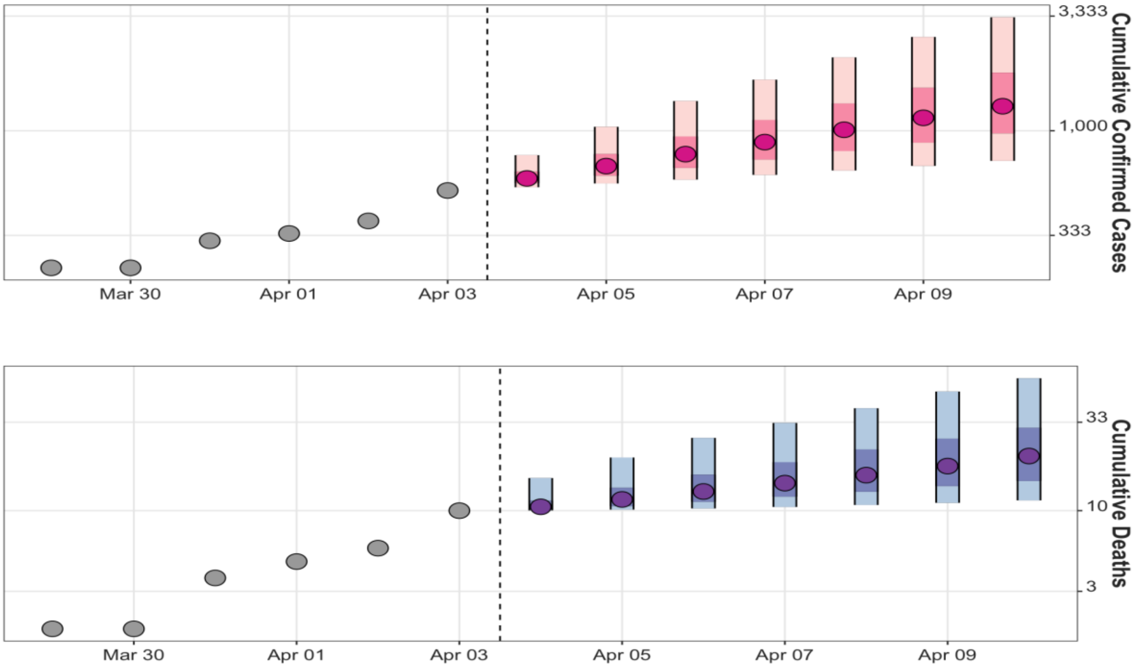


Figure 16: Cumulative Confirmed Cases and Deaths (adapted from LANL COFFEF Cases and Deaths Forecasts, 2020)

John Hopkins COVID Inpatient Risk Calculator (CIRC) Model

Adding a death state to the SIHR state model, the John Hopkins CIRC model "incorporates more than 20 demographic and clinical variables available at hospital admission to predict the likelihood of a patient progressing to severe disease or death within 7 days of patient arrival" ("COVID-19: Risk calculator predicts progression, death among hospitalized patients," 2020). CIRC uses a set of patient risk factors associated with COVID-19 disease ("Johns Hopkins researchers publish COVID-19 'Prediction model'," 2020) to forecast disease outcomes for Patient Type A to F may progress over time while in a healthcare unit or hospital (Figure 17a and 17b). Patient Type risk factor considered include the patients' age, body mass index (BMI), the health of the lung, chronic illnesses, vital signs, and the severity of infected case symptoms when it is admitted to the hospital. Model estimates may aid hospital decision making about patient handling, allocation of resources, interventions, and patient safety.

Patient	Description	Cumulative Incidence of Severe Disease or Death		
		2 Days	4 Days	7 Days
A	81-year-old Black woman with diabetes and hypertension; BMI, 35 kg/m²; respiratory rate, 32 breaths/min; febrile; high CRP level; D-dimer level > 1 mg/L	80%	92%	96%
B	69-year-old Black man with diabetes, coronary disease, and hypertension; BMI, 38 kg/m²; respiratory rate, 23 breaths/min	28%	41%	50%
C	47-year-old Black man with diabetes and hypertension; BMI, 34 kg/m²; respiratory rate, 18 breaths/min; febrile; detectable troponin level	18%	27%	32%
D	79-year-old White man with a CCI of 0; BMI, 24 kg/m²; respiratory rate, 19 breaths/min; afebrile; detectable troponin level	10%	15%	18%
E	60-year-old White woman with a CCI of 0; BMI, 28 kg/m²; respiratory rate, 18 breaths/min; afebrile	6%	9%	11%
F	39-year-old Latinx man with a CCI of 0; BMI, 23 kg/m²; respiratory rate, 18 breaths/min; afebrile	3%	5%	5%

Figure 17-A: John Hopkins COVID-19 Prediction Model (adapted from Garibaldi et al., 2020)

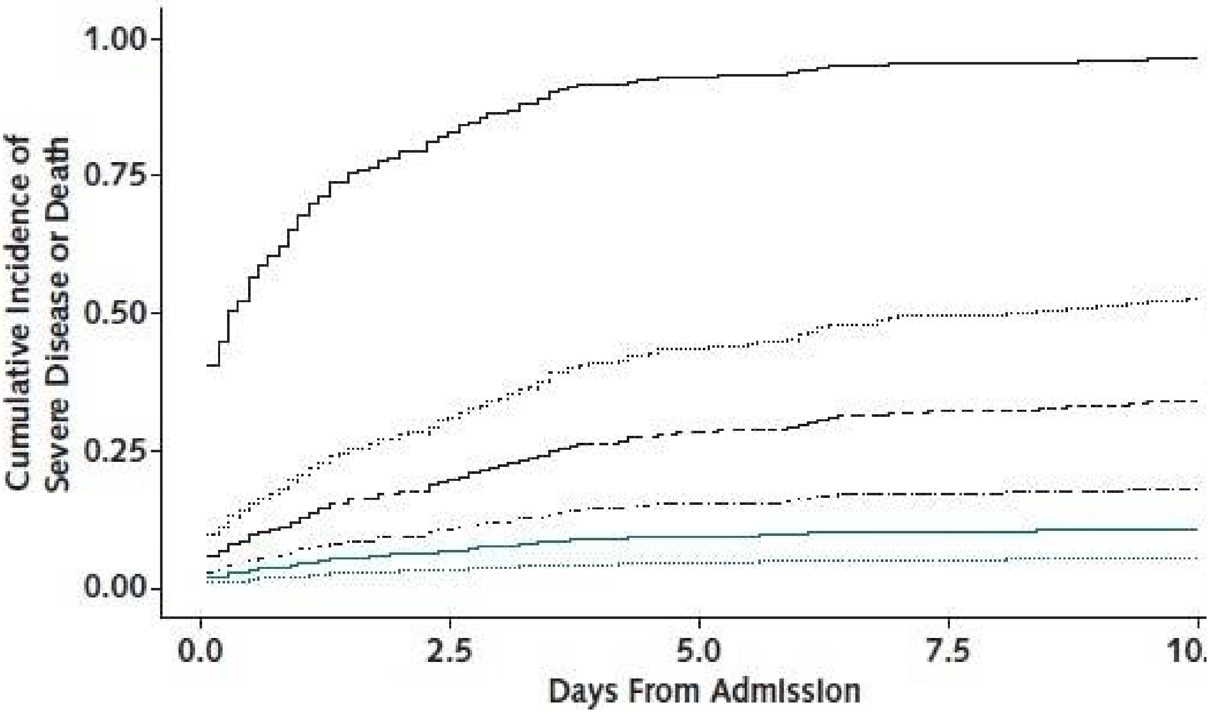


Figure 17-B: Projections for Patient Type A to F Over Time (adapted from Garibaldi et al., 2020)

Traditional System Dynamics Models

Disease transmission, population behavior and vulnerability, and government directed preventive interventions all theoretically change disease outcomes in different nations. The radical differences in the number of infections between nations is highlighted by John Hopkins COVID-19 infection tracking data shown in Figure 18. Theoretical intervention models attempt to explain these differences.

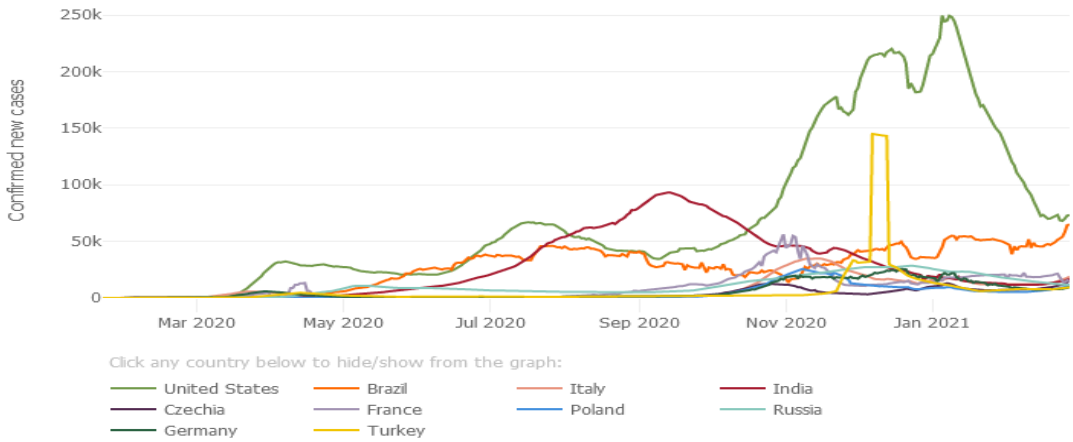


*Figure 18: John Hopkins Daily Confirmed New Cases for the Current 10 Most Affected Countries (**https://coronavirus.jhu.edu/data/new-cases download Feb 28 2021**)*

Modeling the dynamics of disease transmission, population behavior and vulnerability, and preventive interventions (Anderson et al., 2020) attempts to project disease outcomes beyond simple statistical trend extrapolation. Modeling of the dynamics of systems is not new (Forrester, 1958). System dynamics models are applied across a broad-spectrum of applications including pandemics, sexual behavior, smoking, exercising, and use of seat belts (DeJoy, 1996). A fundamental objective of applying systems dynamics for pandemic application is to model, simulate, and quantify interventions and outcomes with the hope of limiting disease spread and impacts (Simpson, 2015; Samui, 2020). During the COVID-19 pandemic, objectives included identifying behavioral intervention measures to "slow the spread" (Fauci, 2020) and gain time to develop vaccines to counter the disease. At least three theoretical intervention models attempt to capture the dynamics of human behavior in light of life threating events have been put forth including the Health Belief Model (HBM), (Champion, Skinner, and others, 2008) , the Theory of Planned Behavior (TPB) (Ajzen, 1991), and the Protection Motivation Theory (PMT) (Okuhara, Okada, & Kiuchi, 2020) (Table 3). These approaches may be integrated together for deeper work, such as integrating TPB theory and PMT theory to gain insights into both behaviors and intentions of individuals (Wang et al., 2019).

Table 3: Traditional System Dynamics Models

Health Belief Model (HBM)
Theory of Planned Behavior (TPB)
Protection Motivation Theory (PMT)

Health Belief Model (HBM)

In the 1950s, social scientists in the United States created the HBM Model (Figure 19) (Champion, Skinner, and others) to explain human behavioral response to outbreaks and associated health protocols. Influences on human behavior included demographics and psychological variables acting through perceived susceptibility to the health threat, perceived severity of the health threat, health motivation, perceived benefits to taking the prescribed action, and perceived barriers to taking actions. This model helps categorize factors that motivate actions of people to act appropriately. Limitations of the model include the need to design data collection methods suited to various communities impacted by the disease.

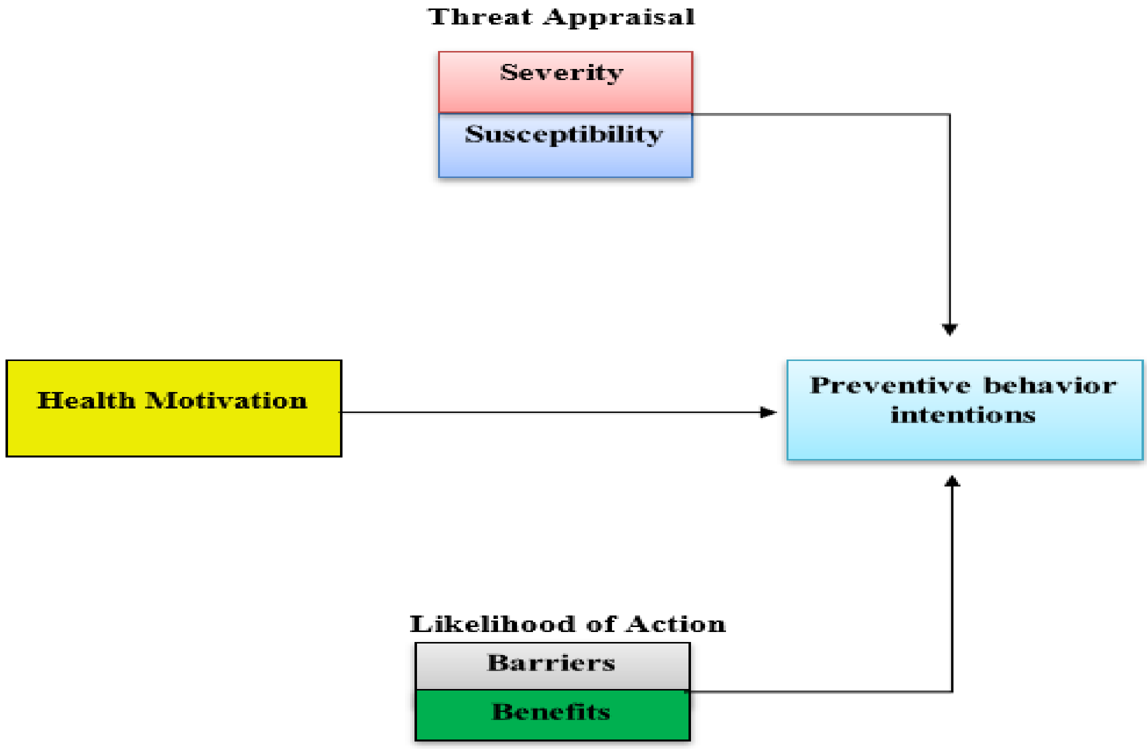


Figure 19: HBM Model (adapted from Lipman & Burt, 2017)

Theory of Planned Behavior (TPB theory)

TPB theory (Figure 20) was introduced by Icek Ajzen in 1985. Though not specifically developed for modeling proposed behavioral interventions during a pandemic, TPB theories of learning, expected value, and cohesiveness may still be applied (Ajzen, 1985). In its simplest form, individuals evaluate a proposed behavior. If the individuals feel that their actions are important to others and wish to do so, then they are more likely to do the proposed behavior. A high association of behaviors and subjective norms with behavioral intent and behavior has been verified in several studies, such as COVID-19 pandemic studies (Bae & Chang, 2020) (Yastica et al., 2020). Limitations of the model include the need to design data collection methods suited to individuals and relate those actions to impacts of the disease.

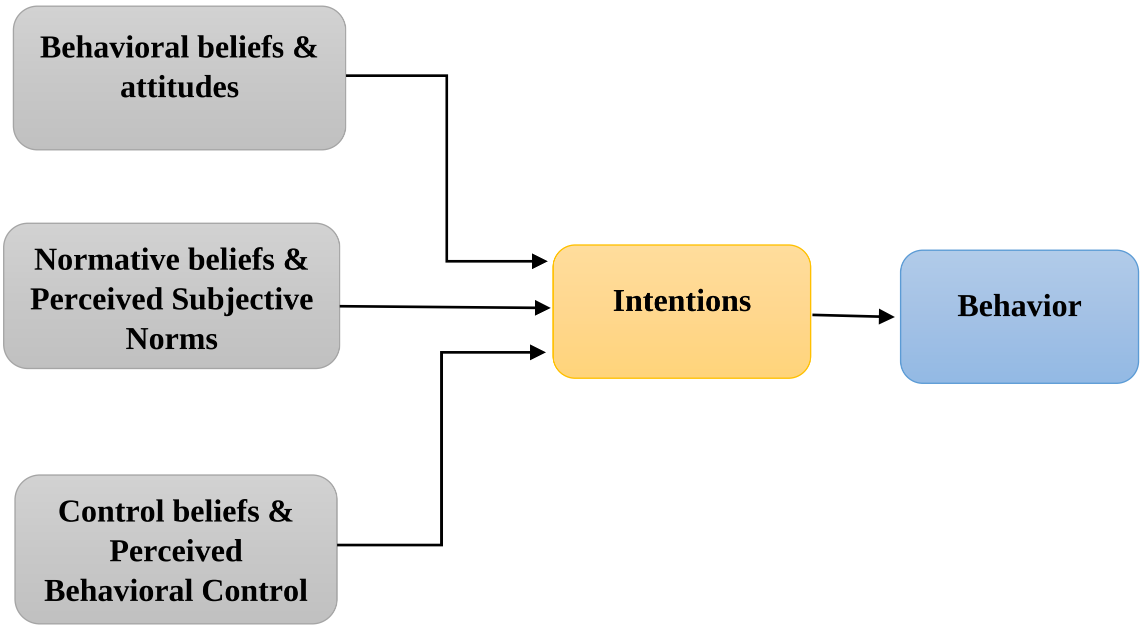


Figure 20: Theory of Planned Behavior (adapted from Ajzen, 1985)

Protection Motivation Theory (PMT)

PMT (Figure 21) addresses how people cope with and makes decisions under stress (Rogers, 1975). PMT draws on factors similar to HBM including: perceived severity of a threatening event; perceived probability of the occurrence or vulnerability; efficacy of the recommended preventive behavior, and perceived self-efficacy (Okuhara, Okada, & Kiuchi, 2020). Moreover, these factors may be grouped more generally in terms of appraisal and the coping appraisal. Threat appraisal involves perceived vulnerability, severity, and level of fear arousal while the coping appraisal involves response efficacy, self-efficacy, and perceived response-cost (Ling, Kothe, & Mullan, 2019; Floyd, Prentice-Dunn, & Rogers, 2000). Complicating interventions such as controlling weight to prevent heart disease or stroke is the possibility that the intervention may have a side effect that causes another problem (Wang et al., 2019). Furthermore, interventions may involve a secondary intervention such as using a medicine to reduce the risk of heart attack (Minnesota heart disease and stroke prevention connection, April 2011).

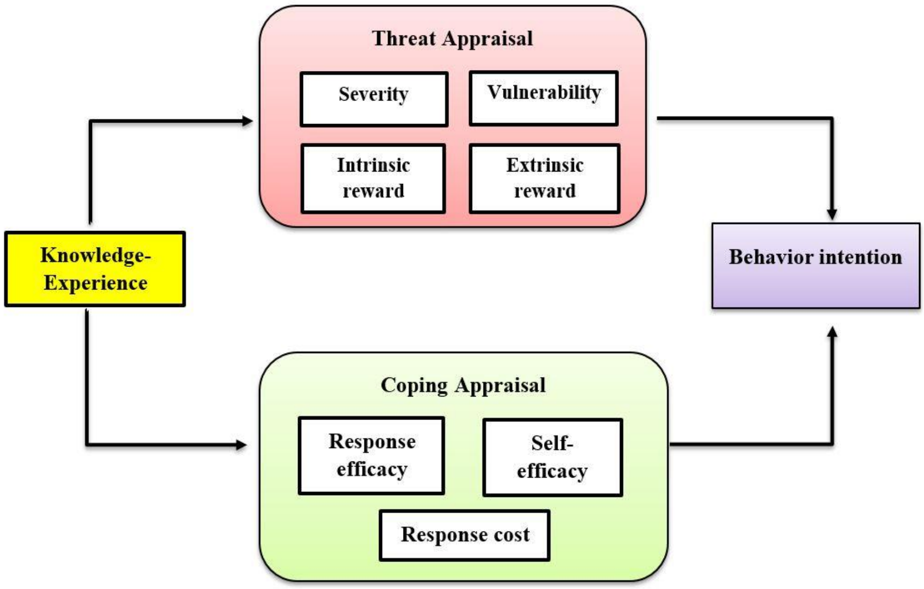


Figure 21: Protection Motivation Theory (adapted from Xiao et al., 2014)

<u>Chapter 1 Summary</u>

Chapter 1 provides the basic knowledge and background about the epidemiological models, where the traditional models as Epidemiological state models, statistical forecast models, and traditional system dynamics models have been discussed. Epidemiological state models focus on the possible states that public might go through during the virus spread, such as susceptible, exposed, infected, isolated, and recovered or dead. In addition, the statistical forecast models have been built based on the state models where the forecasting models aim to forecast the numbers of states based on the given data or recorded data by heath agencies, such as forecasting the number of infected cases, deaths cases, and recovered cases…. etc. Moreover, the traditional system dynamics models have been developed to explore how the causes and effects relationships can explore the change in the public behavior or opinions. For example, they are used to study and analyze the personal' benefits and barriers of following the healthy interventions which resulted in more understanding for the possible public response and providing much accurate analysis. Lacking of individuals' differences, human feelings, and opinions data caused limitations for the efficiency of these models. Thus, the concepts of these models have been developed to include additional human factors, advanced technologies, and more sufficient data resources to build new models with higher capabilities and less limitations. Chapter 2 explores how the traditional models' concepts have been used as concrete for building the advanced modeling approaches.

CHAPTER TWO: A LITERATURE REVIEW

Developing the Epidemiological Models and the Used Advanced Approaches: A Literature Review.

Over the last several decades, human behavior, machine learning techniques, computer software, and data analytics approaches have been developed and included into in the epidemiological modeling in order to improve models' efficiency and accuracy. Furthermore, the data sources have been increased significantly, which provides sufficient amount of data that could be extracted and analyzed with less cost and time. All of these developments cause a significant improvement in epidemiological modeling and simulation field. This chapter discusses these developments and techniques as well as related concepts, principles, and goals.

Agent-based and Multiagent Systems Modeling

Agent-based and Multiagent systems approaches approach modeling behavior from the bottom up rather than the top down. Both techniques attempt to model entity autonomous behavior and go beyond latent factors associated with an entity to discretionary factors left to the entity. In the case of behavior in the presence of an infectious disease includes interventions such as wearing a mask or separating six feet apart. These agent responses are based on some individual entity considerations based on the fore mentioned theories and include such fears as getting infected, benefits of following intervention protocols, or social or organization responses with or against the agent.

Agent-based and Multiagent systems share basic artificial intelligence techniques in an attempt to address forementioned theories by modeling "person(s), firm(s), machine(s), or software" that generate the actions or interactions (Agents in artificial intelligence," 2019) but they do differ.

"Agent-based modeling (ABM) is a technique that allows us to explore how the interactions of heterogeneous individuals impact on the wider behavior of social/spatial systems" (Crooks at el, 2018) with "autonomous and pro-active actors, such as human-centered systems" (Siebers & Aickelin, 2008) within the real world (Bonabeau, 2002). One goal of agent-based modeling is to achieve a clearer understanding of the relation between micro-interaction and emergent macroscale behavior. By modeling the entity level as an autonomous decision-maker, each agent individually assesses its situation and makes decisions based on a set of rules (Zohdi, 2020), and may execute various behaviors suited to itself (Bonabeau, 2002). ABM aims to explore insights into the collective behavior of agents that follow simple rules which means there are limitations on the strength and capabilities of the ABM approach (Niazi at el, 2011). In pandemics spread studies, an ABM approach "can capture the dynamics of disease spread combined with the heterogeneous mixing and social networks of agents" (Hunter et al., 2018). By characterizing the disease transmission rates, agents, and their environment, it is possible to create more real model that generate real scenarios for the pandemics trends (Hunter et al., 2017).

Modeling a population as a family of individual agents within a multi-agent society notionally is more appealing for pandemic applications then basic ABM approach (Parsons & Wooldridge, 2002). Advantages include recognition and implementation of levels of social or organizational control over entity self-regulation. Other advantages include representation of availability and access to channels of communications by agents as well as the nature, content and level of trust of information disseminated on those channels (Parsons & Wooldridge, 2002). Additionally, heterogeneous and various intelligent agents when interact with each other to decide upon and achieve a specific goal (Siebers & Aickelin, 2008). Furthermore, the level of interaction between intelligent agents may be represented with various levels of cooperation or competitiveness (Siebers & Aickelin, 2008).

MAS techniques can handle large complex systems (Seddari & Redjimi, 2013). More directly, MAS techniques are able to solve problems that are hard or impossible for an agent-based models to solve (Alkhateeb et al., 2010). For pandemics spread studies, MAS's analytical power rests in representing social and organizational controls as well as aspects of communications between individuals and between individuals and organizations (Vyklyuk et al., 2021). Table 4 identifies the agent-based and Multi-agent models discussed herein.

Table 4: Agent-based and Multiagent-based Models

COVID-19 A gent-based Simulator (COVASIM)
Social Distancing (SD)
Susceptible, Exposed, Infected, Recovered, agent- based (SEIR-ABM)
Frias-Martinez (FM)
University of Texas at Austin's COVID-19 (UT COVID-19)
The developed Multi Agent Susceptible, Infected, Recovered (DMAS-SIR)

COVID-19 Agent-based Simulator (COVASIM)

COVASIM model (Figure 22 and 23), was developed form SEIRD model to predict pandemic trends, the effects of interventions on the disease spread, and estimating the required resources (Kerr et al., 2020). To explain, COVASIM model takes into consideration multiple factors that makes the model more real, such as "country-specific demographic information on age structure and population size; realistic transmission networks in different social layers, including households, schools, workplaces, long-term care facilities, and communities; age-specific disease outcomes, and interhost viral dynamics, including viral-load-based transmissibility" (Kerr et al., 2020). Moreover, based on Figure 22 shows the structure of the model, where the agents health status change based on their situation, while figure 23 shows the agents contact with each other in houses, schools, or workplaces. Furthermore, the model studies the effects of multiple non-pharmaceutical interventions on the pandemic trend as shown in the model coding lines (Figure 24), so the model able to show the susceptibility, severity, and the threats to get infect in case of ignoring the heath interventions, and also it shows the benefits of following the health interventions, which lead to decrease the cases numbers. Thus, these outputs would explain why agents follow or do not follow the interventions policies these policies as explained previously in the assumed behavioral perceptions of the HBM model. According to Figure 25 "Synthetic population networks for households (top), schools (middle), and workplaces (bottom). Age-specific contact matrices are shown on the l eft, while actual connectivity patterns f or a 127-person subsample of a population of 1 0,000 individuals are shown on the right. All individuals are present in the household network, including some with no household connections. A subset of these individuals, including teachers, are present in the school network (circles); another subset is present in workplace networks (squares); some individuals are in neither school nor work networks (triangles)" (Kerr et al., 2020).

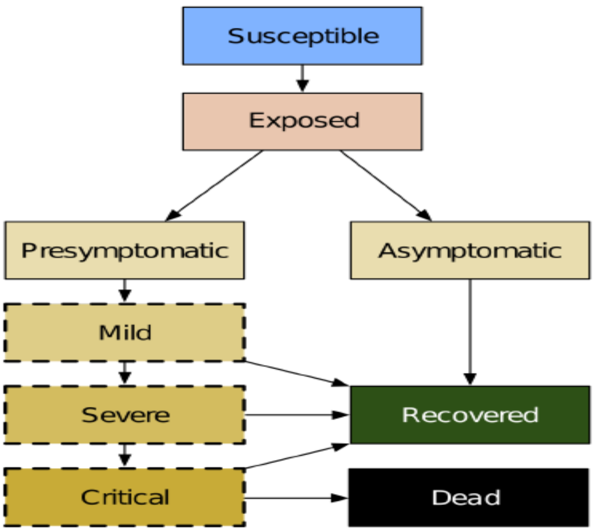


Figure 22: COVASIM Model Structure (adapted from Kerr et al., 2020).

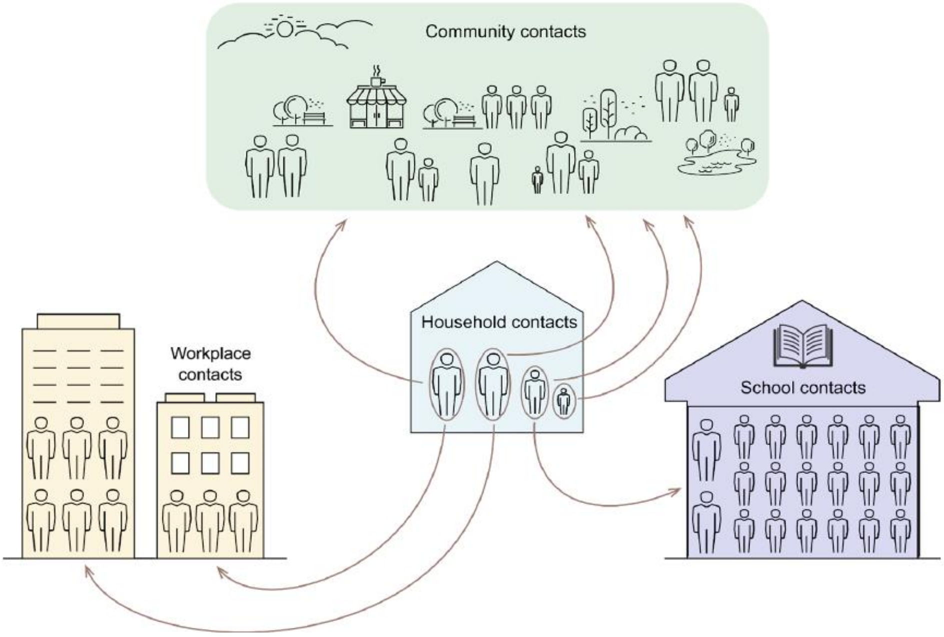


Figure 23: Schematic Diagram of Contact Networks

```python
# Define the interventions
trace_probs = dict(h=0.9, s=0.7, w=0.7, c=0.3) # Probability that a contact in each layer will be traced
trace_time = dict(h=0, s=1, w=1, c=3)          # Time required to trace contacts in each layer
interventions = [
    cv.clip_edges(start_day='2020-03-26', end_day=None,         change={'s':0.0}),             # Close schools
    cv.clip_edges(start_day='2020-03-26', end_day='2020-04-10', change={'w':0.7, 'c':0.7}),    # Reduce work and community
    cv.clip_edges(start_day='2020-04-10', end_day='2020-05-05', change={'w':0.3, 'c':0.3}),    # Reduce both further
    cv.clip_edges(start_day='2020-05-05', end_day=None,         change={'w':0.8, 'c':0.8}),    # Partially reopen
    cv.test_prob(start_day='2020-05-20', symp_prob=0.10, symp_quar_prob=0.8, test_delay=2),    # Testing
    cv.contact_tracing(start_day='2020-04-20', trace_probs=trace_probs, trace_time=trace_time) # Contact tracing
]
```

Figure 24: Codes of Intervention Part in the Model

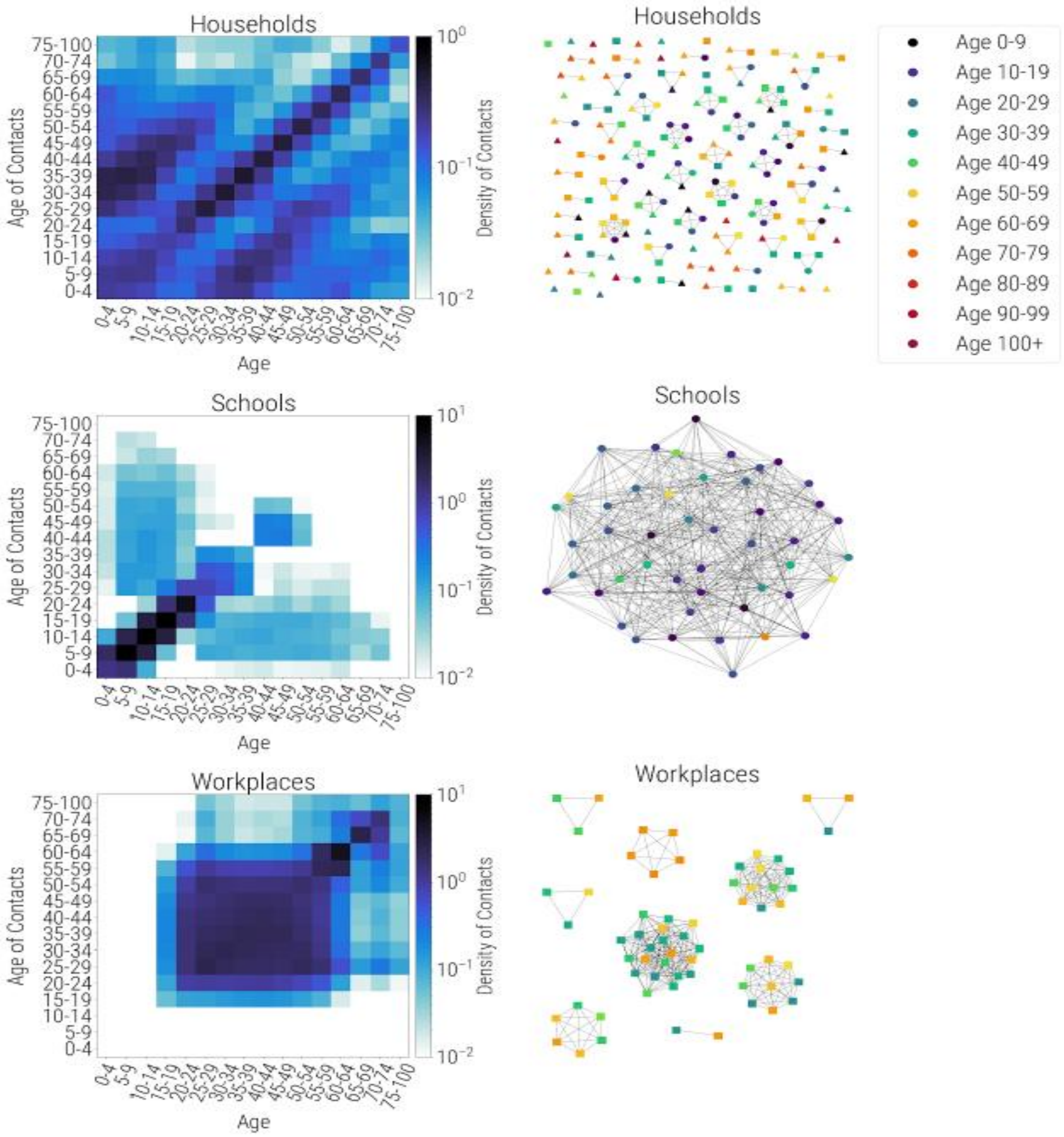


Figure 25: Agents' Contacts Networks

Social Distancing (SD) Model

A developed SD model (Figure 26) was created by **Netlogo** programming software based on SEIRD to quantify the effect of social distancing on the spread of COVID-19 disease, where different levels of social distancing policies were tested (Daghriri and Ozmen, 2021). So, it seeks to quantify the optimal level of social distancing that should be applied and followed by citizens in the United States. Model results show that the optimal level of applying social distancing strategies should be equal or higher than 80% to flat the curve.

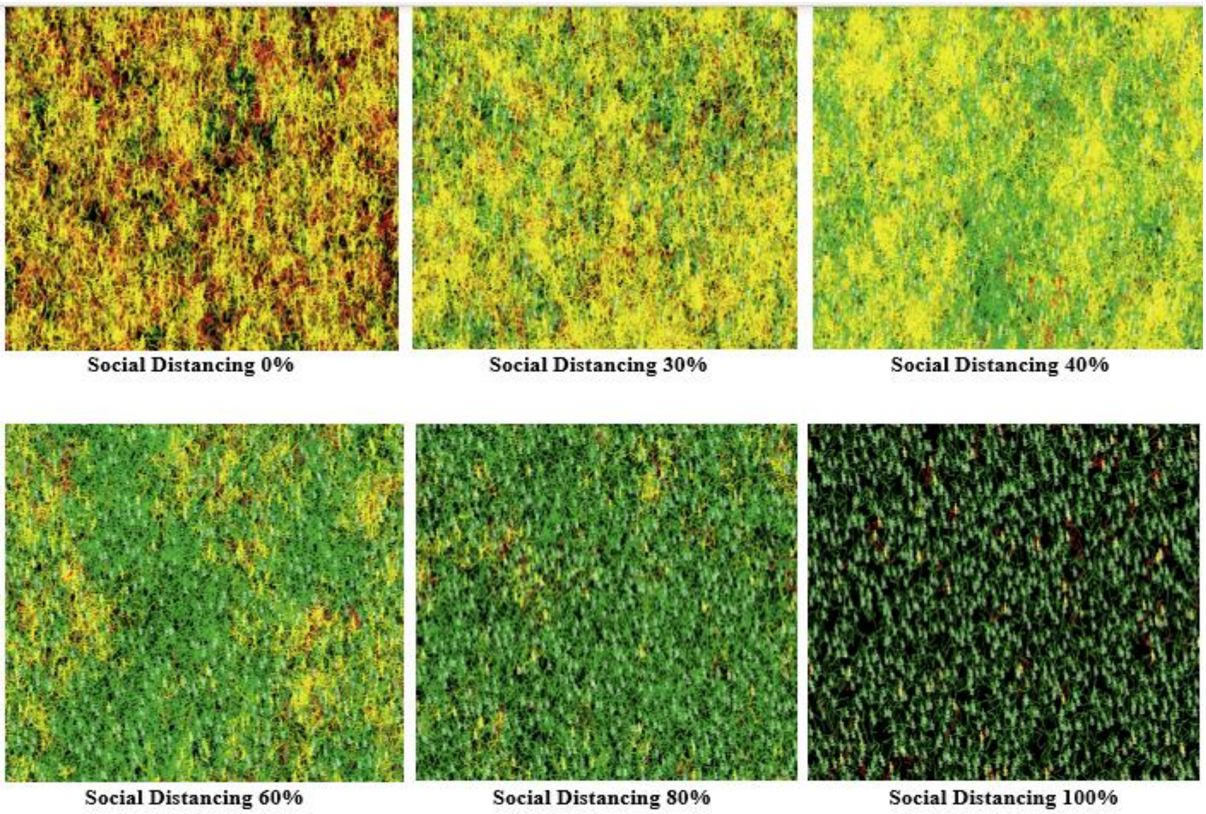


Figure 26: Social Distancing Model (adapted from Daghriri and Ozmen, 2021)

Frias-Martinez (FM) Model

According to a study that discussed integrating the agent-based models with social networks (Frias-Martinez et al., 2011), FM model (Figure 27) was proposed to capture the social interactions and the patterns of human mobility the derived from call detailed records. Moreover, the suggested method was used to study the 2009 H1N1 epidemic in Mexico and to evaluate the effect of government interventions on virus spread(Frias-Martinez et al., 2011), and to predict people intentions to move. Also, it involves a previous SIR model study results (Cruz-Pachecon et al, 2009). To explain, the mobile phone networks are designed using a series of mobile towers called Base Transceiver Stations (BTS) which connect our cell phones to the networks. Therefore, the call detail Records (CDR) databases are generated by making or receiving phone calls or using a service via the connected mobile phone. The simulations shown that applying limited mobility reduced the total number of persons infected by the virus by 10% and delays the pandemic peak by two days. Figure 27 shows the difference with and without applying interventions so that there is a significant difference.

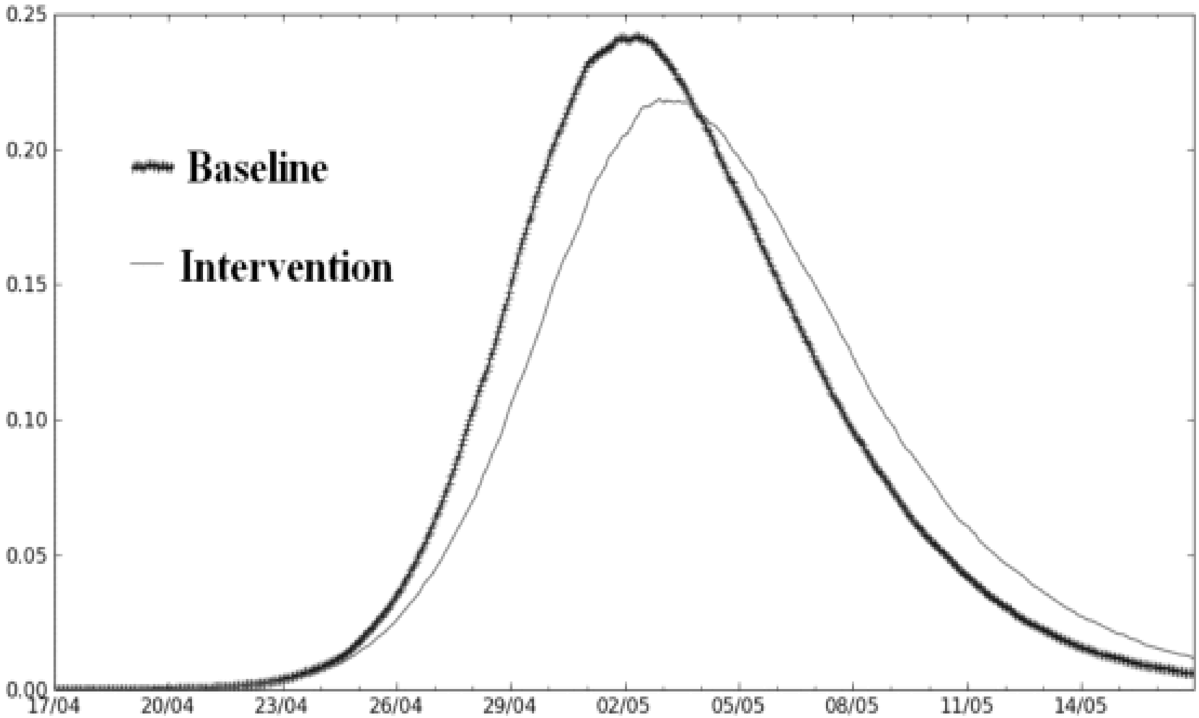


Figure 27: "Fraction of Infected Agents Over Time. These Curves are an Average of All Simulation Runs". (adapted from Frias-Martinez et al., 2011)

University of Texas at Austin's (UT COVID-19) Model

UT COVID-19 model (Figure 28) has been proposed by researchers University of Texas at Austin's that predicts COVID-19 cases rates based on SEIR model (Tec et al. 2020), and it quantifies the effects of social distancing intervention on disease spread through mobility traces, and to forecast the first wave of COVID-19 deaths in the United States

(Woody et al., 2020). It was incorporated with social networks to analyze the obtained data from mobile-phones GPS traces as same as FM model (Frias-Martinez et al., 2011), which allows predicting the effects of social distancing behavior on the disease spread curve. To explain, locations of mobile phones are inferred by SafeGraph company based on the daytime and overnight locations over time, where SafeGraph is a company that collects anonymized positions data from a variety of applications to gain information into physical locations. Furthermore, it was compared with the IHME model estimations, and it more accuracy in predicting the daily deaths per day. (Woody et al., 2020).

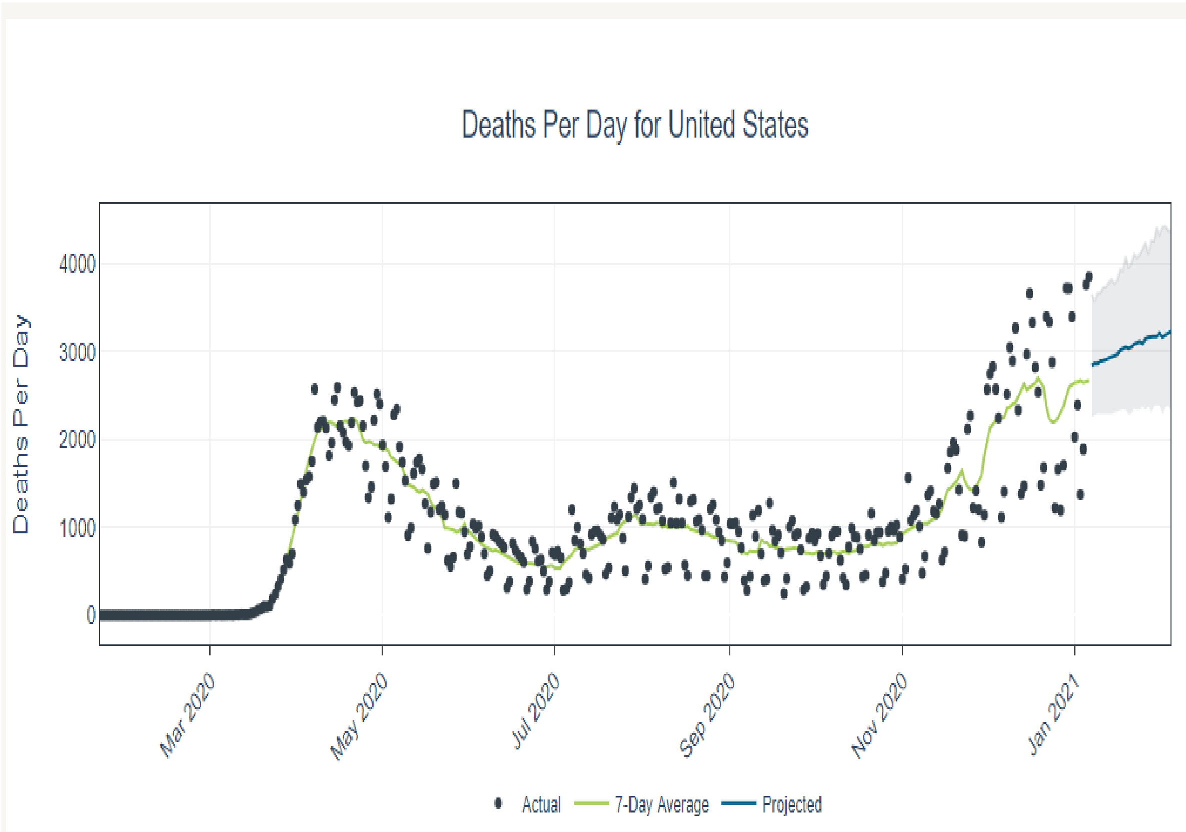


Figure 28: University of Texas at Austin's Model (adapted from "US dashboard," 2020)

Susceptible, Exposed, Infected, Recovered, Agent- based (SEIR-ABM) Model:

SEIR-ABM model (Figure 29) uses social interactions and the patterns of human mobility derived from call information records as same as FM and UT COVID-19 model in order to model the spread of viruses accurately (Silva et al., 2020). The suggested method was used to study the 2009 H1N1 epidemic in Mexico and to evaluate the effect of government interventions on virus spread.

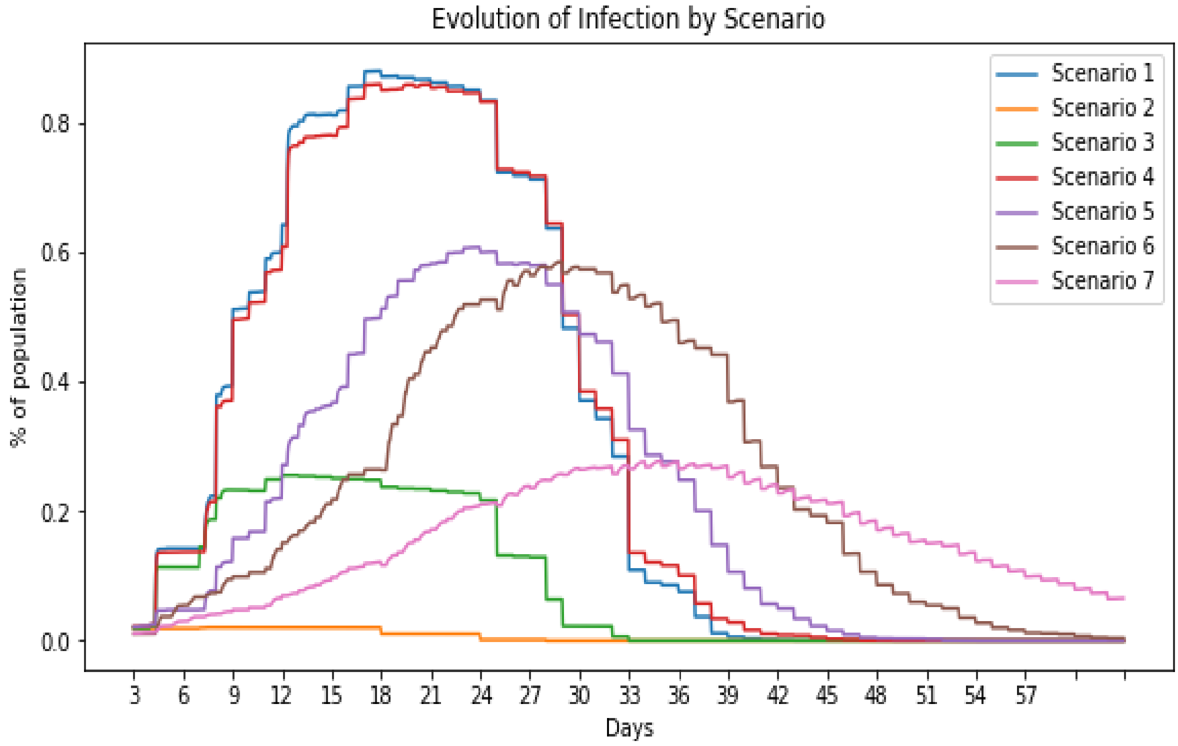


Figure 29: (SEIR-ABM) Model (adapted from Google Colaboratory, 2020)

The Developed Multi Agent Susceptible, Infected, Recovered (DMAS-SIR) Model

According to study (Vyklyuk et al., 2021), DMAS-SIR model (Figure 30) was developed through improving the classical SIR model, where new factors were added to the model , such as "incubation period, people's keeping a safe distance when moving, simulated quarantine, isolation, visiting public places such as supermarkets, parks, churches, schools, gyms, model transport, construction sites, gyms, etc." (Vyklyuk et al., 2021). So, it includes more parameters that could affect the accuracy of predictions of COVID-19 pandemic trends. Furthermore, the multiagent system approach was used to improve the model based on mobile cellular automata. In addition, the model proposed ways to improve the behavioral rules and interactions of agents, and it was carried out in countries including Slovakia, Turkey and Serbia, where the results showed that the model is precisely aligned with real data (Vyklyuk et al., 2021).

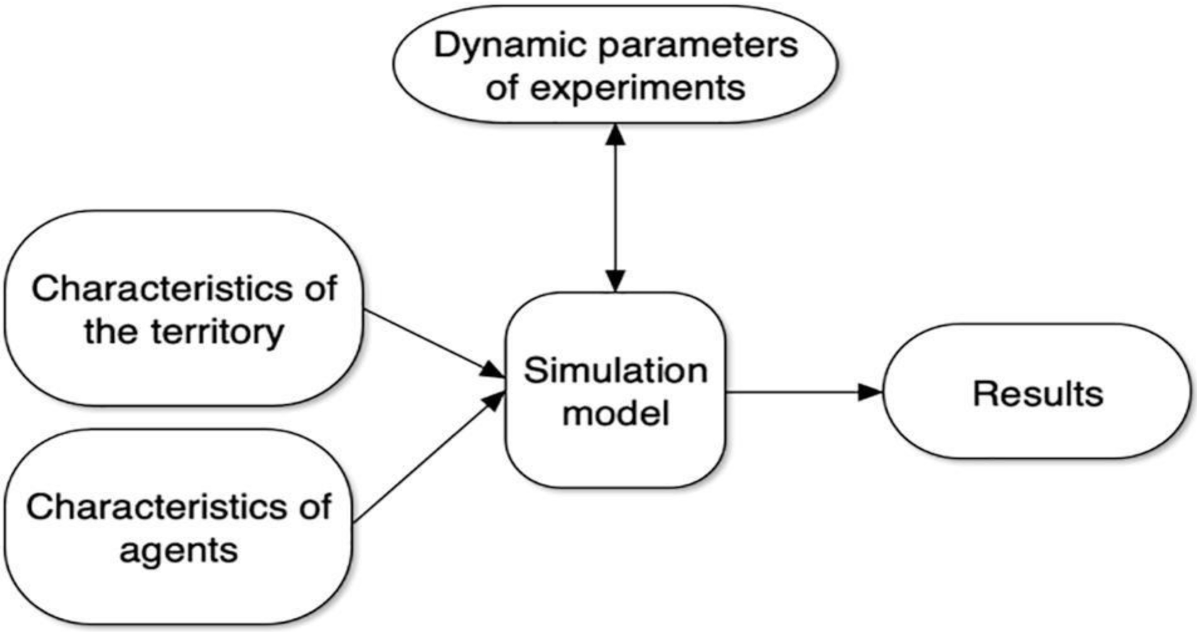


Figure 30: Functional Diagram of a Simulated Model of Virus Spread (adapted from Vyklyuk et al., 2021)

According to Figure 28 (Vyklyuk et al., 2021), The input variables are divided into two categories: territory characteristics (area, shape, number of agents) and agent characteristics (a group of agents with distinct characteristics). During the simulation experiment, you can change the features of the agents in real time, add new ones, and delete old ones, as well as save and restore data. In addition, the modeling process involves modeling the isolation, quarantine, public places with close contact, public places with the gathering of people, safe distance maintenance, regions, and international transfer.

Values description	Values	Base simulation	Simulations
Distance between people	d_{real}	2 m	4 m – Ex2
Days of illness	T_{real}	24	
Disease duration range (days)	t_{il}	7	
Number of agents	N	1000	
Probability of infection	p	0,1	0,05 – Ex3
Probability of death	D	0,05	
Percentage of people in quarantine	P_q	0	0,8 – Ex4
Incubation period (days)	T_{real}^{inc}	10	
Probability of visiting a public place	p_{sup}	0	0,001 – Ex5
Probability of symptoms presence	p_{sym}	0,8	0,5 – Ex7
Isolation of patients		No	Yes – Ex6,7
The number of sick agents to start isolation	I_{max}	10	
There is a public gathering place		No	Yes – Ex8
Maintain a safe distance		No	Yes – Ex9

Figure 31: Coefficients of the Basic Model and their Change in Sensitivity Experiments (Vyklyuk et al., 2021).

Also, from Figure 31, nine simulations experiments were run based of the level and strictness of the applied interventions and their levels. So, the results in Figure 32 have shown that how the interventions with different levels play significant role in the COVID-19 pandemic trend and disease spread.

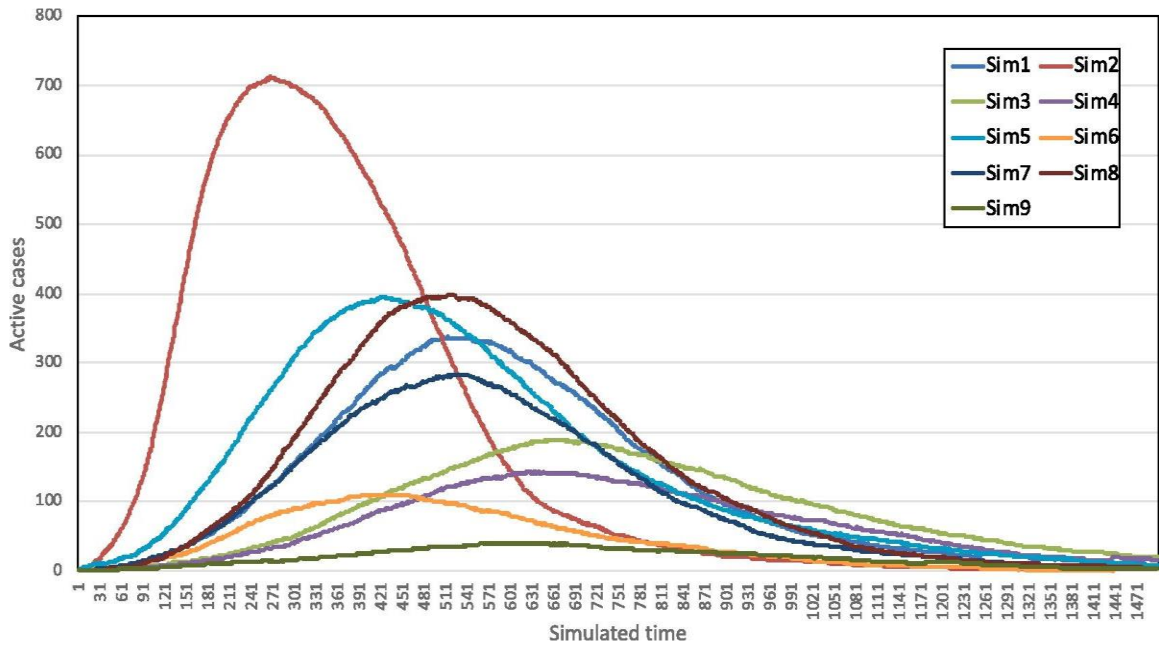


Figure 32: Dynamics of the Active Cases, Patients' Number for Different Nine Simulations (adapted from Vyklyuk et al., 2021).

Supervised Machine Learning Models

Supervised Machine Learning model is defined as "techniques or algorithms that bind previous and current dataset with the help of labeled data to predict future events. The learning process begins with a dataset training process and develops targeted activity to predict output values" (Muhammad et al., 2020).

Hybrid Models

These models employ different types of models into one model in order to design a new model that is stronger and has more features. For example, statistical techniques could be used to model disease parameters which are then used in epidemiological models to forecast cases rates. Majority of epidemiological state models are developing to become hybrid models, such as Youyang Gu [YYG] model (Giattino, 2020), (Gu, 2020), which uses a machine learning method based on the epidemiological state models, SEIR model.

Table 5: Hybrid Models

Youyang Gu COVID-19 (YYG)
Deep transfer learning (DTL)
University of Virginia Biocomplexity Center PatchSim COVID-19 (UVA COVID-19)
Institute for Health Metrics and Evaluation COVID-19 (IHME COVID-19)
Massachusetts Institute of Technology COVID-19 (MIT University COVID-19)

The Youyang Gu COVID-19 (YYG) Model

YYG model (Figures 33 & 34) applies the strength of artificial intelligence to the traditional infectious disease model. A simulator based on the SEIR model was developed to simulate the COVID-19 epidemic in each area. The parameters of this simulator are then trained using machine learning methods that aim to minimize the difference between the predicted and actual outputs. Data of reported deaths by each region are used to estimate the potential confirmed deaths (CDC, 2020) (Giattino, 2020).

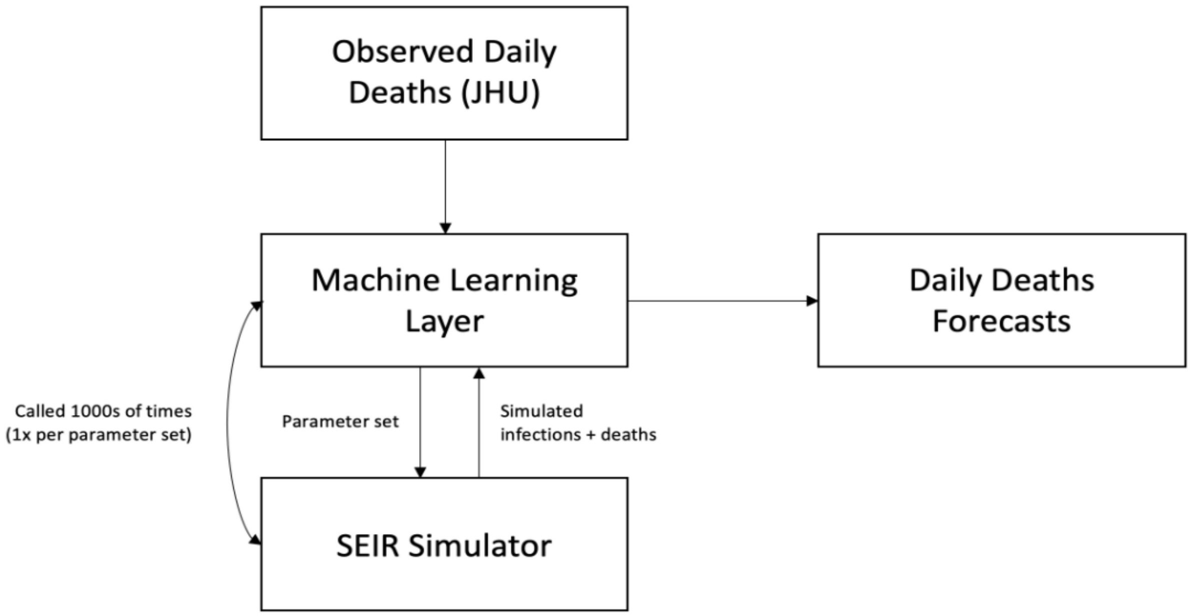


Figure 33: Machine Learning and SEIR Simulator Part (adapted from "Model details," 2020)

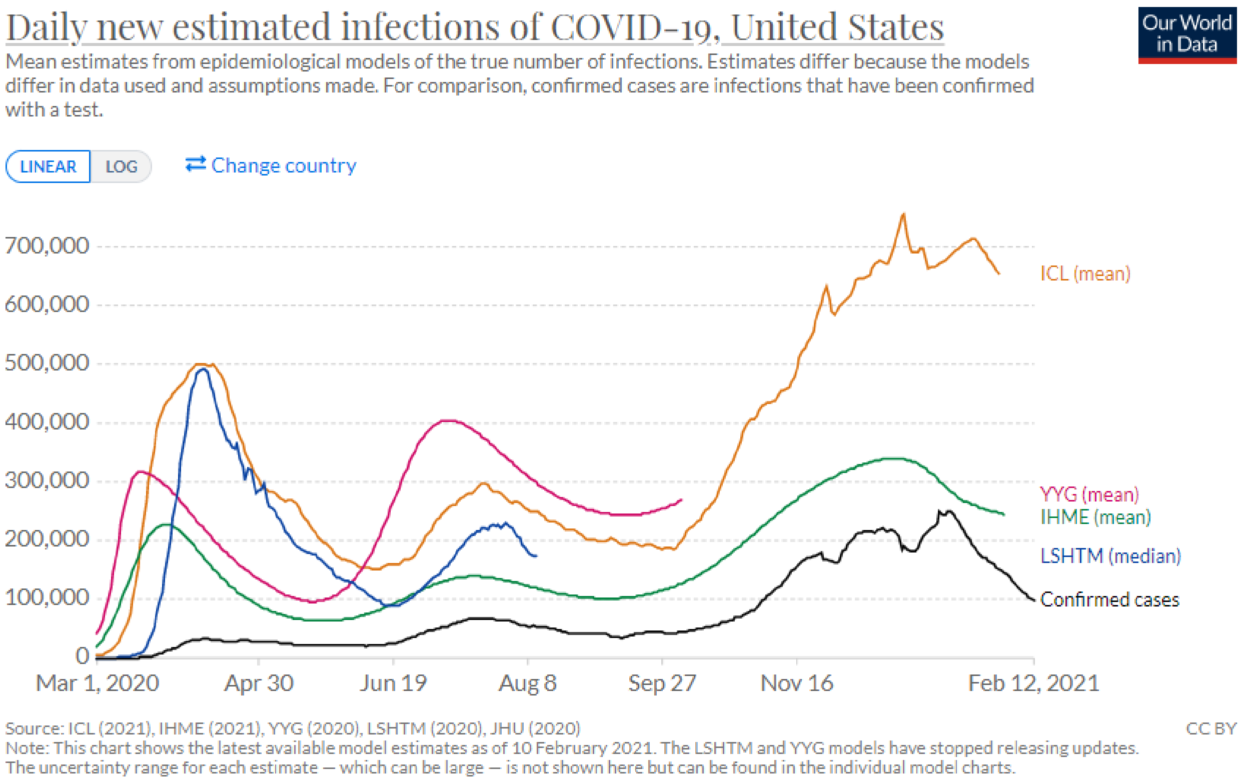


Figure 34: Youyang Gu's Model ("How epidemiological models of COVID-19 help us estimate the true number of infections," 2020)

The Deep Transfer Learning (DTL) Model

DTL model (Figure 35) modelers aimed to identify those who do not wear masks to minimize COVID-19 transmission and spread (Loey, Manogaran, Taha, & Khalifa, 2020). Since the use of face masks has shown that the spread rate of COVID-19 is reduced, the authors have built a new method for recognizing the statuses of wearing face masks among people. They have been able to identify three types of face mask-wearing situations. The types are right face mask-wearing, improper face mask-wearing, and no face mask-wearing. The new hybrid model involved two parts: A- deep and classical machine learning for face mask detection. B- transferring learning (**ResNet 50**) as feature extractor. So, they were combined in a new model which is better and more accurate.

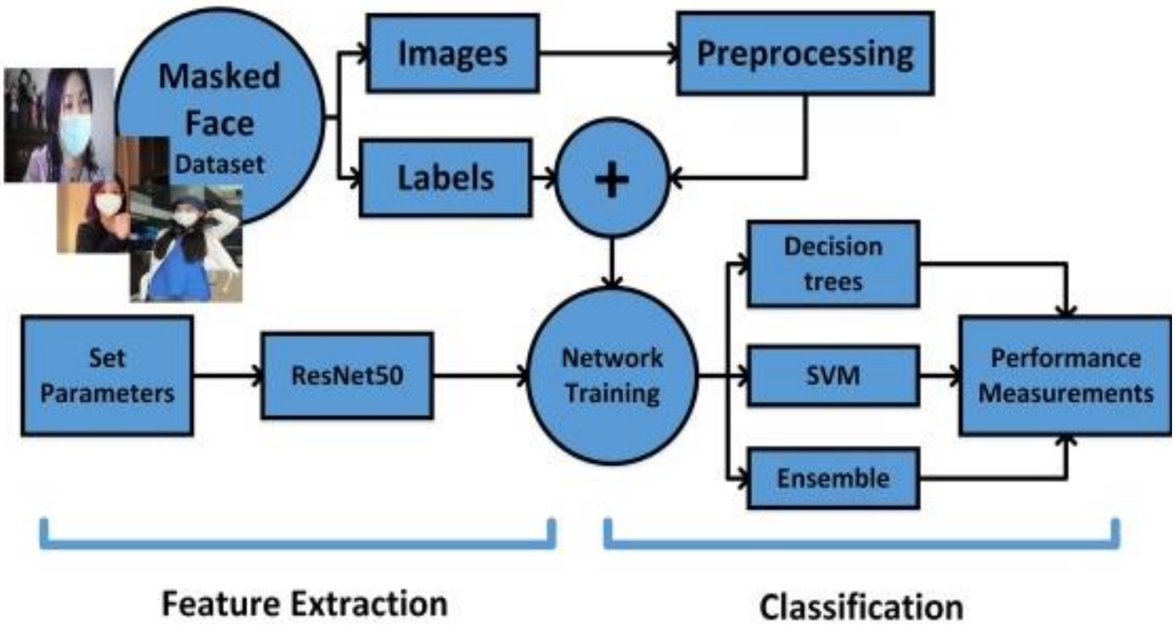
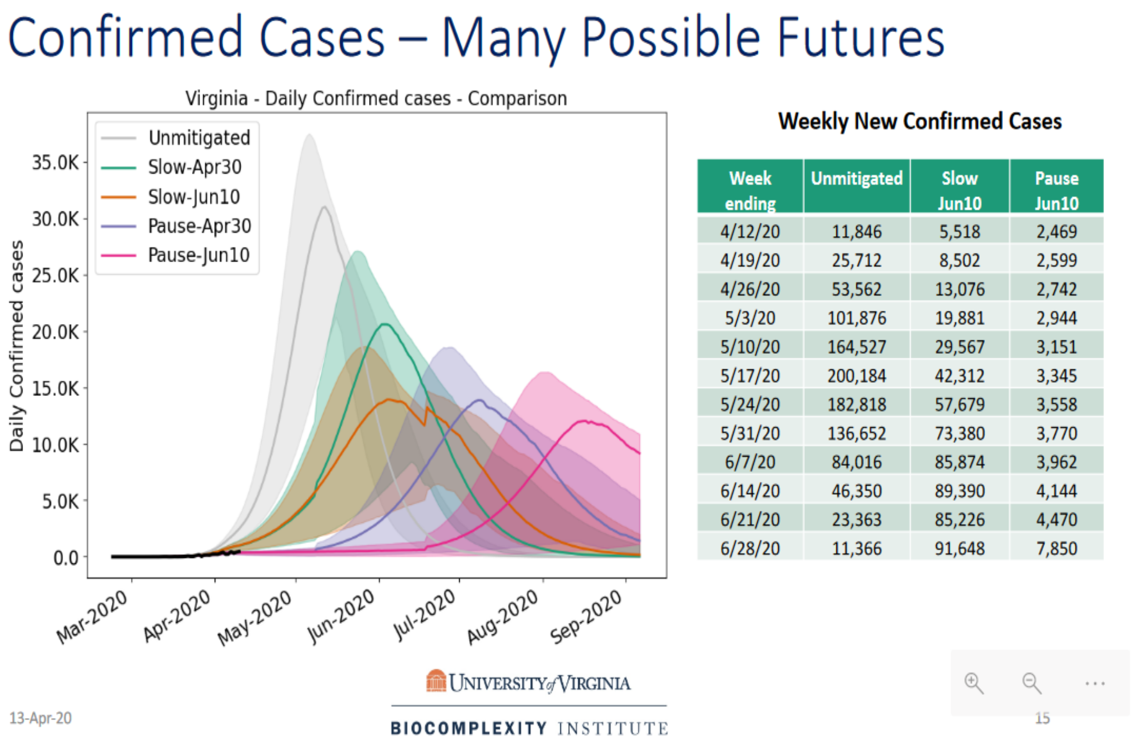


Figure 35: The Deep Transfer Learning Model (adapted from Loey, Manogaran, Taha, & Khalifa, 2020)

UVA COVID-19 model (Figure 36) is an extension for SEIR model, and it is used to support both federal departments and the state of Virginia (Price & Propp, 2020). Based on modeling team previous experience, they had refrained from making longer-term predictions instead of relying on short-term estimates. The model is used in the Projection Selection method, where a series of counterfactual scenarios are created based on-the-ground response and surveillance data, and the best fits are chosen based on historical results. Also, they identify the future potential situations, and serve to establish a fair narrative of previous trajectories, and retrospective comparisons are used for measures such as 'cases avoided by doing X.' These forecasts are updated weekly based on stakeholder reviews and reporting updates. Moreover, we can say that [UVA] Biocomplexity Center PatchSim model is a hybrid model since it includes the mobility tracing as a factor in the model (Virginia Department of Health, 2020).



Week ending	Unmitigated	Slow Jun10	Pause Jun10
4/12/20	11,846	5,518	2,469
4/19/20	25,712	8,502	2,599
4/26/20	53,562	13,076	2,742
5/3/20	101,876	19,881	2,944
5/10/20	164,527	29,567	3,151
5/17/20	200,184	42,312	3,345
5/24/20	182,818	57,679	3,558
5/31/20	136,652	73,380	3,770
6/7/20	84,016	85,874	3,962
6/14/20	46,350	89,390	4,144
6/21/20	23,363	85,226	4,470
6/28/20	11,366	91,648	7,850

Figure 36: UVA Model (adapted from "Virginia-specific model puts coronavirus peak in late summer," 2020)

Institute for Health Metrics and Evaluation COVID-19 (IHME COVID-19) Model

IHME COVID-19 model (Figure 37 & 38) has been established based on SEIR model in response to demands from the University of Washington School of Medicine and other US healthcare systems and state governments to assess whether COVID-19 cases rates would exceed their capacity to care for patients (Jewell, Lewnard, & Jewell, 2020). This model uses a hybrid modeling approach to produce its forecasts by cooperating both demographic forecasts intervention models and agent- based model. where the intervention of social distancing policy is tested through individuals' phone mobility traces, which is related to agent-based modeling approach. The model is updating frequently with new data and information, and it forecasts the demand for hospitals care, daily and cumulative cases and deaths due to COVID-19, rates of infection and testing, mobility and of social distancing, and using masks data on the disease spread, which are grouped by country and state for selected locations ("COVID-19 resources," 2020).

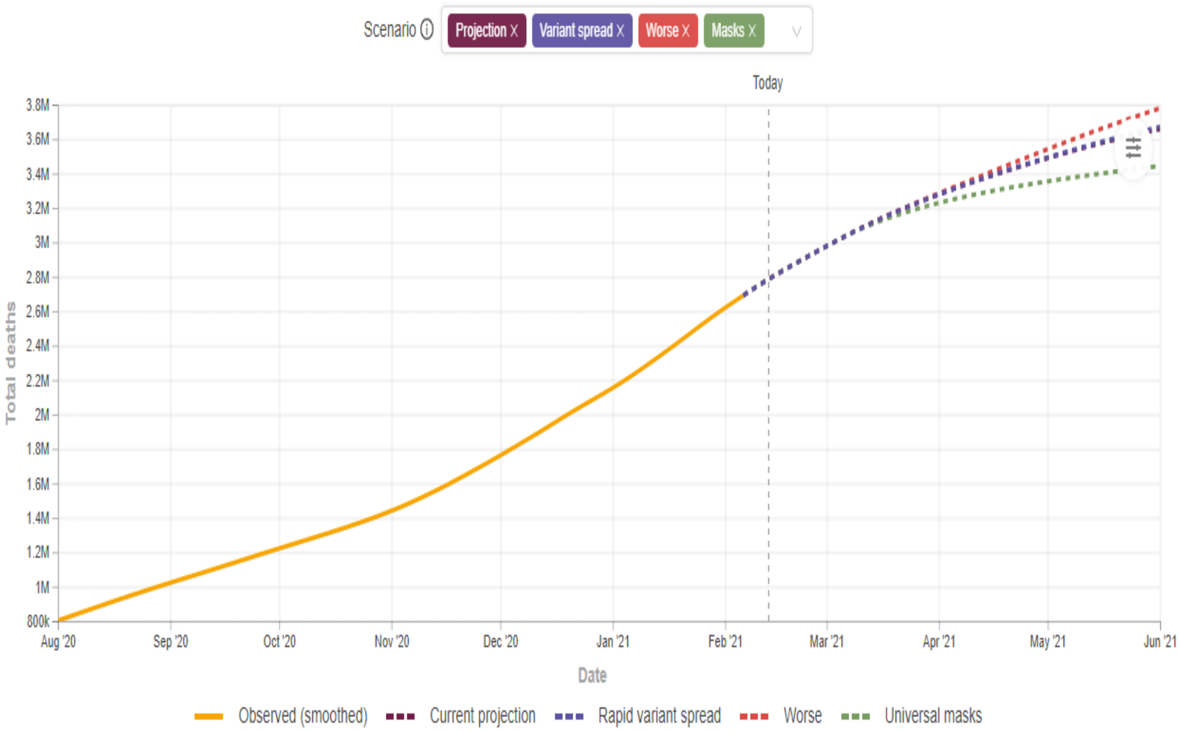


Figure 37: IHME Model ("IHME creates COVID-19 projection tool," 2020)

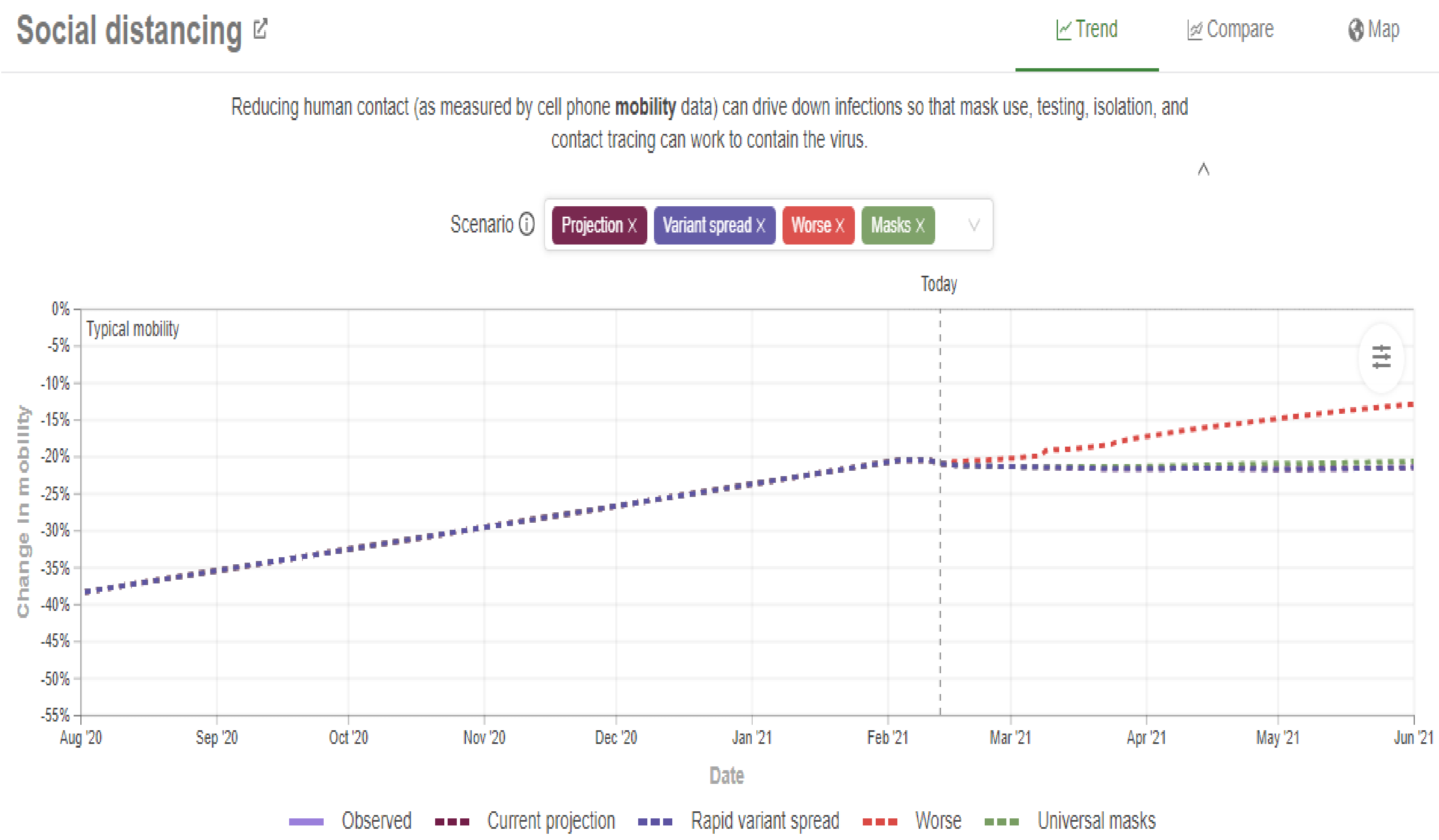


Figure 38: IHME Model – Social Distancing Part (adapted from "IHME creates COVID-19 projection tool," 2020)

Massachusetts Institute of Technology COVID-19 (MIT University COVID-19) Model

MIT has developed a model (Figure 39) that uses data from the Covid-19 pandemic in combination with the neural network ("Model quantifies the impact of quarantine measures on COVID-19's spread," 2020.) to assess the effectiveness of quarantine steps and to help forecast the distribution of the virus (Gallagher, 2020). Most models used to forecast disease transmission follow what is known as the SEIR model, which classes people into 'susceptible,' 'exposed,' 'infectious,' and 'recovered.' (Dandekar and Barbastais, 2020) improved the SEIR model by training a neural network to detect the number of infectious individuals that are under quarantine and therefore no longer transmission the infection to others.

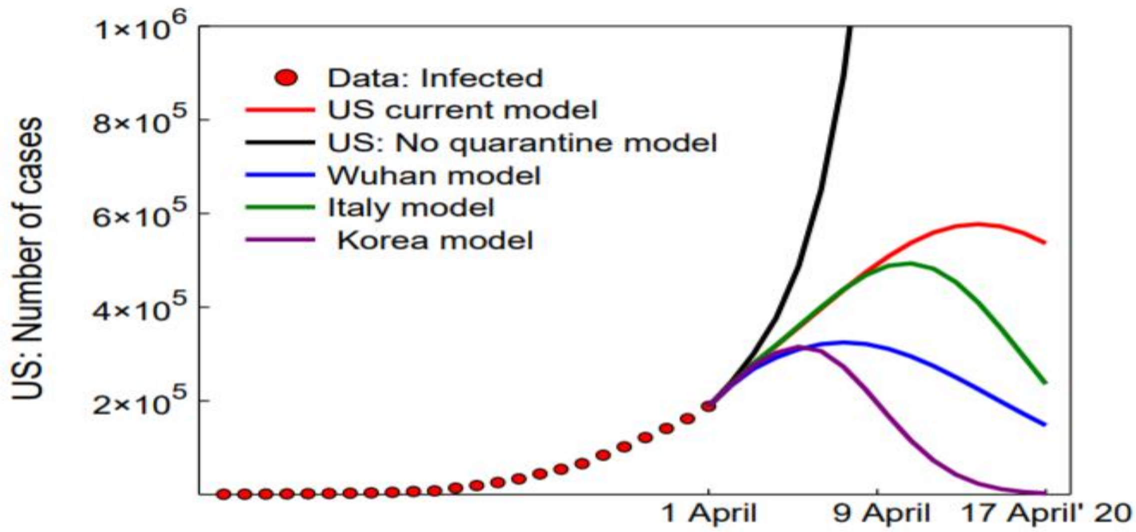


Figure 39: MIT Model (adapted from Gallagher, 2020)

<u>Social Media and Pandemics</u>

Social media can be used to communicate infection diseases outbreak alerts efficiently, and are important for public medical knowledge, promote people to follow the health behavioral (Collinson & Heffernan, 2014), and provide a strong resource of valid data. Infection diseases pose unknown risks to the public who also obtain the information from conventional media and the social media platforms (Freberg, Palenchar, & Veil, 2013). The question is how these Infection diseases are represented and conveyed in media format, and how that impact their decision making and risk management behavior (Slovic, 2001). Social media is considered as an influential factor that provides information for citizens rapidly and in a standard time. However, user-generated knowledge that is published on social media about infectious diseases is not always exact or useful and may involve rumors, misinformation, and theories on conspiracy (Obar & Wildman, 2015). Therefore, The World Health Organization calls social media to be more proactive to disseminate health messages to journalists, physicians, and the public, especially to counteract misinformation about infectious diseases (World Health Organization, 2012).

In recent outbreaks of infectious diseases, social networking websites have become the direct source for people to learn about the disease and share knowledge in real time with their families, friends, and neighbors. While researchers are studying more and more the role of the social media during pandemics, it is important to study thoroughly how the use of the social media can influence affect, understand, and regulate the affective reactions of the public (Xue et al.,2020). In addition, theoretical study continues to constrain population responses to infectious disease outbreaks. A study indicated that communicators and policymakers can take greater note of the role of emotions during outbreaks of infectious diseases (OFOGHI, MANN, & VERSPOOR, 2015). The results were reassuring people that they use social media not only to communicate accurate information but also to express their emotions on public health crises, and how these crises affect their view and response to a crisis (Alexander, 2013). To avoid outbreaks spread, the population should take necessary interventions through efficient communications (Guidry, Jin, Orr, Messner, & Meganck, 2017). Furthermore, decision-makers use social media to express either their voices and directions during a crisis or to interpret and assess people's thoughts on the event in the opposite way. Indeed, recent studies have used social media to provide the views of people, to handle them, to examine them, and to gain some insights. For example, sentiment analysis is a common method of analysis of social media content to identify the feelings of the public regarding a particular subject or event. Moreover, social media and networks contents could be translated to useful data for studying and analyzing the disease spread and impacts rates. So, we concluded that the integration between social media and disease predictive models involves two sides. The first side is using media to provide us with useful data for analysis, while the second side is using media to change individual behaviors and increase their awareness regarding health issues.

Social Media, Social Networks, and Models

In this part, we would discuss studies that incorporated using social media and social networks into *Epidemiological State models, Statistical Prediction models, Theoretical interventions models, Agent-based models, and Hybrid models.*

Epidemiology State Models and Social Media

Some epidemiology state models were integrated with social media in order to improve the outcomes of these models, such as Exposed, Infected, Hospitalized (EIH) model which was integrated with social media to study the effects of social media on the model output, where social media could motivate public behavior to follow the preventive measures, resulting in reducing the infected, exposed, or hospitalized cases numbers (Tchuenche and Bauch 2012; Cui et al. 2008; Liu et al. 2007). To explain, population tend to not modify their actions as easily in the early stages of the outbreak if the number of infected cases is small, only when infected cases numbers are increased, the public behavior change to protect themselves (Lu et al. 2017). According to Figure 40: e^{-mI} parameter reflects the media effects, formula $\left(1 - \frac{I}{(m+I)}\right)$ represents the human behavior, letter **I** represents the infected cases numbers, and **m** represents the media function. As shown in figure, there are two different values of media function (**m=0.2 on top** and **m=2 at the bottom)** are tested (Sooknanan & Comissiong, 2020). Figure 40 with its two parts shows that incorporating social media into the infection model impacts public behavior significantly, which leads to reduce the disease spread.

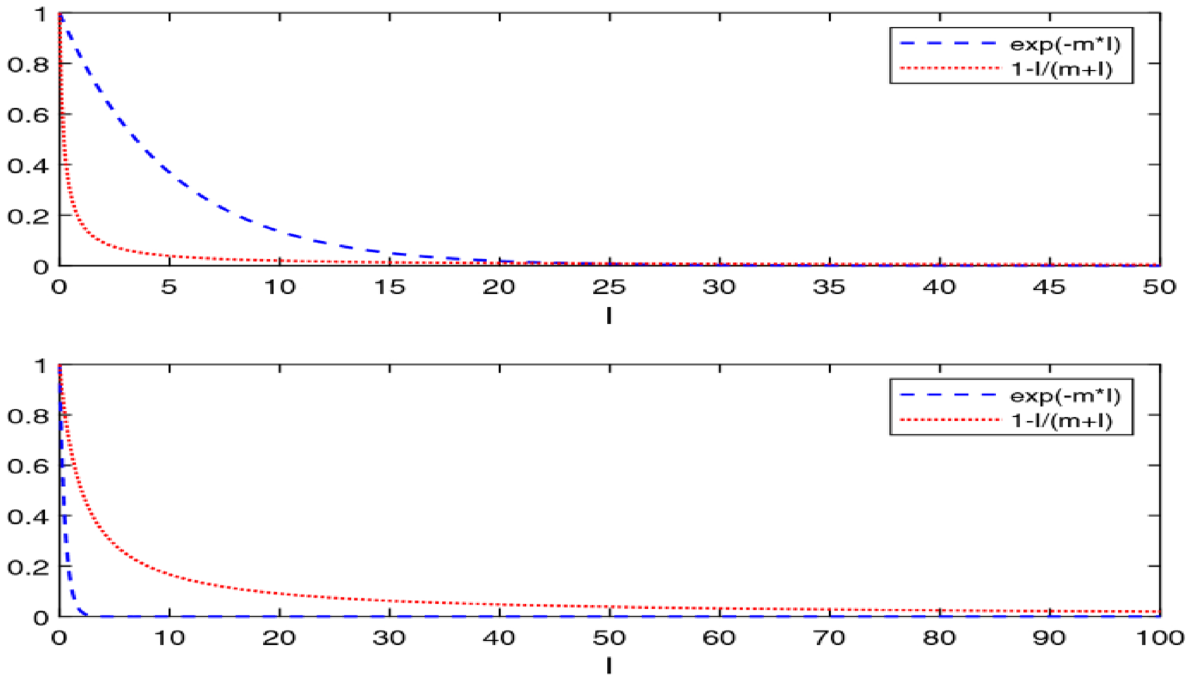


*Figure 40: Infectious Model, Values of Media Function (**m**=0.2 on top and **m**=2 at the bottom)*

Furthermore, the SEI model (Figure 41) was incorporated with different three levels of social media coverage (a= media coverage level), where the results have proven that increasing media influence can reduce the rates of suspectable and infected cases significantly (Lu, Wang, Liu, & Li, 2017)

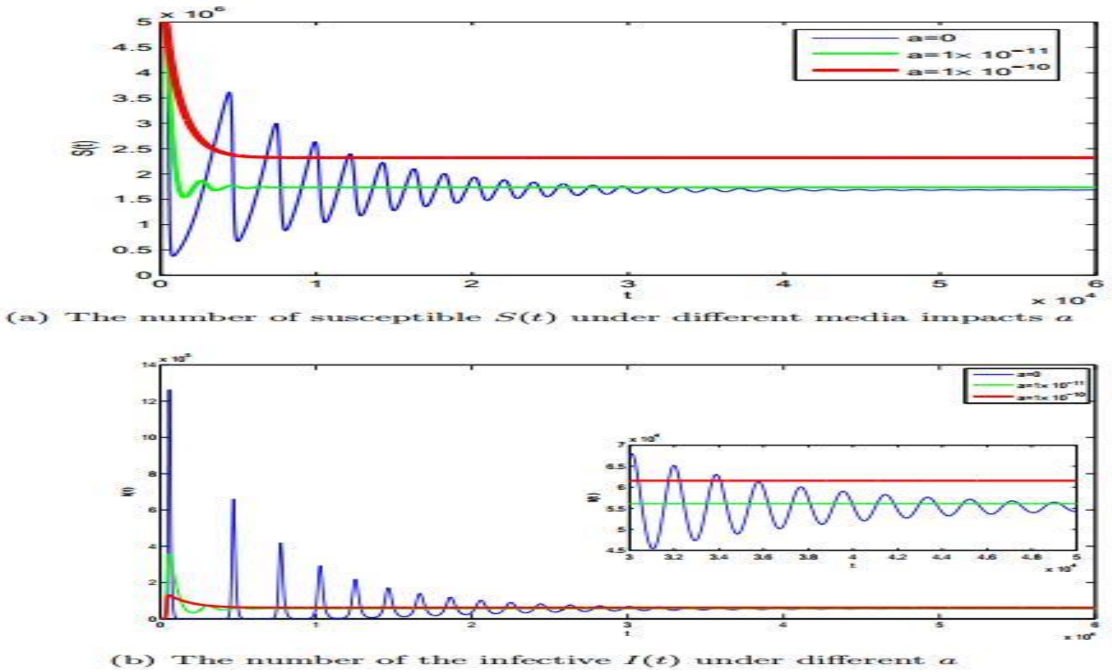


Figure 41: SEI Model: "Effects of Media Impact a on the Value of S(t), I(t) under Different Media Impacts" (adapted from Lu, Wang, Liu, & Li, 2017).

In another study by Tchuenche and Bauch (Tchuenche & Bauch, 2012), they have developed a *SIHR model* which incorporates a signal process that catches the media coverage effect. They state that the outbreak cannot be stopped by media attention, but It may control the spread of the virus. Their results regarding infected and hospitalized cases numbers under media coverage are shown in Figure 42, where the model was tested with two conditions: without media coverage and with media coverage. Moreover, the results show that media coverage can reduce the infected and hospitalized individuals dramatically, which supports what has been shown in figure 38 previously.

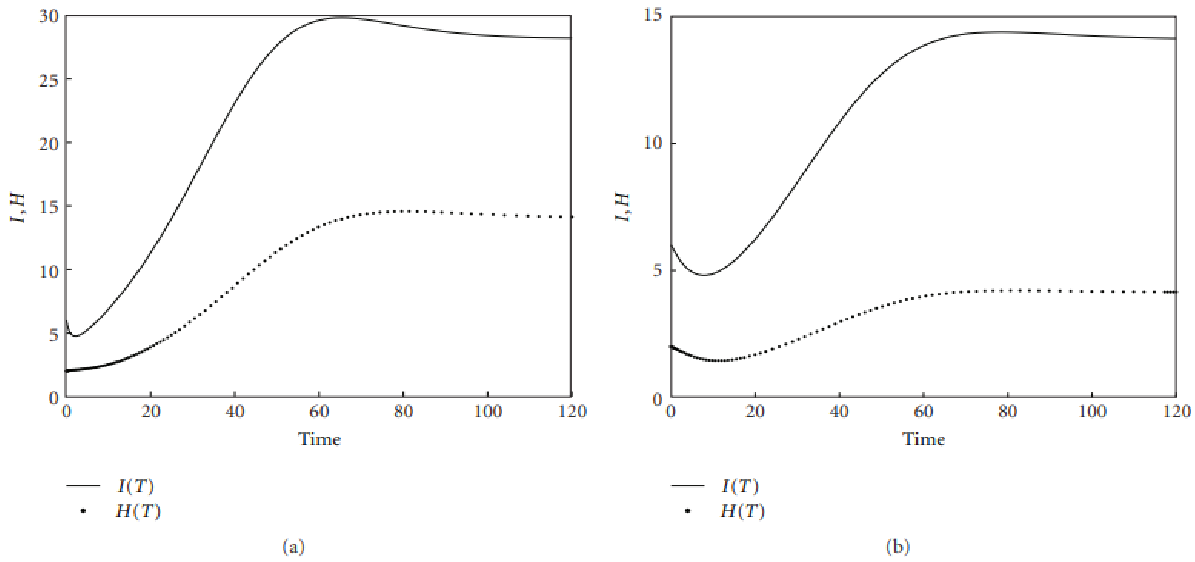


Figure 42: Numbers of Infected and Hospitalized Individuals under: (a) without Media Coverage, and (b) with Media Coverage, there is a Substantial Reduction in the Amount of. (adapted from Tchuenche & Bauch, 2012)

<u>Statistical Prediction Models and Social Media</u>

ARIMA model was incorporated with Twitter and google to use their data in forecasting process into a recent study (Samaras, García-Barriocanal, & Sicilia, 2020). The aim of this study was specifically to collect evidence as to which data source type leads to better results Twitter or Google? Data was acquired from the Internet on a computer that gathers data in real time for 23 weeks. Influenza data from Google and Twitter have been collected in Greece and compared with influenza data from the European Center for Disease Prevention and Control. Data was analyzed with the ARIMA model, which calculated estimates on a weekly basis. The results of this study show that outcomes from Twitter are significantly better than Google as shown in Figures 43 and 44, where Figure 43 shows that Twitter and the Europe Center for Disease Prevention and Control curves are close to each other. On the other hand, epidemiological models are not recommended to study the virus spread.

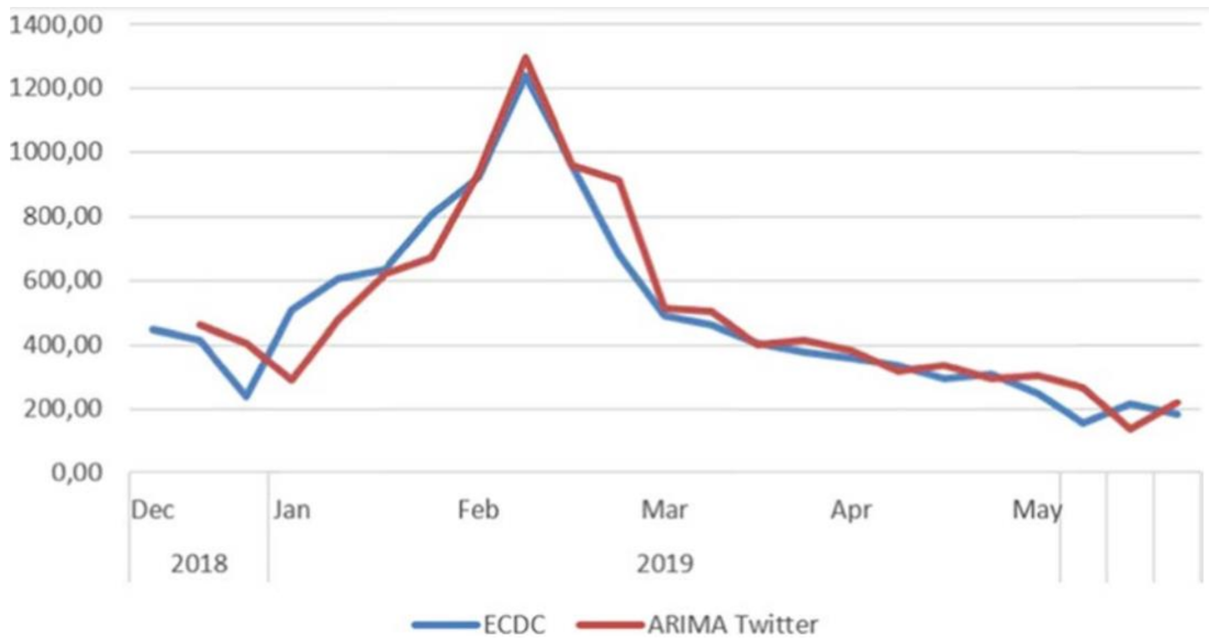


Figure 43: Integrating ARIMA Model with Twitter and Validating Outcomes with ECDC

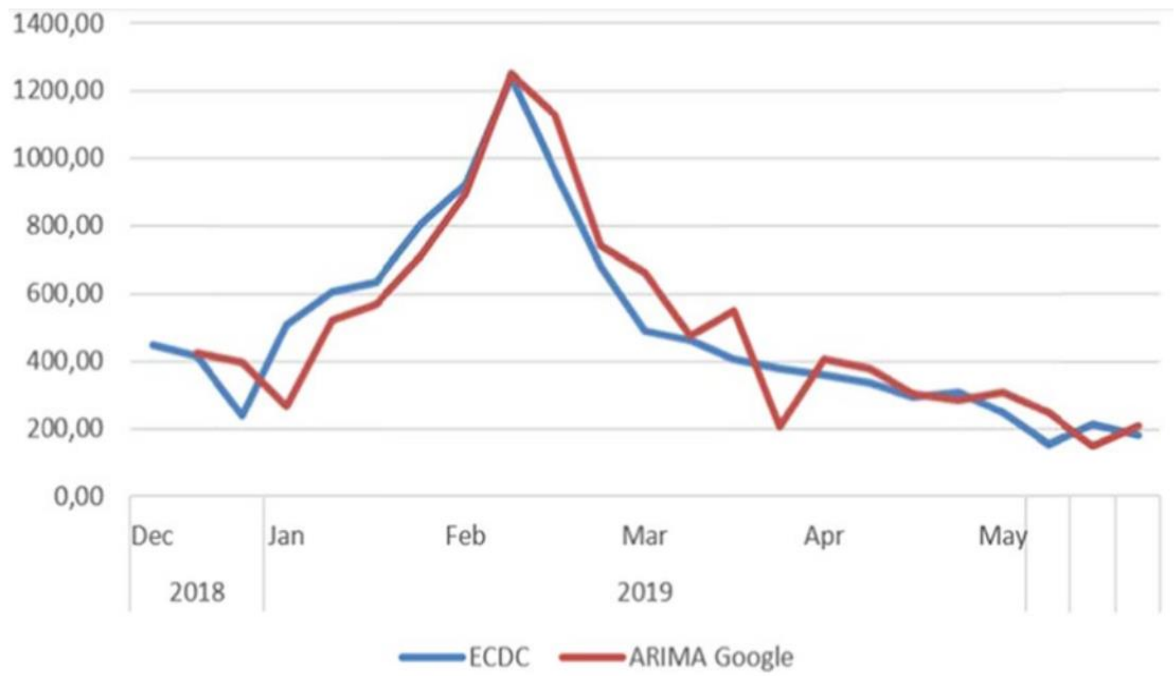


Figure 44: Integrating ARIMA Model with Google and Validating Outcomes with ECDC

Table 6: It Shows the Statistical Calculations and Compares them with ECDC Data, we can see that there are Differences Between Results of **ARIMA Google and ARIMA Twitter**

Table 6: Results of ARIMA Google and ARIMA Twitter

n	Statistics	ARIMA Google	ARIMA Twitter
1	R	0.933	0.943
2	sig.	0.00000000026	0.00000000005
3	MAPE	21.358	18.742
4	MEAN	503.37	505.03
5	Mean df	14.73	16.39
6	Mean df (%)	4.40%	4.74%
7	Predicted cases	11,522.46	11,558.94
8	Real cases	11,368.68	11,368.68
9	Peak	1,248.68	1,298.12
10	Peak df%	0.56%	4.54%

Theoretical Interventions Models, and Social Media

In a published study (Raamkumar, Tan, & Wee, 2020) about integrating social media into a common health promotion model, which is called Health beliefs model (HBM), where data was collected by analyzing the public comments on Facebook COVID-19 posts that published from three agencies: the Singapore Health Ministry, the Public Health in England (PHE), and Centers for the Prevention of Diseases, respectively, then classify them according to the Health beliefs model (HBM). Comments made on social distancing were labeled manually by a yes / no flag in all four HBM constructs. As we mentioned previously, these four constructs are perceived susceptibility, perceived severity, perceived benefits, and perceived barriers (Champion, Skinner, and others}. A recurrent unit-based recurring neural network models with a selected data set of 16,752 responses are trained and validated for text classification

(Figure 45). The model was evaluated with precision and binary cross-entropy losses. For checking the results of the MOH case study classification, specificity, sensitivity, and balance precision were employed.

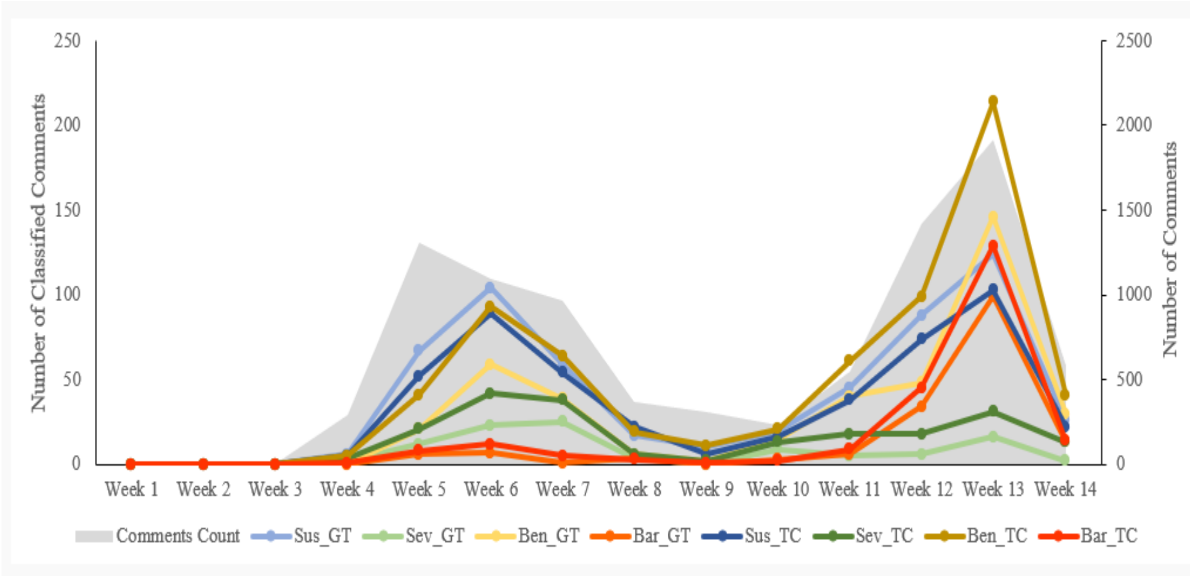


Figure 45: A Recurrent Unit-based Recurring Neural Network Model (adapted from Raamkumar, Tan, & Wee, 2020)

"Classification of Ministry of Health comments with Health Belief Model constructs. The primary x-axis is for the classified comments count for the Health Belief Model constructs, while the secondary x-axis is for the total comments count. Sus refers to perceived susceptibility, Sev refers to perceived severity, Ben refers to perceived benefit, and Bar refers to perceived barrier. Suffixes GT and TC refer to ground truth and text classification, respectively". (Raamkumar, Tan, & Wee, 2020)

<u>Agent-based Models and Social Networks</u>

Integrating agent-based models with social networks is commonly used in epidemiology field to overcome unrealistic expectations of homogenous blending that is used in conventional disease-depletion models on a differential basis (Rahmandad & Sterman, 2008). To explain, the transmission of a disease is influenced directly by people's behaviors and their social interaction (El-Sayed, Scarborough, Seemann, & Galea, 2012). Therefore, focusing on considering different behaviors among individual plays a significant role in exploring the disease transmission. Moreover, social networks are not only included in models that can spread epidemics, but also allow social factors such as their tendency to vaccine and compliance with hygiene, which can influence health outcomes (Will, Groeneveld, Frank, & Müller, 2020).

As we discussed previously, both FM and UT COVID-19 models are agent-based models, where they were designed based on the same idea and concept (using phone network to trace individuals' mobility). Therefore, we can say that the integration of the agent-based modeling approach and the social networking into these two models was done in the beginning when designing these models, and not as happened with other types of models that were previously designed and then the features of social media platforms or social networks were added to them. In general, these two models are an excellent source for obtaining some data that will benefit decision-makers in increasing or decreasing the percentage of preventive interventions.

Hybrid Models and Social Media

Twitter and Vaccination Prediction Model

In recent years, social media platforms became a significant resource of data and analysis. Content within tweets posts provide data to predict public behavior, beliefs, or opinions regarding specific events, personalities, or subjects. In terms of epidemiology, Sattar & Arifuzzaman, 2021 used Twitter and machine learning algorithms to capture and identify public sentiment toward vaccines based on around 1.2 million tweets collected across five weeks of April–May 2021 (Figure 44), They used their data and analysis of trends in population sentiment (Figure 45) to project that around 62.44% and 48% of the US population will get at least one dose of vaccine and be fully vaccinated, respectively, by the end of July 2021. The percentage of people with one dose and people fully vaccinated are 57.53% (164.45 million) and 49.53% (190.98 million) respectively as of July 31 2021 ("United States: Coronavirus pandemic country profile," 2021) indicating that their sentiment indicators for one dose were too optimistic while their two dose sentiment was pessimistic. These errors may indicate a limitation of Twitter-based sentiment analysis. Specifically, if Twitter usage is related to the ebb and flow of fears and the willingness of individuals to socially engage or adopt technology (discussed below) or some combination, then Twitter usage data of those fearful or more technology prone today will NOT be consistent with population behavior in the future who have different about vaccines vs the virus or do not express their fears on Twitter. Thus in the case of a first dose in a two dose sequence, where Sattar & Arifuzzaman over shot first vaccine administration by 5% (62.5 predicted vs 57.5 outcome), one could theorize the overshoot was due to either vaccination reaching a demographic of the population that (a) shunned social engagement and therefore is NOT represented on Twitter or (b) feared the vaccination more than the virus or (c) some combination. If social engagement on Twitter is viewed as use of a technology, Twitter would become a marginal predictor of population behavior due to large portions of population not using Twitter as discussed by Rogers with respect to any technology. If all demographics of people are equally engaged in Twitter, then better search terms would need to be designed to accurately identify levels of resistance for the first dose vaccination. If vaccination is itself viewed as a technology, then willingness of people to accept a technology is well established by Rogers as shown below in Figure 46. A similar logic may be applied to therapeutics for the vaccine. Broad availability of effective therapeutics lessens the fatality rate and hence lessens fear of the virus among the population.

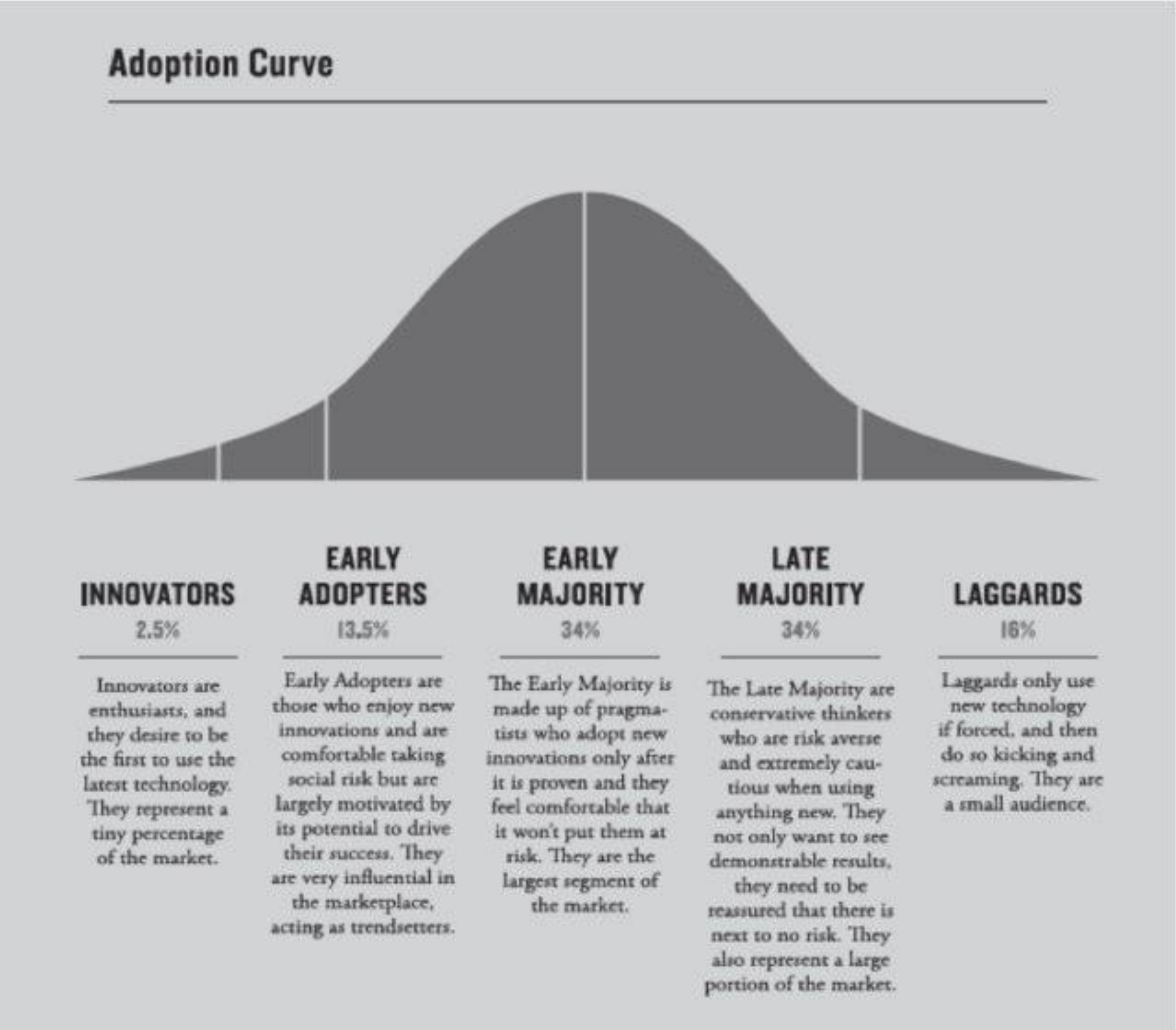
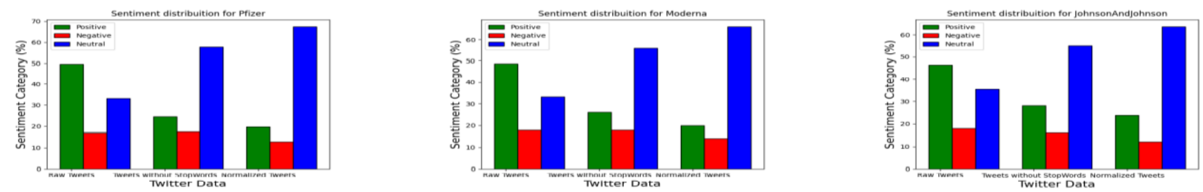


Figure 46: Rogers Technology Adoption Curve (adapted from Rogers, Everett M., Diffusion of Innovations, 5th edition, 2003)

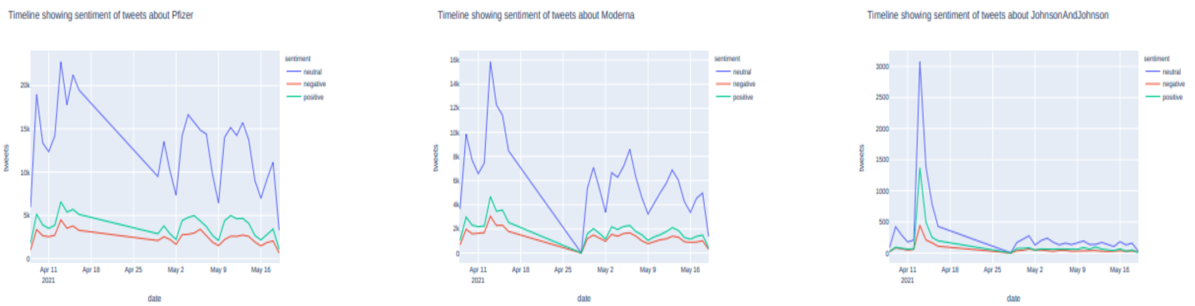
In general, the logical extension of Rogers Curve is the behavior of the first 50% of the population will not likely correspond to the behavior of the last 50% with the last 16% being quite resistant. For vaccination vs virus, the logical consequence of applying the behavior of the first 50% to the last 50% would be the over shoot in predicted first dose vaccination rates, which is what happened (62.5 predicted vs 57.5 outcome). At the same time, if administration of the first dose reduces fears in individuals about the vaccine, then behavior would change more positively toward accepting the second dose. The drop in fear would explain the under shoot of second dose vaccinations predictions by Sattar & Arifuzzaman (48% predicted vs 49.5% actual).

To explain Sattar & Arifuzzaman approach, the extraction step focused on tweets that includes vaccination keywords, such as, Pfizer, Johnson and Johnson, Moderna, Sonic, Etc... Analysis classifies public opinions about COVID-19 vaccine sentiment as positive, negative, or neutral providing sufficient insights to provide projections of vaccination intentions for the users in specific location and specific timeframe. These search terms may need to be expanded to include additional scales to measure strength of sentiment. In additions, the following forecasting models were used to predict the numbers of partially and fully vaccinated people, (a) SVM (b) KNN (c) Linear Regression (d) Random Forest (e) M5 model tree (f) Gaussian (g) Multilayer Perceptron. Regarding validating the forecasting results to the real-world data, the prediction of fully vaccinated models were closed to the real recorded data according to two of used machine learning models (SVM & Multilayer Perceptron model) as shown in Figures 47 & 48. Moreover, the predictions of partially vaccinated numbers were closed to real-world data for three machine learning models (SVM, Linear regression, & Multilayer Perceptron model) as shown in Figures 48, 49, and 50.



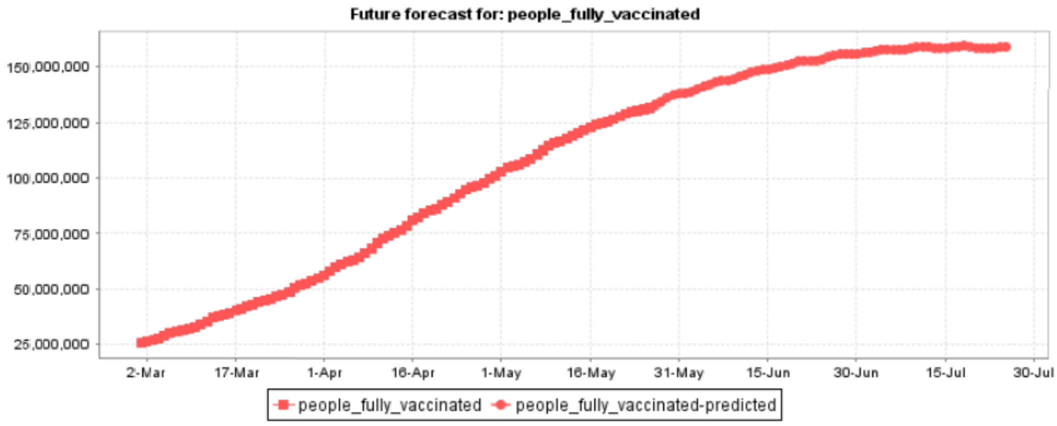
(a) Pfizer (b) Moderna (c) Johnson & Johnson

Figure 47: Sentiment Analysis of Users' Opinions



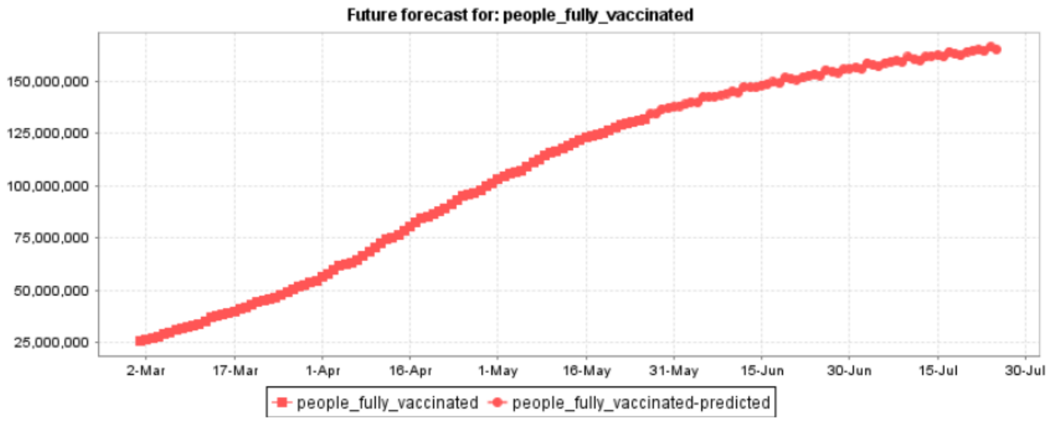
(a) Pfizer (b) Moderna (c) Johnson & Johnson

Figure 48: Sentiment Analysis Curve of Users' Opinions during April 2021- May 2021



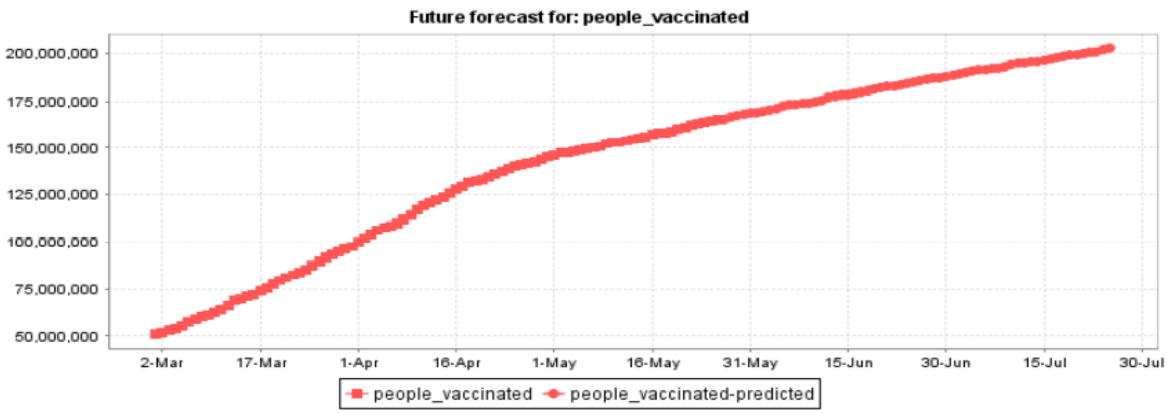
(a) SVM

Figure 49: The Predicted Fully Vaccinated Numbers in US by SVM Model



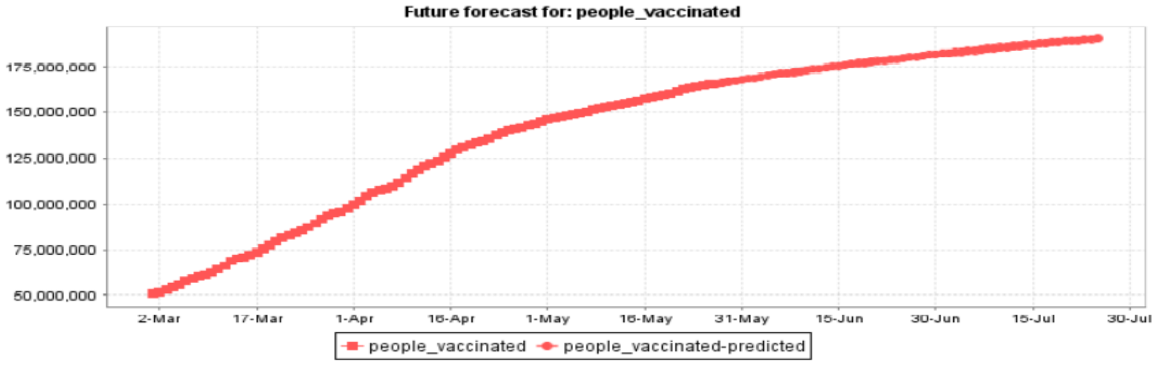
(g) Multilayer Perceptron

Figure 50: The Predicted Fully Vaccinated Numbers in US by Multilayer Perceptron Model



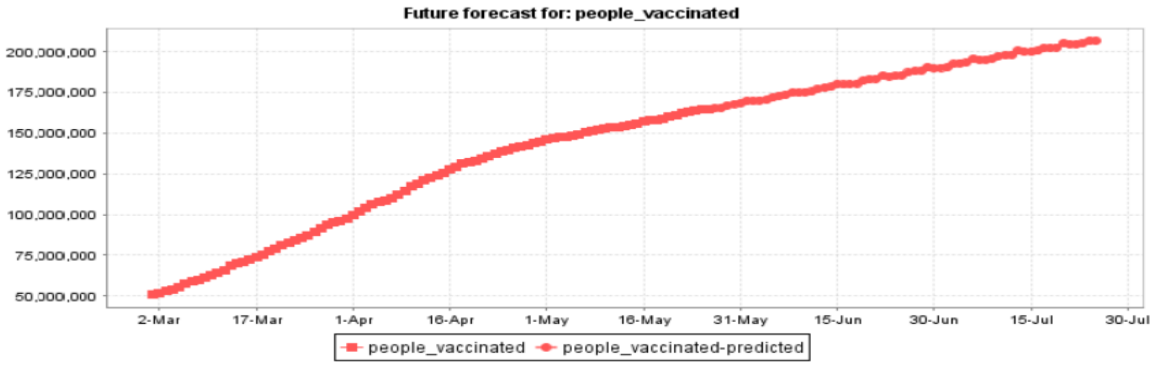
(a) SVM

Figure 51: The Predicted Partially Vaccinated Numbers in US by SVM Model



(c) Linear Regression

Figure 52: The Predicted Partially Vaccinated Numbers in US by Linear Regression Model



(g) Multilayer Perceptron

Figure 53: The Predicted Partially Vaccinated Numbers in US by Multilayer Perceptron Model

Past successful integration of social media into Epidemiological models promises future improvements to their efficacy but the field is still in the Innovators stage of the Roger's Technology Adoption Curve Lifecycle (Rogers, 1995) and clearly less defined than the science of epidemiology itself (Matthews & Proctor, 2019; Matthews & Proctor, 2021).

Based on Rogers research, it will take considerable research successes for epidemiological models to diffuse from the Innovators stage, where researchers who are enthusiastic about new AI and Hybrid Model technologies, to the Early Adopter stage as

Rogers indicates users will want to form a solid opinion of the technology before they vocally support it. Working in favor of further diffusion, social media offers rich real-time data and data mining of social media may uncover changes in human behavior more quickly than traditional data collection methods. An obvious challenge will be data validation, though even hoaxes promoted on the Internet can result in adverse social behavior in the short term (Bond, 2021). Early identification of behavior -whether based on fact or fiction - will help models rapidly adapted to reflect that change and provide information to responsible authorities to counter falsehoods and inappropriate behavior. Similarly early identification of behavioral changes may contribution to improved communications, protocols, and programs promoted by governments, organizations, societies, or individuals. In the longer term, improved model efficacy can speed the implementation of appropriate and timely interventions and reduce the disease spread or delay it. As for the models themselves, the extent of validation of epidemiological models using social media may be estimated using official sources.

This review indicates social media may help researchers and modelers increase the level of efficacy of *Epidemiological State Models*, *Statistical Predictions Models*, *Theoretical Intervention Models, Agent-Based Models,* and *Artificial Intelligence And Hybrid Models* but further work is needed. Increasing efficacy infers increasing public confidence in epidemiological model outputs and related protocols and policies. Increasing public confidence in the models, protocols, and policies infers more stable and compliant behavior pattern that if scientifically based will decrease the pandemic spread, promote public health protocols, and support vaccine implementation plans. Gleaning rationale for behavioral choices, such as vaccine hesitancy, from public commentary expressed through social media channels has yet to be reported in the literature. This research gap results in a failure to avail modelers of quantifiable and articulated social media sources of feedback useful for rapidly modifying or refining pandemic vaccination plans. Based on our review, with the exception of vaccination modeling, we conclude that social media platforms have proven their ability improve epidemiological state models, statistical predictions models, and theoretical intervention models. For the COVID19 complex situation with divergent infection rates, inconsistent national and state health practices, and compounding international and intranational demographical divergent beliefs and behaviors, we conclude that various forms of agent modeling combined with social and traditional media data sourcing may produce the most efficacious models and contribute to study vaccination acceptance. Regarding vaccination hesitancy, Prediction modeling may be used to analyze Twitter users' opinions and quantify their hesitancy, which provides significant insights to aid decision-makers in making policies.

Given research limitations, we plan to study and analyze Twitter users' opinions about vaccinations after Delta variant came to the world, and other events such as booster shot and children vaccination approval. Outcomes of the research are expected to further add insights and hopefully contribute to benchmarks on model efficacy for COVID-19 vaccination prediction models. While limited in scope, the opportunity to gather data during this crisis cannot be missed as the pandemic will pass. Once passed and if insights are not gathered and documented, then the world will be less prepared when the next pandemic strikes.

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

Introduction

Lack of public confidence in the efficacy of models, effectiveness of protocols, safety of vaccines or therapeutics, and appropriateness of policies confounds pandemic spread predictions, undermines public health protocols, and frustrates vaccine implementation. Gleaning rationale for behavioral choices, such as vaccine hesitancy, from public commentary expressed through social media channels may provide quantifiable and articulated sources of feedback useful for rapidly modifying or refining pandemic spread predictions, health protocols, vaccination offerings, and policy approaches. Based on the prior literature review, machine learning and sentiment analysis of Twitter-based commentary may not only provide a snapshot of vaccine hesitancy among a target population but provide trends supporting rationale for those trends greater than that obtainable through statistical extrapolation of past vaccination data (Sattar & Arifuzzaman, 2021). In addition, sentiment analysis combined with machine learning techniques extracts and analyzes large amounts of data in a shorter time with less cost than competing manual surveys or interviews.

This research proposes to utilize these powerful and novel methodologies within the field of epidemiology with the limited scope of investigating the research questions and hypotheses listed below in the body of this chapter. In general, this research investigates public sentiment related to individual: risk toward the virus and vaccines, vaccination state, acquired immunity state, broadly available outpatient therapeutics (e.g. Pfizer Novel COVID-19 pill) that are effective against the virus, and governmental and corporate policies and vaccine mandates during the current COVID-19 pandemic in the United States for the period June 1, 2021 to March 31, 2022. Search terms include categorical response factors on sentiment for: risk, variant type, vaccine type, vaccine booster availability, acquired and natural immunity, individual vaccine state, pill-form therapeutics, FDA emergency use authorizations, full FDA approvals, and government and corporate policies and mandates. Beyond the scope of this research are the impacts on population behavior arising from therapeutics associated with hospital or clinic admissions. Analysis includes descriptive, parametric and non-parametric statistical assessment of relationships between factors and possible trends. Sentiment outcomes will be qualitatively compared to published polling on public sentiment to understand the degree of confidence decision makers may have in this technical approach to sentiment analysis.

Our pre-test of the methodology involved the Delta variant. As the Delta variant began in June 2021 to spread across the United States with higher transmission rates and more severe symptoms than earlier COVID-19 virus variants, the numbers of infected and deaths cases increased dramatically peaking in Aug 2021. Our preliminary research investigated effects of fear of the Delta variant on population vaccination rates in the United States by analyzing Twitter sentiments data and the relationship between the two. Research questions include, have vaccination and virus sentiments changed due to the Delta CV19 virus variant spread and its increased risk profile? Tweet analysis results discussed below show sufficient insights into users' attitudes and opinions regarding COVID-19 vaccines during Delta cases peak in Aug 2021. In addition, the vaccination rates have been increased at the same time compared to June and July 2021 according to CDC data, which confirmed that fear of Delta variants has played a significant role in increasing vaccinations. Thus, we concluded that the methodology was effective to explore the relationship between public fears of Delta variant spread and the target populations' intentions towards taking COVID-19 vaccines.

Moreover, in Oct 2021, Twitter activity regarding vaccinations increased significantly due to the FDA booster shot approval and children vaccination approval. Public opinions expressed on Twitter about these approvals appeared to be more neutral than positive or negative. Greater neutrality or ambivalence, implies less urgency among the population to get booster shots or vaccinate their children than seen to get vaccinated during the Delta wave.

Based on the success of the preliminary research, the methodology may be applied to other epidemiological incremental or stretch topics. Incremental analysis includes sentiment change due to changes in CDC vaccination and population state categories, FDA approval status (e.g. emergency use vs full approval) of the vaccine, FDA approvals status for booster shots and FDA approval status of different age groups among children. Another significant area of sentiment analysis of public commentary that arose in October is the nature and scope of vaccine mandates by the Federal government. Analysis proposed on this topic relates vaccination hesitancy to the Rogers Adoption Curve. As the Rogers Adoption Curve identifies 16% of the overall U.S. population will resist adoption of a new technology, given that a new vaccine is a new technology, then one may infer 16% of the United States population will resist vaccination without mandates. This novel hypothesis relates to the vaccinated proportion of the FDA emergency use eligible population to the long standing 16% resistance level associated with the Rogers curve. Testing the hypotheses of equality of proportionality would entail binomial tests and confidence intervals. A likely alternative hypothesis of inequality between proportions is that vaccine hesitancy arises due to: (1) latent laggard tendency identified by Rogers; (2) plus proportion of the population that perceives they have derived or natural immunity acquired by surviving the virus; and (3) plus fears about severe and long-lasting side effects of vaccines on certain demographics segments of the population. That infers that the proportion of the vaccine eligible U.S. population categorized as a vaccine hesitancy segment will exceed the laggard segment sizes observed by Rogers. Thus we expect to reject the null hypothesis of equality. Lastly, on November 5, 2021, Pfizer announced a Novel COVID-19 Oral Antiviral Treatment Candidate Reduced Risk of Hospitalization or Death by 89% in Interim

Analysis of Phase 2/3 EPIC-HR Study (https://investors.pfizer.com/investor-news/press-release-details/2021/Pfizers-Novel-COVID-19-Oral-Antiviral-Treatment-Candidate-Reduced-Risk-of-Hospitalization-or-Death-by-89-in-Interim-Analysis-of-Phase-23-EPIC-HR-Study/default.aspx). "Today's news is a real game-changer in the global efforts to halt the devastation of this pandemic. These data suggest that our oral antiviral candidate, if approved or authorized by regulatory authorities, has the potential to save patients' lives, reduce the severity of COVID-19 infections, and eliminate up to nine out of ten hospitalizations," said Albert Bourla, Chairman and Chief Executive Officer, Pfizer. A 90% reduction in hospitalization will greatly reduce the fear of the virus and hence motivation to get a vaccine. Since Pfizer plans to submit the data as part of its ongoing rolling submission to the U.S. FDA for Emergency Use Authorization (EUA) as soon as possible, the likely availability of a Pfizer anti-viral pill during the course of this research is likely to result in considerable Twitter discussion, changed attitudes about vaccines, changed population behavior, and changed policies. Particular impacts of anti-viral therapeutics associated with hospitalizations are not tracked by available CDC data is beyond the scope of this research.

Fear Impacts and Vaccine Hesitancy

Essential to understanding incremental research hypotheses is understanding fear factors among the target population. Fear factor can impact the public opinions regarding talking the vaccine significantly, where it could increase or decrease the desire to be vaccinated based on type of fear. Thus, we classified fear levels into two types or directions as shown below:

A- Fear of a Covid-19 variant (A motivation to take the vaccine).

B- Fear of side effects of a vaccine (A motivation to reject the vaccine).

Hypothesized fear of disease and its effects on accepting Covid 19 vaccine or rejecting it are shown below as direct and inverse relationships. **Figure 54** shows how the acceptance level of the vaccine increases directly with the increase in the fear of transmission of infection or the frightening effects of the disease. However, **Figure 55** shows how the fear of the side effects of taking the vaccine will lead to an increase in the fear of the vaccine and thus increase the resistance to it in a direct relationship and inverse relationship with the acceptance of vaccination. Fear rules:

A- As fear of disease increases, vaccine acceptance increases.

C- As fear of vaccine side effects increase, vaccine acceptance decreases

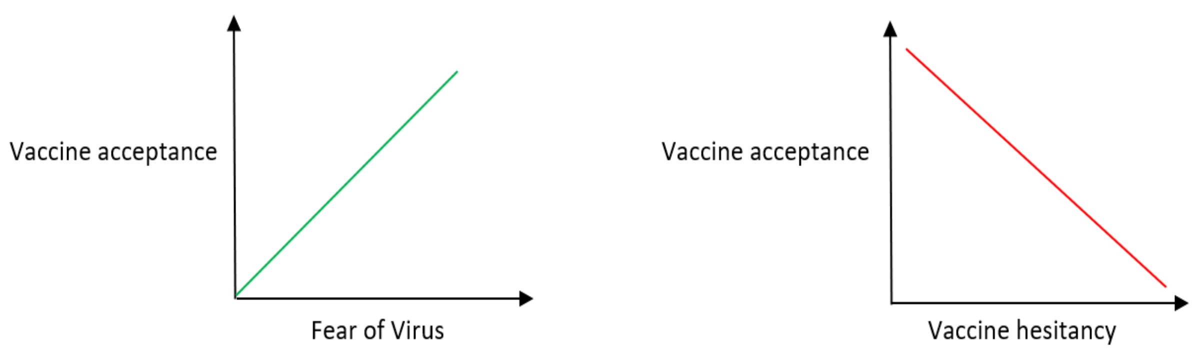


Figure 54: Fears and Vaccinations Relationships

During the period June 1 to Oct 31 2021, virus fear and vaccine acceptance were inversely related to one another with acceptance being based on decisions of an individual. Complicating vaccination status is the expiration of the vaccine effectiveness and the need for booster shots to maintain vaccine effectiveness. On September 22, 2021, the FDA approved boosters in the United States (**https://www.fda.gov/news-events/press-announcements/fda-authorizes-booster-dose-pfizer-biontech-covid-19-vaccine-certain-populations**).

On November 4, 2021 the Biden administration announced vaccination mandate policies. (**https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/04/fact-sheet-biden-administration-announces-details-of-two-major-vaccination-policies/**). This obviously disrupts prior relationship and trends between vaccine fear and vaccine acceptance. Further, on November 5, 2021, Pfizer announced their "game changing" oral antiviral (pill) that reduces the threat of hospitalization or death by 89%. Over the course of this research and should the emergency use be approved by the FDA, the likelihood of a drop in fear of the virus may result in a significant change toward vaccinations and boosters.

Vaccination Topics, Extraction, and Validation

To extract the users tweets from Twitter regarding the vaccination topics during the selected timeframe, preliminary research went through the steps shown in Figure 57. Thus, the basic events and topics that happened during the timeframe were investigated sentimentally by analyzing Twitter users' opinions. Then, the Twitter outcomes were compared to our vaccination model and CDC vaccination datasets to associate and confirm the effectiveness and efficiency of the selected methodology.

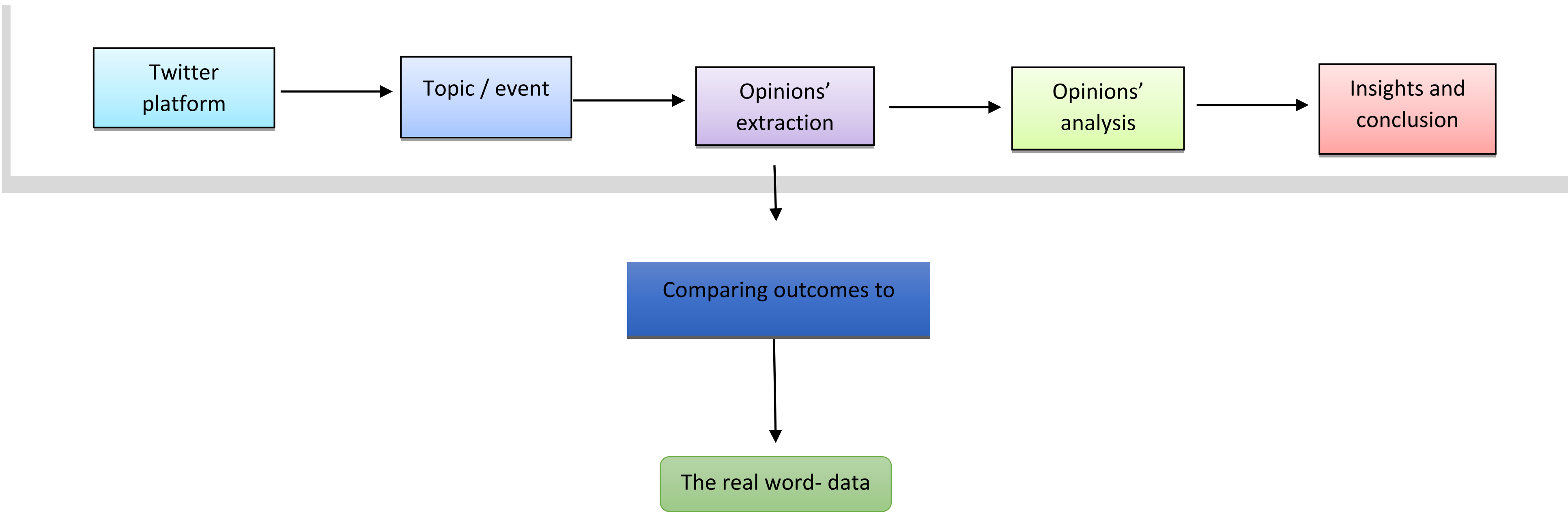


Figure 55: Vaccination Topics, Extraction, and Validation

Preliminary Research findings

In order to measure the vaccine hesitancy and if the Delta variant emergent affect the vaccination intentions in the USA, the tweets about Pfizer, Moderna, and Johnson and Johnson were extracted, preprocessed, and analyzed. So, we aimed to find out if the sentiment about these vaccines have been changed during the timeframe (June 1st, 2021 – Oct 31 2021) due to Delta variant emergent, increasing in cases and deaths rates, in addition COVID-19 vaccines and booster doses approvals. Thus, the preliminary results were able to answer to the following research question:

RQ: Is vaccine sentiment data extracted from Twitter data sensitive to change in Vaccine fear level among the public?

HA0: Positive vaccine sentiment data extracted from Twitter data is not sensitive to change in vaccine hesitancy level among the public for the time period snapshots June, August, October 2021

HA1: Positive vaccine sentiment data extracted from Twitter data is sensitive to change in vaccine hesitancy level among the public For the time period snapshots June, August, October 2021

HB0: Negative vaccine sentiment data extracted from Twitter data is not sensitive to change in vaccine hesitancy level among the public for the time period snapshots June, August, October 2021

HB1: Negative vaccine sentiment data extracted from Twitter data is sensitive to change in vaccine hesitancy level among the public for the time period snapshots June, August, October, October 2021

Methodology Selected

Posted tweets on Twitter reflect people' opinions, ideas, and attitudes towards a particular event by exploring the degrees of acceptance, rejection, or neutrality. In addition, some techniques are used to analyze comments on social media platforms such as Twitter in order to get some insights into public behaviors. Moreover, a previous proposed study (Sattar & Arifuzzaman, 2021) was able to measure vaccination hesitancy, so it shown the public intentions to be vaccinated in different countries during April- May 2021. The study was conducted by extracting Twitter users' tweets and analyzing them by using sentiment analysis approach. In addition, it aimed to forecasting vaccinations future trends in the United Sates for June -July 2021 by supervised machine learning forecasting model. However, it was conducted before Delta variants has spread widely in the States and the last events regarding booster shots and children vaccination approvals. Thus, in our study, we aimed to study effects of Delta variants on the vaccination rates, where Delta Variants has increased the fears levels which affect the public opinions about importance of vaccination. In addition, the booster shot, and children vaccination have been approved recently which resulted in higher public attentions on Twitter platform. Therefore, the quantitative research methodology was selected to do our study and find the answers for the research questions.

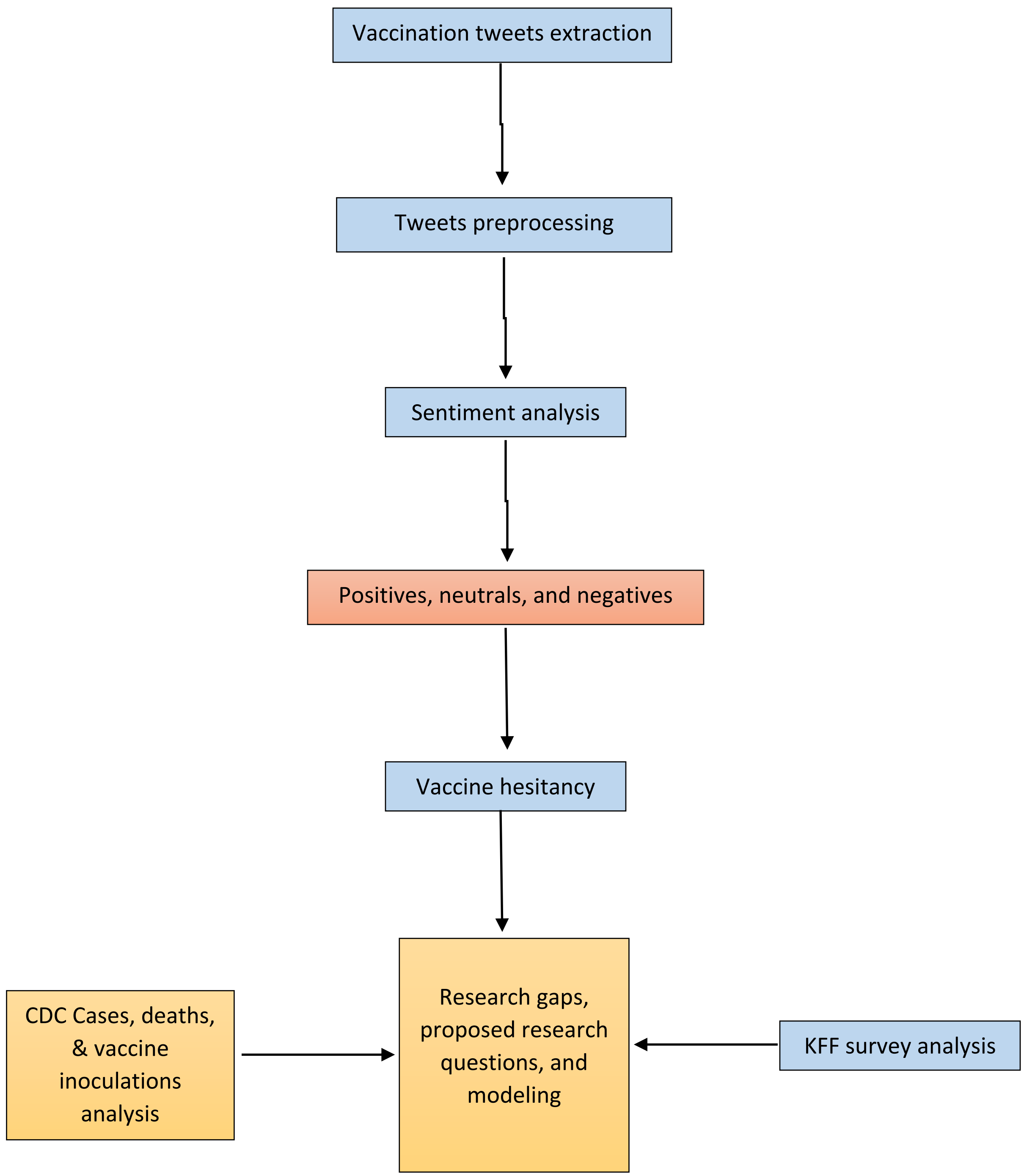


Figure 56: Preliminary experiment methodology

Data Collection

In our study, we used two sources for data collection process: where we used Twitter platform to extract users' tweets and analyze them to measure the public opinions regarding the selected topics. In addition, we used CDC vaccination database and included into our work to confirm the outcomes of Twitter.

Twitter Data

Users' tweets were extracted through Tweepy library tool which is a useful tool that extracted a large number of tweets in short time. Also, the extraction techniques focus on the tweets that include the queries keywords that we are interested in. Furthermore, the location of Twitter users and time frame were included into extraction process. In this work, we focused on the United States users and extracted their tweets about the selected vaccination topics and keywords. To explain, we aimed to study and analyze the vaccine hesitancy levels for three COVID-19 vaccines in the United States (Pfizer, Johnson & Johnson, and Moderna), and selected time frame from June 2021 to Oct 2021. Furthermore, the selected timeframe includes some events, such as Delta variant spread, booster shots approvals, and children vaccination approvals. Thus, the extracted data was useful to study and analyze the public opinions about the vaccination during the events.

CDC Data

To compare and confirm our results, we used CDC vaccination dataset. Thus, the Twitter users' attentions and positive sentiments changes have shown significant raise in Aug and Oct

2021, where CDC dataset has shown significant increase in vaccination numbers in Aug 2021 too. In addition, following the booster shots approval in Oct 2021, the booster vaccinations numbers have shown significant numbers. Also, the children vaccination approval has pulled a lot of public attentions, and the coming period will show the vaccination data.

To build extract and analyze our own Twitter dataset, we used the extraction, preprocessing and classifying steps based on (Rustam et al., 2021) work, which is shown in Table 7 and Figure 55, where the dataset was built by using Tweepy library, Vader library, Jupyter Notebook and Python programming language.

Table 7: Twitter Data Extraction Steps

Step	Description
Vaccination tweets extraction	Use of the technique and the Tweepy library to extract tweets from the Twitter API relevant to vaccinations (Mushtaq et al., 2022; Morshed et al., 2021)
Tweets preprocessing	Use of the technique to reduce tweet content into sentiment components (Ramachandran & Parvathi, 2019; Aljedaani et al., 2020;Monkey, 2020; Rsutam et al, 2021)
Sentiment analysis	Use of the (Devika et al., 2016; Yadav & Vishwakarma, 2019; Dang et al., 2020; Pipis 2022) technique and the Vader library to assign a sentiment score to each relevant tweet.

Twitter Data Extraction Step

Since Twitter platform is considered as the most common social media community among other platforms (Ruz et al., 2020), we used it to build our own dataset. Moreover, we also aspired to use specific time scales of notable events that occurred from the releasing the vaccines to the present day in order to study how these events affected people's desire to be vaccinated. Moreover, we planned to use root tweets for our analysis, so no retweets included. In order to extract the tweets regarding our subject, we worked on the word queries which lead machine learning model to extract tweets that include these words. For our work, some words were used, such as COVID-19, vaccination, vaccines, fear, vaccine acceptance, hesitancy, side effects, and interventions. We used Tweepy library to extract the tweets that includes the words and queries (Morshed et al., 2021), where “Tweepy is an open-source Python package that gives you a very convenient way to access the Twitter API with Python” (Python, 2019). To give a clear idea about Twitter API, where API stands for Application Programming Interface. Twitter API “lets you read and write Twitter data. Thus, you can use it to compose tweets, read profiles, and access your followers' data and a high volume of tweets on particular subjects in specific locations” (Fontanella, 2021).

Timeframe and Queries Keywords of Preliminary Research

We specified a timeframe (**1st June 2021 – 31st Oct 2021**) to extract tweets that include the queries keywords from the United States in order to study and analyze effects of fears of Delta variants spread on vaccination rates. In addition, we aimed to investigate opinions about the booster shots and children vaccina approvals events. Thus, we focused on extracting sample of tweets that includes the keywords that are shown in Table 8.

Table 8: Timeframe and Queries Keywords

Vaccine	Keywords	Timeframe
Pfizer-BioNTech vaccine	Pfizer, Pfizer-BioNTech, BioNTechpfizer	June 2021 – 31Oct 2021
Johnson & Johnson's COVID-19 Vaccine	Johnson & Johnson, Johnson and Johnson, Janssen, Janssen	June 2021 – 31 Oct 2021
Moderna vaccine	Moderna, Moderna_tx, Moderna-NIAID, NIAID, NIAID-Moderna	June 2021 – 31 Oct 2021

Twitter Data Preprocessing Step

In this step, the tweets were preprocessed since is an important step, which affect the learning models accuracy significantly. Thus, the stopwords, usernames, link punctuations and numeric values from tweets had been removed. (Rsutam et al, 2021)

Twitter Data Analysis Process Step

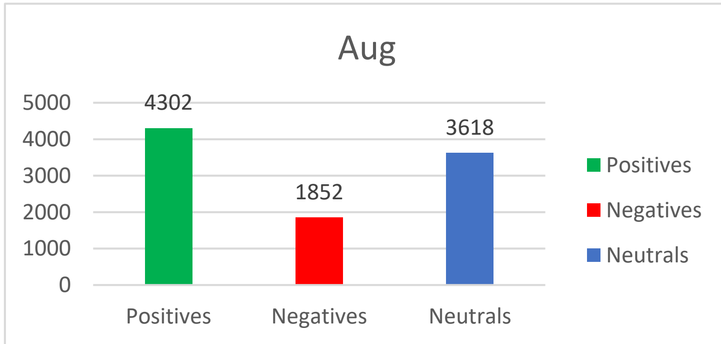
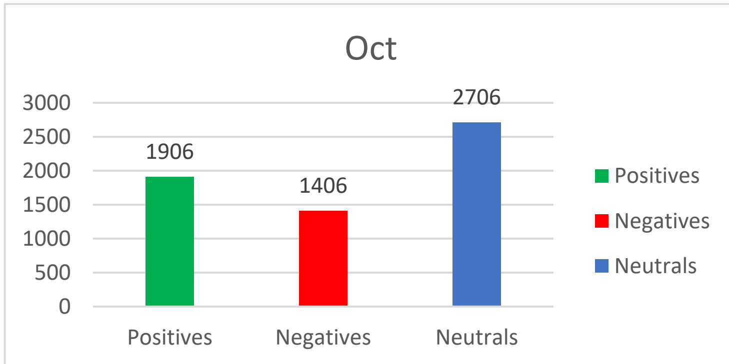
For our dataset analysis process, sentiment analysis was used in order to find out the sentiment scores for the users' opinions, where the scores are classified into positives, negatives, and neutral. Sentiment Analysis is used to measure and analyze people's opinions, beliefs, emotions and classify them into positives or negatives, or neutrals. To give a clear idea about the sentiment analysis, it is "also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. To execute the sentiment analysis and quantify the sentiment scores, the text-blob library tool (toolkit) was used, where Vader library is “a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more” ("TextBlob: Simplified text processing — TextBlob 0.16.0 documentation," 2021).

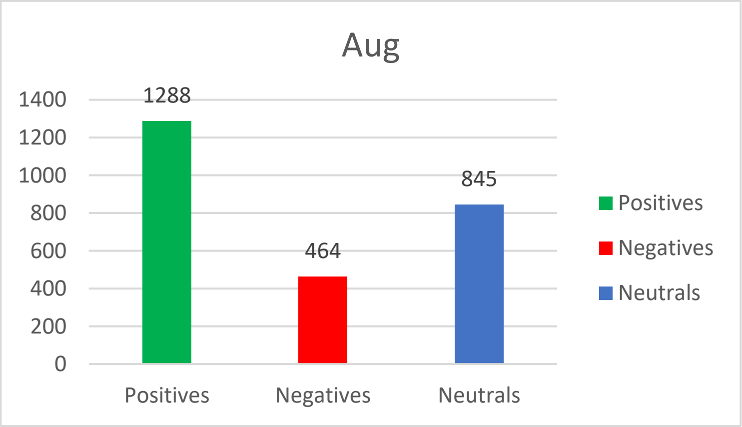
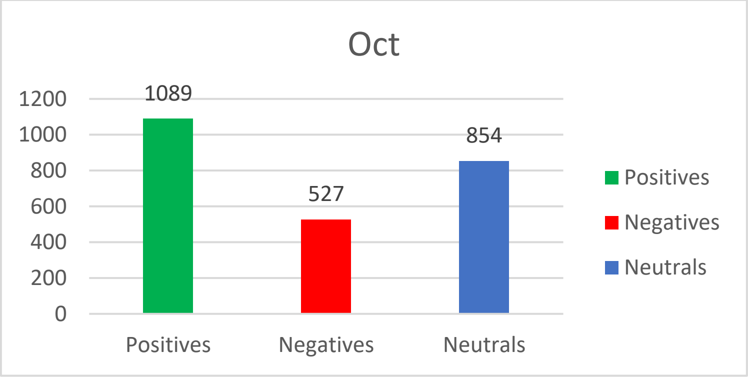
For extraction experiment, we extracted the tweets that are related to the keywords and three vaccines, where table 9 shows the no of tweets for each vaccine.

Table 9: No of extracted tweets

Vaccine	No of extracted Tweets
Pfizer-BioNTech vaccine	29017
Johnson & Johnson’s COVID-19 Vaccine	12885
Moderna vaccine	17632

Second experiment was run to do the sentiment analysis by Vader tool, where Vader uses the lexicon-based approach to execute the sentiment analysis. The sentiment analysis results are shown in the following Figures 59, 60, & 61.

Vaccine	Vader sentiments
Pfizer	 

Johnson & Johnson	 

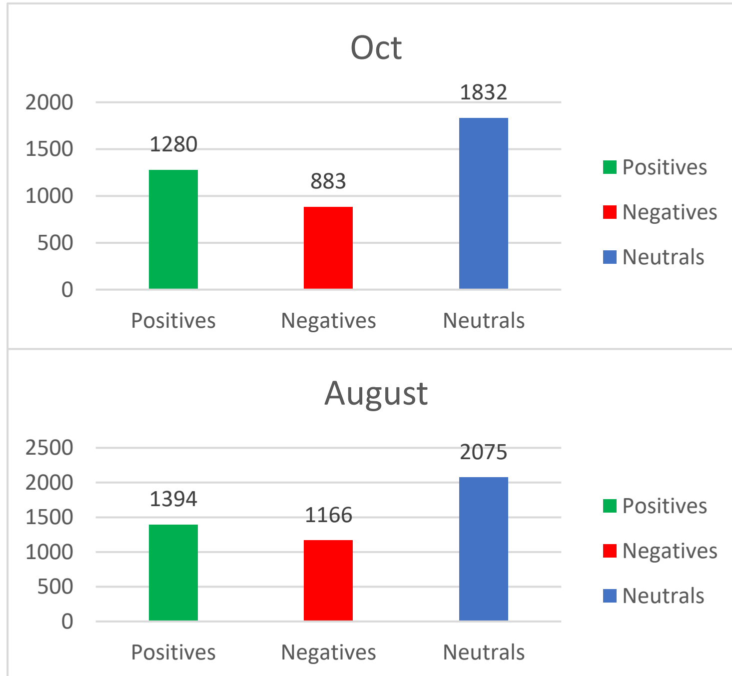
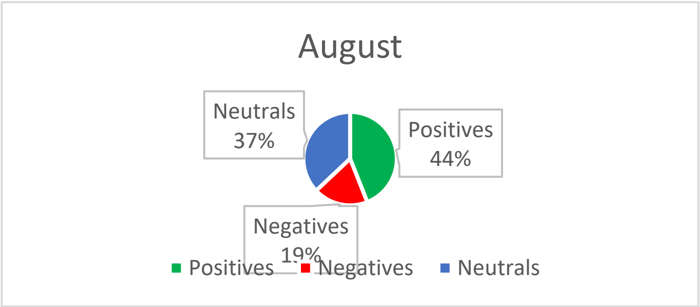
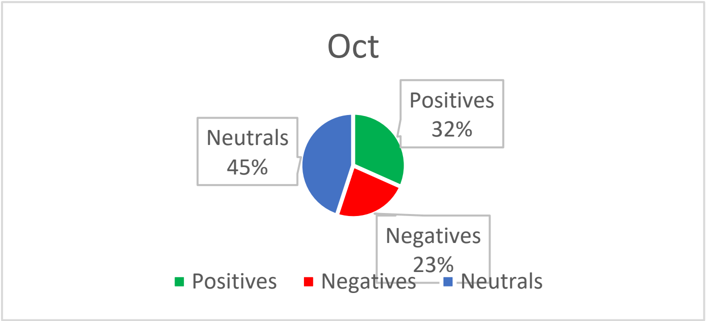
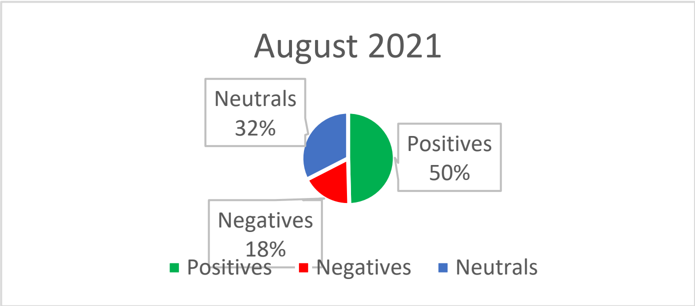
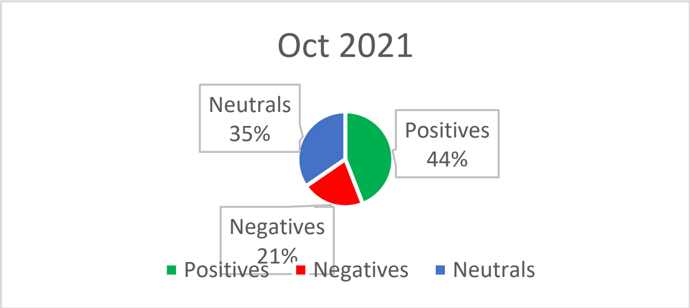
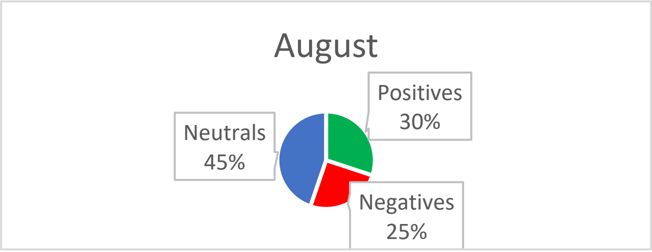
Moderna	

Figure 57: Sentiments Classifications Numbers

Vaccine	Vader sentiments percentages
Pfizer	 

Johnson & Johnson	 
Moderna	

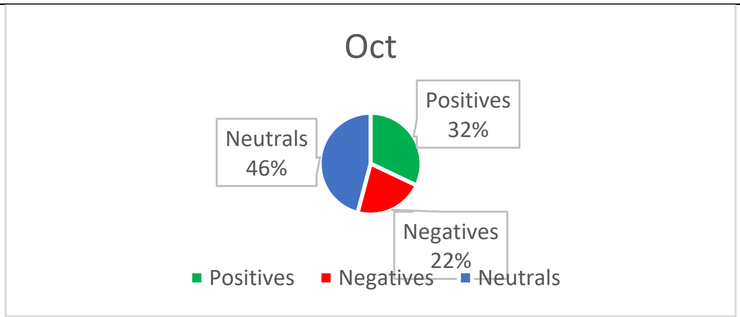
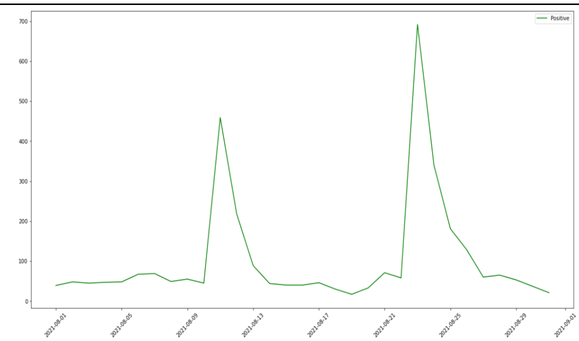
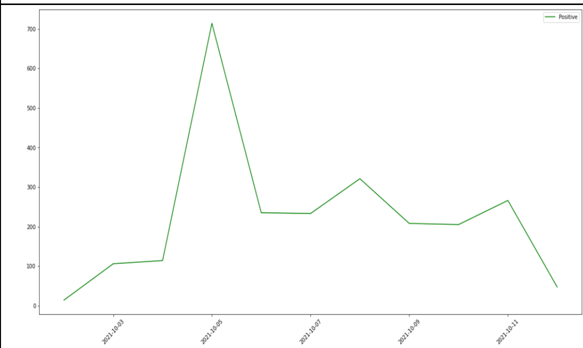
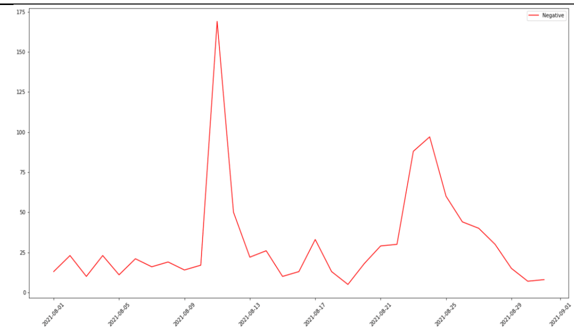
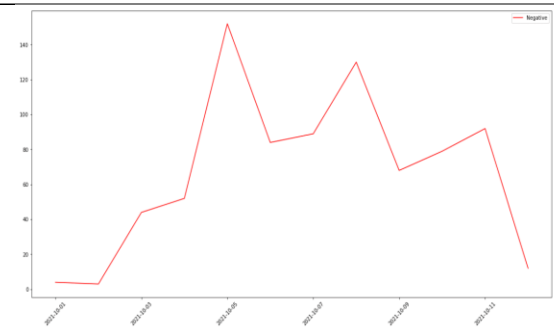
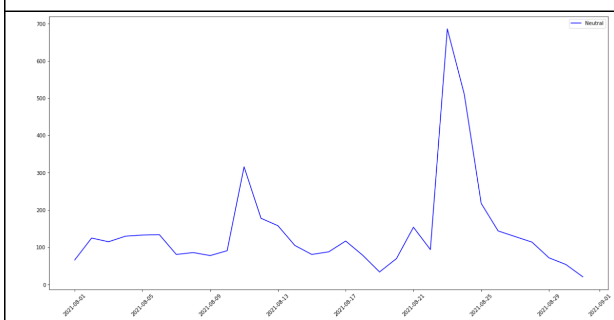
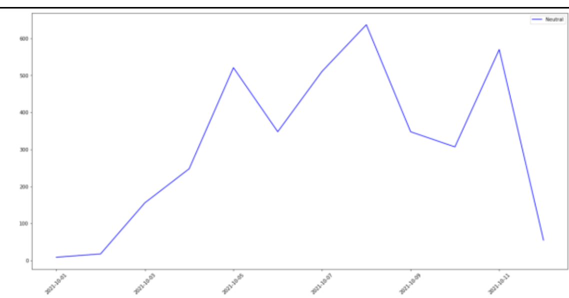
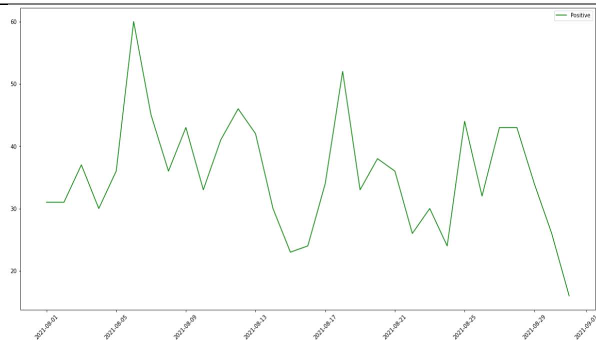
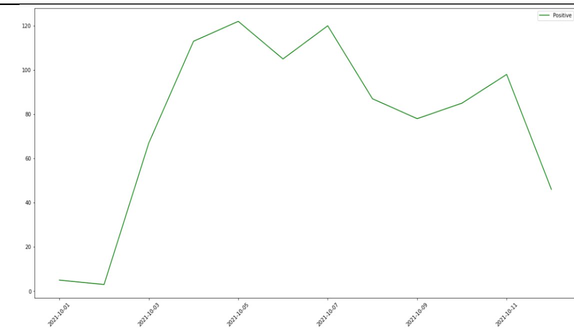
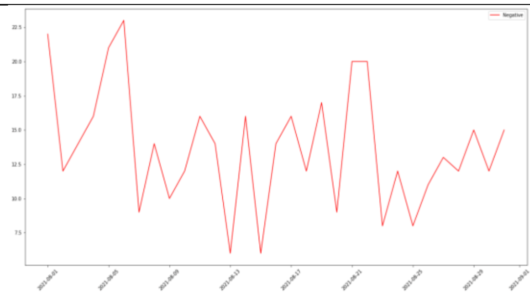
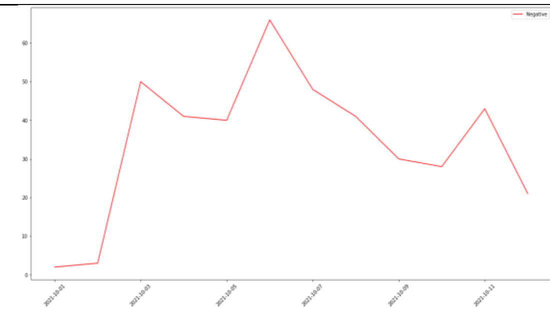
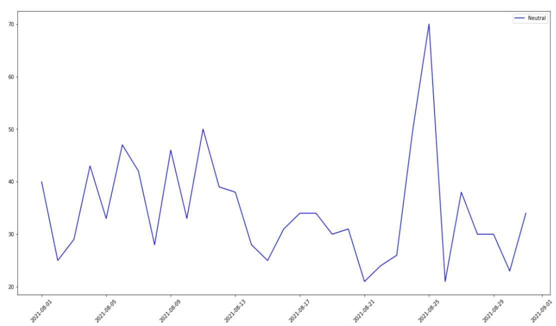
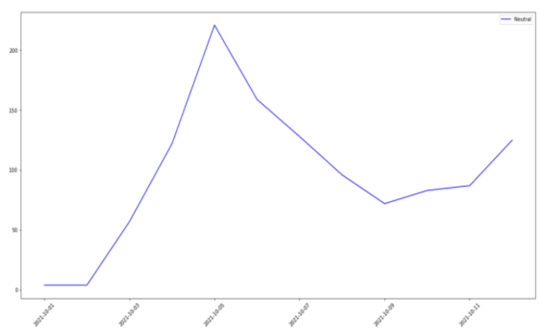


Figure 58: Sentiments Classifications Percentage

Vader sentiments analysis curve

	August 2021	October 2021
Pfizer		
		
		

<table>
<tr><td rowspan="3">Johnson & Johnson</td><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

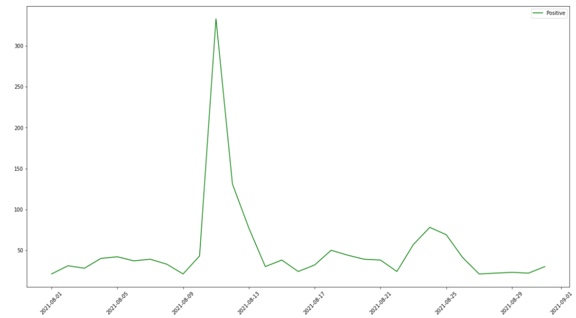
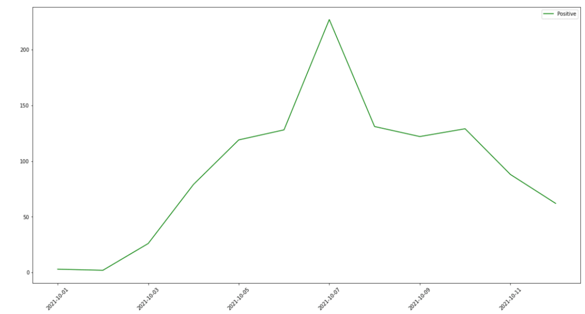
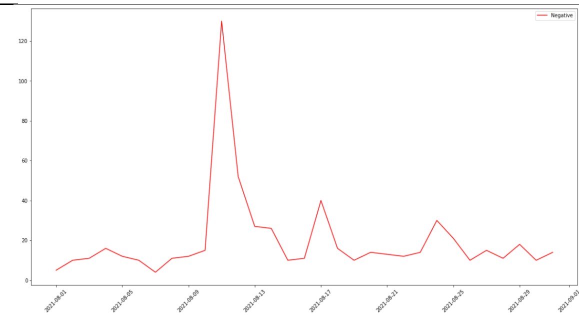
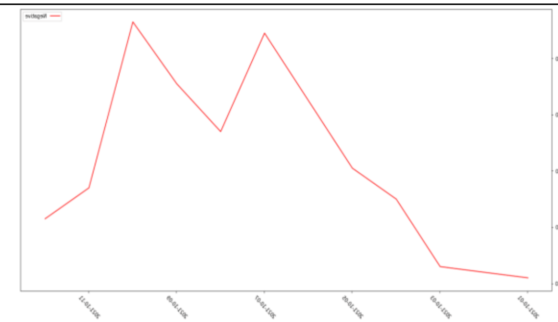
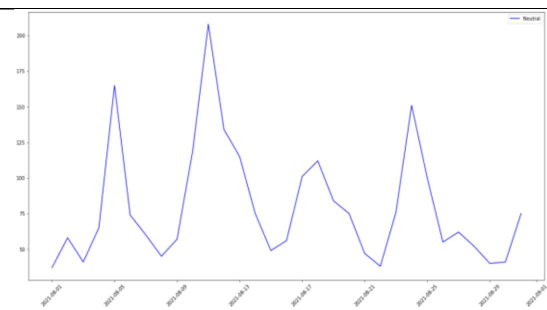
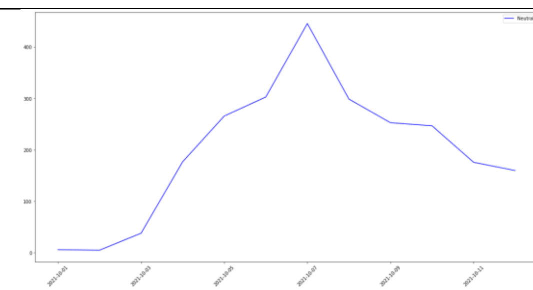
Moderna		
		
		

Figure 59: Sentiment Analysis Curves

The Figures above show the results for sentiment experiments of three vaccines, where the tweets classified by Vader library tool, so the tweets were classified into positives, negatives, or neutrals, where the figures show the numbers and percentages of classified tweets form June 2021 to Oct 2021. Based on the extracted tweets, the numbers of tweets show that Pfizer and Moderna have higher attentions among Twitter users, while Johnson and Johnson had the lowest attention due to the accidents events and hold in April 2021 which resulted in lack of confidence among public. According to Figures 59 and 60, the total counted tweets have shown that more neutrals opinions than positives and negatives during the whole timeframe, but Figure 61 shows that the sentiment change curve has been changed to positive significantly for both Pfizer and Moderna vaccine in Aug 2021 which confirm the effectiveness of delta variant fears on reducing vaccine hesitancy.   Even though Johnson and Johnson vaccine got the lowest attention by Twitter users, the highest opinions are positives, where the experts indicated that the vaccine is acceptable among a segment of population because it is only one dose “One and done “(Weiland, 2021). In addition, the most public opinions about Pfizer and Moderna were turned to neutrals in Oct 2021 because the topics were different from Aug 2021. To explain, the discussed topics in Aug 2021 were about Delta variants and vaccinations, while the recent topics in Oct 2021 were about booster shots and children vaccination. Moreover, Johnson and Johnson vaccine got a high public attention too in Oct 2021, where the majority opinions are neutrals. Comparing between Johnson and Johnson topics in Aug 2021 and Oct 2021, there is a huge difference that explains people had not focused on this vaccine in Aug 2021, where they had focused on it in Oct 2021. To explain, lack of confidence in Johnson and Johnson caused people to not be focused on the vaccine topics during the Delta variant spread peak, while they have shown some interest recently regarding it due to the booster shot approval.

Discussion

The preliminary research studied effects of Delta variant spread on vaccination hesitancy levels in the United States. Fear is considered the most factor that affect the vaccine hesitancy, where significant segments of population are afraid of side effects of vaccines. However, the fear and concerns directions were changed after Delta variants has started spreading in the United States with a high transmission rates and cases (Figures 62). So, the fear levels of Delta variants infection and its harsh symptoms became higher than the fear levels of vaccines side effects. According to CDC vaccination dataset, the vaccination rates have been increased significantly in Aug 2021 compared to June and July 2021 as shown in Figures 63 and 64. In addition, the rates of taking the second dose have been increased too in parallel at the same time as shown in Figure 65, which indicates to the effectiveness of Delta variant fears on increasing the vaccination intentions of taking the vaccine among the public in US.

In order to associate Delta variant fears with affect public opinions about vaccination, we used the sentiment analysis approach to A) Find out if the public attentions and fears of Delta variant have been shown significantly on Twitter or not? B) Study the public opinions about booster and children vaccination topics. Thus, we built our datasets about the three vaccines in the United States (Pfizer, Johnson and Johnson, Moderna), where the users' tweets extracted and analyzed. The datasets were preprocessed and classified sentimentally, where Vader library tool was selected to do the sentiment analysis.

To investigate the vaccination hesitancy levels, we analyzed the public sentiments change over the selected time frame. Thus, we have shown how the variants has increased levels of fears of infections, which resulted in reducing the vaccine hesitancy among population around USA. Figure 63 shows the results of sentiments analysis curves (June to Oct 2021) which more public attentions regarding vaccinations in Aug 2021. Furthermore, Delta variant spread was started by June 2021 and caused higher rates of cases during Aug 2021 (Figure 61) which motivated public to be vaccinated better than getting infected of Delta variants with harsher symptoms than COVID-19. So, the sentiment analysis experiments show that there was a raise in public attention about the vaccines in Aug 2021 and more positives opinions than previous months.

In addition, there was increasing in the real vaccination rates at same period based on CDC data (Figure62, 63, and 64). Moreover, there was a huge increase again in the number of posted tweets during Oct 2021, where some events have been discussing recently as boosters shot and children vaccination approval news. Figure 65 shows a significant increasing in boosters doses vaccination rates after the FDA approval according to CDC data. So, these two topics got a lot of public attentions with high levels of neutral opinions, so the hesitancy levels regarding the booster shot and vaccinating the children still high and need to be reduced through providing the right interventions and promotion policies.

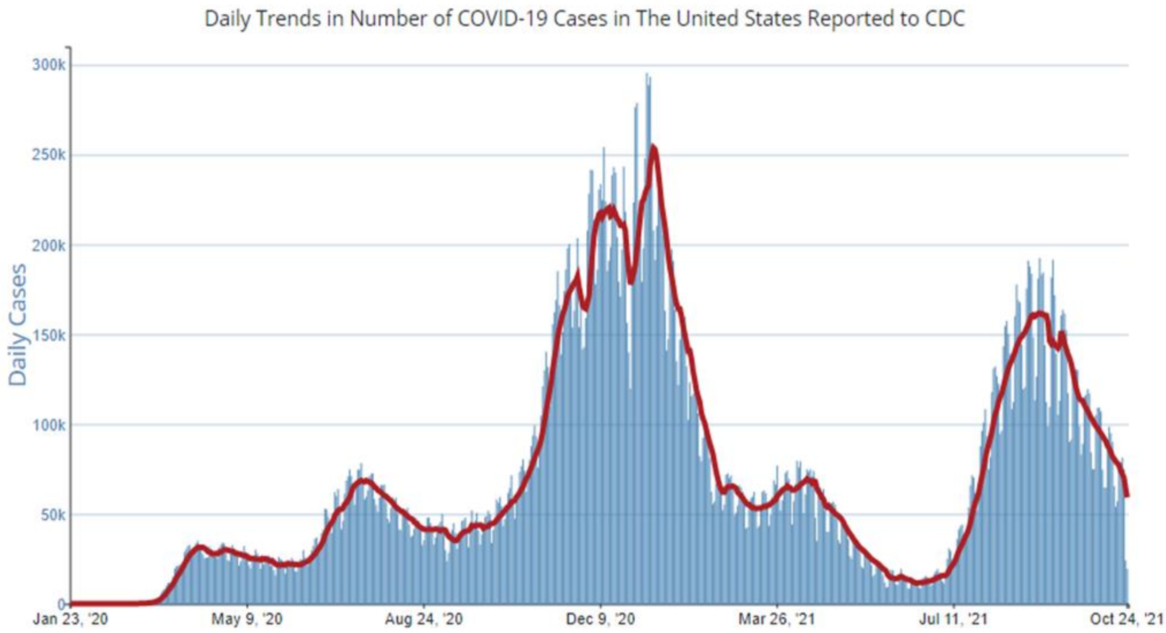


Figure 60: Number of Infected Cases in USA (adapted from CDC, Oct 2021)

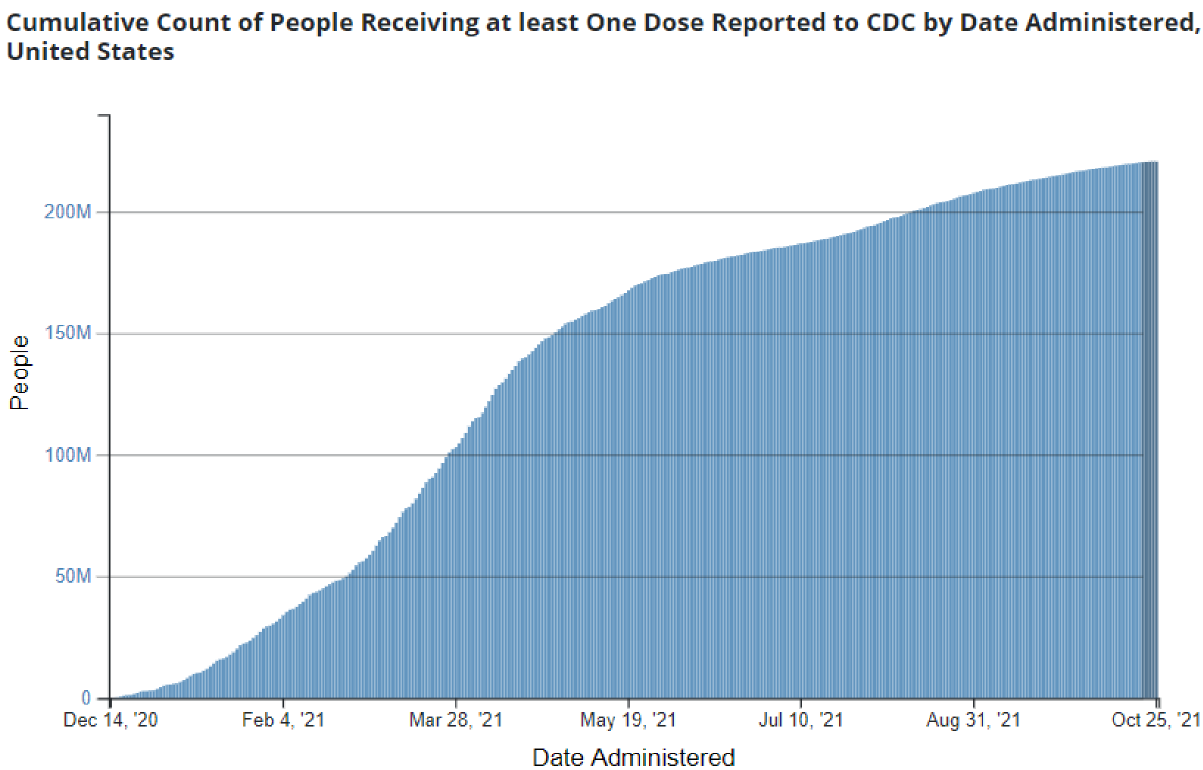


Figure 61: Cumulative Account of Partially Vaccinated People (adapted from CDC, Oct 2021)

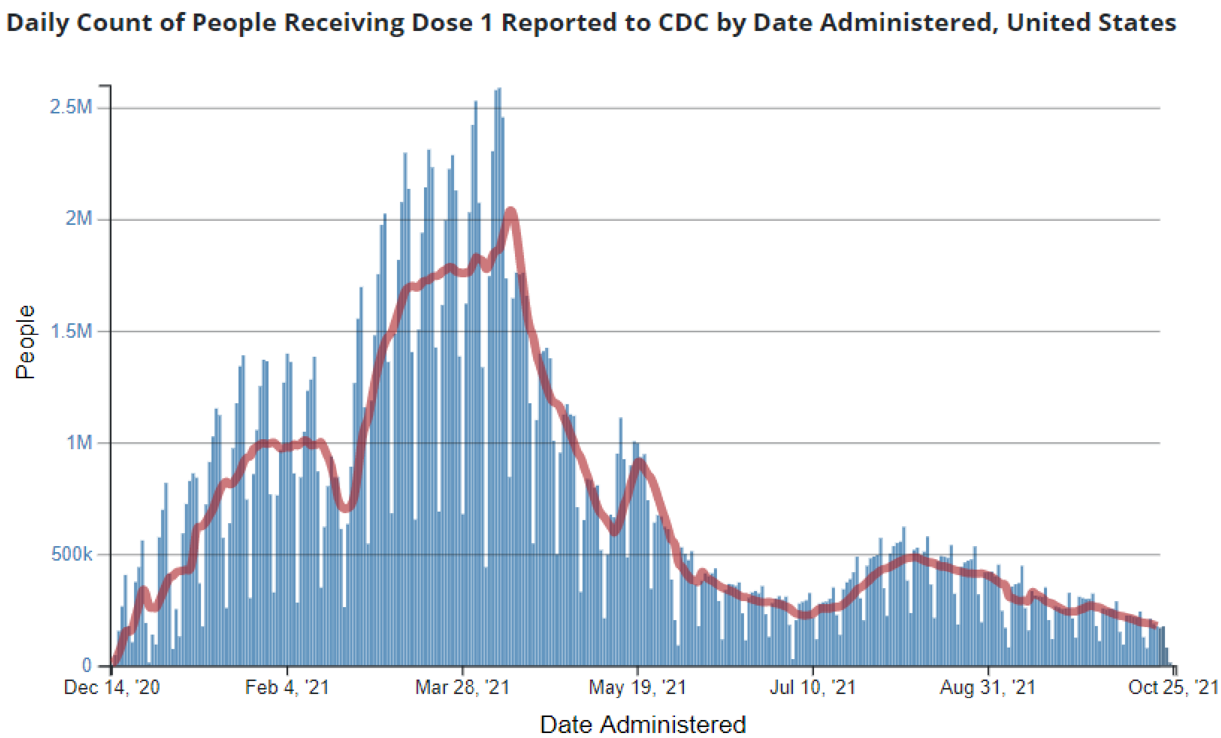


Figure 62: Partially Vaccinated Numbers (adapted from CDC, Oct 2021)

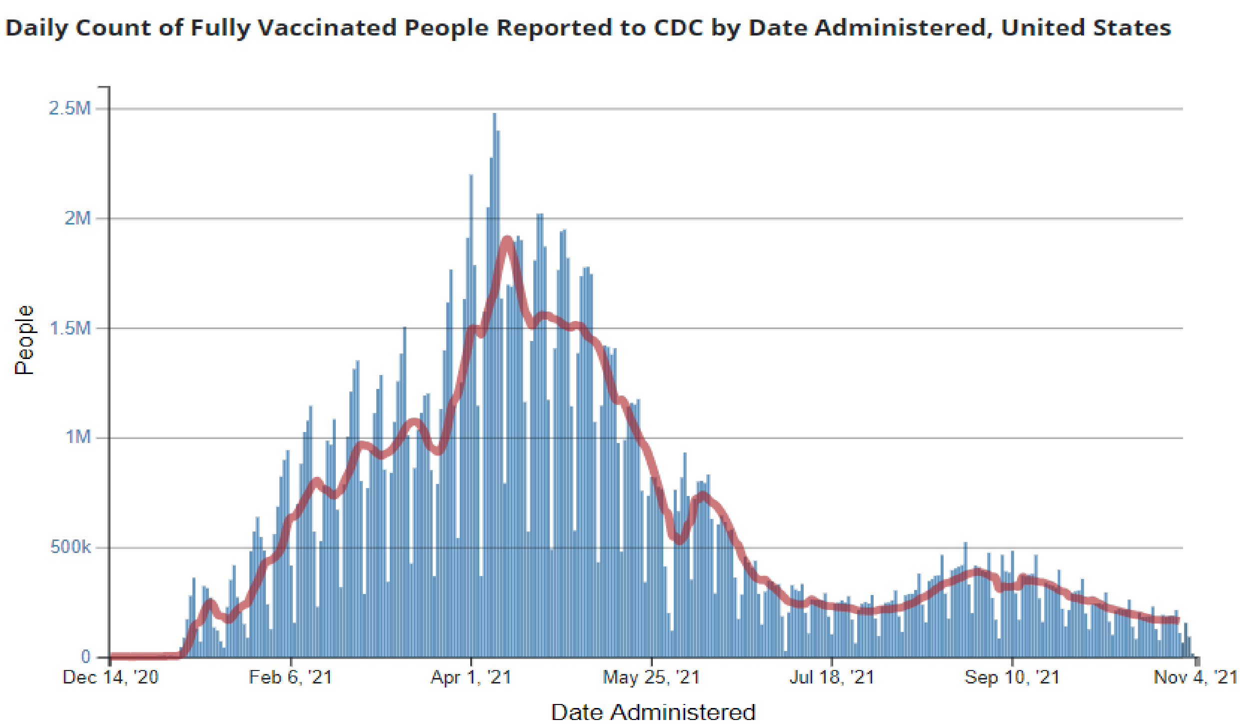


Figure 63: Fully Vaccinated Numbers (adapted from CDC, Oct 2021)

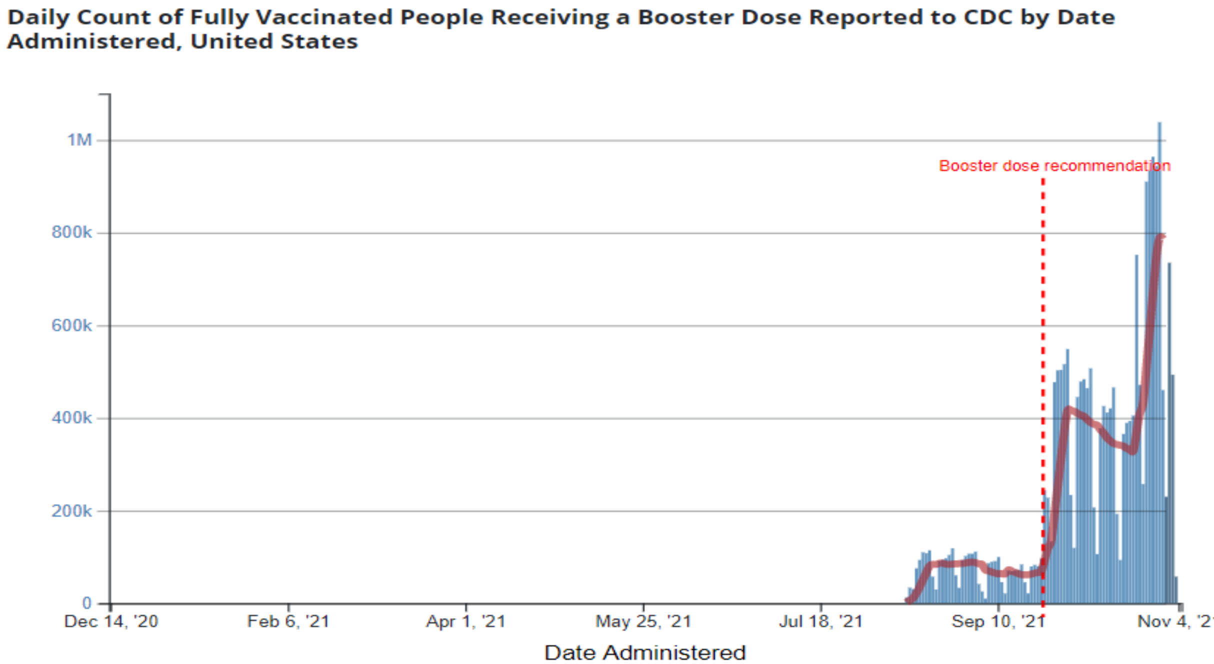


Figure 64: Number of People who Took Booster Shot (adapted from CDC, Oct 2021)

Preliminary results contributions and limitations

During the first experiment, we prove that vaccination tweets and sentiment analysis is useful approach to measure the vaccine hesitancy in the USA. In addition, CDC dataset and KFF survey indicated that changes in the vaccination inoculation has been increased dramatically due to the fears of virus, in addition to COVID-19 vaccine approvals in Aug 2021, Pfizer booster dose approvals in Sep 2021, and Moderna booster approvals in Oct 2021. However, no statistical analysis had been conducted to test the correlations or regressions between tweets and CDC vaccination inoculation dataset. In a word, preliminary results measured the vaccination hesitancy and public opinions about the vaccines in the USA, but it didn't test the level of statistical association between CDC vaccination datasets and Twitter datasets. Thus, we could not build any predictive models that can be used to predict daily vaccination inoculation in the USA.

Proposed Research:

Fears of vaccine side effects prevented a significant segment from taking the COVID-19 vaccines, but once the delta variant spread among the population with a high-risk level, the fears direction has been started to change. Thus, we extracted a dataset that includes tweets about the three vaccines with focusing on the United States users and conducted the sentiment analysis to measure vaccine hesitancy among the nation.

To investigate the effects of Delta variant spread on the vaccination rates in the United States, and the public opinions about the booster shots children vaccination approvals, we proposed this research methodology that involves measuring the twitter users' sentiments about these events and compare it with CDC vaccinations datasets. Thus, the proposed research seeks to utilize Twitter data and CDC datasets to study and analyze the fear effects due to Delta variant spread on the vaccination rates in the United States, in addition, it seeks to measure and understand the public opinions about the recent events, such as taking booster shot and children vaccination. We used the supervised machine learning and sentiment analysis approach to conduct our experiments, where these techniques can extract and analyze large amount of data at less time. Therefore, we were able to collect tweets that includes public opinions about three available vaccines for COVID-19 in order to measure the vaccine hesitancy for each vaccine separately. Thus, the results can be used by decision makers to take the right action. Our experiments result show that the vaccine hesitancy decreased correlated with Delta variant spread and cases peak in Aug 2021, which provided significant evidence about public fear impacts on opinions and intentions. Also, CDC vaccination data for Aug – Sep 2021 confirmed that the vaccination rates have been increased significantly. Then, the users' attentions about vaccine and tweets numbers went down again in Sep 2021, where they have been increased again in Oct 2021 as a result of the boosters shot and children vaccine approvals topics, where the majority still neutrals. The initial results for the time frame (June 2021- Oct 2021) prove that the extracted sentiments data from Twitter has shown significant positives correlations between the CDC vaccination rates datasets and the numbers of tweets, public attentions toward vaccines topics, and approval news.

Based on the motivating preliminary results and addition to Emerging Omicron variant after the experiments, we aimed to go deeper and investigate vaccine hesitancy and fears of virus during three deferent phases (Baseline stage, Delta variant stage, and Omicron stage), where each stage had different factors, such as virus severity risk, new vaccination approvals over time, or governmental intervention, which can affect the intentions toward vaccination inoculation in the USA. In addition, taking into consideration the vaccination status (Partially, fully, or fully vaccinated with the first booster). In addition, quantifying the correlations and regressions between selected variables can provide sufficient understanding for the nature and strength of relationships, and building a regression predictive model for the daily inoculations in the USA.

Linking between public sentiments about the vaccination and fears of virus infection can be used to build regression sentiment-based vaccination prediction model to predict daily inoculation in the USA. As we aimed the first, second, and the booster dose into our research, three predictive models will be proposed.

The developed Methodology

The preliminary methodology was useful to measure the vaccine hesitancy of three vaccines in the United States of America during the selected timeframe June 1st, 2021 - Oct 31, 2021. Also, it was useful to explain the significant raise into the CDC datasets (Cases, deaths, and vaccination inoculations) during Delta variant phase which was a strong indicator for a positive correlation between CDC datasets or on CDC datasets versus Twitter data. However, no statical analysis were performed at that time to show the level of correlations and regression between them. In addition, the CDC VOC's (variant of concern) announcements were not included during the preliminary experiment as we focused only on the timeframe that covers June 1st 2021 to Oct 31st 2021, so we didn't include the exact dates for CDC announcements about Delta variant emergent in the USA as variant of concern, and also the study performed before Omicron variant came to the world. Therefore, we aimed to improve the preliminary methodology in order to include additional factors, such as (CDC VOC's and virus different stages:- Baseline stage, Delta variant stage, and Omicron stage), also, to conduct the required statistical analysis for both Twitter and CDC datasets to quantify associations levels between different stages of virus spread, virus severity, virus case level, virus death level, positive and negative sentiments about vaccines, and the first, second, and booster dose inoculation in the USA. Moreover, the improved methodology has been utilized to design and build sentiment-based daily vaccination prediction models in the USA.

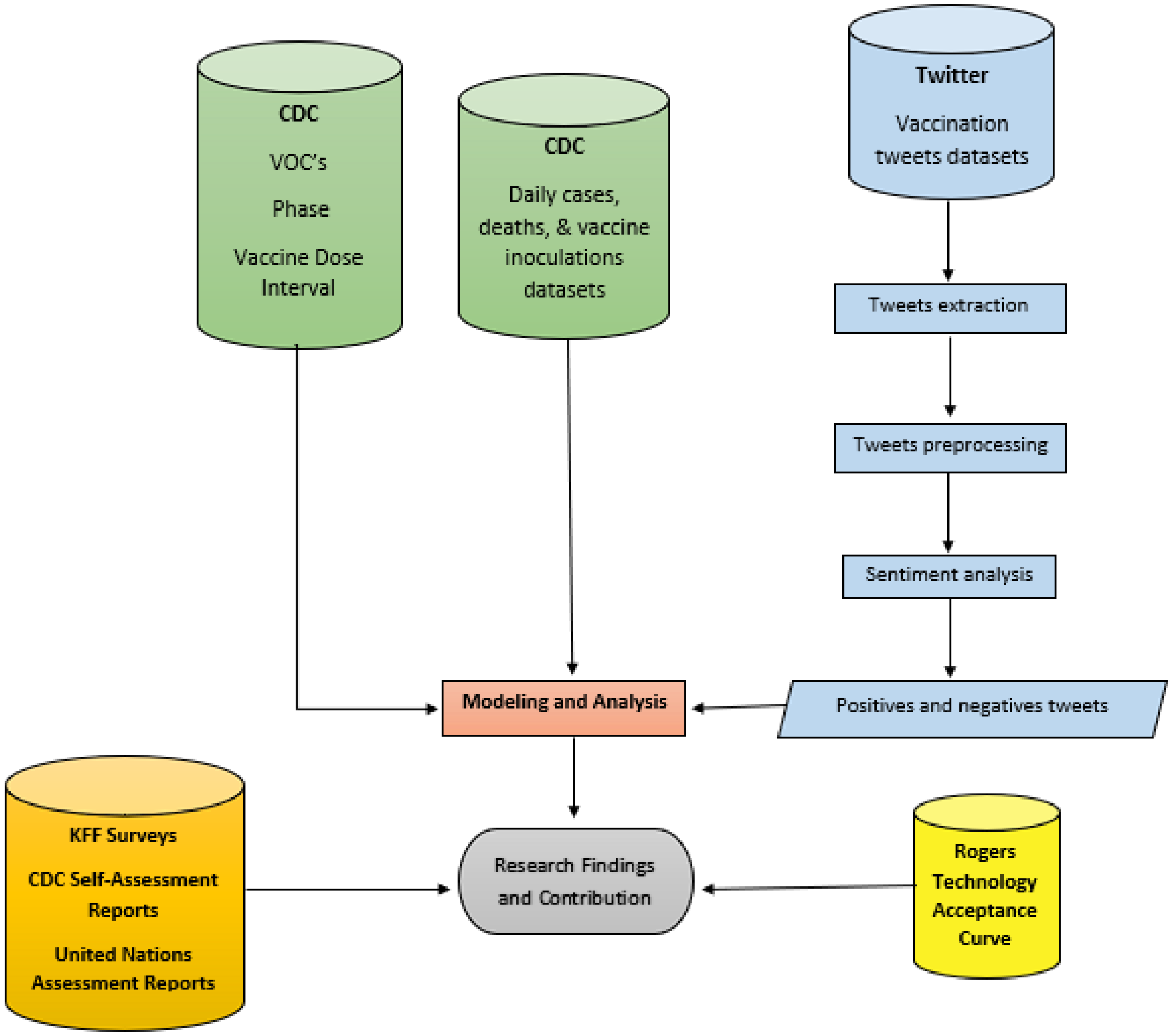


Figure 65: The Developed Research Methodology

- CDC Variant of Concern announcement establish phases (Baseline, Delta, Omicron)
- CDC Vaccine approvals establish Dose Intervals (First Dose, Second Dose, Booster Dose)
- CDC Case, Deaths, & Inoculation Data Collection and Segmentation

- Twitter Data Collection and Sentiment Analysis

- Linear Regression models
- Correlation Analysis
- Lag Time Analysis

To illustrate, this research used these three resources of datasets to investigate hypothesized relationships between health threats of virus variants, vaccination hesitancy and public opinions during the five selected phases. CDC datasets and KFF survey results are published online. This research created the Twitter dataset by extracted data from Tweets using by a crawler and each Tweet analyzed for sentiment as discussed below.

The data and analysis methodology involves determining the level and nature of association of Twitter sentiment data with CDC COVID-19 data from June 1, 2022, to March 31, 2022. Within that time frame, data collection focuses on five phases – Baseline, Delta variant emergence, Adult Booster shot approval, Children Booster shot approval, and Omicron variant emergence (Figure 2). Analysis focuses on determining the level and nature of association of Twitter sentiment about the change in virus variants dominance, availability of vaccinations, change in government mandates on vaccinations, and responses in CDC data. Vaccination types - Pfizer, Moderna, Johnson and Johnson – complicate the analysis.

Figure 66: Research timeframe and phases

Figure 3 shows the relationships between the datasets used to conduct the research and analysis. CDC datasets provide the number of infection cases, deaths, and vaccinated. Twitter datasets provide positive, negative, and neutral tweets about the vaccines. In addition, KFF vaccinations survey outputs provided 3rd party analysis of the public opinions about the vaccination and how the selected phases affected their opinions and intentions.

The proposed research able to:

1- Quantify the nature of association and regressions between COVID-19 cases, deaths, vaccination positive and negative tweets, and the first, second, booster dose.
2- Understanding the effects of vaccine hesitancy, fears of Delta variant, and fears of Omicron variant infection on the vaccination inoculations in the USA.
3- Exploring the effects of vaccination approvals on the vaccination inoculations
4- Quantifying lag times between VOC's, FDA vaccines approvals, and significant raise in vaccinations inoculation.
5- Building a sentiment-based regression predictive model to predict first, second, and booster dose daily vaccination inoculation in the USA

In order to meet the research goals, we the following hypotheses have been investigated:

H1: Virus Daily Change Rate Level Correlated directly to Daily Change Rate Level of Vaccine Inoculations?

H2: Virus Daily Case Level (numbers) Correlated directly to Daily Vaccine Inoculation Level ?

H3: Virus Daily Death RATE Level Correlated directly to Daily Vaccine Inoculation RATE Level ?

H4: Virus Daily Death Level (numbers) Correlated directly to Daily Vaccine Inoculation Level ?

H5a: Positive Sentiment tweets extracted from Twitter are correlated directly to Vaccine Inoculation Level

H5b: Negative Sentiment tweets extracted from Twitter are correlated inversely to Vaccine Inoculation Level

H6a: Positive Sentiment tweets extracted from Twitter are sensitive change to COVID-19 vaccines approvals in the USA

H6b: Negative Sentiment tweets extracted from Twitter are sensitive to change COVID-19 vaccines approvals in the USA

H7a: Positive Sentiment tweets extracted from Twitter are sensitive change FDA VOC in the USA

H7b: Positive Sentiment tweets extracted from Twitter are sensitive change FDA VOC in the USA

H8a: Proportion of the population vaccinated relate to the laggard proportion identified in Rogers' Technology Adoption curve

H8b: Proportion of the population vaccinated does not relate to the laggard proportion identified in Rogers' Technology Adoption curve

H9: The proposed methodology and analyzed datasets are useful to predict daily vaccination inoculation in the USA?

CHAPTER FOUR: DATA, RESULTS, AND ANALYSIS

Analyzing public sentiments towards COVID-19 vaccines and actual vaccinations in the USA

Abstract:

This chapter addresses data and analysis of hypotheses listed in chapter 3, and findings include Delta and Omicron variants' effects on the vaccination inoculations in the USA and how both variants affected public opinions toward vaccinations. Research uses extracted Twitter datasets, CDC datasets, and Kaiser Family Foundation (KFF) Vaccination Monitor surveys to investigate levels of effect of selected factors. Delta and Omicron variants exhibit higher infection numbers compared to the SARS-CoV-2 infection numbers, but the symptoms of the Delta variant are harsher than both SARS-CoV-2 and Omicron variant. Thus, vaccine hesitancy became lower than fears of Delta variant infection which motivated high numbers of people to get vaccinated. In addition, government interventions and vaccination mandates were strong factors that led to increased vaccination inoculation. Previously, (Daghriri, Proctor, & Matthews, 2022) found that the delta variant affected people's opinions regarding vaccination. Therefore, this research aimed to study the effects of FDA booster and kids' vaccination approvals and Omicron variant spread on the vaccination numbers. Chapter four is broken into four major subsections: (1) Overview of Improved Methodology; (2) CDC Datasets; (3) Twitter Datasets; (4) Statistical Analysis of Datasets. (5) Building predictive vaccination models.

Prior and Related Research Findings

Previous review and public sentiment analysis on vaccination hesitancy discussed in Chapters 1 and 2 indicates that significant segments of the U.S. population are hesitant to vaccinate due to their fears of the side effects of the vaccine or because they don't trust the vaccine's effectiveness. However, once the Delta variant emerged with higher infection rates and higher health risks than prior variants, fear of the infection rose above the fear of the vaccine. Additionally, the probability of suffering severe vaccine side effects was lower than probability of the severe health risks associated with of the Delta variant. During the Delta variant phase, Kaiser Family Foundation (KFF) Vaccination Monitor surveys focused on public opinions about the Delta variant and the publics intent to get vaccinated. KFF surveys indicated 39% of the survey participants took their the first dose due to their fear of Delta variant infection. Instead of conducting timely, costly, and often backward-looking surveys, this research aims to extract in real-time sentiment from Twitter Tweets and build on and expand prior Twitter sentiment research. Advantage of the Twitter Tweet extraction methodology is not only its real-time nature but also the ability to obtain large amounts of data in a short time about the topic of interest. This research also investigates hypotheses about the relationships between the health threat associated with a variant and vaccine uptake by the US population. Relationships investigated include correlations between cases, deaths, and vaccination numbers and vaccination uptake rate changes and lags during the Delta and initial Omicron period. Since the level of fear is the dominate factor that increases or decreases vaccination hesitancy, this research classified the impact of fear on vaccine acceptance in two directions:

A- Fear of vaccine side effects, which results in increasing vaccine hesitancy.

B- Fear of Virus infection, which results in decreasing vaccine hesitancy and increasing vaccine acceptance or uptake.

The Figure 1 below shows the fear directions and their impact on vaccine acceptance or uptake.

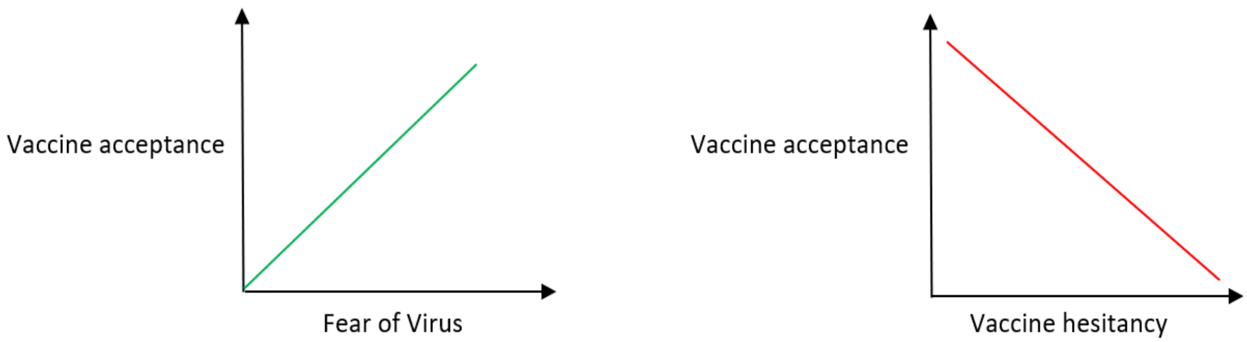


Figure 67: Fears and Vaccinations Relationships

As fear of disease increases, vaccine acceptance or uptake increases, however, as fear of vaccine side effects increase, vaccine acceptance or uptake decreases.

Overview of Data and Analysis Methodology

The research methodology used three primary datasets for the US population.

A- CDC datasets.

B- Twitter users' opinions sentiment analysis.

C- Published national surveys that measure vaccination hesitancy and effective factors on personal vaccination intentions

CDC Datasets

CDC total vaccination status, numbers, and vaccine coverage of the US population are shown in Table 10 below.

Table 10 (Sep 5, 2022)

Vaccinated status	Numbers	Vaccine Coverage
At least one dose	262,908,216	79.2%
Fully vaccinated	224,113,439	67.5%
First booster dose	108,806,974	48.5%

CDC data in Figures 4 (**https://www.washingtonpost.com/nation/2022/02/23/covid-omicron-variant-live-updates/**) and 5 (**https://www.cdc.gov/mmwr/volumes/71/wr/mm7112e2.htm)** below indicates that no vaccination level can stop a virus infection, but vaccination level does impact the level of cases and the level of hospitalizations. Additionally, the figures indicate that vaccination levels were significantly less effective against the Omicron variant than the Delta variant In preventing catching the virus or preventing hospitalization.

CDC data in Figure 6 overlays the number of people in the US from June 1, 2021 to March 31, 2022 that each day contracted the virus (cases) onto the number of people becoming for the first time partially vaccinated (single dose), becoming for the first time fully vaccinated (second dose), and becoming for the first time fully vaccinated with the first booster. First does includes all Pfizer, Moderna, and J&J vaccinations. Second dose includes just Pfizer and

Moderna vaccinations. The case rate rise in July is in response to the proliferation of the Delta virus among the population. The large peak in case numbers beginning in December corresponds to the emergence of the Omicron variant.

CDC data in Figure 7 overlays daily deaths over the same vaccination data for the same time period. The first peak in death rate corresponds to the Delta variant. The second and larger peak corresponds to the Omicron variant.

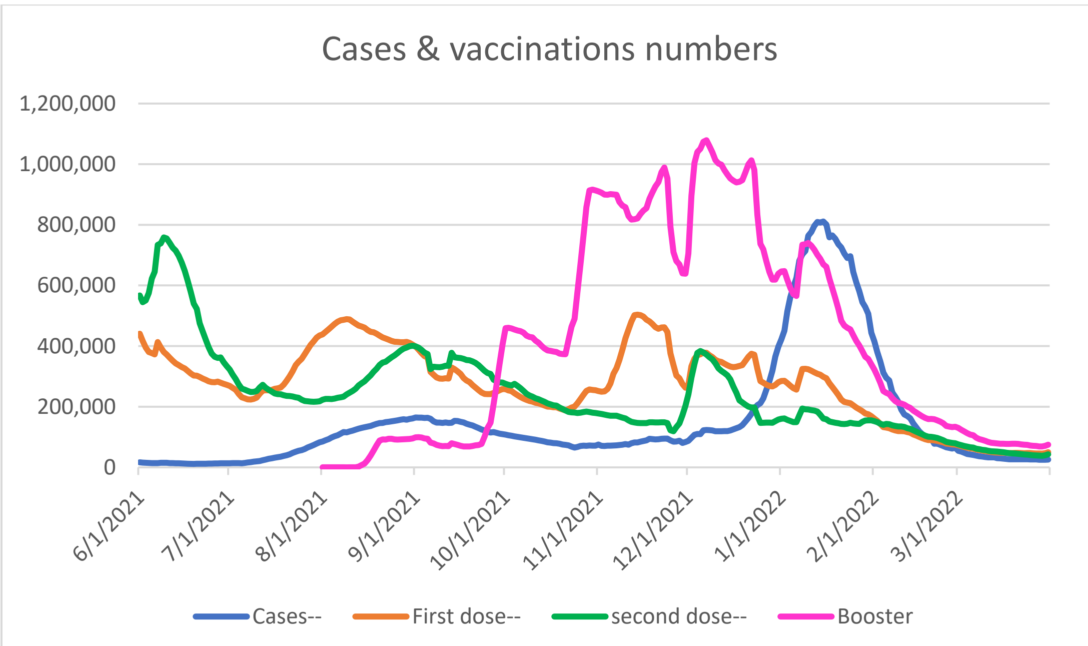


Figure 68: Cases Versus Vaccination Numbers

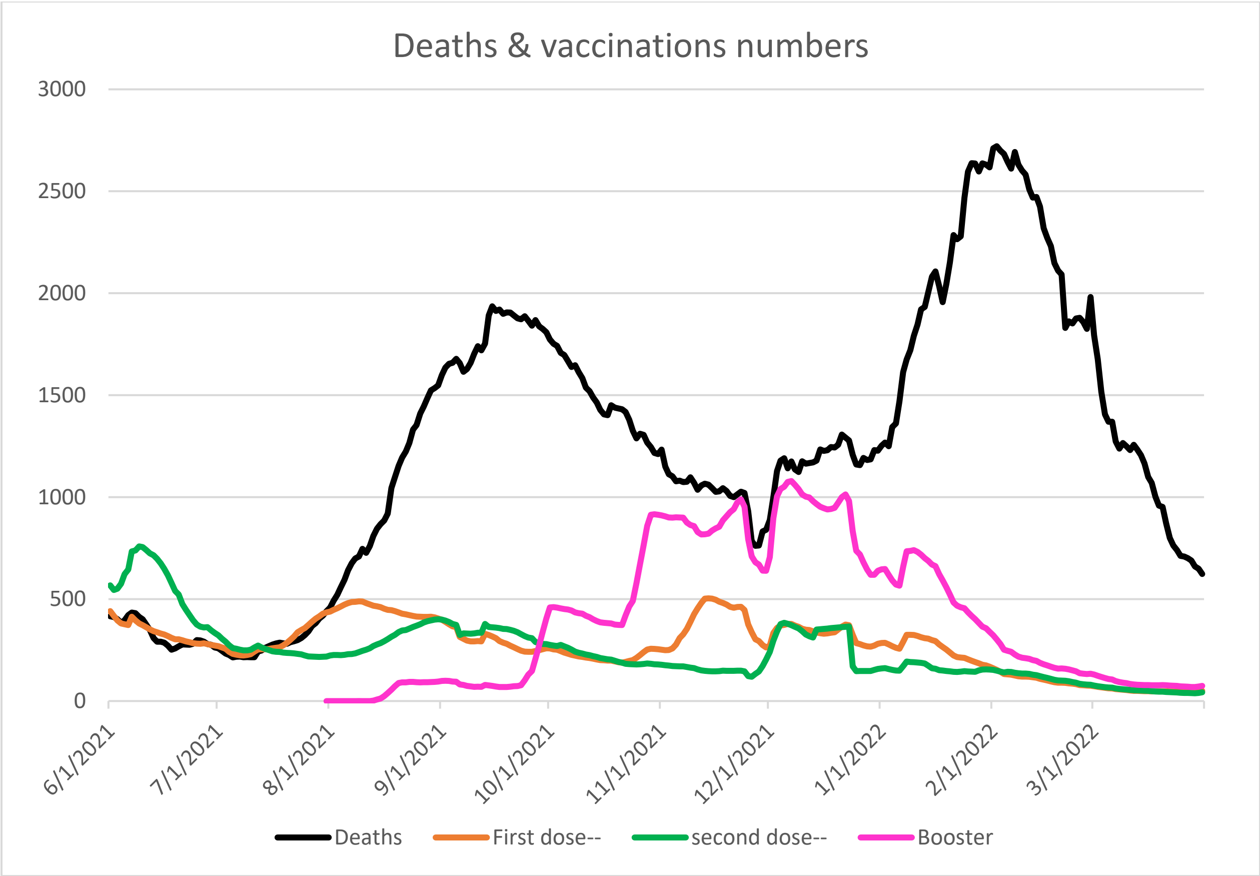


Figure 69: Deaths Versus Vaccination Numbers

Table 11 : Total Monthly Numbers from CDC Datasets

	Cases	Deaths	First dose	Second dose	Booster
Jun-2021	401271.00	9875.00	10063257.00	16835054.00	1756.00
Jul-2021	1144425.00	8772.00	9097556.00	7671358.00	1063.00
Aug-2021	4072822.00	30517.00	13849793.00	9356855.00	1324863.00
Sep-2021	4235747.00	53642.00	8891718.00	10210194.00	3407604.00
Oct-2021	2606552.00	45790.00	6853697.00	6695895.00	15736220.00
Nov-2021	2475824.00	30442.00	11610852.00	4674034.00	25443906.00
Dec-2021	5196793.00	36672.00	10305507.00	7966150.00	28094517.00
Jan-2022	20522153.00	62474.00	7934445.00	4947161.00	17537769.00
Feb-2022	4790507.00	64589.00	3021887.00	3279393.00	5456703.00
Mar-2022	1009461.00	33249.00	1625723.00	1608513.00	2657155.00

Table 12: Monthly Change of Rate in CDC Datasets

Month	Cases	Deaths	First dose	Second dose	Booster
Jun-2021	-	-	-	-	-
Jul-2021	**185%**	-11%	-10%	**-54%**	**-39%**
Aug-2021	**256%**	**248%**	**52%**	**22%**	**124534%**
Sep-2021	**4%**	**76%**	**-36%**	**9%**	**157%**
Oct-2021	**-38%**	**-15%**	**-23%**	**-34%**	**362%**
Nov-2021	**-5%**	**-34%**	**69%**	**-30%**	**62%**
Dec-2021	**110%**	**20%**	**-11%**	**70%**	**10%**
Jan-2022	**295%**	**70%**	**-23%**	**-38%**	**-38%**
Feb-2022	**-77%**	**3%**	**-62%**	**-34%**	**-69%**
Mar-2022	**-79%**	**-49%**	**-46%**	**-51%**	**-51%**

Twitter Datasets

The research methodology extracted US population tweets to understand public opinions about common three vaccines in the US and during the three selected phases. Users' tweets were extracted through Tweepy library tool which is a useful tool that extracted a large number of tweets in short time. Also, the extraction techniques focus on the tweets that include the queries keywords that this research are interested in, also, the extracted tweets with their classifications are attached in the appendixes section. Furthermore, the location of Twitter users and time frame were included into extraction process. The methodology analyzes the vaccine hesitancy levels for three COVID-19 vaccines in the United States (Pfizer, Johnson & Johnson, and Moderna), and selected time frame from June 2021 to 31st March 2022 as proposed in chapter 3. Furthermore, the selected timeframe includes some events, such as Delta variant spread, booster shots approvals, kids' vaccination approvals, and Omicron variant emergence. Furthermore, sentiment analysis analyzes the content and direction of each tweet (Positive, neutral, or negative text). As stated previously, the research builds on prior research by investigating effects of Delta variant spread on people intentions towards COVID-19 vaccination, and does Delta variant with its harsh symptoms increased public fears or not? As seen in the timeline above, June 2021 constitutes a baseline with Delta variant cases spreading highly at the beginning of July and caused high numbers of infection and deaths numbers. To elaborate, sufficient segments of population preferred to keep following social distancing and health interventions rather than taking the vaccines that could cause side effects, or be unsafe, or ineffective. However, fears of Delta variant became higher than fears of vaccine so that selecting the lower level of fears was the only available option for the those who hesitated to get vaccinated.

This research extracted three datasets that include the tweets for the selected vaccines, topics, keywords, and timeframe that this research proposed in chapter 3. Tweets are distributed below in Table 2 below:

Pfizer, followed by Moderna then Johnson & Johnson, has the most extracted tweets, which is more evidence of Pfizer popularity in the USA, and also as demonstrated by the CDC datasets.

For analyzing the tweets and finding out the opinions' direction and polarity, sentiment analysis was used in order to find out the sentiment scores of the users' opinions, where the scores are classified into positives, negatives, and neutrals. Therefore, quantifying the sentiments over the timeframe was a useful tool to show if there was a significant response to the four events or not. The following Figures 8,9, & 10 show the sentiment analysis results of the three datasets respectively over the whole timeframe (June 1st, 2021 – March 31st, 2022).

Twitter Data Extraction, Preprocessing, and Sentiment Analysis:

Data extraction from Twitter focused on United States users and their tweets about inoculation and vaccine hesitancy for the three COVID-19 vaccine types available (Pfizer, Johnson & Johnson, and Moderna) from June 2021 to March 2022. Users' tweets were extracted from the Twitter feed using technique demonstrated by (Pokharel, 2020; Morshed, et al. 2021) and the Tweepy library tool. "Tweepy is an open-source Python package that gives you a very convenient way to access the Twitter Application Programming Interface (API) with Python" (Python, 2019) in order to compose tweets, read profiles, and access your followers' data and a high volume of tweets on particular subjects in specific locations" (Fontanella, 2021).

Table 13: Twitter Data Extraction, Processing, and Sentiment analyzing Steps

Step	Description
Vaccination tweets extraction	Use of the technique and the Tweepy library to extract tweets from the Twitter API relevant to vaccinations (Mushtaq et al., 2022; Morshed et al., 2021). Using the vaccines keywords to extract the tweets that include then by using Tweepy library. In addition, using the geolocation features to extract tweets that were posted by the USA users. Public tweets extracted from Tweeter via its Application Programming Interface (API) was used for the experiment.
Tweets preprocessing	the Retweets and URLs were removed in the preprocessing step, emojis were converted into words, and the dataset was cleaned. We also removed stop words and performed tokenization. Stemming and lemmatization were done as well. (Ramachandran & Parvathi, 2019; Aljedaani et al., 2020;Monkey, 2020; Rsutam et al, 2021)
Sentiment analysis	Using Vader library to classify tweets into positive, neutral, and negatives. (Devika et al., 2016; Yadav & Vishwakarma, 2019; Dang et al., 2020; Pipis 2022) where tweets classifications are defined as following: - **Positive tweets** represent the opinions that are favor and support accepting the vaccines. - **Negative tweets** represent the opinions that are against and hesitating and rejecting the vaccines. - **Neutral tweets** represent the opinions that are not favor or against the vaccines, where tweets can’t represent positive neither negative opinions.

Table 14: Timeframe and Queries Keywords (Extended study)

Vaccine	Keywords	Timeframe
Pfizer-BioNTech vaccine	Pfizer, Pfizer-BioNTech, BioNTechpfizer COVID-19 vaccine, vaccination, dose	June 2021 – 31March 2022
Johnson & Johnson's COVID-19 Vaccine	Johnson & Johnson, Johnson and Johnson, Janssen, Janssen, COVID-19 vaccine, vaccination, dose	June 2021 – 31March 2022
Moderna vaccine	Moderna, Moderna_tx, Moderna-NIAID, NIAID, NIAID-Moderna COVID-19 vaccine, vaccination, dose	June 2021 – 31March 2022

Tweet Sentiment Analysis identified 326,124 tweets that indicated a positive viewpoint of vaccines and 163,716 tweets that indicated a negative viewpoint toward vaccines over the time frame identified in the Figure, where:

Positive tweets represent the opinions that are favor and support accepting the vaccines.

Negative tweets represent the opinions that are against and hesitating and rejecting the vaccines.

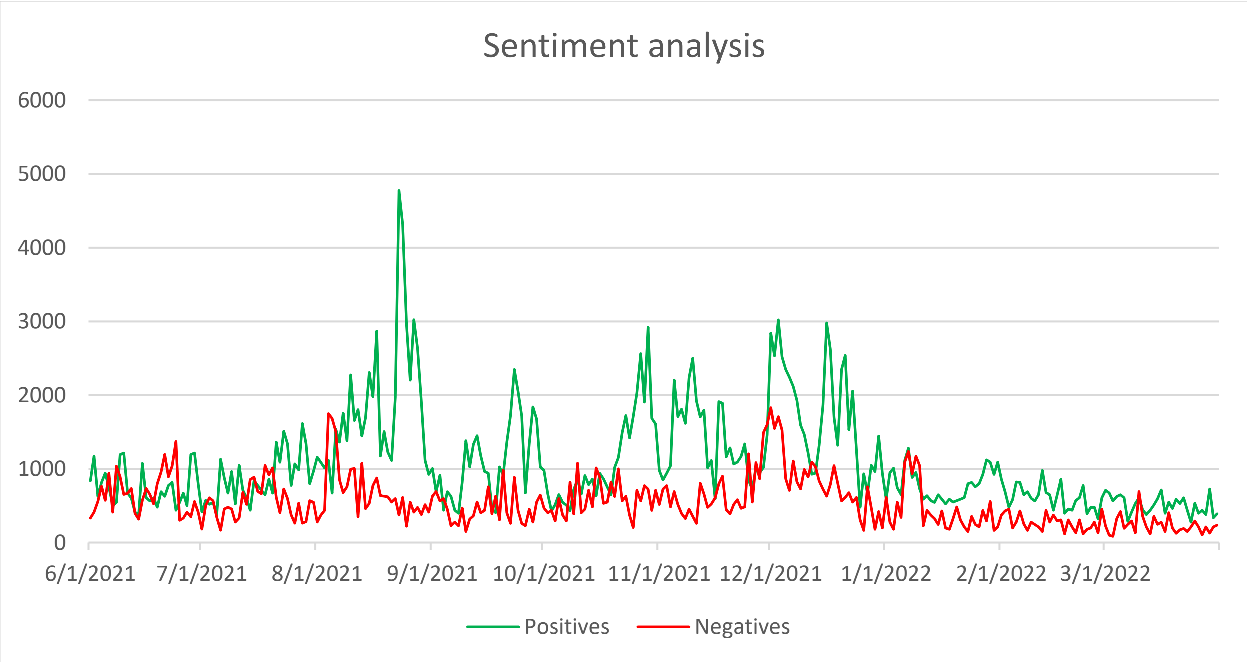


Figure 70: Total Positive and Negative Tweets of the three Vaccines

Statistical Analysis of Datasets

In this part, this research have analyzed both recorded datasets from CDC and extracted tweets from Twitter. Figures compare tweets to both cases, deaths, and vaccinated numbers separately.

In this part, this research conducted our statical analysis and hypotheses testing in order to find out the answers to our research questions. Thus, this research selected separated segments of datasets that represent the different four events. In order to quantify the correlations between the selected variables for each research question, this research used PEARSON test. Also, this research conducted the regression analysis on the chosen segments to investigate the relationships and find out if the regressions statistically significant or not?

CDC vaccination datasets visualization and analysis

The following three Figures below show the First, Second, and Booster dose inoculations in the USA, in addition to virus cases spread and daily change rates in vaccination.

: Red points indicate to the inflection points when the data starts to change dramatically from low levels to high levels.

: Green points indicate to top points and when the data starts to change dramatically from high levels to the low levels.

<u>First, second, and booster dose Figures</u>

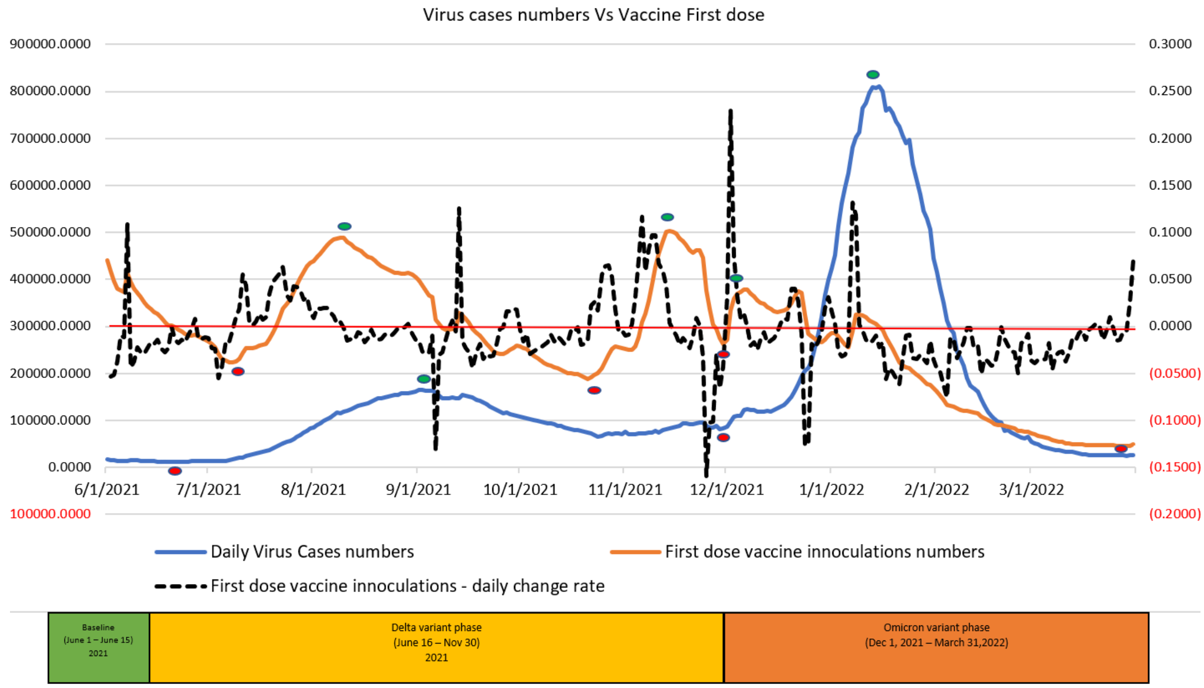


Figure 71: Virus cases numbers Vs Vaccine First dose

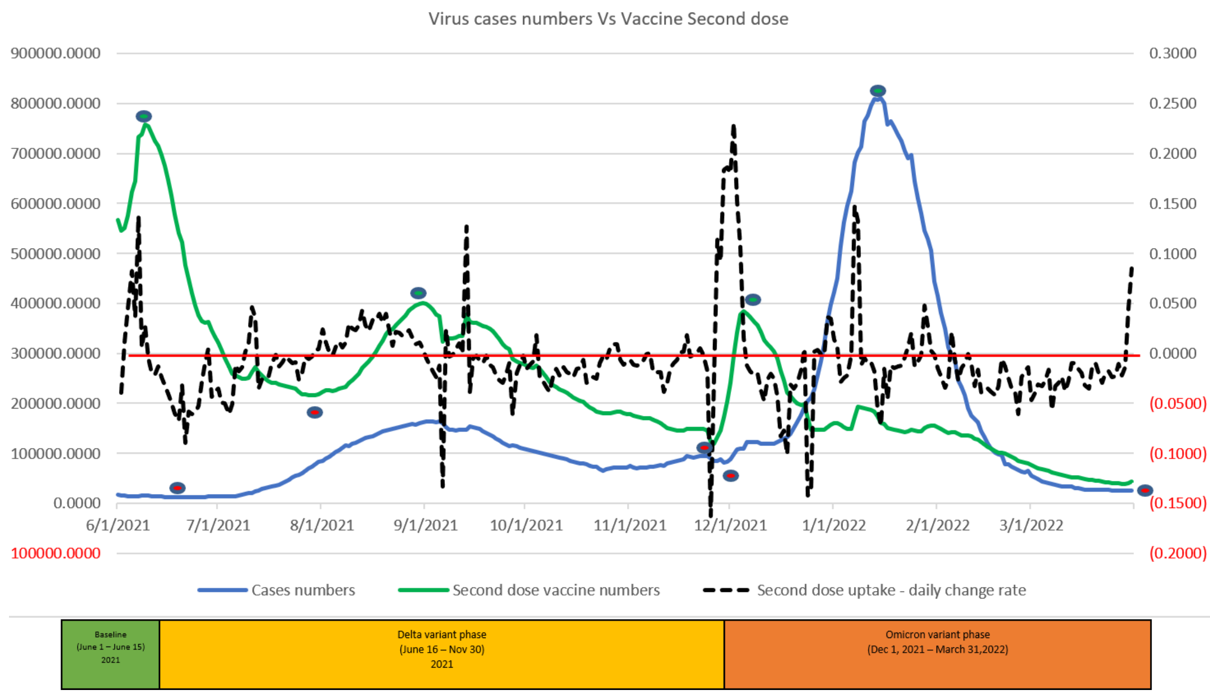


Figure 72: Virus cases numbers Vs Vaccine Second dose

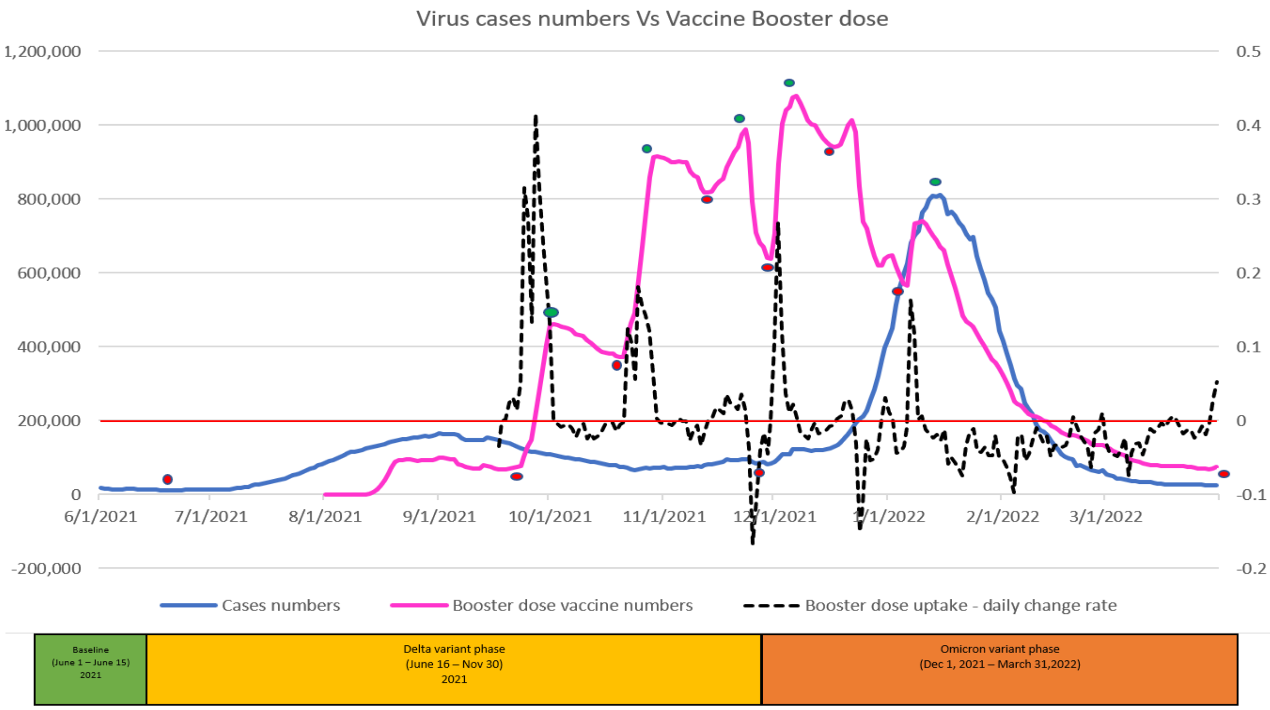


Figure 73: Virus cases numbers Vs Vaccine Booster dose

7-day moving average have been taken for the positive and negative tweets in order to smooth them and to be consistent with CDC datasets that we used into our research (7-day moving average). Also, the positive and negative tweets were collected together (Pfizer, Moderna, and Johnson & Johnson) as same as CDC vaccination inoculation datasets that represent the total daily number of COVID-19 vaccines together. So, the correlations and analysis have been performed on datasets with the same structure and features.

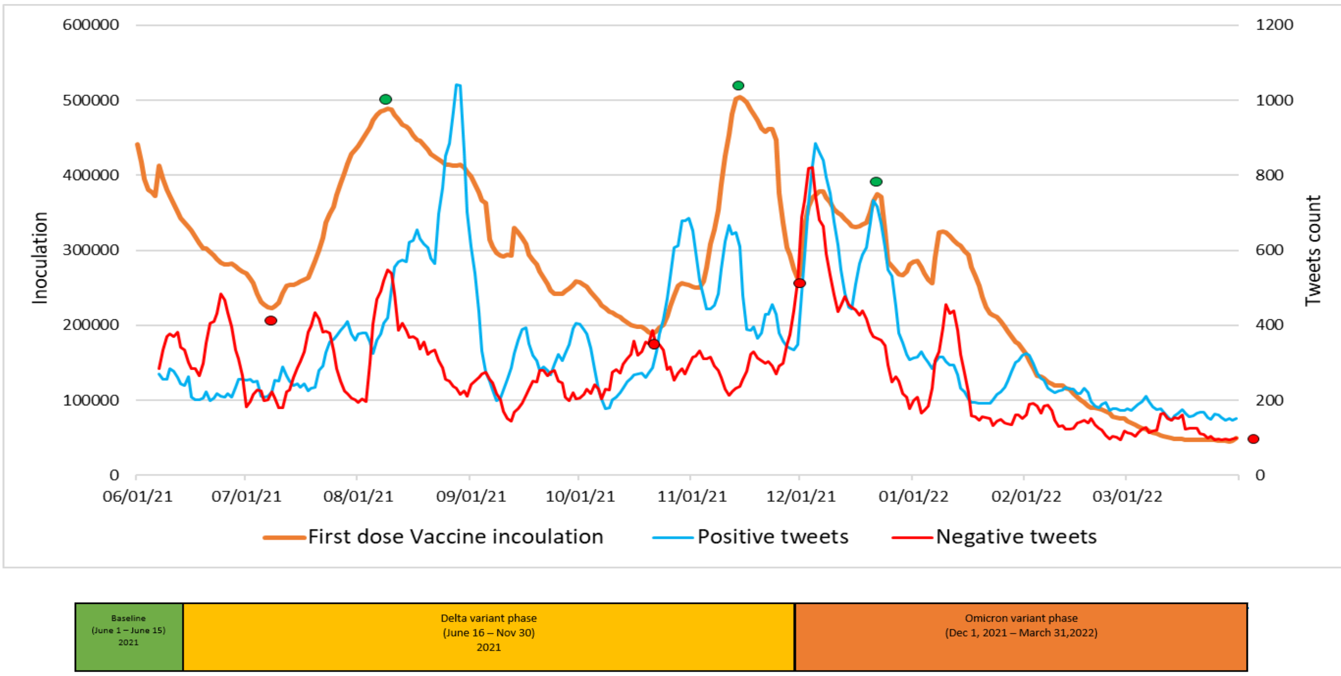


Figure 74: Positive and Negative Tweets toward First dose vaccines and vaccinations over time

Visual analysis of Figures 8 also appears to show and analysis confirms a large positive correlation (0.870) between positive tweets and Second Dose inoculations during the Omicron phase. Paradoxically, a large correlation of 0.892 was observed between negative tweets and Second Dose inoculations during the Omicron phase.

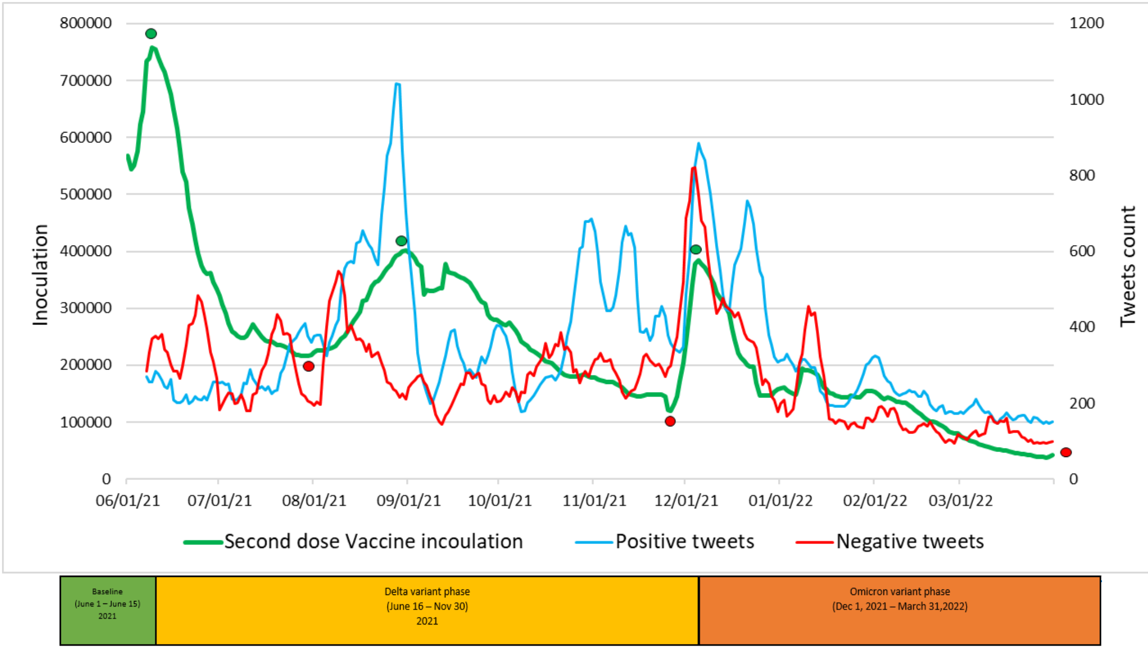


Figure 75: Positive and Negative Tweets toward Second dose vaccines and vaccinations over time

Visual analysis of Figures 9 also appears to show and analysis confirms a large positive correlation between positive tweets and booster dose inoculations during a portion of the Delta phase (0.768) and across the entire Omicron phase (0.868). Paradoxically, a large correlation of 0.845 was observed between negative tweets and Booster inoculations during the Omicron phase.

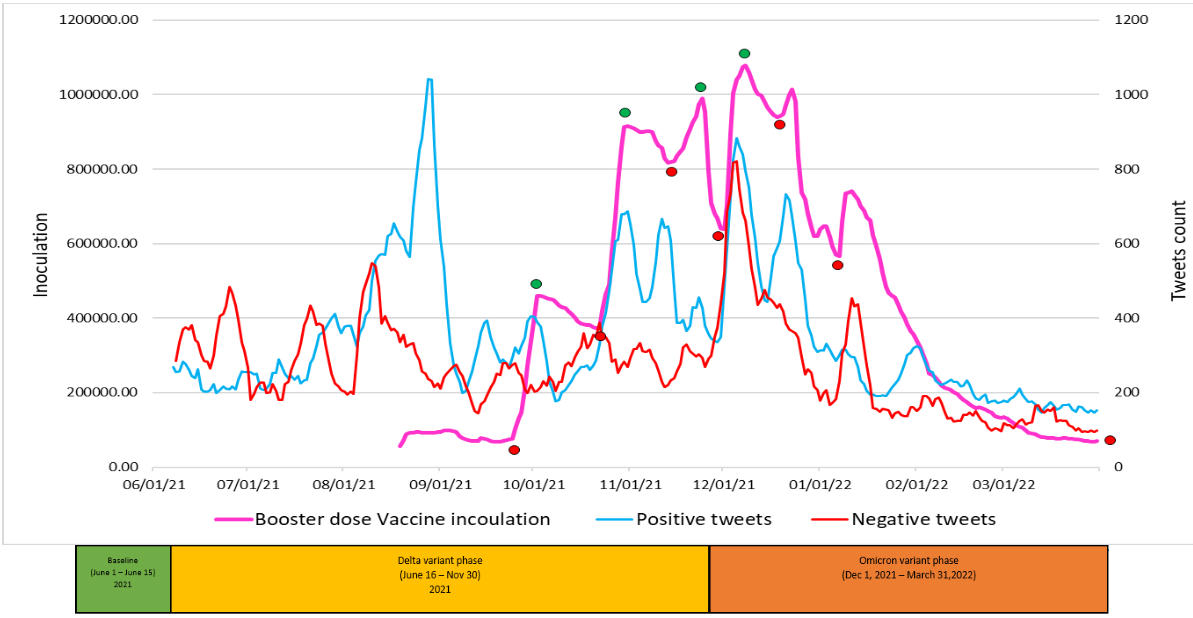
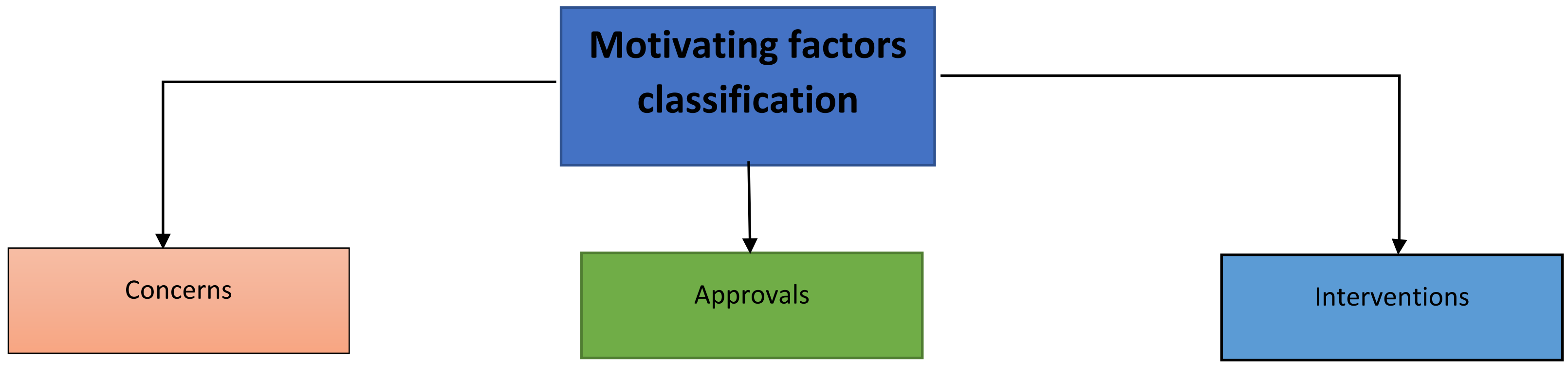


Figure 76: Positive and Negative Tweets toward Booster dose vaccines and vaccinations over time



Motivating factors included into the timeframe	**Concerns factors (CF)**	- Vaccines side effects and holds. - June 2021: Delta variant emergence. - Increasing in Virus and Deaths cases. - Dec 2021: Omicron variant emergence.
	Approvals factors (AF)	- Aug 2021: FDA Approves First COVID-19 Vaccine - Sep 2021: Pfizer Booster approval - Sep 2021: Moderna Booster approval - Oct 2021: Children dose approval
	Interventions factors (IF)	- Nov 2021: Biden mandate plan - Dec 2021: Reducing Incubation period

Figure 77: Motivating factors that affect vaccination levels

Results Analysis

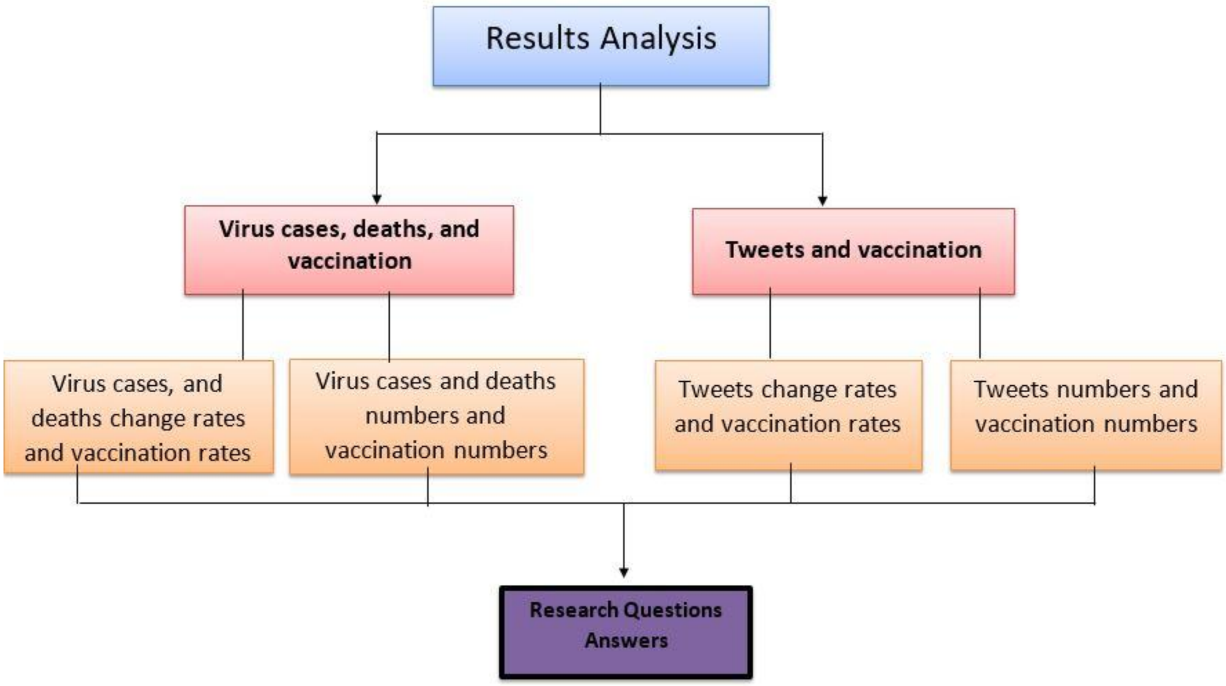


Figure 78: Correlations Analysis diagram

A. Cases Vs Vaccinations inoculations correlations analysis

Hypothesis: Virus Daily Case Level (numbers) Correlated to Daily Vaccine Inoculation Level

Table 15: Virus Daily Case Level Versus Daily Vaccine Inoculation Level

	First Dose	**Second Dose**	**Booster**
Baseline Phase	Accept: Large Correlation (0.766)	Reject: Small Inverse Correlation (-0.42)	N/A[5]
Delta Phase	Accept: Medium Correlation (0.375)	Reject: Negligible Correlation (0.082)	Reject: Large Inverse Correlation (-0.7069)
Omicron Phase	Accept: Medium Correlation (0.465)	Accept: Small Correlation (0.19)	Accept: Medium Correlation (0.323)
Overall	Reject: Negligible Correlation (0.010)[1]	Reject: Small Inverse Correlation (-0.221)[3]	Accept: Small Correlation (0.154)
Notable exceptions:	Delta First Period: Reject: Large Inverse Correlation (-0.872)[2]	Delta Second Period: Accept: Large Correlation (0.943)[4]	Delta Phase had several periods of either large inverse or direct correlations [6]

Conclusion: Case Level is NOT by itself a RELIABLE indicator of Inoculations.

Factors and Correlations interpretations:

1: The overall correlation (0.010) negligible as there are huge inverse and direct correlations between virus case and first dose inoculations over the timeframe.

2: A huge negative correlation during Delta first period (-0.872) since the virus cases increased and vaccine inculcation decreased due to the fears of vaccine side effects.

3: The overall correlation is small inverse (-0.22) as the whole-time frame includes direct and inverse correlations, but more inverse.

4: A large direct correlation (0.943) as the large raised in the virus cases motivated people to get the second dose, in addition to COVID-19 vaccine approval which was additional motivation factor that increased trust in vaccine safety and effectiveness.

5: Booster wasn’t approved at that time.

6: Large Inverse and direct correlations due to Pfizer booster dose approval in Sep 2021 and Moderna booster dose in Oct 2021, which results in huge increase periods and declining periods for both types of doses.

Hypothesis: Virus Daily Change Rate Level Correlated to Daily Change Rate Level of Vaccine Inoculations

Table 16: Virus Daily Change Rate Level Versus Daily Change Rate Level of Vaccine Inoculations

	First Dose	Second Dose	Booster
Baseline Phase	Accept: Small Correlation (0.241)	Accept: Medium Inverse Correlation (0.316)	N/A [2]
Delta Phase	Accept: Medium Correlation (0.427)	Accept: Small Correlation (0.129)	Reject: Negligible Correlation (0.0556) [3]
Omicron Phase	Accept: Medium Correlation (0.438)	Accept: Small Correlation (0.28)	Accept: Medium Correlation (0.396)
Overall	Accept: Medium Correlation (0.436)	Accept: Small Correlation (0.221)	Accept: Small Correlation (0.205)
Notable exceptions:	None	Delta Second Period: Reject: Medium Inverse Correlation (-0.454) [1]	

Conclusion: With two exceptions, Virus Daily Change Rate is by itself a Small to Medium indicator of Inoculation Daily Change Rate

Factors and Correlations interpretations:

1: The inverse correlation (-0.454) due to the overall declining in the virus cases daily change rate and overall raising in second dose vaccine inoculation daily change rate.

2: Booster dose wasn't approved at that time.

3: Negligible correlation (0.0556) as the multiple raising and declining in the same and inverse directions, which results in weak correlations.

B. Deaths Vs Vaccinations inoculations correlations analysis

Hypothesis: Virus Daily Death Level (numbers) Correlated to Daily Vaccine Inoculation Level

Table 17 : Virus Daily Death Level (numbers) Versus Daily Vaccine Inoculation Level

Phase	First Dose	Second Dose	Booster
Baseline Phase	Accept: Large Correlation (0.886)	Reject: Small Inverse Correlation (-0.167)	N/A[3]
Delta Phase	Reject: Small Inverse Correlation (-0.151)	Reject: Negligible Correlation (0.011)	Reject: Large Inverse Correlation (-0.866)
Omicron Phase	Reject: Negligible Correlation (-0.067)	Reject: Negligible Correlation (0.190)	Reject: Small Inverse Correlation (-0.165)
Overall	Reject: Medium Inverse Correlation (-0.360)	Reject: Medium Inverse Correlation (-0.369)	Reject: Medium Inverse Correlation (-0.351)
Notable exceptions:	Delta First & Second Period: Accept: Large Correlation (0.781 & 0.928)[1]	Delta Second & Third Period: Accept: Large Correlation (0.992 & 0.883)[2]	Delta Phase Second, Fourth, & Sixth periods had large correlations (0.978, 0.662, 0.868)[3]

Conclusion: With several Period and three Phase exceptions, Virus Daily Deaths are a Small, Medium, and Large INVERSE indicator of Inoculations. The notion that HIGHER LEVELS of DEATHS would infer LOWER LEVELS of inoculation is counter intuitive UNLESS one considers the demographics of the population available for vaccination. That is to say, where are we on the Johnson Technology Adoption Curve?

Factors and Correlations interpretations:

1: The large direct correlation (0.781 & 0.928) due to the declining in virus deaths and the first dose vaccine inoculation during the Delta first period, and the raising in the virus deaths and first dose vaccine inoculation during the Delta second period respectively.

2- The large direct correlation (0.781 & 0.928) due to the raising in virus deaths and the second dose vaccine inoculation during the Delta second period, then the declining in the virus deaths and second dose vaccine inoculation during the Delta third period respectively.

3- The declining in virus deaths and booster dose inoculation during these notable three periods results in large direct correlations (0.978, 0.662, 0.868).

Hypothesis: Virus Daily Death RATE Level Correlated to Daily Vaccine Inoculation RATE Level

Table 18: Virus Daily Death RATE Level Versus Daily Vaccine Inoculation RATE Level

	First Dose	**Second Dose**	**Booster**
Baseline Phase	Accept: Medium Correlation (0.382)	Accept: Large Correlation (0.551)	N/A[3]
Delta Phase	Accept: Small Correlation (0.271)	Accept: Medium Correlation (0.355)	Accept: Small Correlation (0.188)
Omicron Phase	Accept: Medium Correlation (0.388)	Accept: Medium Correlation (0.473)	Accept: Medium Correlation (0.421)
Overall	Accept: Medium Correlation (0.346)	Accept: Medium Correlation (0.404)	Accept: Small Correlation (0.217)
Notable exceptions:	Delta Second Period: Reject: Negligible Correlation (-0.058)[1]	Delta Second Period: Reject: Negligible Correlation (0.097)[2]	Delta Phase Fourth periods had NEGLIGIBLE correlations (-0.014)[3]

Conclusion: With three exceptions, Virus Daily Death Change Rate is by itself a Small, Medium, &Large indicator of Inoculation Daily Death Change Rate

C. Tweets Vs Vaccination inoculations correlations analysis

Hypothesis: Daily Vaccine Tweets Level (numbers) Correlated to Daily Vaccine Inoculation Level

Table 19: Daily Vaccine Tweets Level (numbers) Versus Daily Vaccine Inoculation Level

	First Dose		**Second Dose**		**Booster**	
Phase	**Positive tweets**	**Negative tweets**	**Positive tweets**	**Negative tweets**	**Positive tweets**	**Negative tweets**
Baseline Phase	Accept: Medium Correlation 0.3472	Reject: Negligible Correlation -0.1838	Accept: Medium Correlation 0.4407	Accept: Large Correlation 0.5359	N/A	N/A
Delta Phase	Accept: Large Correlation 0.6071	Accept: Medium Correlation 0.3582	Reject: Negligible Correlation -0.0375	Reject: Negligible Correlation 0.0226	Accept: Large Correlation 0.7681	Accept: Medium Correlation 0.4413
Omicron Phase	Accept: Large Correlation 0.8043	Accept: Large Correlation 0.7936	Accept: Large Correlation 0.8696	Accept: Large Correlation 0.8918	Accept: Large Correlation 0.8684	Accept: Large Correlation 0.8452
Overall	Accept: Large Correlation 0.6738	Accept: Large Correlation 0.6181	Accept: small Correlation 0.25430	Accept: Medium Correlation 0.4615	Accept: Large Correlation 0.8416	Accept: Large Correlation 0.7384

Conclusion: Overall positive and negative tweets have a large direct correlation with vaccination. To explain, the positive and negative tweets increase around VOC's and vaccine approvals events as a public reaction and opinions with taking into consideration the possible the difference between proportions of positive and negatives that could increase or decrease based on the event. Moreover, tweets declining during next periods after that event, where Twitter users' reaction attentions decline toward that recent event. On the other hand, CDC vaccination data show that vaccination inoculation increased significantly after approvals or VOC's except the Delta variant announcement (VOC's) that took time to do impact, where people started taking that VOC seriously after the dramatic raising in the virus cases which became larger times than baseline cases. So, fears of virus infection motivated a lot of people to get their first dose.

Hypothesis: Daily Vaccine Tweets Level Correlated to Daily Vaccine Inoculation Level

Table 20: Daily Vaccine Tweets Level Versus Daily Vaccine Inoculation Level

	First Dose		**Second Dose**		**Booster**	
Phase	**Positive tweets**	**Negative tweets**	**Positive tweets**	**Negative tweets**	**Positive tweets**	**Negative tweets**
Baseline Phase	Accept: Large Correlation 0.5114	Reject: Large Inverse Correlation -0.8221	Reject: Small Inverse Correlation -0.1082	Accept: Large Correlation 0.7380	N/A	N/A
Delta Phase	Accept: Medium Correlation 0.3887	Reject: Small Inverse Correlation -0.1253	Accept: Small Correlation 0.1751	Accept: Small Correlation 0.1216	Accept: Medium Correlation 0.4068	Reject: Negligible Correlation 0.0219
Omicron Phase	Accept: Medium Correlation 0.4291	Accept: Small Correlation 0.2167	Accept: Medium Correlation 0.4808	Accept: Medium Correlation 0.3183	Accept: Large Correlation 0.5043	Accept: Small Correlation 0.2831
Overall	Accept: Medium Correlation 0.4033	Reject: Negligible Correlation 0.0111	Accept: Medium Correlation 0.3068	Accept: Small Correlation 0.2236	Accept: Medium Correlation 0.9989	Accept: Medium Correlation 0.5027

Conclusion: Overall daily change rate in positive and negative tweets have a medium direct correlation with daily change in booster dose vaccination as booster approval events and large booster dose inoculation that recorded during the timeframe, while the first and dose inoculation rates have medium correlation with positive tweets, but weak correlation with daily change rate in negative tweets.

Johnson Technology Adoption Curve Numbers

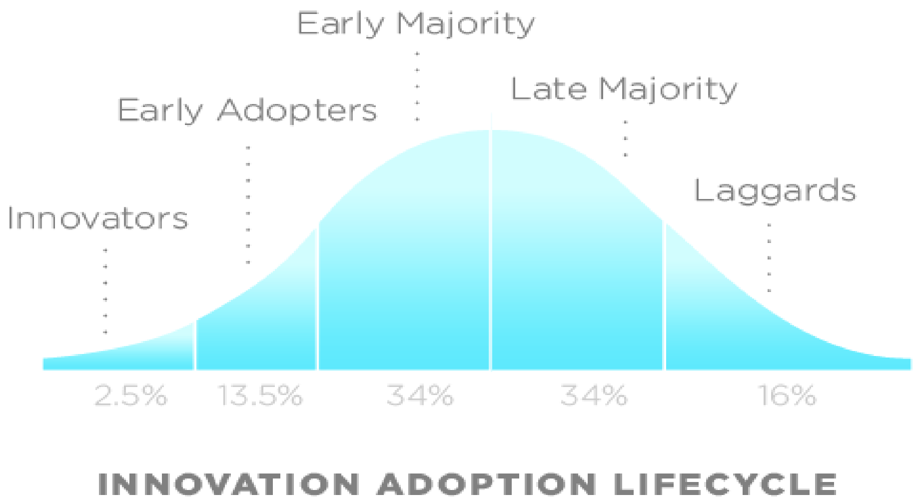


Figure 79: Johnson Technology Adoption Curve

Innovators (2.5% of employees) - These employees are innovators, the first individuals to adopt new technologies in the workplace, they are not afraid to take risks and usually test out different technologies in their personal lives as well.

Early Adopters (13.5% of employees) - These employees are not as risk averse as the innovators; however they care about progressing their career, building a reputation in the company and making an impact. Usually on higher end in the company hierarchy and want to invest in whatever it takes to help the company succeed.

Early Majority (34% of employees) - Typically middle and line managers these employees tend to be slower in the adoption process, they will usually wait and see how a new technology is faring and if the higher ups are adopting it before committing to using it themselves.

Late Majority (34% of employees) - These individuals are usually very skeptical about innovations and new technologies, and will only adopt it after they see a large percentage of the company using it. They are usually older in age and have been at the company for a long period of time.

Laggards (16% of Employees) - These individuals tend to be advanced in age, typically focused on traditions vs. innovations. These Individuals tend to use new technology in one of two scenarios:

There is no other alternative for them to get the job done without using technology

They are being forced to use the technology and would be penalized for not using it.

Table 21: Total vaccinated population

Vaccinated status	May 31, 2021	Nov 30, 2021	March 31, 2022
At least one dose	173,531,874	227,829,618	256,144,043
Fully vaccinated	146,813,131	199,245,522	219,319,838
First booster dose	9,318	34,920,883	100,230,127

Table 22: Vaccination Coverage

Vaccinated status	May 31, 2021	Nov 30, 2021	March 31, 2022	To reach late majority	To reach laggards
At least one dose	52.30%	70.50%	77.10%	6.9%	22.9%
Fully vaccinated	44.20%	61%	66.10%	17.10%	33.9%
First booster dose	0%	18.50%	45.70%	38.3%	54.3%

Table 22 shows the vaccination coverage during the baseline, delta variant, and Omicron variant phase. Also, it shows the difference between three vaccination doses coverages and criteria of Johnson Technology Adoption.

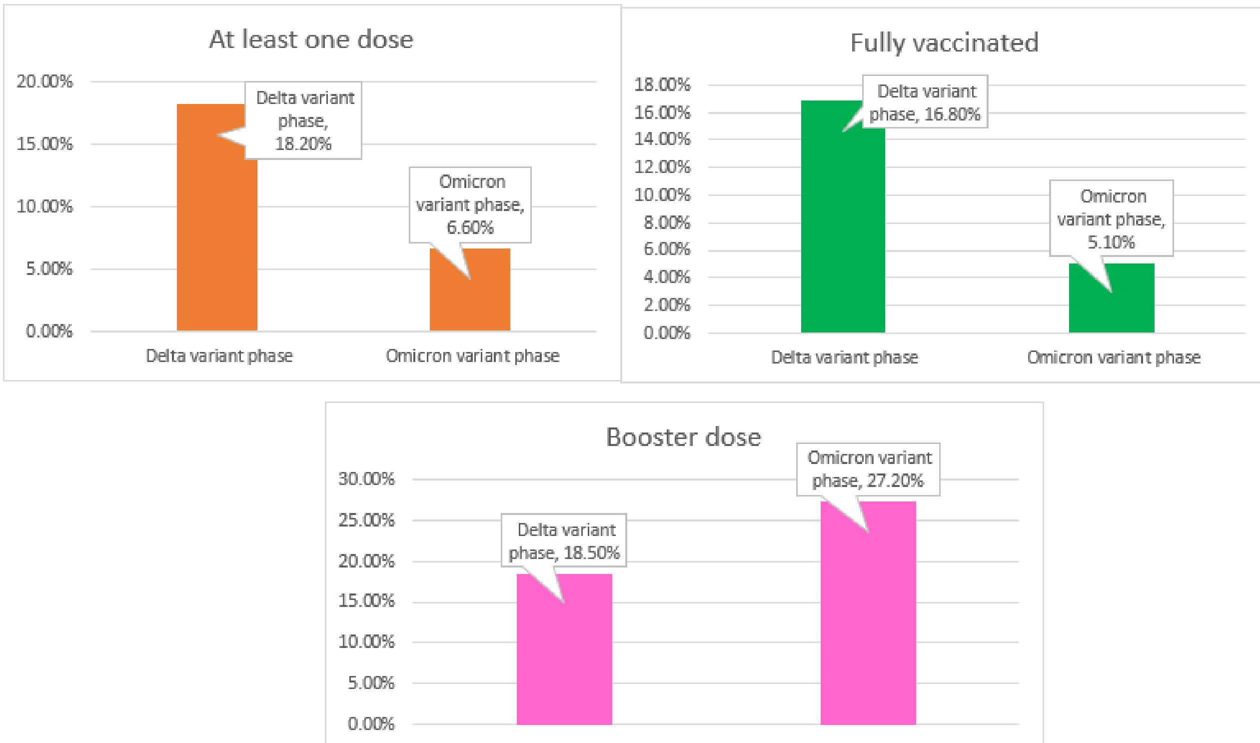


Figure 80: Vaccination Coverage (Delta variant Versus Omicron variant phase)

Three Figures compare the vaccination coverage during Delta phase and Omicron phase, where the first and second dose inoculation were higher during Delta phase, However, booster dose vaccination achieved higher vaccination coverage during Omicron phase. In another word, beside intentions to take the booster dose, a huge segment of partially and fully who got vaccinated during Delta phase, they got their booster dose during Omicron phase.

Modeling and Predicting Daily vaccination inoculation

The Modeling and Analysis involved the fore mentioned linear regressions to predict first dose, second dose, and booster vaccination inoculation. Independent values included a constant associated with dose and variables associated with phase, daily virus cases, daily deaths, daily positive tweets, and daily negative tweets on day **x**. Respective regression equations produced first dose vaccination inoculation for day **x+1**, second dose vaccination inoculation for day **x + 1**, or Booster dose vaccination inoculation for day **x + 1.** Figure below explains the model structure (*Predictive sentiment-based model*).

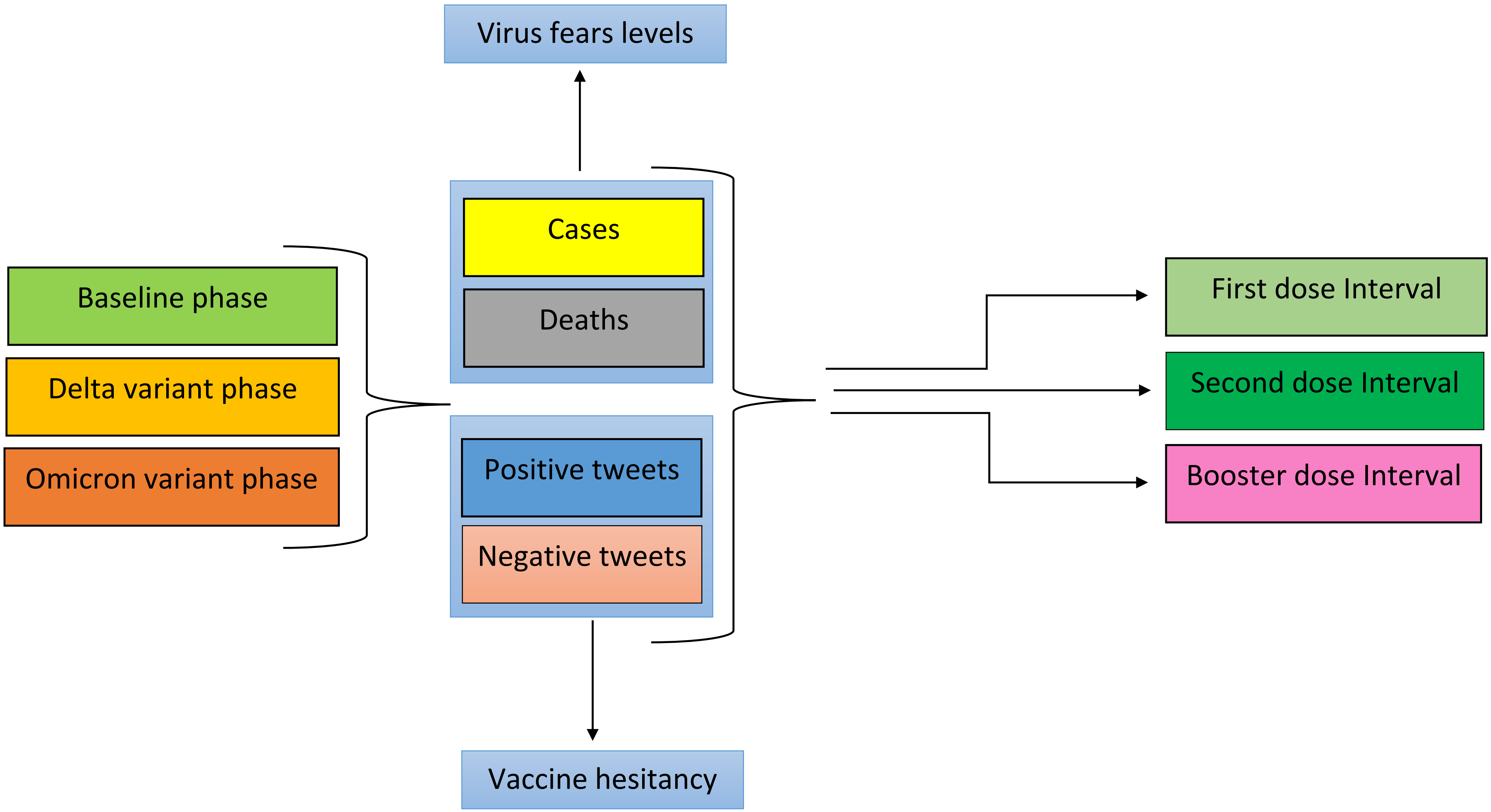


Figure 81: Predictive sentiment-based model

Regression analysis

Regression Equation

First dose Vaccination numbers = 396886 - 113202 Phases + 0.2849 Cases - 50.57 Deaths
+ 327.7 positive tweets numbers
+ 165.3 Negative tweets numbers

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	394879	19871	19.87	0.000	
Phases	-112483	7475	-15.05	0.000	1.44
Cases	0.2845	0.0227	12.55	0.000	1.62
Deaths	-50.54	6.99	-7.23	0.000	1.91
positive tweets numbers	328.4	21.5	15.26	0.000	1.39
Negative tweets numbers	165.0	33.2	4.97	0.000	1.61

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
58205.8	78.25%	77.87%	77.47%

Figure 82: First dose prediction model. P values are significant for all five x variables

Regression Equation

Second dose Vaccination numbers = 521548 - 166129 Phases + 0.0974 Cases + 329.9 Negative tweets numbers

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	521548	30273	17.23	0.000	
Phases	-166129	11339	-14.65	0.000	1.32
Cases	0.0974	0.0318	3.06	0.002	1.27
Negative tweets numbers	329.9	42.6	7.74	0.000	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
92271.8	55.06%	54.60%	53.82%

Figure 83: Second dose prediction model.

P values are significant for all three variables except the positive tweets and deaths variables which have been removed as their P. values were not significant.

Regression Equation

Booster dose Vaccination number = 360234 - 96629 Phases + 0.6428 Cases - 166.9 Deaths + 1221.1 positive tweets numbers + 378.3 Negative tweets numbers

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	360234	68244	5.28	0.000	
Phases	-96629	20638	-4.68	0.000	1.21
Cases	0.6428	0.0520	12.35	0.000	1.55
Deaths	-166.9	21.1	-7.91	0.000	1.56
positive tweets numbers	1221.1	77.8	15.70	0.000	2.24
Negative tweets numbers	378.3	97.3	3.89	0.000	2.25

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
127707	85.29%	84.90%	84.45%

Figure 84: Booster dose prediction model.

P values are significant for all five variables

Note: The dissertation and its findings are summarized in Appendix B, which is formatted as a journal paper and is being considered for publication.

Research Methodology and Public Health Field

The research methodology can be used by heath care professional and decision maker to explore the public behavior and their attentions towards vaccinations or interventions during infectious disease pandemics. Understanding the relationship between the public behavior response levels and the different waves or different virus risky level during pandemics can be used to predict public attentions towards healthy practices, such as vaccination, wearing masks, and following social distancing interventions. Into this research, social media factors were integrated with CDC datasets to perform strong analysis and behavior modeling that predict public response to multiple events. According to the conducted analysis, there is a relationship between CDC announcements and public response to them. As these announcements spread through social media in short time, they can motivate population to react quick to them. However, announcements are not enough factor to motivate people to take vaccine, where these announcements should involve other factors, such as nature of virus risk, infection rates, and deaths rates. In other word, announcement can make a significant impact if the virus is risky, so people are going to take the vaccine. Otherwise, people don't response significantly to the announcement. Announcements are more effective based on the virus risk level, where fears of infections or deaths are the most effective factors that push people to take the vaccines. Even though the Omicron variant wasn't danger as much as Delta variant, people reacted immediately to the Omicron VOC announcement on Nov 30, 2022, because of their past experience with Delta variant wave in the past. For Delta VOC announcement on June 15, 2022, they reacted lately to the announcement after the variant caused a lot of deaths so their experience with Delta variant wave pushed them to react immediately to Omicron VOC and take the vaccine.

In another word, for the pandemics and vaccination modeling, many factors need to be involved as CDC VOC, virus risk level, vaccine acceptance and hesitancy levels, vaccine approvals, and governmental interventions. As all these factors generate a great modeling for the human behavioral and intentions toward vaccination under different waves of pandemic.

APPENDIX A: MODELS AND SOCIAL MEDIA

MODELS AND SOCIAL MEDIA

Table 23 shows a summary of models that we reviewed, and it demonstrates their details with clarifying which modes were integrated with social media and social networks.

1.Model: model name.

2.Type: What is the model type?

3.Goal: What are the model outcomes?

4.Integrating with social media or social networks: Was model Calibrated with social media (SM) or social networks (NW).

5.Platform/application? (The involved platform or application in the model, Facebook, Twitter, Google, …. etc.

6.How social media and social networks was used? Was it added to model to study its effect on the public behaviors, or as a data source (Data Mining)?

Therefore, all these questions and details would be explained in table 23, so that comparison and understanding the differences among the discussed models becomes easier and clearer.

The below table summarizes the presented models in the literature review:

***Social media: SM**

***Social Network: SN**

Table 23: Models Summary

Model	Type	Goal / Predict No of:	Calibrated with SM or SN?	Platform/ Application ?	How was SM or SN used?	Reference
SIR	Epidemiological State models	Susceptible cases Infected cases Recovered cases	No	-----	-----	(Cooper, Mondal, & Antonopoulos, 2020)
SIS	Epidemiological State models	Susceptible cases infected cases	No	-----	-----	(Vargas-De-León, 2011)
SIRD	Epidemiological State models	Susceptible cases infected cases Recovered cases Deaths cases	No	-----	-----	(Fernández-Villaverde & Jones, 2020)
MSIR	Epidemiological State models	Susceptible cases infected cases Recovered cases	No	-----	-----	(Seyoum Desta, 2019)
SEI	Epidemiological State models	Exposed cases Infected cases Recovered cases	Yes/ SM	Media awareness programs	Use media to modify public behavior	(Lu, Wang, Liu, & Li, 2017)
SEIR	Epidemiological State models	Susceptible cases Exposed cases infected cases Recovered cases	No	-----	-----	(Lekone & Finkenstädt, 2006)
SEIS	Epidemiological State models	Susceptible cases Exposed cases infected cases	No	-----	-----	(Fan, Li, & Wang, 2001)

MSEIR	Epidemiological State models	Susceptible cases Exposed cases infected cases Recovered cases	No	-----	-----	(Zhou & Li)
MSEIRS	Epidemiological State models	Susceptible cases Exposed cases infected cases Recovered cases	No	-----	-----	(Mosavi, 2020)
SLIRDS	Epidemiological State models	Susceptible cases Latent cases Infected cases Recovered cases Deaths cases	No	-----	-----	(Bin, Sun, & Chen, C. C. 2019)
EIH	Epidemiological State models	Exposed cases Infected cases Hospitalized cases	Yes/ SM	Media awareness programs	Use media to modify public behavior	(Greenhalgh et al., 2015)
SIHR	Epidemiological State models	Susceptible cases Infected cases Hospitalized cases Recovered cases	Yes/ SM	Media awareness programs	Use media to modify public behavior	(Tchuenche & Bauch, 2012)
SEIQRDV.F	Statistical prediction model	Susceptible cases Exposed cases Infected cases Quarantined cases Recovered cases Vaccinated cases	No	-----	-----	(Ghostine et al., 2021)
DELPHI	Statistical prediction model	Infected cases Hospitalized cases Deaths cases	No	-----	-----	. (COVIDAnalytics, 2020)
ARIMA	Statistical prediction model	Infected cases	Yes/ SM	Google & Twitter	Data source	(Samaras, García-Barriocanal, & Sicilia, 2020)
LANL COFFEF	Statistical prediction model	Infected cases Death cases Predicting time of pandemic peak	No	-----	-----	Loey, Manogaran, Taha, &) Khalifa, 2020)

JHU COVID-19	Statistical prediction model	"Forecast how likely a patient's disease is to worsen while being treated in a hospital and at what point in their care that might happen"	No	-----	-----	(Garibaldi et al., 2020)
HBM	Theoretical Interventions model	Analyzing and predicting population intentions to follow healthy interventions	Yes/ SM	Facebook	Data source	(Champion, Skinner, and others, 2008)
TPB	Theoretical Interventions model	Predicting the human behavior and intentions toward the healthy interventions	No	-----	-----	(Ajzen, 1991)
PMT	Theoretical Interventions model	Explaining how individuals are motivated to act to protect themselves.	No	-----	-----	(Okuhara, Okada, & Kiuchi, 2020)
CCDFD	Agent-based model	Modeling disease dynamics and fear as two interacting contagion processes.	No	-----	-----	(Epstein et al., 2008)
SD	Agent-based model	Testing effects of different levels of social distancing policies on the diseases spread.	No	-----	-----	(Daghriri & Ozmen, 2021)
COVASIM	Agent-based model	Projecting epidemic trends. Exploring the intervention scenarios. Estimating the resources needs.	No	-----	-----	.(Kerr et al., 2020)
COVASIM & Vaccination	Agent-based model	"Effectiveness of a nationwide vaccine campaign in response to different vaccine efficacies, the willingness of the population to be vaccinated, and the daily vaccine capacity under two different federal plans. Studying the interactions between nonpharmaceutical interventions and vaccines.	No	-----	-----	Li & Giabbanelli Giabbanelli, 2021)

FM	Agent-based model	Tracing users' phones and their mobility through network to study effects of government' interventions on virus spread	Yes/ SN	Mobile phones-Calls	Data source	(Frias-Martinez, Williamson, & ,Frias-Martinez, 2011)
COVID-19 agent-based simulation (COVID-ABS)	Agent-based model	Simulating the epidemiological and economic impacts of social distancing policies	No	-----	-----	(Silva et al., 2020).
UT COVID-19 -SD	Agent-based model	Tracing users' phones and their mobility through GPS to study effects of government' interventions on virus spread	Yes/ SN	Mobile phones-GPS traces	Data source	(Woody et al., 2020)
DMAS-SIR model	Multiagent-based model	Susceptible cases Infected cases Recovered cases Quarantine impact Transport restrictions impact Effectiveness of the interventions on the disease spread	No	-----	-----	(Vyklyuk et al., 2021)
YYG	Hybrid model	Infected cases Deaths cases	No	-----	-----	(Giattino, 2020)
DTL	Hybrid model	Processing population' images to detect who wear mask or who not	No	-----	-----	(Loey, Manogaran, Taha, & Khalifa, 2020)
UVA COVID-19	Hybrid model	Effectiveness of the interventions on the disease spread. No of required beds and at hospitals and care units. Trace users' phones and their mobility through GPS.	Yes/ SN	Mobile phones-GPS	Data source	(Price & Propp, 2020) UVA COVID-19 model, 2021))
IHME COVID-19	Hybrid model	Effectiveness of the interventions on the disease spread. Tracing users' phones and their mobility through GPS	Yes/ SN	Mobile phones-GPS	Data source	(Institute for Health Metrics and Evaluation, 2020)

MIT University COVID-19	Hybrid model	Infected cases Deaths cases No of required beds and at hospitals and care units	No	-----	-----	(Gallagher, 2020)
TWV Model	Hybrid model	Studying and analyzing twitter users’ emotions, beliefs, and opinions about vaccination	SM /Yes	Twitter/User tweets s’	source Data	Sattar & Arifuzzaman, 2021

APPENDIX B: MANUSCRIPT

vaccines MDPI

Article

Modeling Behavior and VACCINE HESITANCY using TWITTER-derived US population SENTIMENT during the COVID19 Pandemic to Predict daily VACCINATION INOCULATIONs

Talal Daghriri [1,2*], Michael Proctor [2,3], Sarah Matthews [3] and Abdullateef H Bashiri [4]

1 Department of Industrial Engineering, Jazan University, Jazan 82822, Saudi Arabia
2 Department of Industrial Engineering & Management Systems, University of Central Florida, Orlando, FL 32816, USA
3 Interdisciplinary Modeling and Simulation Program, University of Central Florida, Orlando, FL 32816, USA
4 Department of Mechanical Engineering, Jazan University, Jazan 82822, Saudi Arabia
* Correspondence: tdaghriri@knights.ucf.edu

Citation: Daghriri, T.; Proctor, M.; Matthews, S.; Bashiri, A.H. Modeling Behavior and VACCINE HESITANCY using TWITTER-derived US population SENTIMENT during the COVID19 Pandemic to Predict daily VACCINATION INOCULATIONs. *Vaccines* **2023**, *9*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Received: date
Accepted: date
Published: date

Abstract: Sentiment analysis of social media for predicting behavior during a pandemic is seminal in nature. As an applied contribution, we present sentiment-based regression models for predicting United States COVID19 first dose, second dose, and booster daily inoculations from June 1 2021 to March 31, 2022. Models merge independent variables representing fear of the virus and vaccine hesitancy. Large correlations exceeding 77% and 84% for first dose and booster dose models inspire confidence in the merger of the independent variables. Death count as a traditional measure of fear is a lagging indicator of inoculations while Twitter positive and negative tweets are strong predictors of inoculations. Thus the use of sentiment analysis for predicting inoculations is strongly supported with administrative events being catalysts for tweets. Non-inclusion in the second dose regression model of data occurring before the June 1, 2021 timeframe appear to limit second dose model results to only achieving moderate correlation exceeding 53%. Limiting tweet collection to geolocated tweets does not encompass the entire U.S. Twitter population. None the less, results from Kaiser Family Foundation (KFF) surveys appear to generally support the regression factors common to the First Dose and Booster Dose regression models and their results.

Introduction

In the early stages of the COVID19 pandemic, online contributors put forth predictions through various social media channels with one of the more infamous declarations of an "eradication" phase ending the pandemic in June 2021 (Forrester, 2020). Like so many other commentaries through social media during the pandemic, the prediction proved wrong despite its widespread distribution and possible influence on the public.

From June 2011 to April 2019 researchers including Piedrahita-Valdés et al. 2021 studied the controversial influence on public sentiment and behavior generated through global social media channels with the particular focus on vaccine hesitancy. Other researchers in Korea (Shim et al., 2021), Turkey (Küçükali et al., 2022), India (SV et al., 2022), and United States of America (Ruz et al., 2020,; Daghriri, Proctor, and Matthews, 2022) expanded social media sentiment analysis, in particular using Twitter tweets, as a significant resource of data and analysis to rapidly track and quantified public opinions, beliefs, or behavior regarding pandemic related events, personalities, or subjects including quickly and effectively measuring vaccine hesitancy. Staying abreast of rapidly changing COVID19 events during the first half of 2021, Sattar and Arifuzzaman, 2021; Mushtaq et al., 2022 inferred that similar levels of social media positive and negative sentiment toward vaccines indicated proportionally equal segments of the population were either inclined or not inclined toward getting vaccinated. Declining cases and deaths from January 14th, 2021 until June 23 supported the "eradication" prediction leading to decreasing fear of the virus (CDC, 2021). The US situation changed rapidly with the emergence of Delta variant. The sudden increase in the number of cases, the severity of symptoms as evidenced by increasing deaths, and changing social sentiment appeared to change behavior toward vaccine acceptance and inoculation. Researchers similarly extended vaccine social media sentiment analysis (Satter and Arifuzzaman, 2021; Daghriri, Proctor, and Matthews, 2022).

Statement of contribution

This research starts with two premises, the first being that the level of vaccine acceptance or inoculation is driven by the level of fear of the virus with higher fear of virus threats (e.g. illness, death) theoretically resulting in increasing vaccine acceptance or increasing inoculation (Figure 1a). The second premise is that the level of vaccine acceptance or inoculation is also driven by the level of fear of vaccine side effects resulting in vaccine hesitancy with higher vaccine hesitancy theoretically resulting in decreasing vaccine acceptance and decreasing inoculations (Figure 1b) (Sekizawa et al., 2022).

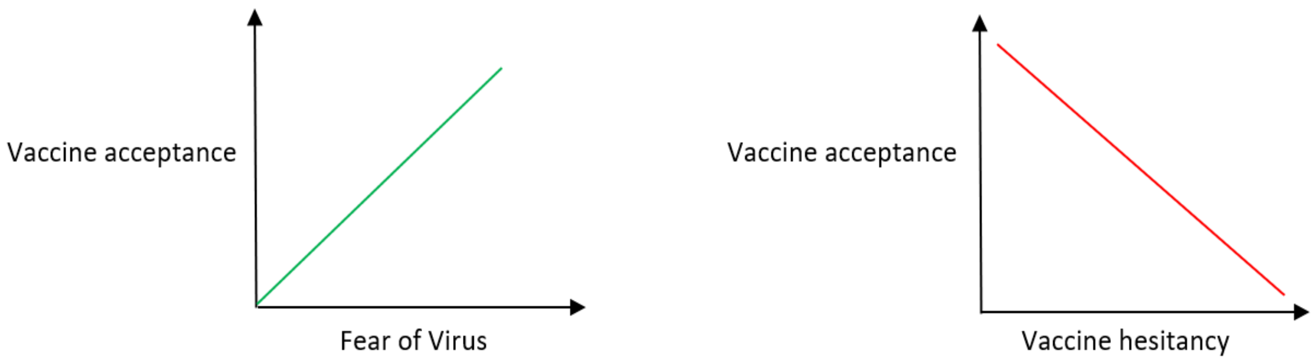


Figure 1. a: Fear of Virus and Vaccinations Acceptance. Figure 1b: Vaccine hesitancy and Vaccine Acceptance.

This research merges these two inconsistent drivers of inoculation behavior into data driven models. The resulting models extend social media sentiment analysis-by applying regression to predicting United States daily vaccine inoculations during the research timeframe of June 1 2021 to March 31, 2022. Spanning this timeframe, this research creates predictive regression models for each of the three vaccine inoculation types - first dose, second dose (fully vaccinated), and booster. The models encompass the three phases of the COVID-19 pandemic in the United States during this timeframe including a portion of the errant COVID 19 "eradication phase" as the Baseline phase, the Delta variant phase, and the Omicron variant phase. The regression research approach uniquely incorporates positive and negative sentiment analysis of virus fear and vaccine hesitancy from Twitter tweets supplemented by CDC data to predict future CDC first dose, second dose, and booster inoculations.

Further, the research indicates the degree to which different fears, factors and levels impact daily vaccine inoculation count by segmenting the pandemic timeframe into phases based on CDC VOC announcements. The research revealed each phase impacted the vaccine inoculation models based on phase characteristics and events and public response to those characteristics and events. Inconsistent correlations between traditional indicators of fear of the virus and traditional indicators of fear of the vaccine were also accompanied by rapid changes in vaccine inoculation trend. Likewise first, second, and booster dose inoculations perceived value vs risks during a phase impacted vaccine inoculation models. In order to more accurately quantify and predict inoculation trends, the research extends Twitter sentiment analysis by classification and quantification of the nature and strength of association between opinions on Twitter and daily vaccination and inoculation spikes. Overall, the regression models provide the means for predicting first, second, and booster dose daily vaccination inoculation in the USA. Regression results are consistent with KFF vaccination opinions surveys and with the Technology Acceptance Curve (Rogers, 1995). Comments on limitations and future research goals are provided.

Materials and Methods

To predict vaccine inoculations in light of fear of the virus and vaccine hesitancy, this research methodology used three primary US population datasets. First, the Center of Disease Control (CDC) identifies virus and variant threats relevant to the U.S. population through virus alerts and subsequent Variant of Concern (VOC) alerts ("Coronavirus disease 2019 (COVID-19)," 2022). The FDA in coordination with the CDC approves vaccines and associate vaccine inoculation guidelines for state and local health agencies. (Disease Prevention & Control - San Francisco Department of Public Health, 2021.). The CDC reports daily virus cases (a traditional measure of the level of threat of becoming sick from the virus), virus deaths (a traditional measure of the level of threat of dying from the virus), and inoculations (a traditional measure of dose acceptance) (CDC, 2020). The CDC also identifies COVID-19 Treatments and Medications used to mitigate virus effects but do not report daily outcomes (CDC, 2023). In terms of vaccine side effects, the CDC also reports “Selected Adverse Events Reported after COVID-19 Vaccinations” but the reports are not a daily occurrence, have significant latency between events and reports, and do not claim to represent a collection of all adverse events (CDC, 2023). As a supplemental measure to account for CDC virus mitigation and vaccine side effect reporting limitations and as demonstrated previously by (Daghriri, Proctor, and Matthews, 2022), Sentiment analysis of Twitter users’ geolocated tweets , while limited (Stechemesser et al., 2022; Sattar & Arifuzzaman, 2021),may be used to identify levels of fear of the virus and levels of vaccine hesitancy in a given population during a pandemic.

Within the June 1, 2022 to March 31, 2022 time frame, data collection focuses on three pandemic phases determined by CDC virus and VOC alerts ("SARS-Cov-2 B.1.617.2 (Delta) variant COVID-19 outbreak ..," 2021; "Coronavirus disease 2019 (COVID-19)," 2022) (Figure 2). The Baseline phase encompasses June 1-15 2021 and represents the state of fear caused by the COVID-19 virus and as mitigated by existing vaccines just prior to the CDC alerting the public on June 15 2022 of the Delta VOC. The Delta variant phase follows the Delta VOC alert and spans the period from June 16th to CDC Omicron VOC issued on November 30, 2022. The Omicron variant phase follows the Omicron VOC alert on December 1, 2021 until the end of the research study on March 31, 2022.

Figure 2. Pandemic Phases within the Research timeframe.

Figure 3 shows the relationships between the datasets used to conduct the research and analysis. Shown on the left side of the figure, the CDC issued virus and VOC alerts while the FDA approved three vaccine doses with the first dose and the second dose available for inoculation throughout the entire research window. The booster dose became available for general inoculations on Sep 22, 2021. Also shown in the center of the figure, the

CDC reported daily and a running total of vaccine inoculations, virus cases, and deaths. To complete the right side of Figure 3, Twitter datasets provide positive, negative, and neutral tweets about the virus and the vaccine. Twitter data extraction is discussed in more detail below. Finally, the bottom of the figure shows that KFF vaccinations survey outputs and Rogers Technology Acceptance Curve population segment descriptions provided 3rd party analysis and benchmarks of public opinions and traditional behaviors that may relate to the research.

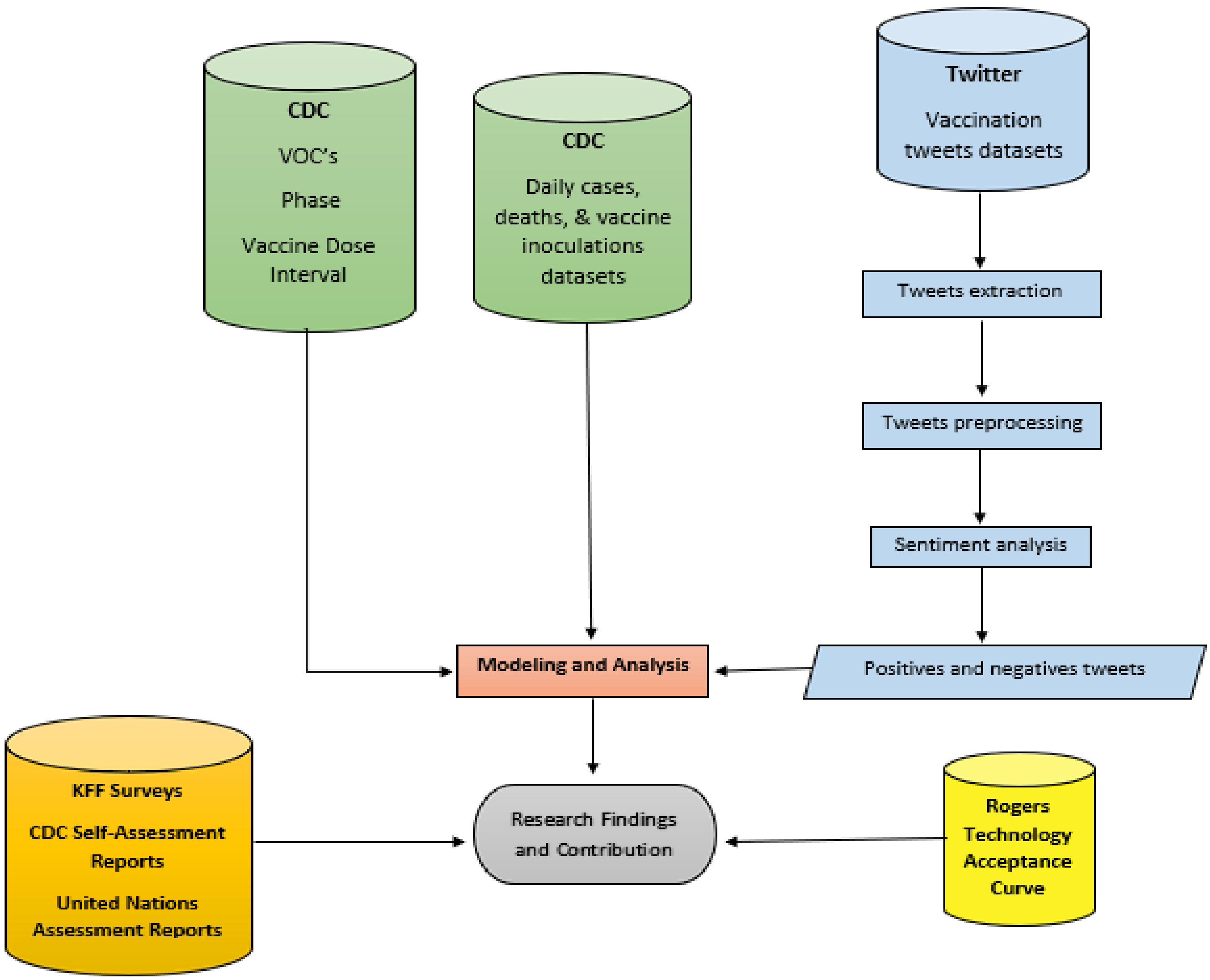


Figure 3: Research Methodology

- CDC Variant of Concern announcement establish phases (Baseline, Delta, Omicron)
- CDC Vaccine approvals establish Dose Intervals (First Dose, Second Dose, Booster Dose)
- CDC Case, Deaths, & Inoculation Data Collection and Segmentation

- Twitter Data Collection and Sentiment Analysis

- Linear Regression models
- Correlation Analysis
- Lag Time Analysis

Twitter Data Extraction:

Twitter data may be extracted based on user location and demographics that exists on their profiles. Data extraction from Twitter focused on United States users and their tweets about inoculation and vaccine hesitancy for the three COVID-19 vaccine types available (Pfizer, Johnson & Johnson, and Moderna). Users' tweets were extracted from the Twitter feed using the technique described in Table 1 below and demonstrated by (Pokharel, 2020; Morshed, et al. 2021). We built and coded our Tweepy library crawler following (Morshed et al., 2021) approach to extract tweets containing words identified in Table 2. "Tweepy is an open-source Python package that gives one a very convenient way to access the Twitter Application Programming Interface (API) with Python" (Python, 2019) "in order to compose tweets, read profiles, and access your followers' data and a high volume of tweets on particular subjects in specific locations" (Fontanella, 2021).

Step	Description
Vaccination tweets extraction	Use of the technique and the Tweepy library to extract tweets from the Twitter API relevant to vaccinations (Mushtaq et al., 2022; Morshed et al., 2021). Using the vaccines keywords to extract the tweets that include then by using Tweepy library. In addition, using the geolocation features to extract tweets that were posted by the USA users. Public tweets extracted from Tweeter via its Application Programming Interface (API) was used for the experiment.
Tweets preprocessing	the Retweets and URLs were removed in the preprocessing step, emojis were converted into words, and the dataset was cleaned. We also removed stop words and performed tokenization. Stemming and lemmatization were done as well. (Ramachandran & Parvathi, 2019; Aljedaani et al., 2020;Monkey, 2020; Rsutam et al, 2021)
Sentiment analysis	Using Vader library to classify tweets into positive, neutral, and negatives. (Devika et al., 2016; Yadav & Vishwakarma, 2019; Dang et al., 2020; Pipis 2022) where tweets classifications are defined as following: - **Positive tweets** represent the opinions that are favor and support accepting the vaccines. - **Negative tweets** represent the opinions that are against and hesitating and rejecting the vaccines. - **Neutral tweets** represent the opinions that are not favor or against the vaccines, where tweets can't represent positive neither negative opinions.

Table 1. Twitter Data Extraction, Processing, and Sentiment analyzing Steps.

Vaccine	Keywords	Timeframe
Pfizer-BioNTech vaccine	Pfizer, Pfizer-BioNTech, BioNTechpfizer, vaccine, vaccination, dose	June 1st, 2021 – 31 March 2022
Johnson & Johnson's COVID-19 Vaccine	Johnson & Johnson, Johnson and Johnson, Janssen, Janssen, vaccine, vaccination, dose	June 1st, 2021 – 31 March 2022
Moderna vaccine	Moderna, Moderna_tx, Moderna-NIAID, NIAID, NIAID-Moderna, vaccine, vaccination, dose	June 1st, 2021 – 31 March, 2022

Table 2. Vaccination keywords used to extract the Twitter datasets.

Modeling construction:

Modeling involved correlations and linear regressions to predict daily first dose, second dose, and booster vaccination inoculations. Regression independent terms included a constant associated with each dose type and variables associated with phase, daily virus cases, daily deaths, daily positive tweets, and daily negative tweets on day x. Respective regression equation outputs included first dose vaccination inoculation for day x+1, second dose vaccination inoculation for day x + 1, or Booster dose vaccination inoculation for day x + 1.

Results

CDC Results and Analysis:

As context to the research, on May 31, 2021, the day prior to the start of this research, the CDC reported that 173,531,874 (52.3% of a 331,800,000 population) had received first dose inoculations. 146,813,131 (44.2%) had received a second dose inoculation, making them "fully vaccinated" at the time. Since the Booster shot had NOT been authorized for the general public on May 31, 2021, only 9,318 people had received the Booster representing 0% of the population.

The graphic displays in Figures 4, 5, and 6 summarize daily virus cases and death counts and respectively first, second, and booster doses reported by the CDC during the duration of the research. Beneath the graphic display of daily CDC data reported, the colored bar indicates the three phases of the pandemic identified by the CDC through their initial COVID 19 virus alert and subsequent Delta VOC and Omicron VOC alerts as discussed in Methods and Material section above.

Within each figure, a red dot indicates a trough inflection point and a green dot indicated a peak inflection point for a given curve. Common to all three figures, are two virus case peaks that precede lagging death count peaks, corresponding respectively to the Delta and Omicron phases. Inoculations ups and downs for a given dose infers, respectively, rising level of virus fear or rising level of vaccine hesitancy among the remaining populations for each dose type.

Common to all three inoculation curves is an overall decline toward zero for new inoculations at the end of the research timeframe. At the conclusion of the research, the CDC reported 256,144,043 (77.1%) first dose, 219,319,838 (66.1%) second dose, and 100,230,127 (45.7%) booster inoculations. The declining inoculation percentages with each dose type as well as the failure of any of the dose types to exceed 78% coverage raises the notion of population segments with different levels of acceptance of new technology (e.g. different vaccine doses) as identified in the Technology Acceptance Curve (Rogers, 1995), where acceptance would be expressed as a sentiment. The variability in virus fear and vaccine hesitancy for each inoculation type (first, second, booster) varies significantly and is discussed in the next section in terms of correlations with objective virus case and death count measures. The nature and degree of change in inoculation behavior attributed to changing social sentiments are deferred to the Twitter Sentiment Results and Analysis section below.

Correlations between Inoculations and Virus Cases and Death Counts:

A visual inspection of Figure 4 indicates a rapid drop in First Dose inoculations, virus cases, and virus deaths during the Baseline phase, at that time supporting the "eradication" theory. For the Baseline phase, correlation analysis confirmed large correlations between declining First Dose inoculations and declining death counts (0.886) and declining virus cases (0.766). In terms of the two theoretical premises, the correlations support the notion that the overall declining fear of sickness from the virus and declining fear of death from the virus resulted in declining inoculations among the remaining undosed population.

Despite the June 15, 2021 CDC Delta VOC alert and contrary to the theoretical expectation of an increase in the fear of the virus a VOC might cause, the public appeared to ignore the VOC as inoculations continued to drop even after the VOC. Inoculations also continued to drop inversely to rising virus cases (-0.872), until the inoculation trend reversed at an inflection point 23 days (July 8) into the Delta phase coincident with, not preceded by, an increase in the death count. As the VOC alert, the long increasing virus cases, and the existence of a preceding increase in death count do not appear causal to the change in inoculation trend, trend change is discussed below in Twitter Sentiment Results and Analysis section. Between the inoculation trough and subsequent July 8, 2022 peak, large correlations were observed between rising first dose inoculations and rising virus cases (0.987) as well as rising death counts (0.928). After the first peak and as also discussed in the Twitter Sentiment section below, inoculation behavior, inconsistently correlated with virus cases and death count, peaked before the Thanksgiving Holiday and again before the Christmas Holidays.

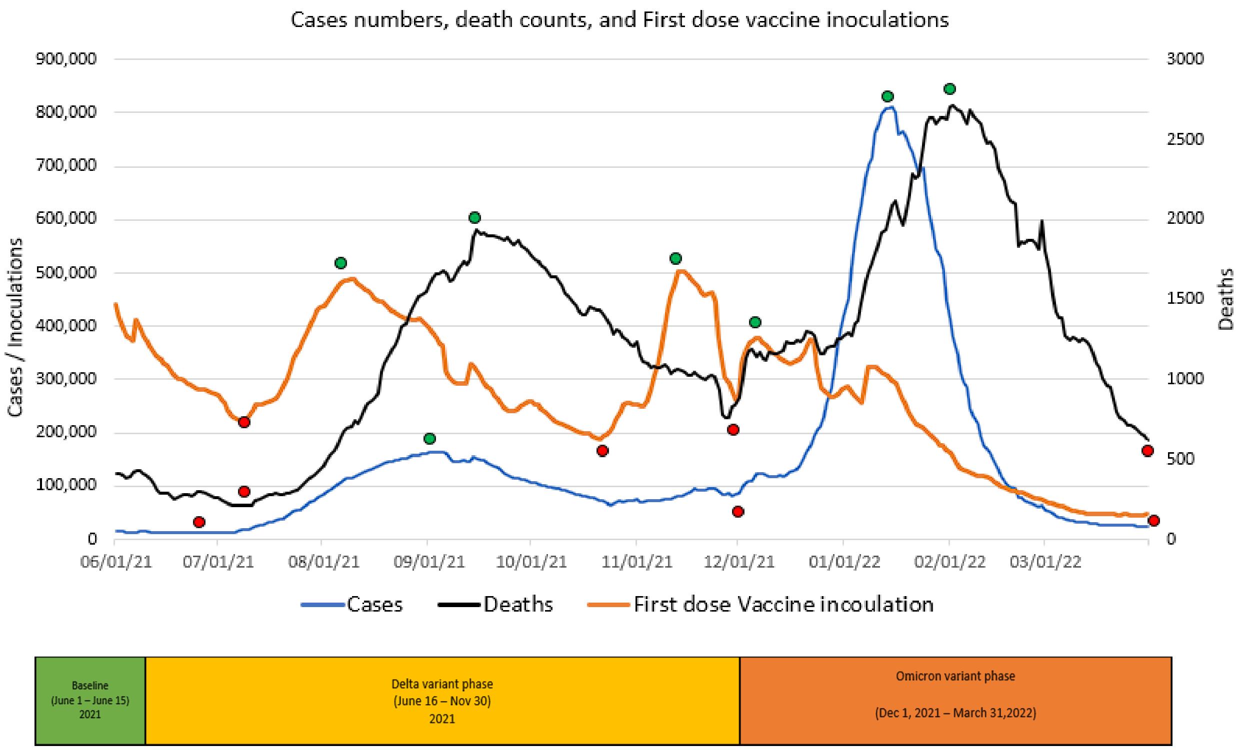


Figure 4. First Dose vaccine inoculations versus Virus Daily Cases.

A visual inspection of Figure 5 reveals a Second Dose peak of 758,476 on June 9, 2021 in the midst of the Baseline phase when virus cases and death count were both going down. This inverse correlation during the first part of this phase infers that a large segment of the population sought to be "fully vaccinated" despite dropping virus cases and death counts. After the initial inoculation peak, Second Dose inoculations fell until July 29, 2021, even though this was well past the June 18, 2021 virus trough and subsequent increase in virus cases. The inversely correlated precipitous drop in inoculations with rising virus cases potentially manifests higher levels of vaccine hesitancy among the remaining unvaccinated population segments. From July 29, 2021 to the August 31, 2021 inoculation peak, rising Second Dose inoculations were highly correlated (0.943) with rising virus cases. After the August 31, 2021 peak, Second Dose inoculations varied due to sentiment factors discussed below.

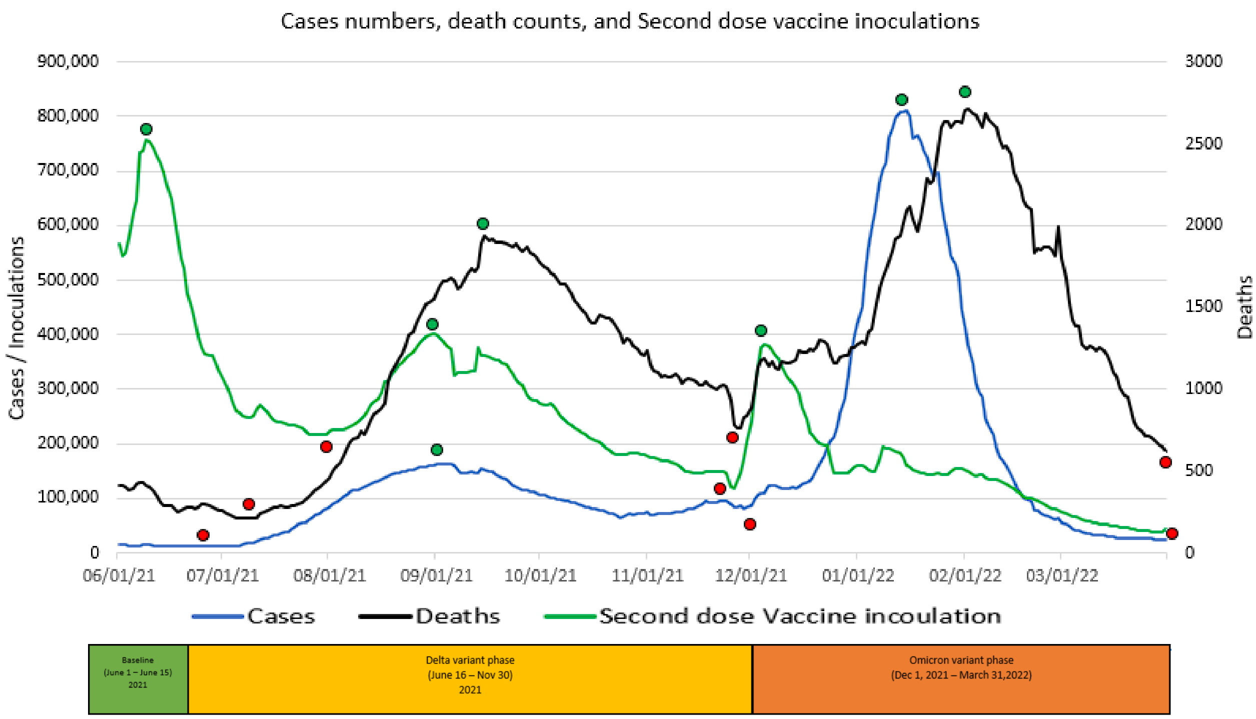


Figure 5. Second Dose vaccine inoculations versus Virus Daily Cases.

A visual inspection of Figure 6 reveals the vast majority of the Booster Dose inoculations occurred entirely within the research timeframe and experienced multiple peaks with the highest daily peak of 1,078,908 Booster Dose inoculations occuring on December 7, 2021 alone. Factors driving Booster Dose inoculations are discussed in the Sentiment section below.

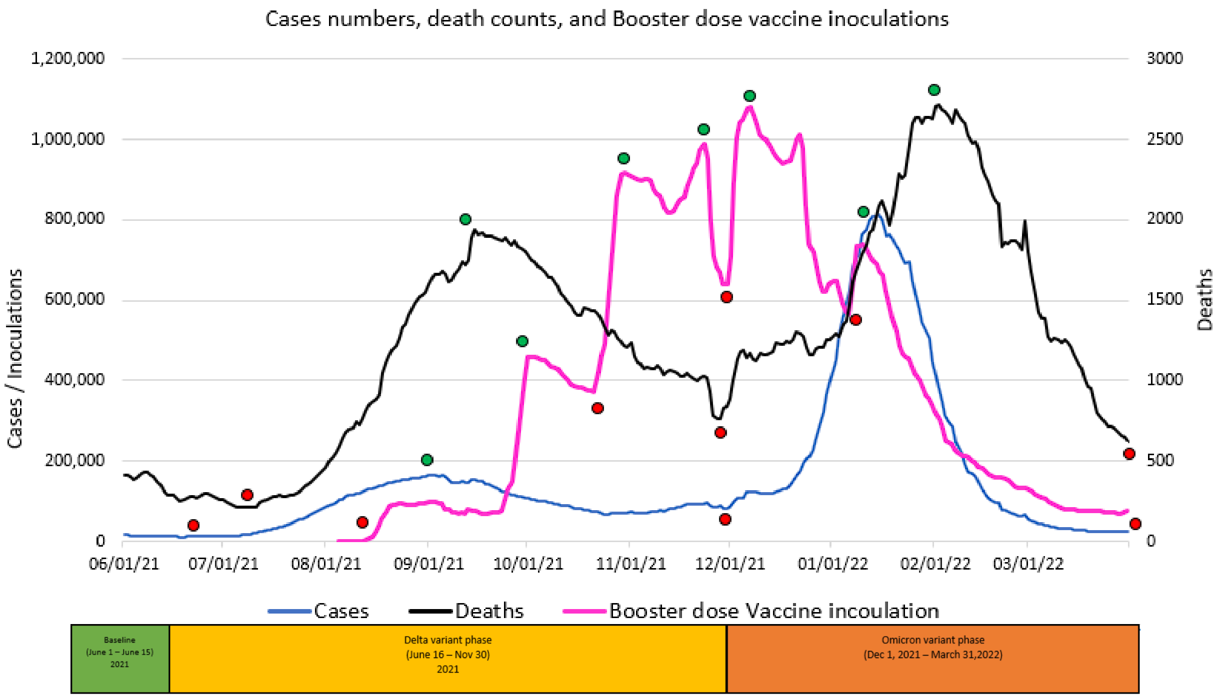


Figure 6. Booster Dose vaccine inoculations versus Virus Daily Cases.

Twitter Sentiment Results and Analysis:

over the entire research time frame, 949,529 tweets have been classified sentimentally. Sentiment analysis identified 326,124 tweets that indicated a positive viewpoint toward vaccines and 163,716 tweets that indicated a negative viewpoint toward vaccines, while 459,689 tweets indicted neutral sentiments toward vaccines.

Visual analysis of Figure 7 appears to show, and analysis confirms large positive correlations between Twitter positive tweets and First Dose inoculations across the entire Delta variant phase (0.607) and across the entire Omicron variant phase (0.804). Catalysts for tweet activity include CDC, FDA, and other Biden administration announcement events. Beginning in July 2021 positive tweets increased slowly until the end of August 2021 when there was a significant jump in positive tweets. Positives tweet levels spiked again at Sep 2021 with approval events for the Pfizer booster dose for regular use, and again in Oct 2021 coincident with approval events associated with the Moderna booster and Children first dose vaccination approval. Negative tweets cited recent press reports on vaccines side effects, which likely increased the level of vaccine hesitancy sentiment. None the less, paradoxically, a large correlation of 0.794 was observed between negative tweets and First Dose inoculations during the Omicron phase.

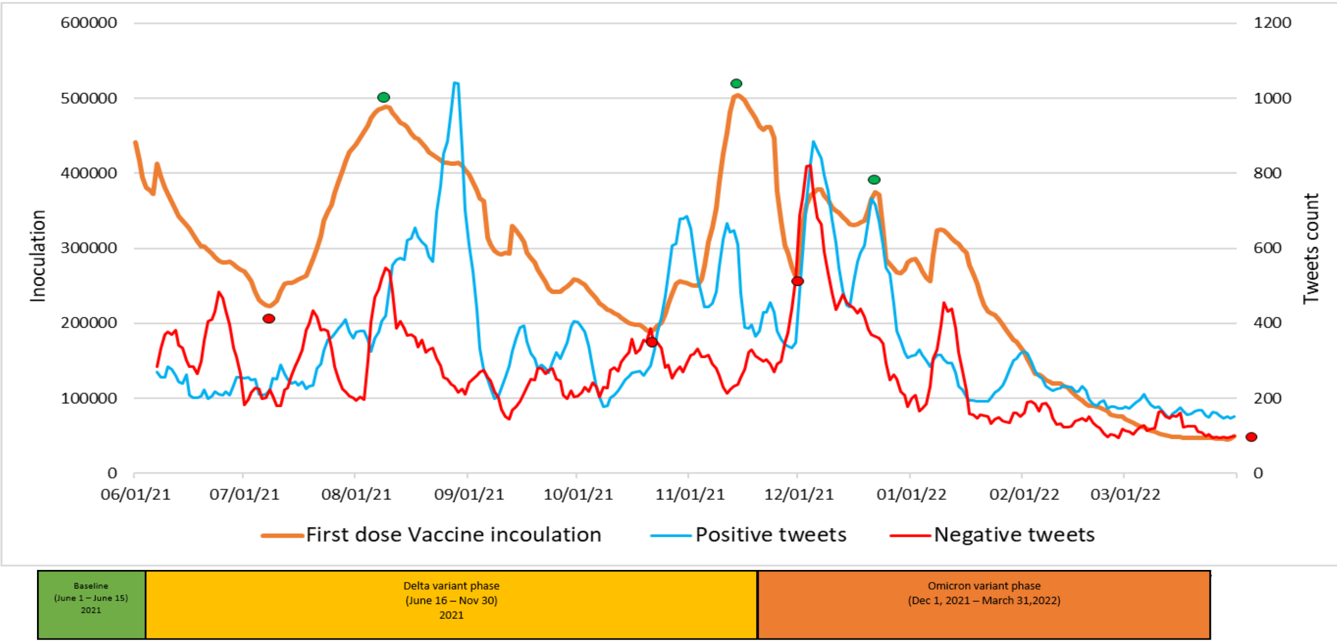


Figure 7. Positive and Negative Tweets toward First dose vaccines and vaccinations over time.

Visual analysis of Figures 8 also appears to show and analysis confirms a large positive correlation (0.870) between positive tweets and Second Dose inoculations during the Omicron phase. Paradoxically, a large correlation of 0.892 was observed between negative tweets and Second Dose inoculations during the Omicron phase.

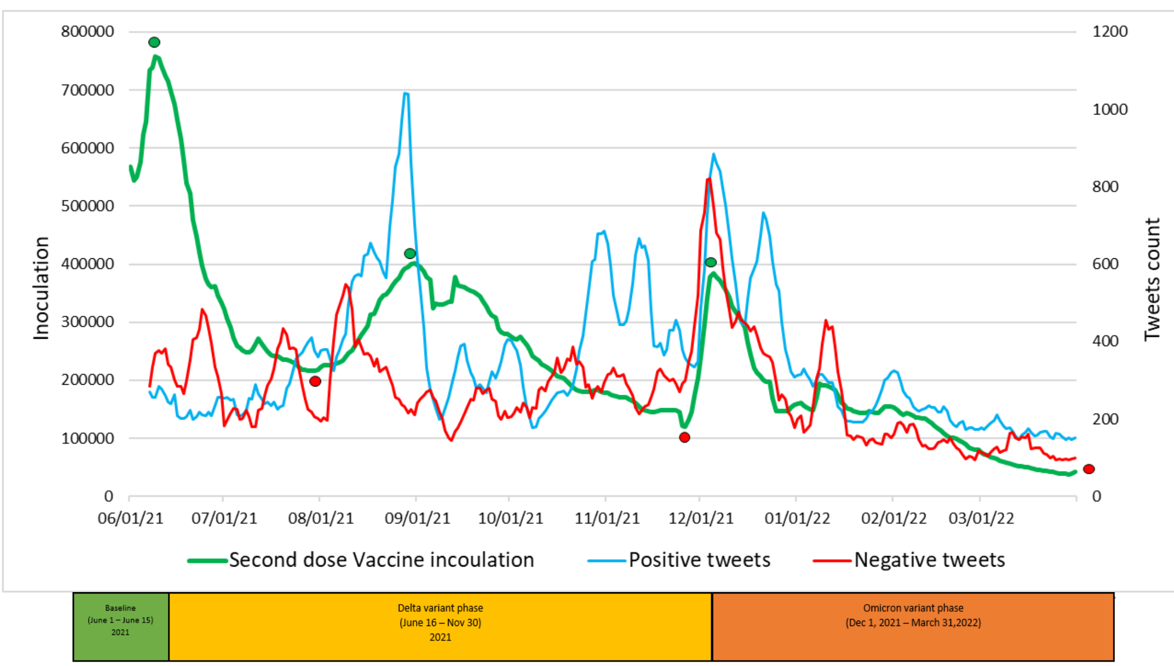


Figure 8. Positive and Negative Tweets toward Second dose vaccines and vaccinations over time.

Visual analysis of Figures 9 also appears to show and analysis confirms a large positive correlation between positive tweets and booster dose inoculations during a portion of the Delta phase (0.768) and across the entire Omicron phase (0.868). Paradoxically, a large correlation of 0.845 was observed between negative tweets and Booster inoculations during the Omicron phase.

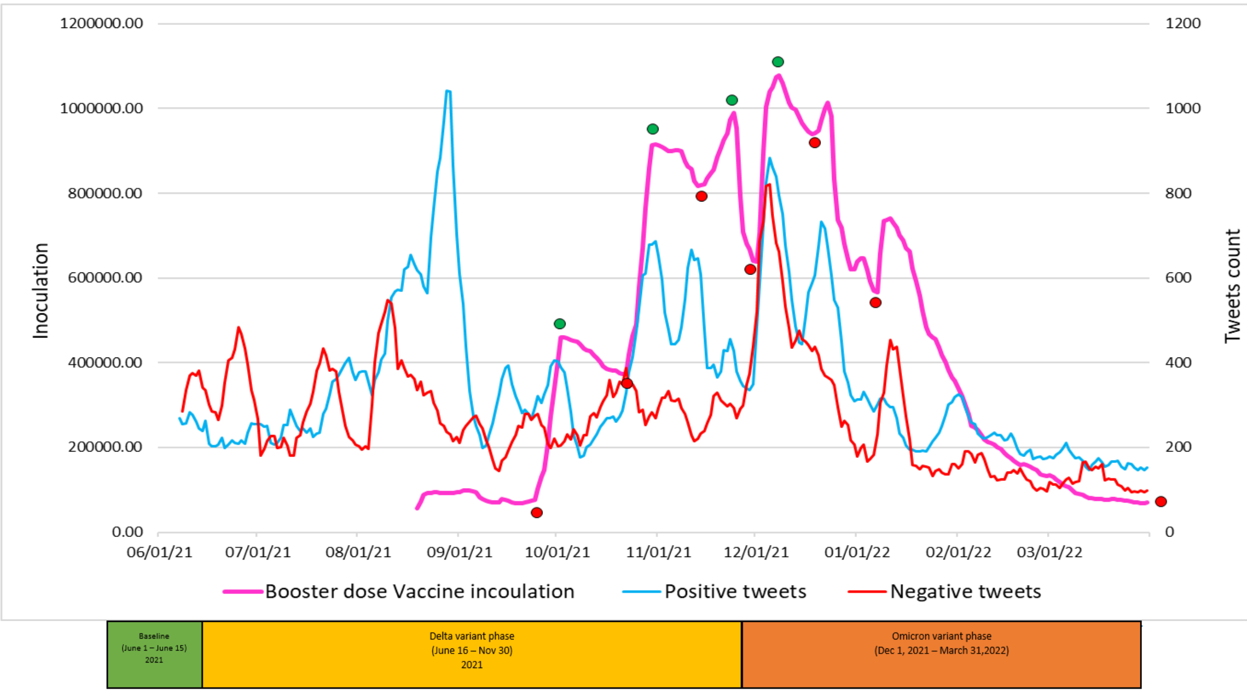


Figure 9. Positive and Negative Tweets toward Booster dose vaccines and vaccinations over time.

Regression models outcomes

As indicated in the Methods section, regression models for predicting daily first dose, second dose, and booster vaccination inoculations quantify independent terms and their coefficients. Terms include a constant associated with each dose type and variables associated with phase, daily virus cases, daily deaths, daily positive tweets, and daily negative tweets on day x. The respective regression equation output first dose vaccination inoculation for day x+1 (Figure 10), second dose vaccination inoculation for day x + 1 (Figure 11), or Booster dose vaccination inoculation for day x + 1 (Figure 12).

First (Figure 10) and Booster Dose (Figure 12) inoculation models both had large predictive R-squares of 77.47% and 84.45% respectively. First Dose and Booster Dose inoculation models had in common: all factors of interest, each factor was statistically significant in predicting daily inoculations in both models, and each corresponding factor was directionally the same in each model. Death counts for both inoculation models actually had a negative impact on inoculation counts, inferring death counts were a lagging indicator. Differences between the First and

Booster inoculation models emerge with sentiment variable coefficients. Combined sentiment variables (Positive and Negative Tweets) impact Booster Dose inoculation model 3.23 times as great as the same variable impacts the First Dose model. Of note, during the research timeframe, CDC and Biden administration announcement events occurred more often in association with the First and Booster Doses, with only one event associated with the Second Dose (Figure 11). Tweets increase around such events with positive tweet spikes observed in association with First and Booster Dose events. Negative tweets increased with smaller spikes than positive tweets during events except during the Baseline phase, where the negative tweets were at the same level of positive tweets.

Regression Equation

First dose Vaccination numbers = 396886 - 113202 Phases + 0.2849 Cases - 50.57 Deaths + 327.7 positive tweets numbers + 165.3 Negative tweets numbers

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	394879	19871	19.87	0.000	
Phases	-112483	7475	-15.05	0.000	1.44
Cases	0.2845	0.0227	12.55	0.000	1.62
Deaths	-50.54	6.99	-7.23	0.000	1.91
positive tweets numbers	328.4	21.5	15.26	0.000	1.39
Negative tweets numbers	165.0	33.2	4.97	0.000	1.61

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
58205.8	78.25%	77.87%	77.47%

Figure 10. First dose prediction model. P values are significant for all five x variables.

The Second Dose inoculation regression model (Figure 11) had a moderate predictive R-square of 53.82%. Of immediate note between the three regression equations is the 31.4% and 44.8% respective differences between the 396,886 First and 360,234 Booster Dose equation constants and the 521,548 Second Dose constant. The significantly higher Second Dose constant highlights the fore mentioned importance of being "fully vaccinated" for significant segments of the population. Being "fully vaccinated" drives behavior for this segment of the population even to the point of getting the Second Dose in the Baseline phase despite numerous, although later proven false, indications of a waning pandemic. Similarly, the much larger negative coefficient of the "phase" variable in the Second dose equation infers a rapid drop off in inoculations across phases highlighting this population segment's urgency during the Baseline phase in completing the series. Interestingly, neither death count nor positive Tweets had a significant impact of Second Dose inoculations. Cases made a modest contribution toward inoculations. Negative Tweets paradoxically made a significant contribution to inoculation, inferring resoluteness of or perhaps even indifference or even defiance toward negative social media by this population segment for getting the Second Dose and becoming "fully vaccinated".

Regression Equation

Second dose Vaccination numbers = 521548 - 166129 Phases + 0.0974 Cases + 329.9 Negative tweets numbers

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	521548	30273	17.23	0.000	
Phases	-166129	11339	-14.65	0.000	1.32
Cases	0.0974	0.0318	3.06	0.002	1.27
Negative tweets numbers	329.9	42.6	7.74	0.000	1.05

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
92271.8	55.06%	54.60%	53.82%

Figure 11. Second dose prediction model. P values are significant for all three variables except the positive tweets and deaths variables which have been removed as their P. values were not significant.

For the booster dose model, the positive tweets are three times negative tweets. Drivers of the differences were: Pfizer booster approval in Sep 2021 and Moderna & J.J booster dose approvals in Oct 2021. In addition, the booster dose interval constant was lower in part due to the booster never being authorized for children but also indicates a lower inclination on the part of subjects to get the booster than observed with either the first dose or second dose.

Regression Equation

Booster dose Vaccination number = 360234 - 96629 Phases + 0.6428 Cases - 166.9 Deaths + 1221.1 positive tweets numbers + 378.3 Negative tweets numbers

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	360234	68244	5.28	0.000	
Phases	-96629	20638	-4.68	0.000	1.21
Cases	0.6428	0.0520	12.35	0.000	1.55
Deaths	-166.9	21.1	-7.91	0.000	1.56
positive tweets numbers	1221.1	77.8	15.70	0.000	2.24
Negative tweets numbers	378.3	97.3	3.89	0.000	2.25

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
127707	85.29%	84.90%	84.45%

Figure 12. Booster dose prediction model. P values are significant for all five variables.

Discussion

Kaiser Family Foundation Surveys & CDC Self-Assessment Report

Results from Kaiser Family Foundation (KFF) surveys indicate major reasons for getting the COVID-19 vaccine during our research timeframe were: increase in cases due to the Delta variant (39%), concern about hospitals filling up (38%), and knowing someone that got seriously ill or died from COVID-19 (36%). These findings appear to generally support the regression factors common to the First Dose and Booster Dose regression models and contribute to their large predictive R-square ("Surging delta variant cases, hospitalizations, and deaths are biggest drivers of recent Uptick in U.S. COVID-19 vaccination rates," 2021). For this research, inconsistencies are most notable between KFF and the moderately predictive, Second Dose regression model. Specifically and most notable is the absence of the death factor from the list of Second Dose regression factors. The moderate, rather than large, predictive R-square of the Second Dose regression model is likely an artificiality of the limitations of the research timeframe. Specifically, for the large predictive R-squares for the First Dose and Booster Dose regression models, the vast majority or the entire total of inoculations taken within our timeframe, occurred outside the Baseline phase. For the Second Dose regression model a much larger percentage of Second Dose inoculations taken during our timeframe occurred during the Baseline phase. Thus motivation before the Baseline Phase are more likely to have driven Second Dose inoculation behavior than driven either First Dose or Booster Dose inoculations.

The CDC has done numerous self-assessment reports of their role in the COVID-19 pandemic. (Meng, et al 2022) studied the Second Dose phenomenon and in part found behavior differed by population segments driven largely by age. Specifically, "Compared with first-dose recipients 18–39 years of age, recipients 40–64 and >65 years of age were less likely to have missed a second dose. Persons in older age groups had more time to complete their primary series, given the prioritization when COVID-19 vaccine first became available. Older adults also are at higher risk for severe COVID-19 illness and may have been more motivated to become fully vaccinated (14,15)." If Meng, et al findings are accurate, his findings support the notion that the most likely factor adversely impacting the Second Dose regression model resulting in a moderate rather than large predictive R-square was older segments of the populations disproportionately getting inoculations during the Baseline phase. As Meng et al also further indicate, older population segments may have been driven to become "fully vaccinated" with the Second Dose by fear of "severe illness" or death. Since the pandemic began in the United States in March 2020 and without knowledge of the Delta and Omicron VOCs , this fear would have been accumulated well before the Baseline phase and before our research timeframe. Future research that included data reaching back to March 2020 would likely significantly change Second Dose regression model factors and weights to be more in line with First Dose and Booster Dose regression models. Additionally, future research may improve the predictive R-square of the regression equations by including a variable for population segments, possibly identified by age and other demographics that may impact vaccine acceptance.

A KFF survey also solicited feedback on sentiment about vaccine side effect concerns and the resulting vaccine hesitancy (Personal Concerns About COVID-19 Vaccination) https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-february-2021/. For the Booster dose, KFF identified significantly different sentiments among vaccinated and unvaccinated population segments. Specifically, 78% of the vaccinated respondents indicated that the Booster dose "shows that scientist are continuing to find ways to make vaccines more effective" https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-september-2021/ . In contrast, 71% of the unvaccinated respondents indicated that the Booster dose, "shows that the vaccines are not working as well as promised." KFF also indicated that the Omicron variant only motivated about 12% of the unvaccinated to get their first dose while 87% remained unconvinced. ("KFF COVID-19 vaccine monitor: Early omicron update," 2021). In contrast, among vaccinated adults who had not gotten a booster, 54% indicated the Omicron variant made it “more likely” to “get a booster shot” while 46% disagreed.

Considering Vaccine Acceptance rates in light of Rogers Technology Acceptance Curve

Given the fore mentioned segmentation of the population, this research would be remiss not to acknowledge similarities between population acceptance of and hesitancy toward vaccines and Rogers Technology Acceptance Curve. Specifically, the Rogers Technology Acceptance Curve identifies five segments the US population and characterizes 2.5% as Innovators in accepting technology, 13.5% as the Early Adopters, 34% as the Early Majority, 34% as the late Majority, and 16% as Laggards. Laggard traits include skeptical, resistance to change, and wary of accepting new technology. Assuming each inoculation is a new technology experiencing the Rogers estimates of acceptance, laggards would represent that segment of the population predisposed toward vaccine hesitancy. This assumption infers that approximately 84% of the population might voluntarily accept a First Dose inoculation within the timeframe of the pandemic. As of Feb 7, 2023 and remarkably consistent with the Rogers estimate of 84% acceptance, the CDC reports 85.5% of the U.S. population 5 years of age or greater are inoculated with one dose (CDC, 2023). If a Second Dose inoculations is viewed as another voluntary acceptance challenge, then Rogers estimation of 84% of the First Dose recipients would yield an estimate of 70.6% of the total population will receive a Second Dose inoculation. As of Feb 7, 2023 and remarkably consistent with the Rogers estimate of 70.6% acceptance, the CDC reports 73.2% of the U.S. population 5 years of age or greater are inoculated with one dose (CDC, 2023). The respective 1.5% and 2.6% higher observed inoculation rates over that estimated by Rogers Curve may be due to vaccination mandates imposed by government and/or employers. Unfortunately Rogers statistical booster dose estimation is confounded by replacement of the original booster with the updated (bivalent) booster dose. Further, as of Feb 7, 2023, the CDC only reports updated (Bivalent) Booster data, not the original booster data discussed herein. Using the March 31, 2022 CDC reported 100,230,127 (45.7%) original booster inoculations and applying Rogers 84% acceptance rate to the 70.5% who actually accepted the Second Dose inoculations yields an expected 59.3% acceptance rate for Booster inoculations among the total population. Clearly the observed 45.7% acceptance rate is well below the 59.3% estimated rate. Future research may reveal the reasons for the inoculation shortfall but the shortfall is likely due to the timeframe limitations of the experiment or timeframe limitations due to replacement of the booster with the updated booster but may also be due to rising vaccine hesitancy or fewer mandates.

While we believe our sentiment analysis approach is sound and the results accurate to the degree cited above, manipulation by government agencies or large corporations may have an impact on social media content and may therefore limit sentiment analysis where such manipulation occurs. Social media content impacts sentiment. Revelation of content manipulation will likely undermine confidence in content found on social media and thereby undermine the value of social media sentiment analysis. As an example, with the acquisition of Twitter by Elon Musk a number of independent journalists investigated governmental and industrial manipulation of social media and released reports termed the "Twitter Files". Among those investigations, David Zweig released the 40-tweet Twitter Files report titled, "How Twitter Rigged the Covid Debate" (Mills, 2022). Zweig identified in that report "that both the Biden and Trump administrations pressured Twitter and other social-media platforms to elevate content that fit their narratives and to suppress information that didn't". One of the more important and concerning finding was actual interference in free speech by silencing of Alex Berenson, a critic of the Biden administration COVID policies. Specifically, "Berenson's Twitter account was suspended hours after Biden alleged that social-media companies were "killing people" for allowing vaccine misinformation. Berenson later sued and eventually settled with Twitter." Further, Lee Fang revealed how "the pharmaceutical industry lobbied social media to shape content" related to the COVID vaccine (Flood, 2023). Even Pfizer board member Dr. Scott Gottlieb flagged tweets questioning COVID vaccines (Wulsohn, 2023). An August 2021 email Gottlieb sent to Twitter's senior public policy manager Todd O'Boyle flagging a tweet written by former Trump administration official Dr. Brett Giroir is but one example. Giroir had written "It's now clear #COVID19 natural immunity is superior to #vaccine immunity, by ALOT. There's no scientific justification for #vax proof if a person had prior infection." "This is the kind of stuff that's corrosive," Gottlieb told O'Boyle. "Here he draws a sweeping conclusion off a single retrospective study in Israel that hasn't been peer reviewed. But this tweet will end up going viral and driving news coverage." According to Berenson, O'Boyle forwarded Gottlieb's email to Twitter's "Strategist Response" team, writing "Please see this report from the former FDA commissioner." Giroir's tweet was later slapped with a "misleading" label and blocked any ability to like or share the tweet, telling Twitter users "Learn why health officials recommend a vaccine for most people."

Besides directed manipulation of social media content by governments and industry, this research revealed that CDC and Biden administration announcements and events precipitated considerable Twitter tweets. The nature of the tweet response directly impacted the regression equations. Thus two recent United Nations reports stating "scientist in China, the US and the UK have been accused of deliberately covering up the origins of the coronavirus outbreak" (King, 2023) may also undermine social media sentiment analysis in the future by creating further distrust. Epidemiologists Colin Butler, from the National Centre for Epidemiology and Population Health in Canberra, Australia, and Delia Randolph, from the University of Greenwich in London were responsible for the reports. Both concluded that high-risk experiments being carried out in the Chinese city of Wuhan were shrouded in a cloak of suspicious secrecy, deception, and conflicts of interest. They argued that this was 'implemented not only by China but also by Western funding agencies and influential Western scientists'.

While the sentiment analysis in this research did not consider the undermining of the confidence of the fair, accurate, and equitable publishing of social media content, future research must consider how potential social media content manipulation impacts behavior.

References

https://www.forbes.com/sites/forrester/2020/04/09/four-phases-of-the-coronavirus-pandemic/

SV, P., Lorenz, J. M., Ittamalla, R., Dhama, K., Chakraborty, C., Kumar, D. V., & Mohan, T. (2022). Twitter-based sentiment analysis and topic modeling of social media posts using natural language processing, to understand people's perspectives regarding COVID-19 booster vaccine shots in India: Crucial to expanding vaccination coverage. *Vaccines*, *10*(11), 1929. https://doi.org/10.3390/vaccines10111929

Twitter-based sentiment analysis and topic modeling of social media posts using natural language processing, to understand people's perspectives regarding COVID-19 booster vaccine shots in India: Crucial to expanding vaccination coverage. (2022, November 15). MDPI. https://www.mdpi.com/2076-393X/10/11/1929/htm

Evolution of select epidemiological modeling and the rise of population sentiment analysis: A literature review and COVID-19 sentiment illustration. (2022, March 9). MDPI. https://www.mdpi.com/1660-4601/19/6/3230

Sekizawa, Y., Hashimoto, S., Denda, K., Ochi, S., & So, M. (2022). Association between COVID-19 vaccine hesitancy and generalized trust, depression, generalized anxiety, and fear of COVID-19. *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-021-12479-w

Piedrahita-Valdés, H., Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J. R., Guillem-Saiz, J., Sicilia-Montalvo, J. A., & Machío-Regidor, F. (2021). Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines*, *9*(1), 28. https://doi.org/10.3390/vaccines9010028

Coronavirus disease 2019 (COVID-19). (2022, August 31). Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/variants/index.html

(n.d.). Disease Prevention & Control - San Francisco Department of Public Health. https://www.sfcdcp.org/wp-content/uploads/2021/12/Health-Alert-SARS-CoV-2-Virus-Variant-of-Concern-B.1.1.529-Omicron-SFDPH-FINAL-12.10.2021.pdf

CDC. (2020, March 28). COVID data tracker. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#datatracker-home

SARS-Cov-2 B.1.617.2 (Delta) variant COVID-19 outbreak ... (2021, July 14). Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/70/wr/mm7028e2.htm

Aljedaani, W., Javed, Y., & Alenezi, M. (2020). LDA categorization of security bug reports in chromium projects. Proceedings of the 2020 European Symposium on Software Engineering. https://doi.org/10.1145/3393822.3432335

VaderSentiment. (2022). PyPI. https://pypi.org/project/vaderSentiment/

Yadav, A., & Vishwakarma, D. K. (2019). Sentiment analysis using deep learning architectures: A review. Artificial Intelligence Review, 53(6), 4335-4385. https://doi.org/10.1007/s10462-019-09794-5

Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. Electronics, 9(3), 483. https://doi.org/10.3390/electronics9030483

Devika, M., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. Procedia Computer Science, 87, 44-49. https://doi.org/10.1016/j.procs.2016.05.124

Ramachandran, D., & Parvathi, R. (2019). Analysis of Twitter specific Preprocessing technique for tweets. Procedia Computer Science, 165, 245-251. https://doi.org/10.1016/j.procs.2020.01.083

https://necromuralist.github.io/Neurotic-Networking/posts/nlp/01-twitter-preprocessing-with-nltk/

Monkey, C. (2020, July 3). Twitter Preprocessing with NLTK. Neurotic Networking. https://necromuralist.github.io/Neurotic-Networking/posts/nlp/01-twitter-preprocessing-with-nltk/

Mushtaq, M. F., Fareed, M. M., Almutairi, M., Ullah, S., Ahmed, G., & Munir, K. (2022). Analyses of public attention and sentiments towards different COVID-19 vaccines using data mining techniques. Vaccines, 10(5), 661. https://doi.org/10.3390/vaccines10050661

Küçükali, H., Ataç, Ö., Palteki, A. S., Tokaç, A. Z., & Hayran, O. (2022). Vaccine hesitancy and anti-vaccination attitudes during the start of COVID-19 vaccination program: A content analysis on Twitter data. Vaccines, 10(2), 161. https://doi.org/10.3390/vaccines10020161

Shim, J., Ryu, K., Lee, S. H., Cho, E., Lee, Y. J., & Ahn, J. H. (2021). Text mining approaches to analyze public sentiment changes regarding COVID-19 vaccines on social media in Korea. International Journal of Environmental Research and Public Health, 18(12), 6549. https://doi.org/10.3390/ijerph18126549

Surging delta variant cases, hospitalizations, and deaths are biggest drivers of recent Uptick in U.S. COVID-19 vaccination rates. (2021, September 28). KFF. https://www.kff.org/coronavirus-covid-19/press-release/surging-delta-variant-cases-hospitalizations-and-deaths-are-biggest-drivers-of-recent-uptick-in-u-s-covid-19-vaccination-rates/

KFF COVID-19 vaccine monitor: Early omicron update. (2021, December 29). KFF. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-early-omicron-update/

Coronavirus disease 2019 (COVID-19). (2022, August 31). Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/variants/index.html

https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-early-omicron-update/

Rogers, E. M. (1995). Diffusion of innovations (4th ed.). New York: Free Press.

https://www.cdc.gov/coronavirus/2019-ncov/your-health/treatments-for-severe-illness.html

https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/adverse-events.html

https://wwwnc.cdc.gov/eid/article/28/8/22-0557_article

Lu Meng, Neil Chandra Murthy, Bhavini Patel Murthy, Elizabeth Zell, Ryan Saelee, Megan Irving, Hannah E. Fast, Patricia Castro Roman, Adam Schiller, Lauren Shaw, Carla L. Black, Lynn Gibbs-Scharf, LaTreace Harris, and Terence Chorba, CDC, 2022

https://news.yahoo.com/twitter-files-platform-suppressed-valid-200142382.html https://www.foxnews.com/media/new-twitter-files-reveal-pharmaceutical-industry-lobbied-social-media-over-covid-vaccine-content

https://www.foxnews.com/media/twitter-files-pfizer-board-member-dr-scott-gottlieb-flagged-tweets-questioning-covid-vaccine

https://euroweeklynews.com/2023/01/29/damning-reports-claim-massive-cover-up-china-us-uk-over-origins-coronavirus/

https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-people-booster-percent-total

Sattar, N. S., & Arifuzzaman, S. (2021). COVID-19 vaccination awareness and aftermath: public sentiment analysis on Twitter data and vaccinated population prediction in the USA. Applied Sciences, 11(13), 6128.

Stechemesser, A., Levermann, A., & Wenz, L. (2022). Temperature impacts on hate speech online: evidence from 4 billion geolocated tweets from the USA. The Lancet Planetary Health, 6(9), e714-e725.

REFERENCES

1. He, D., Zhao, S., Li, Y., Cao, P., Gao, D., Lou, Y., & Yang, L. [2020]. Comparing CO He, D ID-19 and the 1918–19 influenza pandemics in the United Kingdom. International Journal of Infectious Diseases, 98, 67-70.

2. Squazzoni, F., Polhill, J. G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., ... & Gilbert, N. [2020]. Computational models that matter during a global pandemic outbreak: A call to action. Journal of Artificial Societies and Social Simulation, 23[2].

3. Correia, S., Luck, S., & Verner, E. [2020]. Fight the Pandemic, Save the Economy: Lessons from the 1918 Flu. Federal Reserve Bank of New York, Liberty Street Economics [March 27]. https://libertystreeteconomics. newyorkfed. org/2020/03/fight-the-pandemic-save-the-economy-lessons-from-the-1918-flu. html.

4. Cooper, I., Mondal, A., & Antonopoulos, C. G. [2020]. A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons & Fractals, 139, 110057

5. Brauer, F., Driessche, P. V., & Wu, J. (2008). Mathematical epidemiology. Springer Science & Business Media. (Brauer, Driessche, & Wu, 2008, p. xx)

6. Raamkumar, A. S., Tan, S. G., & Wee, H. L. [2020]. Use of health belief model–based deep learning classifiers for covid-19 social media content to examine public perceptions of physical distancing: Model development and case study. JMIR public health and surveillance, 6[3], e20493.

7. Ajzen, I. [1991]. The theory of planned behavior. Organizational behavior and human decision processes, 50[2], 179-211.

8. Wang, Y., Liang, J., Yang, J., Ma, X., Li, X., Wu, J., ... & Feng, Y. [2019]. Analysis of the environmental behavior of farmers for non-point source pollution control and management: An integration of the theory of planned behavior and the protection motivation theory. Journal of environmental management, 237, 15-23.

9. Vargas-De-León, C. [2011]. On the global stability of SIS, SIR and SIRS epidemic models with standard incidence. Chaos, Solitons & Fractals, 44[12], 1106-1110.

10. Fernández-Villaverde, J., & Jones, C. I. [2020]. Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities [No. w27128]. National Bureau of Economic Research.

11. Jiang, D., Yu, J., Ji, C., & Shi, N. [2011]. Asymptotic behavior of global positive solution to a stochastic SIR model. Mathematical and Computer Modelling, 54[1-2], 221-232.

12. Seyoum Desta, B. (2019). Modified model and stability analysis of the spread of hepatitis B virus disease. American Journal of Applied Mathematics, 7(1), 13. doi:10.11648/j.ajam.20190701.13

13. Fan, M., Li, M. Y., & Wang, K. [2001]. Global stability of an SEIS epidemic model with recruitment and a varying total population size. Mathematical Biosciences, 170[2], 199-208. (Fan, Li, & Wang, 2001, p. xx)

14. Zhou, Y., & Li, Y. Application of MSEIR Model of Secondary Infection in Pest Control.

15. Zohdi, T. I. [2020]. An agent-based computational framework for simulation of global pandemic and social response on Planet X. Computational Mechanics, 1-15.

16. Bin, S., Sun, G., & Chen, C. C. (2019). Spread of infectious disease modeling and analysis of different factors on spread of infectious disease based on cellular automata. International journal of environmental research and public health, 16(23), 4683.

17. Woody, S., Tec, M. G., Dahan, M., Gaither, K., Lachmann, M., Fox, S., ... & Scott, J. G. (2020). Projections for first-wave COVID-19 deaths across the US using social-distancing measures derived from mobile phones. medRxiv.

18. Samaras, L., García-Barriocanal, E., & Sicilia, M. A. [2020]. comparing Social media and Google to detect and predict severe epidemics. Scientific Reports, 10[1], 1-11.

19. Sooknanan, J., & Comissiong, D. M. G. [2020]. Trending on social media: integrating social media into infectious disease dynamics. Bulletin of Mathematical Biology, 82[7], 1-11.

20. Bonabeau, E. [2002]. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the national academy of sciences, 99[suppl 3], 7280-7287.

21. Silva, P. C., Batista, P. V., Lima, H. S., Alves, M. A., Guimarães, F. G., & Silva, R. C. [2020]. COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos, Solitons & Fractals, 139, 110088.

22. Freberg, K., Palenchar, M. J., & Veil, S. R. [2013]. Managing and sharing H1N1 crisis information using social media bookmarking services. Public relations review, 39[3], 178-184.

23. Slovic, P. E. [2001]. The perception of risk. Earthscan publications.

24. Champion, V. L., & Skinner, C. S. [2008]. The health belief model. Health behavior and health education: Theory, research, and practice, 4, 45-65.

25. Tchuenche, J. M., & Bauch, C. T. [2012]. Dynamics of an infectious disease where media coverage influences transmission. ISRN biomathematics, 2012.

26. Lu, X., Wang, S., Liu, S., & Li, J. [2017]. An SEI infection model incorporating media impact. Mathematical Biosciences & Engineering, 14[5&6], 1317.

27. Cui, J., Sun, Y., & Zhu, H. [2008]. The impact of media on the control of infectious diseases. Journal of dynamics and differential equations, 20[1], 31-53.

28. El-Sayed, A. M., Scarborough, P., Seemann, L., & Galea, S. [2012]. Social network analysis and agent-based modeling in social epidemiology. Epidemiologic Perspectives & Innovations, 9. doi:10.1186/1742-5573-9-

29. Will, M., Groeneveld, J., Frank, K., & Müller, B. [2020]. Combining social network analysis and agent-based modelling to explore dynamics of human interaction: A review. Socio-Environmental Systems Modelling, 2, 16325-16325.

30. Frias-Martinez, E., Williamson, G., & Frias-Martinez, V. [2011, October]. An agent-based model of epidemic spread using human mobility and social network information. In 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing [pp. 57-64]. IEEE.

31. Collinson S, Heffernan JM [2014] Modelling the effects of media during an influenza epidemic. BMC Public Health 14[1]:376

32. Price, C. C., & Propp, A. M. A Framework for Assessing Models of the COVID-19 Pandemic to Inform Policymaking in Virginia.

33. Brown, G., & Ozanne, M. (2019). Statistical Models for Infectious Diseases: A Useful Tool for Practical Decision-Making. The American journal of tropical medicine and hygiene, 101(1), 1–2. https://doi.org/10.4269/ajtmh.19-0354

34. Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2020). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement, 167, 108288.

35. Cacciapaglia, G., Cot, C., & Sannino, F. (2020). Second wave COVID-19 pandemics in Europe: a temporal playbook. Scientific reports, 10(1), 1-8.

36. Kermack, W. O., & McKendrick, A. G. (1991). Contributions to the mathematical theory of epidemics—III. Further studies of the problem of endemicity. Bulletin of mathematical biology, 53(1-2), 89-118.

37. Lekone, P. E., & Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. Biometrics, 62(4), 1170-1177.

38. Mosavi, A. (2020). COVID-19 Pandemic Prediction for Hungary; a Hybrid Machine Learning Approach.

39. Okuhara, T., Okada, H., & Kiuchi, T. (2020). Predictors of staying at home during the COVID-19 pandemic and social lockdown based on protection motivation theory: A cross-sectional study in Japan. Healthcare, 8(4), 475. doi:10.3390/healthcare8040475

40. Forrester, J. W. (1958). Industrial Dynamics. A major breakthrough for decision makers. Harvard business review, 36(4), 37-66.

41. Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Caution warranted: using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic.

42. Giattino, C. (2020, August 24). How epidemiological models of COVID-19 help us estimate the true number of infections. Retrieved from https://ourworldindata.org/covid-models

43. Obar, J. A., & Wildman, S. S. (2015). Social media definition and the governance challenge-an introduction to the special issue. Obar, JA and Wildman, S.(2015). Social media definition and the governance challenge: An introduction to the special issue. Telecommunications policy, 39(9), 745-750.

44. Rahmandad, H., & Sterman, J. (2008). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. Management Science, 54(5), 998-1014. doi:10.1287/mnsc.1070.0787\

45. Lipman, S. A., & Burt, S. A. (2017). Self-reported prevalence of pests in Dutch households and the use of the health belief model to explore householders' intentions to engage in pest control. PLOS ONE, 12(12), e0190399. doi:10.1371/journal.pone.0190399

46. Guidry, J. P., Jin, Y., Orr, C. A., Messner, M., & Meganck, S. (2017). Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public relations review, 43(3), 477-486.

47. Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. Journal of medical Internet research, 22(11), e20550.

48. Simpson, V. (2015). Models and theories to support health behavior intervention and program planning. *Health and Human Sciences, 2*(8), 45-47.

49. Greenhalgh, D., Rana, S., Samanta, S., Sardar, T., Bhattacharya, S., & Chattopadhyay, J. (2015). undefined. Applied Mathematics and Computation, 251, 539-563. doi:10.1016/j.amc.2014.11.091

50. Wang, X., Pasco, R. F., Du, Z., Petty, M., Fox, S. J., Galvani, A. P., ... & Meyers, L. A. (2020). Impact of social distancing measures on coronavirus disease healthcare demand, central Texas, USA. Emerging infectious diseases, 26(10), 2361.

51. Cascini, F., Hoxhaj, I., Zaçe, D., Ferranti, M., Di Pietro, M. L., Boccia, S., & Ricciardi, W. (2020). How health systems approached respiratory viral pandemics over time: a systematic review. BMJ global health, 5(12), e003677.

52. Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. The Lancet, 395(10228), 931-934.

53. Garibaldi, B. T., Fiksel, J., Muschelli, J., Robinson, M. L., Rouhizadeh, M., Perin, J., ... & Gupta, A. (2020). Patient Trajectories Among Persons Hospitalized for COVID-19: A Cohort Study. Annals of internal medicine.

54. COVIDAnalytics. (2020). COVIDAnalytics. https://www.covidanalytics.io/projections

55. https://youyanggu.com/blog/six-months-later

56. US dashboard. (2020). UT COVID-19 Modeling Consortium. https://covid-19.tacc.utexas.edu/dashboards/us/

57. DeJoy, D. M. (1996). Theoretical models of health behavior and workplace self-protective behavior. Journal of Safety Research, 27(2), 61-72.

58. Hernandez, S. (2020). This Is How a Group Linked to Betsy DeVos Is Organizing Protests to End Social Distancing, Now With Trump Support.

59. COVID-19 resources. (2020, June 23). Institute for Health Metrics and Evaluation. https://www.healthdata.org/covid

60. Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H. (2020). Forecasting of COVID19 per regions using ARIMA models and polynomial functions. Applied Soft Computing, 96, 106610.

61. As coronavirus cases rise, Fauci says public health measures are the way to slow spread. (2020, October 19). WINK NEWS. https://www.winknews.com/2020/10/19/as-coronavirus-cases-rise-fauci-says-public-health-measures-are-the-way-to-slow-spread/

62. Siebers, P. O., & Aickelin, U. (2008). Introduction to multi-agent simulation. In Encyclopedia of decision making and decision support technologies (pp. 554-564). IGI Global.

63. Niazi, M., & Hussain, A. (2011). Agent-based computing from multi-agent systems to agent-based models: a visual survey. Scientometrics, 89(2), 479-499.

64. Alkhateeb, F., Al Maghayreh, E., & Aljawarneh, S. (2010, January). A multi agent-based system for securing university campus: Design and architecture. In 2010 International Conference on Intelligent Systems, Modelling and Simulation (pp. 75-79). IEEE.

65. Seddari, N., & Redjimi, M. (2013, March). Multi-agent modeling of a complex system. In 2013 3rd International Conference on Information Technology and e-Services (ICITeS) (pp. 1-6). IEEE.

66. Agents in artificial intelligence. (2019, August 6). GeeksforGeeks. **https://www.geeksforgeeks.org/agents-artificial-intelligence/**

67. Vyklyuk, Y., Manylich, M., Škoda, M., Radovanović, M. M., & Petrović, M. D. (2021). Modeling and analysis of different scenarios for the spread of COVID-19 by using the modified multi-agent systems–Evidence from the selected countries. Results in physics, 20, 103662.

68. Bae, S. Y., & Chang, P. J. (2020). The effect of coronavirus disease-19 (COVID-19) risk perception on behavioural intention towards ‘untact’tourism in South Korea during the first wave of the pandemic (March 2020). Current Issues in Tourism, 1-19.

69. Parsons, S., & Wooldridge, M. (2002). Game theory and decision theory in multi-agent systems. Autonomous Agents and Multi-Agent Systems, 5(3), 243-254.

70. Cruz-Pacheco, G., Duran, L., Esteva, L., Minzoni, A. A., Lopez-Cervantes, M., Panayotaros, P., ... & Ruiz, I. V. (2009). Modelling of the influenza A (H1N1) v outbreak in Mexico City, April-May 2009, with control sanitary measures. Eurosurveillance, 14(26), 19254.

71. Hunter, E., Mac Namee, B., & Kelleher, J. (2018). An open-data-driven agent-based model to simulate infectious disease outbreaks. PloS one, 13(12), e0208775.

72. Hunter, E., Mac Namee, B., & Kelleher, J. D. (2017). A taxonomy for agent-based models in human infectious disease epidemiology. Journal of Artificial Societies and Social Simulation, 20(3).

73. Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Hart, G., Rosenfeld, K., ... & Klein, D. J. (2020). Covasim: an agent-based model of COVID-19 dynamics and interventions. *medRxiv*.

74. Crooks, A., Malleson, N., Manley, E., & Heppenstall, A. (2018). Agent-based modelling and geographical information systems: a practical primer. Sage.

75. (2020, July 16). Centers for Disease Control and Prevention. **https://www.cdc.gov/**

76. Protests, policing, and COVID-19. (2021, April 13). National Governors Association. **https://www.nga.org/center/publications/protests-policing-covid-19/**

77. Ghostine, R., Gharamti, M., Hassrouny, S., & Hoteit, I. (2021). An extended seir model with vaccination for forecasting the covid-19 pandemic in saudi arabia using an ensemble kalman filter. Mathematics, 9(6), 636.

78. Li, J., & Giabbanelli, P. (2021). Returning to a Normal Life via COVID-19 Vaccines in the United States: A Large-scale Agent-Based Simulation Study. JMIR Medical Informatics, 9(4), e27419.

79. Epstein, J. M., Parker, J., Cummings, D., & Hammond, R. A. (2008). Coupled contagion dynamics of fear and disease: mathematical and computational explorations. PLoS One, 3(12), e3955.

80. *One in five still shun vaccine*. (2021, April 14). Monmouth University Polling Institute. **https://www.monmouth.edu/polling-institute/reports/monmouthpoll_us_041421/**

81. Nottmeyer, L. N., & Sera, F. (2021). Influence of temperature, and of relative and absolute humidity on COVID-19 incidence in England-A multi-city time-series study. Environmental research, 196, 110977.

82. Thomas, C., & Nagchoudhury, S. (2020, October 15). India's peak holiday season set to send COVID-19 cases surging. U.S. **https://www.reuters.com/article/health-coronavirus-india-festivals-holid-idUSKBN2700M9**

83. Maxouris, C., Hanna, J., & Paul Vercammen And Dakin Andone. (2021, January 10). After holidays, U.S. COVID-19 cases and deaths climb to all-time highs. CP24. **https://www.cp24.com/world/after-holidays-u-s-covid-19-cases-and-deaths-climb-to-all-time-highs-1.5260840?cache=kcfnyoei%3FclipId%3D89619**

84. Rogers, Everett M. (2003) Diffusion of innovations. New York : Free Press 5th ed.

85. Samui, P., Mondal, J., & Khajanchi, S. (2020). A mathematical model for COVID-19 transmission dynamics with a case study of India. Chaos, Solitons & Fractals, 140, 110173.

86. Kukulka, A. (2021, April 30). COVID-19 vaccine hesitancy among conservatives is highest, but highlighting vaccine benefits could combat that, public health experts say. chicagotribune.com. **https://www.chicagotribune.com/suburbs/post-tribune/ct-ptb-vaccine-rollout-feat-st-0502-20210430-qk6cxdjc2vbm3gel6gxq5g55se-story.html**

87. Edwards, K. M., Hackell, J. M., Committee on Infectious Diseases, & Committee on Practice and Ambulatory Medicine. (2016). Countering vaccine hesitancy. Pediatrics, 138(3).

88. Daghriri, T., & Ozmen, O. (2020). Quantifying the Effects of Social Distancing on the Spread of COVID-19. Available at SSRN 3696983. (Accepted by International Journal of Environmental Research and Public Health, under preparing to be published online)!

89. (2020). Virginia Department of Health. **https://www.vdh.virginia.gov/**

90. Model quantifies the impact of quarantine measures on Covid-19's spread. (Gallagher, 2020).

91. Dandekar, R., & Barbastathis, G. (2020). Neural Network aided quarantine control model estimation of global Covid-19 spread. arXiv preprint arXiv:2004.02752.

92. UVA COVID-19 model. (2021). Virginia Department of Health. **https://www.vdh.virginia.gov/coronavirus/covid-19-data-insights/uva-covid-19-model/**

93. Matthews, S. D., & Proctor, M. D. (2021). Public health informatics, human factors and the end-users. Health Services Research and Managerial Epidemiology, 8, 233339282110122. **https://doi.org/10.1177/23333928211012226**

94. Bond, S. (2021, May 13). Just 12 people are behind most vaccine hoaxes on social media, research shows. NPR.org. **https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-Twitters-ability-to-curb-vaccine-hoaxes**

95. **https://coronavirus.jhu.edu/vaccines/international**

96. **https://coronavirus.jhu.edu/vaccines/us-states#race-to-vaccine-us-2021-05-18**

97. Perrinp, A., & Atske, S. (2021, April 3). 7% of Americans don't use the internet. Who are they? Pew Research Center. **https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/**

98. **https://abcnews.go.com/Politics/risks-unvaccinated-americans-weighing/story?id=77319410**

99. Auxier, B., & Anderson, M. (2021, April 7). Social media use in 2021. Pew Research Center: Internet, Science & Tech. **https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/**

100. Sattar, N. S., & Arifuzzaman, S. (2021). COVID-19 vaccination awareness and aftermath: Public sentiment analysis on Twitter data and vaccinated population prediction in the USA. Applied Sciences, 11(13), 6128. **https://doi.org/10.3390/app11136128**

101. Biden mandates vaccines for workers, saying, 'Our patience is wearing thin'. (2021, September 10). The New York Times - Breaking News, US News, World News and Videos. **https://www.nytimes.com/2021/09/09/us/politics/biden-mandates-vaccines.html**

102. WADMAN, M. (2021.). Having SARS-Cov-2 once confers much greater immunity than a vaccine—but vaccination remains vital. Science | AAAS. **https://www.science.org/content/article/having-sars-cov-2-once-confers-much-greater-immunity-vaccine-vaccination-remains-vital**

103. Erman, M., & Manojna, M. (2021, June 10). Heart inflammation in young men higher than expected after Pfizer, Moderna vaccines -U.S. CDC. Reuters. **https://www.reuters.com/world/us/cdc-heart-inflammation-cases-ages-16-24-higher-than-expected-after-mrna-covid-19-2021-06-10/**

104. Neel, J. (2021, August 31). Vaccination rates rose in August as COVID cases surged due to the delta variant. NPR.org. **https://www.npr.org/2021/08/31/1032994955/vaccination-rates-rose-in-august-as-covid-cases-surge-due-to-the-delta-variant**

105. Magness, P. W. (2020, November 11). Opinion. WSJ. **https://www.wsj.com/articles/case-for-mask-mandate-rests-on-bad-data-11605113310**

106. Zanin, M., & Papo, D. (2020). Travel restrictions during pandemics: A useful strategy? Chaos: An Interdisciplinary Journal of Nonlinear Science, 30(11), 111103. **https://doi.org/10.1063/5.0028091**

107. United States: Coronavirus pandemic country profile. (2021). Our World in Data. **https://ourworldindata.org/coronavirus/country/united-states#what-share-of-the-population-has-received-at-least-one-dose-of-the-covid-19-vaccine**

108. Muhammad, L. J., Algehyne, E. A., Usman, S. S., Ahmad, A., Chakraborty, C., & Mohammed, I. A. (2021). Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. SN computer science, 2(1), 1-13.

109. Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine learning techniques for sentiment analysis: A review. Int. J. Multidiscip. Sci. Eng, 8(3), 27.

110. Alorini, G., Rawat, D. B., & Alorini, D. (2021, June). LSTM-RNN Based Sentiment Analysis to Monitor COVID-19 Opinions using Social Media Data. In ICC 2021-IEEE International Conference on Communications (pp. 1-6). IEEE.

111. Kumar, K. N., & Uma, V. (2021). Intelligent sentinet-based lexicon for context-aware sentiment analysis: Optimized neural network for sentiment classification on social media. The Journal of Supercomputing, 1-25.

112. Arti, Dubey, K. P., & Agrawal, S. (2019). An opinion mining for Indian Premier League using machine learning techniques. 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). **https://doi.org/10.1109/iot-siu.2019.8777472**

113. Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis. PLOS ONE, 16(2), e0245909. **https://doi.org/10.1371/journal.pone.0245909**

114. Arti, Dubey, K. P., & Agrawal, S. (2019). An opinion mining for Indian Premier League using machine learning techniques. 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). **https://doi.org/10.1109/iot-siu.2019.8777472**

115. Morshed, S. A., Khan, S. S., Tanvir, R. B., & Nur, S. (2021). Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis. Journal of Urban Management, 10(2), 155-165. **https://doi.org/10.1016/j.jum.2021.03.002**

116. Python, R. (2019, June 10). How to make a Twitter bot in Python with Tweepy – Real Python. Python Tutorials – Real Python. **https://realpython.com/Twitter-bot-python-tweepy/**

117. Fontanella, C. (2021, February 5). How to get, use, & benefit from Twitter's API. HubSpot Blog | Marketing, Sales, Agency, and Customer Success Content. **https://blog.hubspot.com/website/how-to-use-Twitter-api**

118. What is sentiment analysis (opinion mining)? - Definition from WhatIs.com. (2019, March 19). SearchBusinessAnalytics. **https://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining**

119. TextBlob: Simplified text processing — TextBlob 0.16.0 documentation. (n.d.). TextBlob: Simplified Text Processing — TextBlob 0.16.0 documentation. **https://textblob.readthedocs.io/en/dev/**

120. Ippolito, P. P. (2019, October 11). Feature extraction techniques. Medium. **https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be**

121. A gentle introduction to the bag-of-Words model. (2019, August 7). Machine Learning Mastery. **https://machinelearningmastery.com/gentle-introduction-bag-words-model/**

122. First U.S. adverse COVID-19 vaccine reaction reported in Alaska healthcare worker. (2021, January 5). KXAN Austin. **https://www.kxan.com/news/coronavirus/first-u-s-adverse-covid-19-vaccine-reaction-reported-in-alaska-healthcare-worker/**

123. Coronavirus disease 2019. (2021, April 23). Centers for Disease Control and Prevention. **https://www.cdc.gov/media/releases/2021/fda-cdc-lift-vaccine-use.html**

124. COVID-19 vaccination. (2021, April 16). Centers for Disease Control and Prevention. **https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/JJUpdate.html**

125. Quick introduction to bag-of-Words (Bow) and TF-IDF for creating features from text. (2020, December 23). Analytics Vidhya. **https://www.analyticsvidhya.com/blog/2020/02/quick-introduction-bag-of-words-bow-tf-idf/**

126. Chatterjee, S. (2021, April 7). What is feature extraction? Feature extraction in image processing. GreatLearning Blog: Free Resources what Matters to shape your Career!. **https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/**

127. One and done: Why people are eager for Johnson & Johnson's vaccine. (2021, March 5). The New York Times - Breaking News, US News, World News and Videos. **https://www.nytimes.com/2021/03/04/health/covid-vaccine-johnson-and-johnson-rollout.html**

128. Sattar, N. S., & Arifuzzaman, S. (2021). COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA. Applied Sciences, 11(13), 6128.

Pfizer's novel COVID-19 oral antiviral treatment candidate reduced risk of hospitalization or death by 89% in interim analysis of phase 2/3 EPIC-HR study. (2021). Pfizer Inc. – Investors Overview. **https://investors.pfizer.com/investor-news/press-release-details/2021/Pfizers-Novel-COVID-19-Oral-Antiviral-Treatment-Candidate-Reduced-Risk-of-Hospitalization-or-Death-by-89-in-Interim-Analysis-of-Phase-23-EPIC-HR-Study/default.aspx**

130. FdA authorizes booster dose 171fizerzer-biontech COVID-19 vaccine for certain populations. (2021, September 22). U.S. Food and Drug Administratio**n. https://**www.fda.gov/news-events/press-announcements/fda-authorizes-booster-dose-pfizer-biontech-covid-19-vaccine-certain-populations

131. Fact sheet: Biden administration announces details of two major vaccination policies. (2021, November 4). The White House. **https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/04/fact-sheet-biden-administration-announces-details-of-two-major-vaccination-policies/**

132. Non parametric data and tests (Distribution free tests). (2021, January 22). Statistics How To. **https://www**.statisticshowto.com/probability-and-statistics/statistics-definitions/parametric-and-non-parametric-data/

133. Goyal, A. (2021, June 8). Hypothesis testing | Parametric and non-parametric tests. Analytics Vidhya. **https://www.analyticsvidhya.com/blog/2021/06/hypothesis-testing-parametric-and-non-parametric-tests-in-statistics/**

134. Daghriri, T., Proctor, M., & Matthews, S. (2022). Evolution of select epidemiological modeling and the rise of population sentiment analysis: A literature review and COVID-19 sentiment illustration. International Journal of Environmental Research and Public Health, 19(6), 3230. doi:10.3390/ijerph19063230

135. https://www.washingtonpost.com/nation/2022/02/23/covid-omicron-variant-live-updates/

136. https://www.cdc.gov/mmwr/volumes/71/wr/mm7112e2.htm

137. https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/

138. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-early-omicron-update/

139.

140. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-september-2021/

141.

142. https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/#mandates

143. https://www.hhs.gov/coronavirus/covid-19-vaccines/index.html

144. https://www.cnn.com/2021/08/09/health/covid-19-pandemic-timeline-fast-facts/index.html
https://academic.oup.com/ofid/article/7/7/ofaa258/5865318

145. https://www.mdpi.com/2227-9032/9/9/1110