MATHEMATICAL MODELLING OF THE SPREAD OF SARS-COV-2 VARIANTS IN SOUTH AFRICA AND BRAZIL

YUAN LIU

PhD

THE HONG KONG POLYTECHNIC UNIVERSITY

2024

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF APPLIED MATHEMATICS

MATHEMATICAL MODELLING OF THE SPREAD OF SARS-COV-2 VARIANTS IN SOUTH AFRICA AND BRAZIL

YUAN LIU

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPH

April 2024

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

_____Yuan Liu_____ (Name of student)

Dedication

To my families, friends and loves.

Abstract

The novel coronavirus that caused this pandemic has infected millions of people and resulted in over a million deaths. Similar to the HIV, Zika virus, Ebola virus, and many influenza strains, the novel coronavirus had already evolved from animals to humans before causing widespread destruction. Our ongoing battle against it continues.

As a Ph.D. student who concentrates on epidemic mathematical modeling, I'm trying to make some scientific contributions to the world's fight against this intractable infectious disease. My aim with this research is to aid these marginalized regions in promptly mitigating the detrimental effects caused by the COVID-19 virus. I have conducted research on various aspects, including the variations in infection fatality rates and transmission dynamics among different strains of the novel coronavirus, as well as the heterogeneity of in-hospital mortality rates in underdeveloped regions. And I aspire to utilize my theoretical knowledge to assist humanity in overcoming this formidable new virus.

We studied the reduction in the infection fatality rate of the Omicron (B.1.1.529) variant compared with previous variants in South Africa. This research work started when the Omicron variants just emerged, and South Africa was the first place where the virus wreaked havoc. Before we did this work, some previous studies have shown that the variant has enhanced immune evasion ability and transmissibility and reduced severity. In this study, we developed a mathematical model with time-varying transmission rate, vaccination, and immune evasion. We fit the model to report case and death data up to February 6, 2022, to estimate the transmissibility and infection fatality ratio of the Omicron variant in South

Africa. As a result, we found that the high relative transmissibility of the Omicron variant was mainly due to its immune evasion ability, whereas its infection fatality rate substantially decreased by approximately 78.7% (95% confidence interval: 66.9%, 85.0%) concerning previous variants.

Another research focuses on the transmissibility of all COVID-19 variants that have been spread in South Africa. The COVID-19 pandemic caused multiple waves of mortality in South Africa, where three genetic variants of SARS-COV-2 and their ancestral strain dominated consecutively. In this research, we fit a state-of-the-art mathematical modeling approach to estimate the time-varying transmissibility of SARS-COV-2 and the relative transmissibility of Beta, Delta, and Omicron variants. As a result, the transmissibilities of the three variants were about 73%, 87%, and 276% higher than their preceding variantsThe transmissibility of the Omicron variant is substantially higher than that of previous variants.

In addition to South Africa, we examined the regional variations in COVID-19 in-hospital mortality in Brazil using a multivariate mixed-effect Cox model applied to national inpatient data from February 27, 2020, to March 15, 2022. We compared mortality risks between vaccinated and unvaccinated patients, adjusting for age, state, ethnicity, education, and comorbidities. Our analysis showed age as the primary risk factor for death. Illiterate patients (hazard ratio: 1.63, 95% CI: 1.56-1.70) faced higher risks compared to those with higher education. Common comorbidities, such as liver disease (HR: 1.46, 95% CI: 1.34-1.59) and immunosuppression (HR: 1.32, 95% CI: 1.26-1.40), increased mortality risk. States like Sergipe (HR: 1.75, 95% CI: 1.46-2.11), Roraima (HR: 1.65, 95% CI: 1.43-1.92), Maranhão (HR: 1.57, 95% CI: 1.38-1.79), Acre (HR: 1.44, 95% CI: 1.12-1.86), and Rondônia (HR: 1.26, 95% CI: 1.10-1.44) in the north and northeast had higher mortality risks. Vaccination did not significantly reduce mortality, with varying effectiveness between

Sinovac and AstraZeneca across regions. The study highlights regional differences in mortality, the limited impact of vaccination, and the role of social inequality. Ethnic and regional disparities suggest uneven interactions within communities may influence epidemic spread, and vaccine efficacy varies by region.

Publication arising from the thesis

List of articles used in the thesis:

1. Liu, Y., Yu, Y., Zhao, Y., & He, D. (2022). Reduction in the infection fatality rate of Omicron variant compared with previous variants in South Africa. *International Journal of Infectious Diseases*, *120*, 146-149.
2. Yu, Y., Liu, Y., Zhao, S., & He, D. (2022). A simple model to estimate the transmissibility of the Beta, Delta, and Omicron variants of SARS-COV-2 in South Africa. *Math. Biosci. Eng*, *19*(10), 10361-10373.
3. Liu, Y., Wang, K., Yang, L., & He, D. (2022). Regional heterogeneity of in-hospital mortality of COVID-19 in Brazil. *Infectious Disease Modelling*, *7*(3), 364-373.
4. Liu, Y., Feng, A., Zhao, S., Wang, W., & He, D. (2022). Large-scale synchronized replacement of Alpha (B. 1.1. 7) variant by the Delta (B. 1.617. 2) variant of SARS-COV-2 in the COVID-19 pandemic. *Mathematical Biosciences and Engineering*, *19*(4), 3591-3596.

List of other published articles during the course of the study:

1. Song, H., Fan, G., Liu, Y., Wang, X., & He, D. (2021). The second wave of COVID-19 in South and Southeast Asia and the effects of vaccination. *Frontiers in medicine*, *8*, 773110.
2. Musa, S. S., Tariq, A., Yuan, L., Haozhen, W., & He, D. (2022). Infection fatality rate and infection attack rate of COVID-19 in South American countries. *Infectious diseases of poverty*, *11*(02), 42-52.

CURRICULUM VITAE

Academic qualifications of the thesis author, Mr. Liu Yuan:

- Received the degree of Bachelor of Science from Miami University, May, 2019.
- Received the degree of Master of Science from Georgetown University, May, 2021.

Acknowledgments

Research work is an interesting and challenging endeavor, yet an individual's capacity is inherently limited. Here, I extend my deepest gratitude to my chef supervisor, friends, and families who have supported me in various ways over these three years.

First and foremost, I would like to thank my Chief Supervisor, Dr. Daihai He, my advisor, for his patient, friendly, and selfless guidance. From reading research works, collecting data, to writing and publishing papers, Dr. He has provided detailed and patient guidance at every important juncture of my research work over the past three years. Without his selfless and patient assistance, my research work would not have been possible.

I also want to extend profound respect and gratitude to my family, thanking them for their dual support, both financially and spiritually, which has enabled me to have the opportunity to receive an education at a world-renowned institution. It is precisely because of the selfless support and encouragement from my family members that I have been able to focus wholeheartedly on my research work.

Finally, I would like to thank all of my friends and research fellows at Hong Kong Polytechnic University, for their support and love.

Contents

List of Tables

List of Figures

Basic Notations

S	Susceptible
E	Exposed
I	Infectious
R	Recovered
H	Hospitalization
V	Vaccination
$\mathbf{R_0}$	The basic reproduction number
$\boldsymbol{\beta}$	Transmission rate
$\boldsymbol{\sigma^{-1}}$	Mean Latent Period
$\boldsymbol{\gamma^{-1}}$	Mean Infectious Period

Chapter 1

1 Introduction

The sudden outbreak of infectious diseases has been continuously threatening the global public health system. Novel modes of transmission and rapid spread rates can instantly destabilize seemingly robust public health systems, leading to enormous economic losses, psychological trauma, and even endangering lives across different regions and nations. Particularly in economically underdeveloped countries and regions, where primary healthcare facilities have limited capacity to accommodate patients, experience in managing rapidly spreading infectious diseases is limited, and government financial support is constrained, the relatively fragile public health systems become increasingly overwhelmed in the face of widespread transmission of emerging infectious diseases.

In the face of the threat posed by a novel infectious disease, scientific research stands as humanity's best weapon. Researchers in the fields of mathematical epidemiology, bioinformatics, medicine, and others are frontline warriors confronting emerging infectious diseases. They study virus characteristics, vaccine development, transmission rates, mortality rates, and the

substantial economic losses and demographic changes that such diseases may bring to human society from various angles and directions.

As a doctoral student in the field of mathematical epidemiology, I have been dedicated to studying the infection mortality rates of the novel coronavirus in economically underdeveloped regions, particularly in South Africa and Brazil, as well as the transmissibility of different strains and the effectiveness of vaccines. I hope that my research can contribute to the global health community's efforts in combating infectious diseases, especially the novel coronavirus. Through enhancing human understanding of disease dynamics, establishing mathematical models, and conducting computational simulations, I aim to provide valuable insights for public health systems worldwide, especially in economically underdeveloped regions, to help them better prepare and enhance their capacity to combat current and future threats, thus safeguarding the health and well-being of all residents and people around the world.

This chapter first introduces the common characteristics of infectious diseases that have had a profound impact on human history since the last century. Subsequently, it discusses the principles and foundations of mathematical models for infectious diseases and the modern methodology of computational simulation. The third section outlines the overall structure of the thesis. The final section offers prospects for future research in mathematical epidemiology.

1.1 Infectious Diseases: Humanity's Perpetual Shadow

In recent years, the world has witnessed the profound impact of infectious diseases on global health, societal stability, and economic prosperity. From the devastating effects of pandemics like HIV/AIDS, influenza, and most recently COVID-19, to the persistent threat of endemic diseases such as malaria, tuberculosis, and cholera, the complex interplay between pathogens, hosts, and environments underscores the critical need for advanced epidemiological research and intervention strategies. At the forefront of this multidisciplinary endeavour lies mathematical epidemiology, a field that integrates mathematical modelling, statistical analysis, and computational techniques to elucidate the dynamics of infectious diseases and evaluate the effectiveness of control measures. By providing quantitative insights into transmission patterns, population dynamics, and the impact of interventions, mathematical epidemiology plays an indispensable role in guiding public health policies and mitigating the burden of infectious diseases worldwide.

Infectious diseases have long been entwined with the narrative of human existence, leaving an indelible imprint on civilizations and shaping the course of history, and infectious diseases have already been an enormous burden on the whole society and medical system. A new infectious disease will spread among different regions and continents rapidly, mercilessly bringing down the health of individuals, the healthcare system, the stability of social order, and even the lack of foundation supplies: food and drinks. From ancient plagues to contemporary

pandemics, infectious diseases have profoundly impacted societies, altering demographics, social structures, and scientific paradigms. Understanding their historical evolution and contemporary challenges is pivotal in devising effective strategies to mitigate their impact on global health.

In the 21st century, infectious disease-related disciplines have become increasingly important, regardless of whether in developed or developing countries. Issues related to infectious diseases are repeatedly mentioned by global leaders, health policymakers, and various charitable organizations. The focus of high-level attention is not only on scientific research such as vaccine development but also on the impact of various infectious diseases on economic development and political stability. With the continuous increase in financial investment and support for scientific research, people's understanding of infectious diseases has become more comprehensive, and related research has become associated with the basic healthcare infrastructure of many countries and regions.

Some argue that with the development of technology and medical standards, research related to infectious diseases will gradually disappear. However, through an understanding of the history of infectious diseases, the opposite is true. Infectious diseases may become more frequent in the future and have profound impacts on societies worldwide, potentially even affecting human lifestyles (e.g., wearing masks).

Since the last century, the names of certain infectious diseases have repeatedly appeared and to some extent influenced the course of human history (Faria et al. 2014; Keele et al. 2006). For instance, Human Immunodeficiency

Virus (HIV), which initially sporadically spread from non-human primates to humans in the 20th century. It wasn't until the 1980s when a form of immune deficiency disease was discovered within the male homosexual community that this virus began to attract global attention (Gottlieb et al. 1981). HIV infection is one of the leading causes of illness and death worldwide, with the majority of cases concentrated in the South African region (Correction Naghavi et al. 2015). Infections typically occur in adults of working age, thus, to a certain extent, this virus has significantly altered the economic development of many countries and regions. The impact it has had on humanity is challenging to comprehensively quantify (Deeks et al. 2015).

When it comes to infectious diseases, another one that cannot be ignored is the catastrophic influenza A. Even in the 21st century, with significant advancements in technology and healthcare, on average, influenza A causes 20,000 deaths annually in the United States alone (Webster 1998). In the last century, the first global rampage of influenza A occurred in 1918, resulting in at least 20 million deaths worldwide. The second outbreak was in 1957 when the deadly influenza A caused approximately 70,000 deaths in the United States. In 1968, there was a third major flu pandemic, resulting in 35,000 to 40,000 deaths (Fauci 2001). Historically, from a historical perspective, there has been a major flu outbreak approximately every 20 to 40 years. From 1997 to 1998, the spread of H5N1 influenza A in Hong Kong, China served as a warning, indicating that influenza-like infectious diseases pose a threat like a dark cloud looming over the entire human race on Earth (Chan 2002).

Entering the 21st century, in 2002, an outbreak of an unknown cause of atypical pneumonia occurred in southern China, gradually spreading to neighboring countries and regions (Ksiazek et al. 2003). The virus responsible was a coronavirus that caused severe acute respiratory syndrome (SARS). Hong Kong and Beijing were the most severely affected cities at the time (Demmler and Ligon 2003). Fortunately, the epidemic was declared over in July 2003. However, within just one year, more than 8,000 suspected cases were reported in 29 countries, with a mortality rate reaching 10% (Watkins 2018). The initial outbreak of the virus was linked to the consumption of wild animals, and it spread through close contact and droplets (Hui and Chan 2010). Its hallmark was recurrent fever accompanied by hypoxemia and progressive pneumonia, with approximately 50% of patients requiring oxygen supplementation and about 20% needing intubation. Diarrhea and hepatitis were common complications, and some survivors also exhibited persistent pulmonary dysfunction (Watkins 2018).

Like a cycle of fate, in December 2019, a series of acute atypical respiratory diseases occurred again in Wuhan, China. The virus quickly spread from Wuhan to other areas, leading to a global pandemic over the following two years. In approximately two years, it has led to over 5.59 million deaths and resulted in millions of people suffering from multi-system diseases(Young et al. 2022). According to preliminary research, the virus was quickly identified as a novel coronavirus, named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), due to its high homology with the SARS virus that appeared in 2002-2003 (Yuki, Fujiogi, and Koutsogiannaki 2020). Initially thought to be a zoonotic virus, human-to-human transmission of this novel coronavirus was soon

recognized at the beginning of the outbreak (Li et al. 2020). The disease caused by this virus was subsequently named Coronavirus Disease 2019 (COVID-19), and shortly after, the World Health Organization declared it a pandemic. This novel coronavirus primarily affects the respiratory system, but subsequent research has found that it also affects other organ systems to varying degrees. According to initial studies, initial cases presented symptoms of lower respiratory tract infection, including fever, dry cough, and difficulty breathing (Huang et al. 2020). Some patients also experienced headaches, fatigue, vomiting, and diarrhea (Shi et al. 2020). The original strain circulating in Wuhan was highly virulent, with patients progressing from symptom onset to severe hypoxia in just nine days (Huang et al. 2020). The rapid progression of the disease led to a rapid increase in severe cases and deaths worldwide.

Although all viruses undergo mutations over time, most mutations have little impact. However, for the novel coronavirus, several mutations have significantly altered its pathogenicity and transmission capability, changing the severity of the disease, hindering vaccine development and treatment strategy design, increasing the virus's immune evasion, and reducing vaccine effectiveness. According to the World Health Organization's report in 2022, the WHO has focused on five variants: alpha, beta, gamma, delta, and omicron (Young et al. 2022). Different variants have even caused different clinical symptoms; alpha and delta variants increase the risk of severe illness and death, the gamma variant is associated with loss of smell and taste (Luna-Muschi et al. 2022), while the omicron variant may cause milder symptoms (Nyberg et al. 2022).

The Alpha variant (B.1.1.7) was first discovered in the UK in September 2020. Due to the D614G mutation, this variant increased the binding ability of SARS-CoV-2 to the ACE2 receptor, resulting in increased infectivity (Lan et al. 2020). Other mutations in this strain also enhanced the virus's ability to evade antibody detection. According to a report from a medical institution in Madrid, Spain, patients infected with the alpha strain were twice as likely to be admitted to the intensive care unit as those infected with the original strain (Martínez-García et al. 2021). This variant became the dominant strain within four months. The Beta variant (B.1.351) was first discovered in South Africa in May 2020. This variant contains five mutations in the S protein, similar to the alpha variant. The mutations increase the binding affinity to the ACE2 receptor, enhancing virulence and resistance to antibodies (Starr et al. 2020). Due to the enhanced binding ability, the transmissibility and immune evasion of the beta variant are significantly increased. The Gamma variant (P.1) was first discovered in Brazil in November 2020, carried by a traveler returning from Brazil to Japan. Like the previous two variants, the S protein mutation in the gamma variant also increases the binding ability to the ACE2 receptor, infectivity, and lethality (ECDC 2022). Studies have shown that compared to previous variants, the infectivity of the gamma variant increased by 1.7 to 2.4 times (Campbell et al. 2021), and the infection and mortality rates of patients in the main transmission area of the gamma variant, Brazil, increased by 13% (Freitas et al. 2021).

The Delta variant (B.1.617.2) was first discovered in India in October 2020. Unlike previous variants, the delta variant contains the E484Q mutation, which, together with the L452R mutation, significantly increases the affinity of

this variant to the ACE2 receptor, much higher than that of the original strain and other variant strains (Korber et al. 2020). Soon, the Delta variant became the dominant variant worldwide (Torjesen 2021a; Campbell et al. 2021). Research has shown that the reproductive number of the delta variant is 97% higher than that of non-concern variants, approximately three times higher than the three aforementioned variants (Sinha, Tam, and Wang 2021). The rapid spread and reproductive ability of the Delta variant quickly gave it a competitive advantage among variant strains and rapidly became the dominant strain worldwide. More frighteningly, the Delta variant also increased the severity of the disease. In Scotland, the risk of hospitalization from the delta variant was significantly increased compared to the alpha strain (Sheikh et al. 2021). In India, a cross-sectional study found that the risk of death among delta variant patients was approximately 1.8 times higher, and the proportion of symptomatic cases among young people was higher than that of the original strain (Kumar, Asghar, et al. 2021).

The Omicron variant (B.1.1.529) was first discovered in South Africa and Botswana in November 2021. The variant's S protein has more than 30 mutations, with 15 mutations in the RBD (Cao et al. 2021). The emergence of the Omicron strain has led to a surge in infection cases globally. According to data from South Africa, the proportion of infections caused by the Omicron variant rose from 3% to 98% in just two months from October to December 2021 (Wolter et al. 2022). By late December, the number of Omicron infections in most regions of the UK, US, and Europe had doubled or tripled (Young et al. 2022). The variant's transmissibility and replication speed are astonishing, with a replication rate

approximately 70 times faster than the delta variant (Sheikh et al. 2022). Another notable feature is that the Omicron variant has remarkable reinfection capability; according to a study, the reinfection rate of this variant is more than ten times higher than previous variants (Pulliam et al. 2022). Fortunately, the severity of infections caused by the Omicron variant is lower than previous variants. In an early analysis in South Africa, the risk of hospitalization and severe illness among Omicron-infected individuals was lower than that of delta patients (Wolter et al. 2022). Although the Omicron variant appears to cause mild illness, high infection numbers could still lead to high hospitalization and mortality rates. This three-year-long COVID-19 pandemic has once again sounded the alarm for all humanity, as infectious diseases remain an ever-looming cloud, periodically shrouding humanity.

1.2 Mathematical Modelling of Infectious Diseases

When a disease outbreak occurs, scientists immediately seek to understand the new disease. Some scientists eagerly don their lab coats and examine new infectious diseases under microscopes, while others rush to their computers and code. Infectious disease epidemiologists, mathematical biologists, biostatisticians, and other professionals with relevant expertise are all striving to answer these questions: How fast does the infection spread? Which modes of transmission should be blocked to control the virus spread? What is the probability of hospitalization or death for infected individuals?

Theory and rigorous mathematical analysis are the foundations of ecology, evolutionary biology, and epidemiology. The idea of using mathematics to understand the spread of infectious diseases is older than the germ theory of disease itself. Fluid mechanics Daniel Bernoulli designed a model in 1760 to predict the benefits of smallpox vaccination (Bourouiba 2021). on the other hand, Nobel laureate Ronald Ross developed mathematical models advocating for controlling mosquito populations to reduce the transmission of malaria (Kevin Baird 2017).

A mathematical model can be imagined as a microscopic world or system, where the rules of this micro-world and system are precisely formulated, and entities governed by these accurately formulated rules constitute it. Mathematics provides humanity with a precise, clear language for formulating the behavioral rules of such systems and micro-worlds, thereby assisting in the clear construction and description of a hypothetical environment for human society. After the construction of a mathematical model, mathematical analysis, such as utilizing computer technology and algorithms for numerical simulations, can help humanity study the global behavior of the model, thereby obtaining results based on the assumptions made in the mathematical model. It is precisely because of the construction of mathematical models and the numerical simulation techniques of computer science algorithms that humans can truly predict the future behavior of constructed micro-worlds and systems in a relatively low-cost and interpretable manner while sitting indoors. Furthermore, we can explore the diversity of potential outcomes through the formulation and variation of system management rules.

In the field of infectious diseases, mathematical models allow humanity to gain a deeper understanding of the significant impact of an epidemic on a specific region, country, or even the entire global population by constructing micro-scale systems that describe the transmission of pathogens among hosts. Specifically, the construction of mathematical models for infectious diseases depends on factors such as the mode of contact between infectors and susceptible people, the latent period from infection to manifestation, the duration of infectiousness after infection, the transmissibility of the virus, the degree of immunity acquired by susceptible people after infection, the protective efficacy of vaccines if successfully developed, vaccine effectiveness, and the probability of vaccine breakthrough after administration.

Once all human-related factors are incorporated and expressed in the mathematical models of infectious diseases, researchers in this field can predict the expected number of infections during a pandemic, the duration of the epidemic, the peak time of infections, and even forecast the epidemic curve throughout the entire pandemic process. This enables medical institutions, regional and national public health agencies, and the World Health Organization to provide expected case numbers at each time point. In the case of terrifying infectious diseases that may lead to severe illness or death among infectors, factors related to hospitalization and mortality can also be incorporated into the constructed mathematical micro-world of infectious diseases. This allows for the prediction of information related to the number of severe cases and deaths during the pandemic, such as changes in mortality rates, peak mortality predictions, and expected mortality cases at different time points. All of this information provides

significant support to the World Health Organization, governments of various countries, and public health agencies, ultimately benefiting all of humanity.

Through literature search and reference, the first publication in human history discussing epidemic models can be traced back to 1776 (Bernoulli 1766). This ground-breaking paper described a mathematical model developed by the renowned scientist Daniel Bernoulli. The mathematical model was used at that time to analyze the mortality rate caused by smallpox in England. Smallpox, as a highly threatening infectious disease back then, resulted in a mortality rate accounting for one-fourth of the total death rate. Daniel Bernoulli's analysis with his model concluded that vaccinating against smallpox could increase the expected lifespan of newborns by about three years (Bernoulli and Blower 2004). Subsequently, in 1772, another scientist, Lambert, continued to explore Bernoulli's mathematical model, expanding the original model by incorporating parameters related to the age information (Lambert 1772). In 1911, a seminal article on modern mathematical epidemiology was published by Ross, marking the formal entry of the study direction of mathematical epidemiology into a systematic development path (Smith et al. 2012). In Ross's work, he proposed and utilized a mechanistic modelling approach, approximating the discrete-time dynamics of malaria transmission by mosquito-borne pathogens through equations.

In subsequent research, three pioneering papers were published by Kermack and McKendrick, which formally established and determined the compartmental infectious disease models (Kermack and McKendrick 1927, 1932, 1933). The emergence of compartmental infectious disease models laid a solid

foundation for later research on mathematical models of infectious diseases and served as cornerstone models widely applied in the study of mainstream infectious diseases such as influenza, COVID-19, and tuberculosis. The initial work by Kermack and McKendrick discussed large-scale events during the disease transmission cycle, exploring the probability of susceptibility similar to the frequency of contact between susceptible people and infectives. In their work, the rate at which susceptible become infected is denoted as κSI, where S represents the density of susceptible and I represents the density of infectives. In such a system, the recovery rate of infected individuals is λI, considering the possibility that recovered individuals may become susceptible again after recovery, implying that recovered individuals may become susceptible again, with a rate denoted as μR. The κ, λ, μ are analogical constants. Guldberg and Waage proposed the SIR model in 1864, suggesting strong similarities between this mechanistic deterministic representation and the law of mass action. The introduction of the SIR model also assumed homogeneous mixing, overall conservation, and relatively low interaction rates between individuals (Guldberg, Waage, and Lund 1864).

This continuum model describes the dynamic flow of infectious diseases within a population. Humans can establish and study disease evolution models through continuum models, which can also serve as functions of age and time after vaccination, and can be used to investigate the effects of quarantine and isolation on infected populations. For exploring such models, we can employ ordinary differential equations or partial differential equations. In the infectious disease dynamic systems established by these models, the population is divided

into different compartments based on different health statuses, such as susceptible people (S), infectives (I), and recovered people (R). In practical applications, additional compartments can be introduced to control and incorporate other population statuses relevant to policies, such as vaccination (V), hospitalization (H), death (D), and isolation (Q).

S (t), I (t), and R(t) represent their respective proportions in the population at time t, and this system is described by different equations:

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{1.1}$$

$$\frac{dR}{dt} = \gamma I$$

The derivatives dS/dt, dI/dt, and dR/dt represent the rates of change for the proportions S(t), I(t), and R(t) in the population over time. The transmission parameter β quantifies the average number of individuals infected by an infective per unit time, assuming all contacts between infectives and susceptible people lead to transmission. Therefore, diseases with higher infectivity exhibit higher β values. On the other hand, the parameter γ signifies the recovery rate, where $1/\lambda$ denotes the average duration of an infected individual's infectious period. The term $\beta S(t)I(t)$ denotes the total infection rate, indicating the proportion of the population infected per unit time at time t. In scenarios where only a small fraction of the population, represented by $I(t)$, is currently infected, their contacts with susceptible people would result in infecting a fraction $\beta I(t)$ of the population per unit time. However, considering that only

a small fraction $S(t)$ of the population remains susceptible, they would infect $\beta S(t)I(t)$ of the population per unit time.

The ratio β/γ, often referred to as the basic reproduction number R_0, serves as a vital metric for assessing the transmission of a pathogen. R_0 is defined as the average number of individuals an infected person infects within a fully susceptible population during the infectious period.

This foundational infectious disease model forms the basis for many intricate models and offers insightful predictions. Researchers can utilize numerical methods to solve the corresponding differential equations and specify initial values for parameters like beta, gamma, $S(0)$, $I(0)$, and $R(0)$. This allows them to generate epidemic curves, predicting the daily proportion of the population infected by the disease. Furthermore, utilizing historical epidemic data alongside the model enables the prediction of significant parameters such as the epidemic threshold, also known as the basic reproduction number R_0. R_0 represents the average number of infections caused by a single infected individual entering the susceptible population. If $R_0 < 1$, it indicates a rapid decrease in infections, suggesting no epidemic outbreak. Conversely, if $R_0 > 1$, regardless of the initial number of infected individuals, an epidemic will occur. This underscores that during an epidemic, the scale of disease development hinges on the initial proportion of susceptible individuals $S(0)$ and the basic reproduction number $R(0)$, rather than the initial number of infectors. It's important to note that the final scale of an epidemic remains smaller than the initial proportion of susceptible people $S(0)$, signifying the continual presence of an uninfected susceptible population.

In 2009, Gaudart et al. (Gaudart et al. 2009) made advancements to the SIR model by introducing an improved version known as the SIRS model. This was used to approximate the dynamic spread of malaria observed in the Bancoumana region of Mali between 1996 and 2001. Expanding upon the frameworks established by Ross and McKendrick, they integrated demographic data and genetic variations into the model to simulate the transmission of malaria in Mali and the historical plague outbreaks. Gaudart employed Archimedean copulas to link infection risk with biological age (Gaudart et al. 2010). In parallel studies, refined SIR models incorporated continuous age stratifications and the diffusion of human and vector subpopulations within infected regions by incorporating demographic and spatial dynamics. These investigations employed differential or partial differential equations to describe the generalized dynamics of disease propagation within populations.

Within the field of infectious disease modelling, stochastic models are also prevalent. These encompass Markov chain models that operate on discrete and continuous-time scales, focusing on individual-level dynamics. Such models relax the assumptions of mean-field approximations for infinitely large populations and introduce individual behavior variability. A notable example is the Discrete Markov Chain (DMC), where time and states are discretized, and transitions between states occur probabilistically based on predefined rules. For instance, Lekone et al. (Lekone and Finkenstädt 2006) utilized stochastic SEIR models to analyze the spread of the 1995 Ebola outbreak in the Democratic Republic of Congo. Bishai et al. (Bishai et al. 2011) developed a stochastic SIR model incorporating age structure and additional compartments to account for

vaccine heterogeneity. Similarly, Wang et al. (Wang et al. 2012) employed a stochastic SIR-based model to simulate and interpret the multi-wave patterns observed during the North American avian influenza outbreak, considering random interactions between individuals and the viral transmission environment.

When studying and addressing issues related to novel coronavirus, we utilize SEIR-based models for modelling and predicting the spread of infectious diseases. Specifically, in the research work that studies the reduction in the infection fatality rate of the Omicron variant compared with previous variants in South Africa, we used the Susceptible-Exposed-Infectious-Hospitalized-Death-Recovered-Vaccinated (SEIHDRV) model to the observed case and death data in South Africa. Here S, E, I, and R represent the number of susceptible populations, exposed populations, infectious populations, and recovered populations. The delay class V, H which represents vaccination and hospitalization was added between infections and death.

According to different research purposes, the SEIR-based model was used to do the simulation. In our work, the Susceptible-Exposed-Infectious-Hospitalized-Death-Recovered-Vaccinated (SEIHDRV) model was utilized to simulate weekly cases and deaths to study the infection-fatality ratio of patients carrying Omicron variants in South Africa. We simulated weekly cases as $C_{t+\Delta t}$ and as $D_{t+\Delta t}$, and we denoted the weekly reported cases and death cases as $Y_{t+\Delta t}$ and $Z_{t+\Delta t}$. We assumed the weekly reported cases and deaths were followed by negative binomial distributions and connected the reported cases/deaths and

simulated cases/deaths via two negative binomial distributions. Then the log-likelihood function can be defined.

We employ the Iterated Filtering method within a likelihood-based inference framework, specifically utilizing the Partially Observed Markov Process (POMP) approach (Ionides, Bretó, and King 2006; Ionides et al. 2011; He, Ionides, and King 2010b). This allows us to derive maximum likelihood estimates for the unknown parameters of the log-likelihood function. The Iterated Filtering algorithm serves as a valuable tool for inferring maximum likelihood in partially observed dynamical systems. To explore the parameter space, we introduce stochastic perturbations to the unknown parameters. Subsequently, we apply Sequential Monte Carlo, or the particle filter, to the extended model, enabling the selection of parameter values that align better with the observed data. If executed effectively, iterated filtering with progressively reduced perturbations will converge to the maximum likelihood estimate. The POMP package can be executed on high-performance workstations in either parallel or serial mode, facilitating the selection of large-scale models.

1.3 Outline of this thesis

This dissertation encapsulates the culmination of my doctoral research spanning over three years, during which I have closely observed the evolution and progression of the novel coronavirus. As a Ph.D. student specializing in epidemic mathematical modeling, I am gratified to have contributed scientifically to the global efforts against this relentless infectious disease.

My research endeavors have been dedicated to understanding the transmission patterns and mortality impacts of infectious diseases, with a particular focus on the context of COVID-19 within marginalized communities. The core objective of my thesis revolves around aiding these underserved areas in promptly mitigating the adverse consequences induced by the COVID-19 virus. Throughout my investigation, I have explored various facets, including variations in infection fatality rates and transmission dynamics across different strains of the novel coronavirus. Furthermore, my research has delved into the disparities in in-hospital mortality rates prevalent in underdeveloped regions. My

overarching goal is to utilize the theoretical insights gleaned from my studies to significantly contribute to humanity's endeavors in combating this formidable new viral threat.

Chapter 2 of this dissertation delves into our study on the reduction in the infectious disease fatality rate of the Omicron variant compared to previous variants in South Africa. We also estimate the transmissibility of three predominant SARS-COV-2 variants in South Africa, namely the Beta variant, the Delta variant, and the Omicron variant. This research commenced when the Omicron variants first emerged, and South Africa bore the brunt of the virus's impact. In this study, we developed a mathematical model incorporating time-varying transmission rates, vaccination data, and immune evasion characteristics. By fitting the model to reported case and death data up to February 6, 2022, we estimated the transmissibility and infection fatality ratio of the Omicron variant in South Africa. Our findings revealed that the substantially higher relative transmissibility of the Omicron variant was primarily attributed to its immune evasion capabilities, while its infection fatality rate decreased significantly by approximately 78.7% (with a 95% confidence interval of 66.9% to 85.0%) compared to previous variants. In subsequent research, we employed a state-of-the-art mathematical modeling approach to estimate the time-varying transmissibility of SARS-COV-2 and the relative transmissibility of the Beta, Delta, and Omicron variants. Our analysis indicated that the transmissibilities of these three variants were approximately 73%, 87%, and 276% higher, respectively, than their preceding variants. To the best of our knowledge, our model represents the first simple model capable of simulating multiple mortality

waves and the replacement of three variants in South Africa, highlighting the substantially higher transmissibility of the Omicron variant compared to previous variants.

Chapter 3 of the dissertation focuses on the investigation of regional disparities in in-hospital mortality rates of COVID-19 in Brazil. The COVID-19 pandemic has significantly impacted Brazil, with various dominant variants circulating throughout different periods. To explore the regional heterogeneity of in-hospital mortality rates of COVID-19 in Brazil and assess the effects of vaccinations and social inequality, we fitted a multivariate mixed-effect Cox model to a national database of inpatient data in Brazil spanning from February 27, 2020, to March 15, 2022. We compared the in-hospital mortality risks of vaccinated and unvaccinated patients while adjusting for age, state, ethnicity, education, and comorbidities. Furthermore, we analyzed the effects of these variables on in-hospital mortality and conducted stratified analyses across different age groups and vaccine types.

The reason that we did the research work about covid-19 in two different regions since there are notable connections between the research on COVID-19 in Brazil and South Africa:

Emergence of Variants: Brazil and South Africa were key locations for the discovery of the Gamma and Beta variants, respectively. Both variants share similar mutations that increase transmissibility and potential immune escape, suggesting common evolutionary patterns.

Vaccine Effectiveness: Studies from both countries have been crucial in assessing vaccine effectiveness against these variants. While some vaccines

showed reduced efficacy, they still provided protection, leading to further development of boosters and vaccine improvements.

Public Health Responses: Despite different public health strategies, both countries faced challenges like managing rapidly spreading variants. Their experiences offer valuable lessons for other nations in balancing health measures with economic considerations.

Immune Escape and Herd Immunity: Research in both Brazil and South Africa highlighted the potential for these variants to evade immunity, prompting reevaluation of herd immunity and long-term vaccination strategies.

In summary, the studies from Brazil and South Africa are interconnected through their focus on virus variants, vaccine effectiveness, public health policies, and immune response, providing important insights for global pandemic management.

Chapter 4 is the conclusion and discussion of my research work during these three years, over the past two years of the COVID-19 pandemic, large-scale infections have occurred globally, with surges in cases occurring one after another. Economically underdeveloped regions such as Brazil and South Africa have been particularly hard hit by the onslaught of the novel coronavirus. The public health systems in these areas have been severely strained, resulting in incalculable damage to national economies, government stability, and the lives and properties of the people. During the peak of the COVID-19 pandemic, the numbers of severe cases and deaths reached alarming levels, presenting one of the most severe public health crises in recent decades. The continuous mutations of the virus and the emergence of new variants have further complicated the

situation, with several key variants rapidly becoming predominant globally. These variants exhibit high infectivity and immune evasion capabilities, with significant changes in their biological characteristics such as the incubation period and clinical symptoms of carriers. This poses new challenges for epidemic prevention and control, vaccine development, and the formulation of public health measures and policies by governments in different regions and countries. As a researcher in the field of infectious disease mathematics, I hope to contribute to the best of my abilities by using my knowledge to quickly and efficiently simulate and track the trends of different virus strains, providing scientific assistance and support to the global public health system, especially to economically underdeveloped public health institutions in countries like Brazil and South Africa.

In Appendix part, we present findings on a notable synchronized replacement pattern whereby the Alpha variants of SARS-CoV-2 were supplanted by the Delta variant on a large scale. Our analysis suggests that this phenomenon is closely linked to the timing of the Delta variant's emergence and its inherent transmissibility advantage. Notably, the Alpha variant exhibited a tendency to skip certain countries and regions, such as India and its neighboring areas, resulting in a relatively mild initial wave of reported COVID-19 deaths per capita before the Delta variant's incursion.

This research represents a significant discovery made early in my PhD studies and marks the initial publication of my doctoral research journey. It underscores the rapid mutational dynamics of the novel coronavirus at that time and highlights the efficacy of scientific inquiry in understanding and responding

to emerging viral variants. Our study visualizes the global replacement process from the preceding strain to the Alpha variant and subsequently to the Delta variant.

We observed that the Alpha variant's dominance was relatively short-lived, spanning only 3 to 4 months, before being rapidly replaced by the Delta variant in a remarkably synchronized manner across numerous countries and regions. Specifically, the Delta variant first emerged in India and spread outward to neighboring areas before disseminating further globally. Conversely, the Alpha variants, originating in the UK, followed a similar pattern of regional spread before proliferating worldwide.

Our visual analysis reveals a striking simultaneous replacement of the Alpha variant by the Delta variant in multiple countries and regions. In contrast, the earlier substitution of the Alpha variant for the wild strain did not exhibit a similarly synchronized trend. This discrepancy may be attributed to the heightened transmissibility of the Delta variant compared to its predecessors, underscoring the dynamic nature of viral transmission dynamics during the COVID-19 pandemic.

Chapter 2

2 ASSESSING COVID-19 VARIANT DYNAMICS IN SOUTH AFRICA: REDUCED FATALITY RATE OF THE OMICRON VARIANT AND TRANSMISSIBILITY OF THE BETA, DELTA, AND OMICRON VARIANTS.

2.1 Reduction in the infection disease fatality rate of the Omicron variant compared with previous variants in South Africa.

2.1.1 Introduction

The COVID-19 pandemic has persisted for nearly two years since its initial identification in late 2019. According to the World Health Organization (WHO), the global tally has exceeded 260 million reported cases, with over 5 million deaths attributed to the virus. ('https://www.who.int/emergencies/diseases/novel-coronavirus-2019' 2021). Over this period, the virus has undergone multiple mutations, leading the WHO to classify variants into three categories: variants of concern (VOC), variants of interest (VOI), and variants under monitoring (VUM). Notably, four variants of concern—Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), and Delta (B.1.617.2)—have contributed significantly to infections and mortality worldwide. The emergence of the Omicron (B.1.1.529) variant on November 26, 2021, marked its classification as the fifth VOC. (He et al. 2021)

Before the advent of the Omicron variant, South Africa experienced three distinct waves of infections and deaths driven by different variants. The initial wave commenced in March 2020, reaching its peak in July and subsiding by September 2020 (Pulliam et al. 2021). The second wave, associated with the Beta

variant, began in October 2020, initially affecting Nelson Mandela Bay before spreading to Eastern Cape, Western Cape, and Kwazulu-Natal by early December 2020 (Tegally, Wilkinson, Giovanetti, et al. 2021). In May 2021, the emergence of the Delta variant triggered a third wave across South Africa. According to (Abdool Karim and Baxter 2021), The Delta variant swiftly supplanted the Beta variant, leading to a rapid surge in infections peaking in July, during which the Delta variant accounted for 86% of sequenced viruses in the first week of that month.

The emergence of the Omicron variant, which rapidly became the predominant strain in Gauteng, South Africa, was initially detected on November 23, 2021, coinciding with a significant surge in reported infections. By early December 2021, the number of confirmed Omicron variant cases in South Africa exceeded 2,000 (Vaughan 2021). Characterized as highly mutable, the Omicron variant exhibits remarkable genetic diversity with over 30 amino acid substitutions, three deletions, and one small insertion. Notably, 15 of these substitutions are concentrated within the receptor binding domain (RBD), including key mutations such as S371L, S373P, S375F, K417N, N440K, G446S, S447N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, and D614G ('Science Brief: Omicron (B.1.1.529) Variant' 2021). Additionally, the Omicron variant carries the mutation found in other variants of concern, in which a deletion was found at the peak position 60-79. It has three key mutations similar to the Beta variant and Gamma (P.1) variant, which may increase its ability to escape immunity (Poudel et al. 2022). Research by Zhang et al. demonstrated a reduced neutralization capability of convalescent sera from individuals

previously infected with early strains or the Delta variant against the Omicron virus, with neutralizing antibody titers declining significantly (36-fold and 39-fold reduction, respectively). These findings underscore the Omicron variant's potential to evade immunity, posing challenges for vaccine efficacy and natural immunity (Zhang et al. 2021).

Despite its increased transmissibility and immune evasion features, early observations suggest a potential difference in disease severity compared to previous variants. In Tshwane, South Africa, during the peak of the Omicron variant, bed occupancy rates were approximately half those observed during the Delta variant surge, indicating potentially lower hospitalization rates. Furthermore, fewer intensive care unit (ICU) admissions and shorter hospital stays may indicate reduced disease severity associated with the Omicron variant. (Abdullah et al. 2021).

This study aims to develop and apply a comprehensive modelling approach that integrates reported cases and deaths in South Africa to quantify the Omicron variant's impact on the infection fatality rate (IFR) and disease burden in the region. By elucidating the evolving dynamics and implications of the Omicron variant, this research seeks to inform public health strategies and interventions aimed at mitigating the spread and impact of this novel SARS-CoV-2 lineage.

2.1.2 Method

We obtain reported cases, deaths, excess deaths, and vaccination data from the Our World in Data ('Johns Hopkins University CSSE.' 2022; Hannah Ritchie 2020; Mathieu et al. 2021). We obtained aggregated variant proportion data from the Our World in Data, Covariant.org (Hodcroft 2021), and GISAID ((Shu and McCauley 2017b; Khare et al. 2021; Elbe and Buckland-Merrett 2017).

We fit our previously proposed Susceptible-Exposed-Infectious-Hospitalized-Death-Recovered-Vaccinated (SEIHDRV) model to the observed case and death data in South Africa (Lin, Chen, et al. 2021; Lin, Zhao, et al. 2021). We assume that a proportion denoted as ρ of the infections was reported and $\rho <$ 7%. This can be seen from that the total number of reported cases in the country is much smaller than the estimated proportion of the population being infected based on the seroprevalence from serological studies, (Musa et al. 2021). We incorporate the vaccination (fully vaccinated, or second dose) data, and fit our model to the reported cases and deaths. We denote the proportion of the Omicron variant as ω_t. We assume the infection fatality ratio (IFR) of the previous variant is IFR_1, and the IFR of the Omicron variant is IFR_2. Thus the overall IFR of a one-strain model is (1-ω_t) IFR_1+ ω_t IFR_2, i.e., a weighted average of the two IFRs. We assume a vaccine efficacy of 85% against both infection and death. Considering the high seroprevalence in South Africa, we assume that eventually 80-85% of the whole population were infected. We note that the re-infection of the Omicron is high, which means the Omicron variant has high immune evasion ability. The high relative transmissibility of the Omicron variant comes from two sources: namely the enlargement of the susceptible pool due to immune evasion

and the increase in the intrinsic transmissibility. We consider immunity evasion due to Omicron by allowing a proportion of recovered individuals to move to susceptible on Nov 9, 2021, when the Omicron variant was prevalent. We denote the size of the susceptible pool before Omicron evasion as S, then we consider four scenarios: the immune evasion causes the susceptible pool to increase by 0.25*S, 0.5*S, S, and 2*S.

Our model reads:

$$\begin{aligned}
\dot{S} &= -\frac{\beta SI}{N} - \tilde{v}S, \\
\dot{V} &= (1-\eta)\tilde{v}S - \frac{\psi\beta VI}{N}, \\
\dot{E} &= \frac{\beta SI}{N} + \frac{\psi\beta VI}{N} - \sigma E, \\
\dot{I} &= \sigma E - \gamma I, \\
\dot{H} &= \pi\gamma I - \kappa H, \\
\dot{D} &= \theta\kappa H, \\
\dot{R} &= \eta\tilde{v}S + (1-\pi)\gamma I + (1-\theta)\kappa H.
\end{aligned} \tag{2.1}$$

Here V denotes the vaccinated class which contains a proportion (denoted as $1-\eta$) of vaccinated individuals. The rest proportion of the vaccinated individuals (η) enter R class and gain long-term immunity. These vaccinated individuals in V class are susceptible to breakthrough infection. Parameter ψ accounts for the reduced susceptibility of vaccinated individuals. Here we assume

$\psi = 1$ for simplicity, since the effects of ψ and η compensate each other. A proportion (θ) of hospitalized individuals will eventually die, and this proportion decreases as vaccination coverage increases in the form $\theta = (1 - \varepsilon \int_0^t v(s)\, ds)\, \theta_0$. Namely, we assume that the risk of death drops while the vaccination coverage $\int_0^t v(s)\, ds$ increases, and we set $\varepsilon = 0.5$. We note that the vaccination will only be delivered to those who have not yet been vaccinated. Thus the vaccination rate $\tilde{v}(t)$ takes the form $\tilde{v}(t) = v(t)/(1 - \int_0^{t-1} v(s)\mathrm{ds})$, while $v(t)$ is the daily vaccination rate per capita. Here we only consider the fully vaccinated (second dose) data and ignore the effects of the first dose since the impact of the first dose would be overtaken by the impact of the second dose.

Parameter π denotes the risk of hospitalization or severe outcome of infected cases. Since we do not fit hospitalization or severe cases, we cannot estimate π, rather we can estimate the product of π and θ which is the IFR when $\varepsilon = 0$. We previously found that it is convenient to simply assume $\theta_0 = \pi$ without changing the fitting performance. We estimate the $\beta(t)$ which is an exponential cubic spline function (Vetterling et al. 1992) with 12 nodes spanning over the study period. We fix other parameters $\eta = 0.85$ which reflects the high efficacy of vaccine against both infection and deaths. We did not explicitly separate cases from natural infection and breakthrough infection.

Parameter Description	Symbol
Time-varying transmission rate	$\beta(t)$
Reduced susceptibility of vaccinated individuals	ψ
Proportion of fully protected individuals due to vaccination	η
Proportion of susceptible individuals who received COVID-19 vaccine per day	$\tilde{v}(t)$
Infectiousness onset rate	σ
Rate of loss of infectiousness	γ
Removal rate of hospitalized cases	κ
Ratio of severe cases out of all infected cases	π
Proportion of mortality out of severe cases	θ

Table 1. Parameter Description for the model

The mean latent period $\sigma^{-1} = 2$ days, $\gamma^{-1} = 3$ days and $\kappa^{-1} = 12$ days are fixed, such that the mean generation time (ie, sum of mean latent period and mean infectious period) equals 5 days (Tang et al. 2021) and the mean duration from infection to death is 17 days.

We simulate weekly cases $C_{t+\Delta t}$ and deaths $D_{t+\Delta t}$ as

$$C_{t+\Delta t} = \int_t^{t+\Delta t} \rho\sigma E dt \tag{2.2}$$

$$D_{t+\Delta t} = \int_t^{t+\Delta t} \theta\kappa H dt \tag{2.3}$$

We denote the weekly reported cases and deaths as $Y_{t+\Delta t}$ and deaths $Z_{t+\Delta t}$.

We assume them follows Negative-Binomial distribution:

$$Y_{t+\Delta t} \sim Negative\_Binominal\big(mean = C_{t+\Delta t}, variance = C_{t+\Delta t}(1 + \tau C_{t+\Delta t})\big) \tag{2.4}$$

$$Z_{t+\Delta t} \sim Negative\_Binominal(mean = D_{t+\Delta t}, variance = D_{t+\Delta t}(1 + \tau D_{t+\Delta t})) \tag{2.5}$$

Where τ is an over-dispersion parameter which will be estimated, and when it equals to 0, the negative binominal distribution is ging to reduce to Poisson distribution.

The log likelihood functions are:

$$Log\_Likelihood = \sum_{i=1}^{n} \log f(Y_i | Y_{1:(i-1)}, \Theta) \tag{2.6}$$

$$Log\_Likelihood = \sum_{i=1}^{n} \log f(Z_i | Z_{1:(i-1)}, \Theta) \tag{2.7}$$

Where Θ is the set of unknown parameters and $f(Y_i|Y_{1:(i-1)}, \Theta)$ are the conditional densities for Y_i given $Y_{1:(i-1)}$ which can be numerically calculated via Sequential Monte Carlo.

when studying the number of infectious disease cases, there is often spatial clustering. Such data does not meet the independence principle of the Poisson distribution. Usually, the variance in such data is significantly larger than the mean, making it overdispersed. When conducting Poisson regression with such data, the standard errors of the estimated model parameters tend to be underestimated. Therefore, when count data does not conform to the Poisson distribution, particularly in cases of overdispersion, it is more appropriate to use Negative Binomial regression analysis.

Thus we connect the reported cases/deaths and simulated cases/deaths via two Negative Binomial distributions. Thus the log likelihood can be defined (Lin et al. 2018; Zhao et al. 2018). We fit the model to reported cases and deaths via R package POMP (King, Nguyen, and Ionides 2015; He, Ionides, and King 2010a) and report the maximum likelihood estimate for IFR. The Iterated Filtering algorithm serves as a valuable tool for inferring maximum likelihood in partially observed dynamical systems. To explore the parameter space, we introduce stochastic perturbations to the unknown parameters. Subsequently, we apply Sequential Monte Carlo, or the particle filter, to the extended model, enabling the selection of parameter values that align better with the observed data. If executed effectively, iterated filtering with progressively reduced perturbations will converge to the maximum likelihood estimate. The POMP package can be

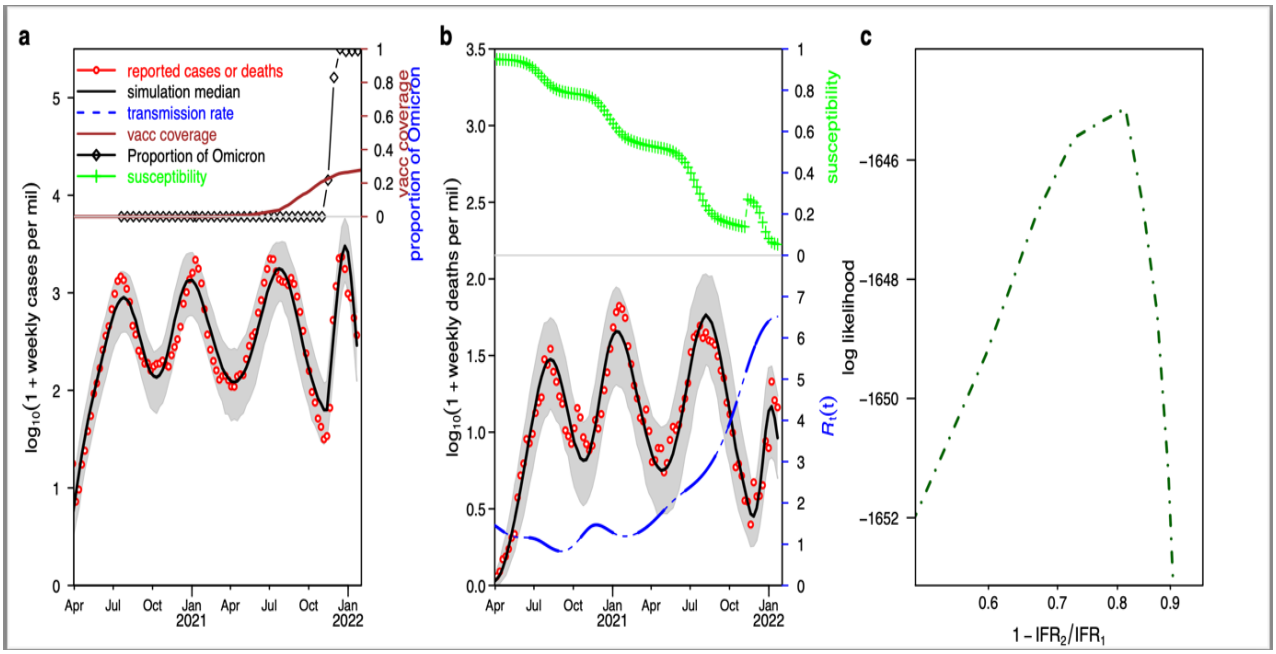
executed on high-performance workstations in either parallel or serial mode, facilitating the selection of large-scale models. This allows us to derive maximum likelihood estimates for the unknown parameters of the log-likelihood function. The 95% confidence interval is defined as the interval of IFR such that the log likelihood of the model given by the data drops by $0.5\chi^2_{0.95,df=1} = 1.92$ from the maximum log likelihood (He, Ionides, and King 2010a).

2.1.3 Result

We found that the COVID-19 cases and death reporting in South Africa were consistent over time. For instance, the reported COVID-19 death was consistently 1/3 of the excess deaths. The raw infection fatality rate (IFR) was consistent over time before the emergence of the Omicron variant. After the emergence of the Omicron variant, the raw IFR seemingly decreased significantly.

In **Figure 2.1**, we show our fitting result of four waves in South Africa. Our model simulations (black curve) match the reported case and death (red circle) reasonably well, with cases in panel (a) and deaths in panel (b). The estimated IFR_1 is about 0.21%, as we discussed we knew the reported death is only 1/3 of excess deaths and it was generally believed that the excess death is a good proxy of the true COVID-19, thus the true IFR could be 0.63% which was well in line with current knowledge of COVID-19 before Omicron (B.1.1.529) variant. We show the proportion of the fully vaccinated population and proportion of the Omicron variant among processed samples in the top panel in Panel a. We show

the overall susceptible level in the top panel in Panel b, where a sudden increase on Nov 9, 2021, can be noticed reflecting the effect of immune evasion due to Omicron variant. Panel (c) shows the log-likelihood profile versus the reduction in IFR. We find that the reduction in IFR is about 78.7% (95% confidence interval: 66.9%, 85.0%) under the scenario that the immune evasion enlarged the susceptible pool by 1-fold. **Table 1** shows the estimated reduction in IFR under four scenarios where immune evasion enlarged the susceptible pool by 0.25-fold, 0.5-fold, 1-fold, and 2-fold. The estimated reduction in IFR is consistent under the four scenarios. In **Figure 2.1**, we show the scenario when the susceptible pool is increased by 1-fold, and in **Figures 2.2, 2.3, and 2.4**, we show the scenario when the susceptible pool is increased by 0.25-fold, 0.5-fold, and 2-fold by immune evasion. The fitting performance is greatly improved, reflected in the maximum log likelihood (MLL), when the immune evasion increased from 0.25-fold to 2-fold of pre susceptible level. Thus the relative transmissibility of the Omicron variant is likely 3-fold of that of the Delta variant. If we assume the immune evasion is at a low level, e.g. 0.25-fold of the pre-susceptible level, then the transmission rate, in units of basic reproductive number, needs to go up to as high as 9.



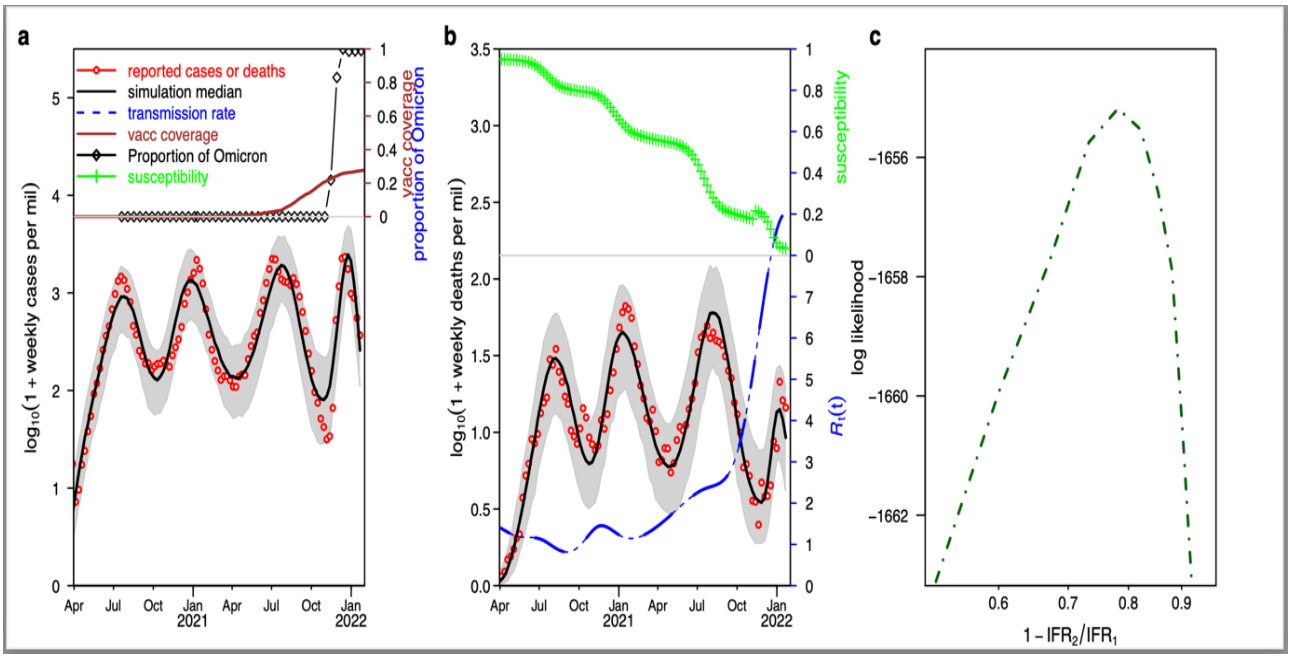
2.1 Fitting model for the 1-fold susceptible pool to reported cases and deaths in South Africa.

The red circle and black curve represent reported cases or deaths, and simulated cases or deaths.

(a) reported cases versus simulated cases, and proportion of fully vaccinated and proportion of Omicron variant among samples processed in the top panel.

(b) reported deaths versus simulated deaths, and simulated susceptible level (green curve at the top), where the susceptible pool doubled on Nov 9, 2021, due to the immunity evasion of the Omicron variant; The dashed blue curve showed the estimated transmission rate in the unit of $\mathcal{R}_t = \frac{\beta(t)}{\gamma}$.

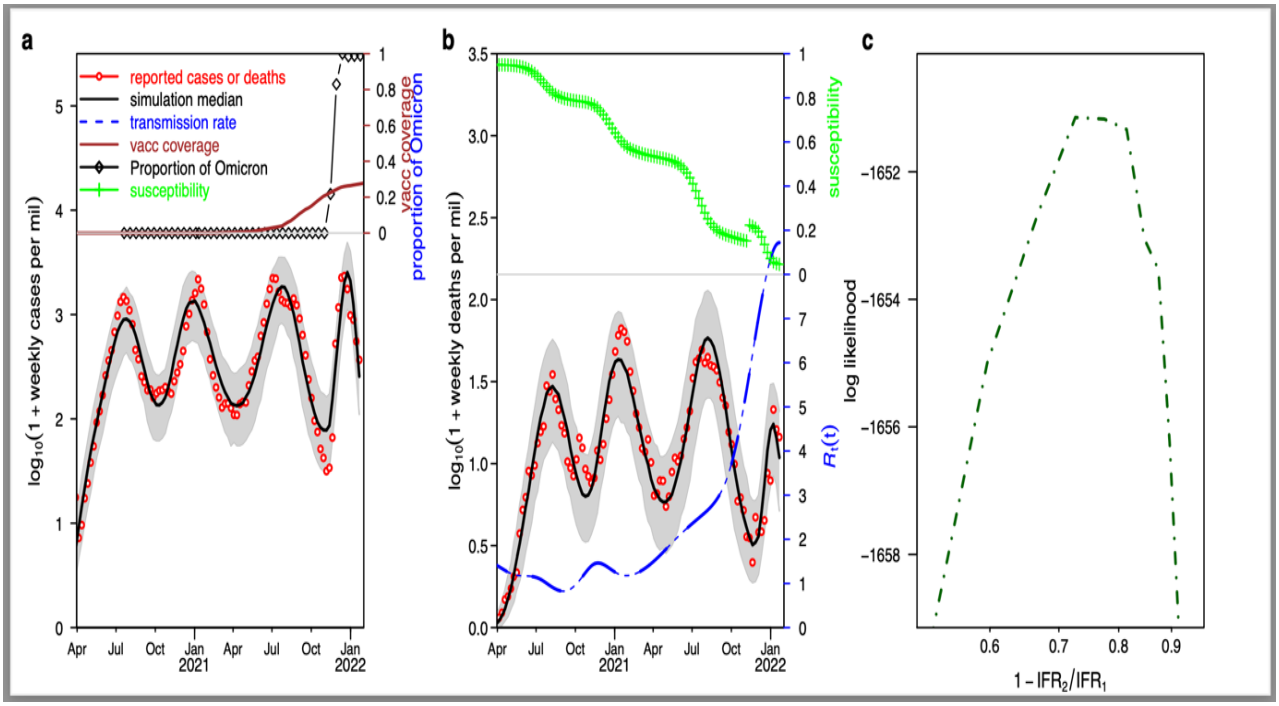
(c) The log-likelihood profile as a function of reduction in IFR before and after Omicron evasion



2.2 Fitting model for the 0.25-fold susceptible pool to reported cases and deaths in South Africa.

The red circle and black curve represent reported cases or deaths, and simulated cases or deaths. For part a, the upper panel contrasts reported cases with simulated cases, alongside the proportions of fully vaccinated individuals and the Omicron variant among samples processed. Part b visualized reported deaths against simulated deaths, accompanied by the simulated susceptible level. Notably, the green curve reflects the doubling of the susceptible pool on Nov 9, 2021, attributed to the immunity evasion of the Omicron variant. Additionally, the dashed blue curve illustrates the estimated transmission rate denoted in $\mathcal{R}_t =$

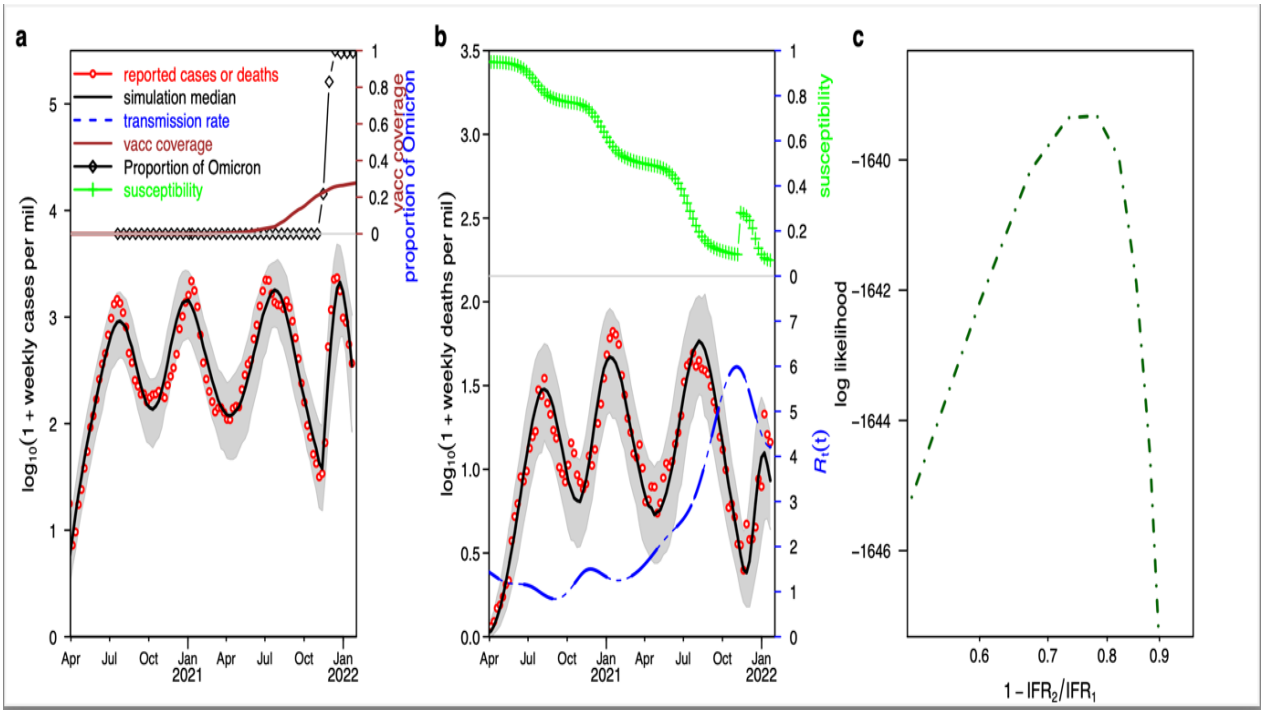
$\frac{\beta(t)}{\gamma}$. Part c shows the log-likelihood profile concerning the reduction in IFR before and after Omicron evasion provides insight into the impact of the variant on disease severity.



2.3 Fitting model for the 0.5-fold susceptible pool to reported cases and deaths in South Africa.

The red circle and black curve represent reported cases or deaths, and simulated cases or deaths. For part a, the upper panel contrasts reported cases with simulated cases, alongside the proportions of fully vaccinated individuals and the

Omicron variant among samples processed. Part b visualized reported deaths against simulated deaths, accompanied by the simulated susceptible level. Notably, the green curve reflects the doubling of the susceptible pool on Nov 9, 2021, attributed to the immunity evasion of the Omicron variant. Additionally, the dashed blue curve illustrates the estimated transmission rate denoted in $\mathcal{R}_t = \frac{\beta(t)}{\gamma}$. Part c shows the log-likelihood profile concerning the reduction in IFR before and after Omicron evasion provides insight into the impact of the variant on disease severity.



2.4 Fitting model for the 2-fold susceptible pool to reported cases and deaths in South Africa.

The red circle and black curve represent reported cases or deaths, and simulated cases or deaths. For part a, the upper panel contrasts reported cases with simulated cases, alongside the proportions of fully vaccinated individuals and the Omicron variant among samples processed. Part b visualized reported deaths against simulated deaths, accompanied by the simulated susceptible level. Notably, the green curve reflects the doubling of the susceptible pool on Nov 9, 2021, attributed to the immunity evasion of the Omicron variant. Additionally, the dashed blue curve illustrates the estimated transmission rate denoted in $\mathcal{R}_t = \frac{\beta(t)}{\gamma}$. Part c shows the log-likelihood profile concerning the reduction in IFR

before and after Omicron evasion provides insight into the impact of the variant on disease severity.

Table 1 Estimated reduction in IFR in four scenarios of immune evasion.		
Immune evasion	**Maximum likelihood estimate of reduction in IFR**	**95% confidence interval**
0.25	0.794	0.683, 0.869
0.5	0.777	0.652, 0.861
1	0.787	0.669, 0.85
2	0.778	0.636, 0.842
IFR, infection fatality ratio.		

Table 2. Estimated reduction in IFR under four scenarios of immunity evasion.

For 1 fold scenarios, the IFR of the Omicron variant is about 78.7% with a 95% confidence interval (66.9%, 85.0%) of that of the previous variant.

2.1.4 conclusion and discussion

In our analysis, we operate under the assumption of consistent case testing and reporting efforts, although the veracity of this assumption may not always hold true. Particularly, with the emergence of a new variant such as the Omicron variant, there might be an intensification of the testing effort, potentially resulting in an underestimation of the Infection Fatality Rate (IFR) attributable to transient effects. Furthermore, our study incorporates various degrees of immunity evasion stemming from the Omicron variant within a single-strain model; however, it is noteworthy to consider the alternative approach of employing a multiple-strain model to further elucidate the dynamics. The temporal scope of our analysis is confined to data available up to Feb 9, 2022, with the overarching objective of providing an early estimate of the IFR specific to the Omicron variant in South Africa. It is crucial to acknowledge that our estimated IFR pertains to a context characterized by a high seroprevalence (infection attack rate) in South Africa, thus may not accurately reflect the intrinsic IFR of the Omicron variant within a population predominantly susceptible to infection.

Drawing a comparative perspective, empirical observations from Hong Kong's fifth wave of the Omicron variant, as of April 4, 2022, reveal a raw Case Fatality Rate (CFR) of 0.69%, marking a significant reduction from the preceding raw CFR of 1.38% recorded up to February 6, 2022 ('Latest situation of COVID-19 (as of 4 April 2022)' 2022). It is imperative to note that this observed reduction

of 50% could potentially be even more substantial, considering the likelihood of higher under-reporting of cases during the fifth wave relative to preceding waves. This reduction in CFR can be attributed in part to intrinsic features of the Omicron variant and in part to the protective effects conferred by vaccination campaigns. Specifically, the raw CFR among the unvaccinated population during the first wave stands at 2.05%, ostensibly surpassing the 1.34% CFR recorded in preceding waves. However, accounting for the considerable underreporting of cases during the initial wave, it is conceivable that the CFR of the Omicron variant among the unvaccinated cohort may actually be substantially lower than that of previous variants. ('Latest situation of COVID-19 (as of 4 April 2022)' 2022).

In summary, we found that the relative transmissibility of the Omicron variant (including due to immunity escaping) could be more than 3-fold higher than previous variants, which is in line with our previous estimate (Yu et al. 2021). Immune evasion is the main reason behind the high relative transmissibility of the Omicron variant. For one-fold scenario, the IFR of the Omicron variant is about 78.7% with a 95% confidence interval (66.9%, 85.0%) of that of the previous variant. For 0.25-fold scenario, the IFR of the Omicron variant is about 79.4% with a 95% confidence interval (68.3%, 86.09%) of that of the previous variant. For 0.5-fold scenario, the IFR of the Omicron variant is about 77.7% with a 95% confidence interval (65.2%, 86.10%) of that of the previous variant. For two-fold scenario, the IFR of the Omicron variant is about 77.8% with a 95% confidence interval (63.6%, 84.20%) of that of the previous variant.

2.2 Estimate the transmissibility of SARS-COV-2 Beta, Delta and Omicron variants in South Africa.

2.2.1 Introduction

The coronavirus disease 2019 (COVID-19) spread rapidly and ravaged the world in a short time. As of December 18, 2021, 271,963,258 confirmed cases and 5,331,019 deaths had been recorded (WHO, https://covid19.who.int/), seriously affecting global public health. The rapid mutation rate of the virus is an important reason for its huge and long-lasting impact. Currently, several variants of the virus have emerged, causing multiple peaks of COVID-19 infection worldwide. In South Africa, the Wild strain, Beta variant, Delta variant, and Omicron variant have emerged and dominated. The newly emerged variant viruses show stronger infectivity and weaken the effectiveness of vaccines (Karim and Karim 2021; Tegally, Wilkinson, Althaus, et al. 2021).

In October 2020, the Beta variant was first discovered in South Africa. It triggered the second wave of outbreaks in the country (Cele et al. 2021). Compared to the Wild strain, the Beta variant spread rapidly in South Africa and increased infectivity and immunity evasion. Additionally, the effect of the

ChAdOx1 nCoV-19 vaccine on the Beta variant was severely weakened (Madhi et al. 2021; Tegally, Wilkinson, Giovanetti, et al. 2021). Therefore, the Beta variant has higher reinfection characteristics than the Wild strain (Chemaitelly, Bertollini, and Abu-Raddad 2021). However, compared to the Beta variant, the Delta variant is much more fatal.

In December 2020, the Delta variant was first detected in Maharashtra, India. It spread rapidly to other countries and regions (Del Rio, Malani, and Omer 2021). Five studies estimated the basic reproductive number, $\mathcal{R}_0$, of the Delta variant and indicated that the range of $\mathcal{R}_0$ of the Delta variant is 3.2–8, with an average value of 5.08 (Liu and Rocklöv 2021), which is significantly higher than those of the Alpha variant (Del Rio, Malani, and Omer 2021) and Wild strain (Liu and Rocklöv 2021). Some studies indicated that the increased replication suitability and decreased sensitivity to neutralizing antibodies of the Delta variant have led to a greatly increased infectivity of the Delta variant (Mlcochova et al. 2021). However, the Omicron variant, as the mutant strain with the most mutation sites currently during the COVID-19 pandemic, seems to have a higher transmission rate, lower vaccine efficiency, and higher reinfection risk (Torjesen 2021b). To explore the transmissibility of the Delta variant, Ito et al. reported the predominance of the Delta variant in the run-up to the July 2021 Olympics in Tokyo, Japan. The authors used a renewal-equation-based model which is different from our model to describe the adaptive evolution of multiple variants in Japan and demonstrated that the Delta variant was more transmissible than its predecessor, with a transmittance 1.4 times higher than that of the Alpha variant (Ito, Piantham, and Nishiura 2021).

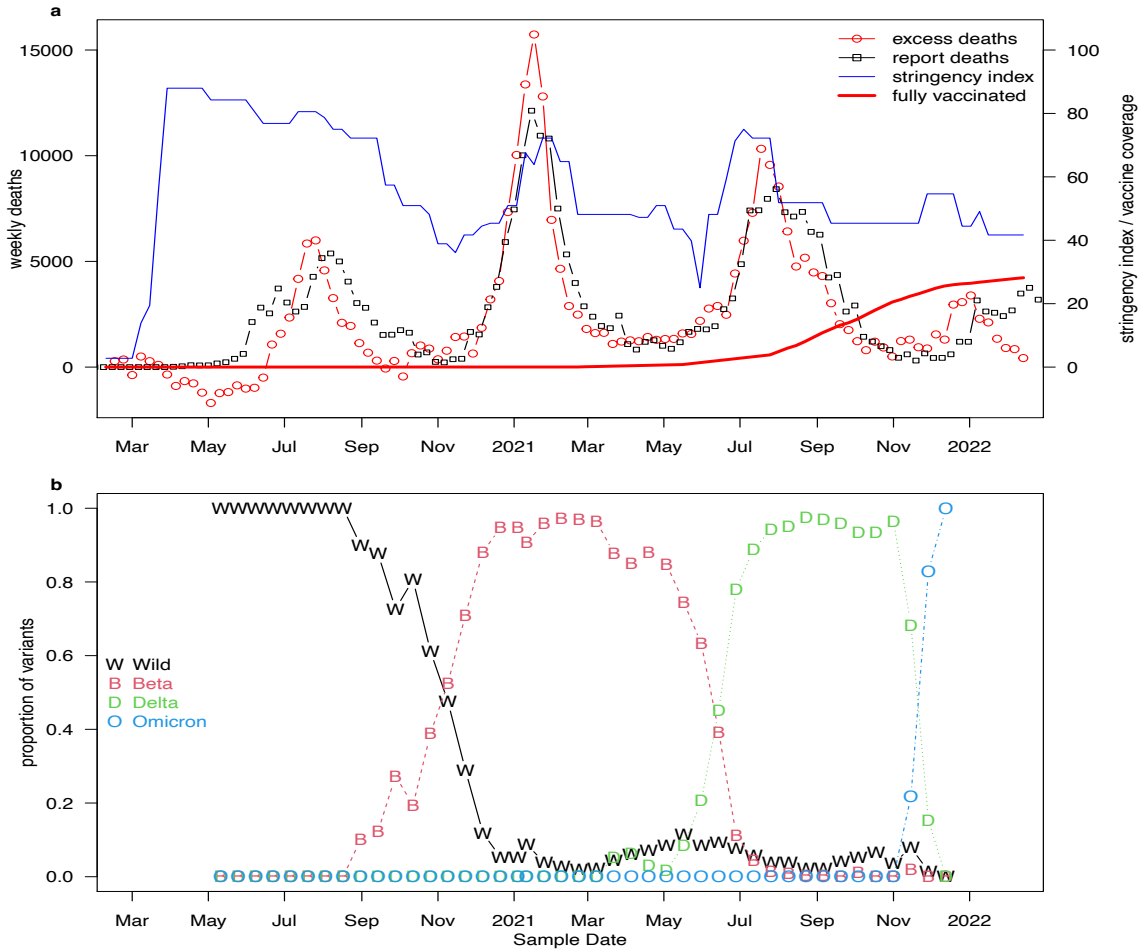
The Omicron variant was first discovered in South Africa on November 9, 2021, and was classified variant of concern by WHO on November 26, 2021 (Gu et al. 2022). As of December 16, 2021, the Omicron variant has existed in 89 countries and regions, and it is spreading at an unprecedented speed. In South Africa, the variant quickly replaced the Delta variant and caused a rapid increase in the number of infections (Dyer 2021). The number of daily cases rose rapidly from 273 cases on November 17, 2021 to above 26,389 cases on December 16, 2021 (WHO, https://covid19.who.int/).

Presently, several studies have been conducted on the Omicron variant. Some of these studies investigated Omicron's vaccine breakthrough rate and antibody resistance through *in vitro* experiments and clinical research. In an *in vitro* experimental study on the SARS-CoV-2 variants, Wilhelm et al. demonstrated that the neutralizing effect of the vaccine against the Omicron variant was severely reduced, compared to the Delta variant (Wilhelm et al. 2021). Furthermore, Zhang et al. demonstrated that the Omicron variant may lead to more obvious evasion of immunity in an *in vitro* study (Zhang et al. 2022). Karim et al. compared the neutralizing titers of the Omicron variant with those of the Victoria, Beta, and Delta variants, and indicated that the Omicron variant will cause more breakthrough infections, which may trigger further infection waves (Karim and Karim 2021; Dejnirattisai et al. 2022). Mohiuddin et al. used the reduction in neutralizing antibody titers to infer vaccine effectiveness. The study reported that the effect of the vaccine on the Omicron variant was severely reduced, the effectiveness of vaccines against severe illnesses was significantly reduced for frail individuals, and the protection against infection, mild illness,

and transmission was almost eliminated (Gardner and Kilpatrick 2021). Kuhlmann et al. revealed that three doses of the mRNA vaccine may not be enough to prevent infection and symptomatic diseases of the Omicron variant based on clinical studies on patients (Kuhlmann et al. 2021). Additionally, Nishiura et al. reported the relative reproductive numbers of Omicron and Delta Variants in South Africa by a mathematical model different from ours(Nishiura et al. 2021). The authors assumed that the effective reproduction number of the Omicron variant, $R_{Omicron}(t)$ was given by multiplying a constant factor k to that of the Delta variant, $R_{Delta}(t)$. This research paper reported that the effective reproduction number of the Omicron variant was estimated to be 4.2 times more than the Delta variant, and 3.3 times more transmissible than the Delta variant.

Some studies used theoretical models to investigate the effects of the Omicron variant. Bai et al. analyzed the population movement data obtained from both flights in South Africa and the Omicron case report data and estimated that the probability of the Omicron variant being introduced into the studied country before November 28, 2021, was higher than 50% (Bai et al. 2021). Kumar et al. studied the spike proteins of the Omicron and Delta variants using several computational tools and a computational saturation mutagenesis model. They found that the Omicron variant has a higher affinity for human angiotensin-converting enzyme 2 (ACE2) receptors than the Delta variant, indicating that the Omicron variant has a higher transmission potential (Kumar, Thambiraja, et al. 2021). In another study on the Omicron variant's infectiousness, vaccine breakthrough, and antibody resistance using an artificial intelligence model, the variant's infectivity was found to be above 10 times higher than that of the Wild

virus or about twice that of the Delta variant. Vaccine breakthrough was twice that of the Delta variant, and antibody resistance had been weakened (Chen et al.). Kuhlmann et al. (Kuhlmann et al. 2021) used the meta-analysis method to predict that after 6 months of initial immunization with mRNA vaccines, the vaccine's efficacy on the symptoms of patients infected by the Omicron variant was estimated to have reduced to about 40%. Additionally, the efficacy on severe diseases had decreased to about 80% (Khoury et al. 2021). Furthermore, the OpenCOVID individual-based model was used to compare the infectiousness, severity, and immune evasion properties of the Omicron and Delta variants. The model indicated that the Omicron variant could become the new dominant variant (Le Rutte et al. 2021).



2.5 Weekly reported excess deaths and reported COVID-19 deaths, stringency index, and vaccination coverage (a) and variant proportion (b) in South Africa.

Figure 2.5.a shows the weekly reported deaths and excess deaths with stringency index and vaccination coverage in South Africa. Data are downloaded from (Hannah Ritchie 2020; Shu and McCauley 2017b; Khare et al. 2021; Elbe and Buckland-Merrett 2017; Mathieu et al. 2021; Hale et al. 2021).

According to the figure, the number of excess deaths was about three-fold that of the reported deaths, the number of weekly deaths varies largely with the stringency index and variant invasion, and vaccination could effectively reduce

the number of deaths. **Figure 2.5.b** shows the trend of infection by the four COVID-19 virus variants over time in South Africa. The Beta variant gradually replaced the wild strain from September 2020 to January 2021, then the Delta variant gradually replaced the Beta variant from May to August 2021. After November, the Omicron variant replaced the Beta variant and became the dominant variant within a month.

In this work, we propose a new model and fit the model to the adjusted COVID-19 deaths and the proportion of variants in South Africa to estimate the relative transmission rates of the Beta, Delta, and Omicron variants, compared to their preceding dominant strains, which were the Wild strain, Beta variant, and Delta variant, respectively.

2.2.2 Materials and Methods

The susceptible-exposed-infectious-hospitalized-recovered-death (SEIHRD) model from our previous studies (Song, Fan, Zhao, et al. 2021; Song, Fan, Liu, et al. 2021) was adopted and extended to simulate the alternative dominance of the Wild strain, Beta variant, Delta variant, and Omicron variant. Based on the assumption that the variant replacement only affected the overall transmission rate of COVID-19 and only transmission changes but no variant changes occurred, the SEIHRD model with a flexible time-varying transmission rate could simulate COVID-19 deaths (or excess deaths) (Musa et al. 2022). To simulate the replacement of the proceeding strain by a variant, an additional set of (EIHRD) equations for the variant were included. Since the replacements occurred thrice in South Africa, one set of the SEIHRD model for the Wild strain and three additional sets of EIHRD models for the Beta, Delta variant, and Omicron variants, amounting to 21 equations, were used. However, one set of SEIHRD and one additional set of EIHRD were sufficient. Therefore, only 8 equations were used if further merging HRD classes for two successive variants. At any moment, at most two strains (or variants) dominated. Therefore, a system of two groups of models was sufficient to simulate the replacement.

The dominance of the variants could be divided into several time intervals:

(1) Before the emergence of the Beta variant, $(SEIHRD)_1$ was used to model the dynamics of the wild strain;

(2) After the emergence of the Beta variant, $(EIHRD)_2$ was used to model the dynamics of the Beta variant;

(3) After the emergence of the Delta variant, by which the Wild stain had almost been replaced, $(SEIHRD)_1$ was reused to model the dynamics of the Delta variant;

(4) After the emergence of the Omicron variant, by which the Beta variant had been replaced, $(SEIHRD)_2$ was reused to model the dynamics of the Omicron variant.

Therefore, only 8 equations were used to simulate the successive replacements of one Wild strain and three variants. This system can be used to model further replacements in principle. To the best of our knowledge, our model is the simplest model for this purpose.

Our model reads as follows:

$$\dot{S} = -\frac{\varepsilon_1 \beta S I_1}{N} - \frac{\varepsilon_2 \beta S I_2}{N},$$

$$\dot{E_1} = \frac{\varepsilon_1 \beta S I_1}{N} - \sigma E_1,$$

$$\dot{E_2} = \frac{\varepsilon_2 \beta S I_2}{N} - \sigma E_2,$$

$$\dot{I_1} = \sigma E_1 - \gamma I_1, \tag{2.8}$$

$$\dot{I_2} = \sigma E_2 - \gamma I_2,$$

$$\dot{H} = \pi\gamma(I_1 + I_2) - \kappa H,$$

$$\dot{D} = \theta\kappa H,$$

$$\dot{R} = (1-\pi)\gamma I_1 + (1-\pi)\gamma I_2 + (1-\theta)\kappa H.$$

If $t \in [0, t_1] \cup [t_2, t_3]$ and $E_1 < 1$, then $E_1 = 1, S = S - 1$. If $t \in [0, t_1]$, $\varepsilon_1 = \varepsilon_2 = 1$, If $t \in [t_1, t_3]$, $\varepsilon_2 = \eta_1$, If $t \in [t_2, t_4]$, $\varepsilon_1 = \eta_1\eta_2$, If $t \in [t_3, t_4]$, $\varepsilon_2 = \eta_1\eta_2\eta_3$.

We simulate weekly deaths $D_{t+\Delta t}$ as

$$D_{t+\Delta t} = \int_t^{t+\Delta t} \theta\kappa H dt \tag{2.9}$$

We denote the weekly reported deaths as $Z_{t+\Delta t}$.

We assume

$$Z_{t+\Delta t} \sim Negative\_Binominal(mean = D_{t+\Delta t}, variance = D_{t+\Delta t}(1 + \tau D_{t+\Delta t})) \tag{2.10}$$

The log likelihood function noted as:

$$Log\_Likelihood = \sum_{i=1}^{n} \log f(Z_i | Z_{1:(i-1)}, \Theta) \tag{2.11}$$

We assumed the transmission rate was an exponential cubic spline function, we used n_β=13, which meant there were 13 nodes in the cubic spline evenly distributed over the study period. We denote these 13 nodes as (t_i, b_i), where i=1 to 13, $t_i = \frac{t_{start}+(\mathrm{i}-1)}{12} * (t_{end} - t_{start})$, b_i were positive values to be estimated via fitting model to data. Given our model and parameter setting (including fixed parameter and unknown parameters) and data, we used standard iterated filtering to achieve the maximum log-likelihood estimates of all unknown parameters

which included these 13 b_i. The transmission rate is shown in **Figure 2.2.2** as a dashed blue curve.

According to **Figure 2.5**, the number of excess deaths was about three-fold that of the reported deaths and excess deaths had negative values. Therefore, reported COVID-19 deaths was multiplied by a factor of 3, resulting in the adjusted COVID-19 deaths.

The plug-and-play likelihood-based inference framework (He, Ionides, and King 2010a) was adopted and implemented in the R package POMP (Ionides, Bretó, and King 2006) to fit the model.

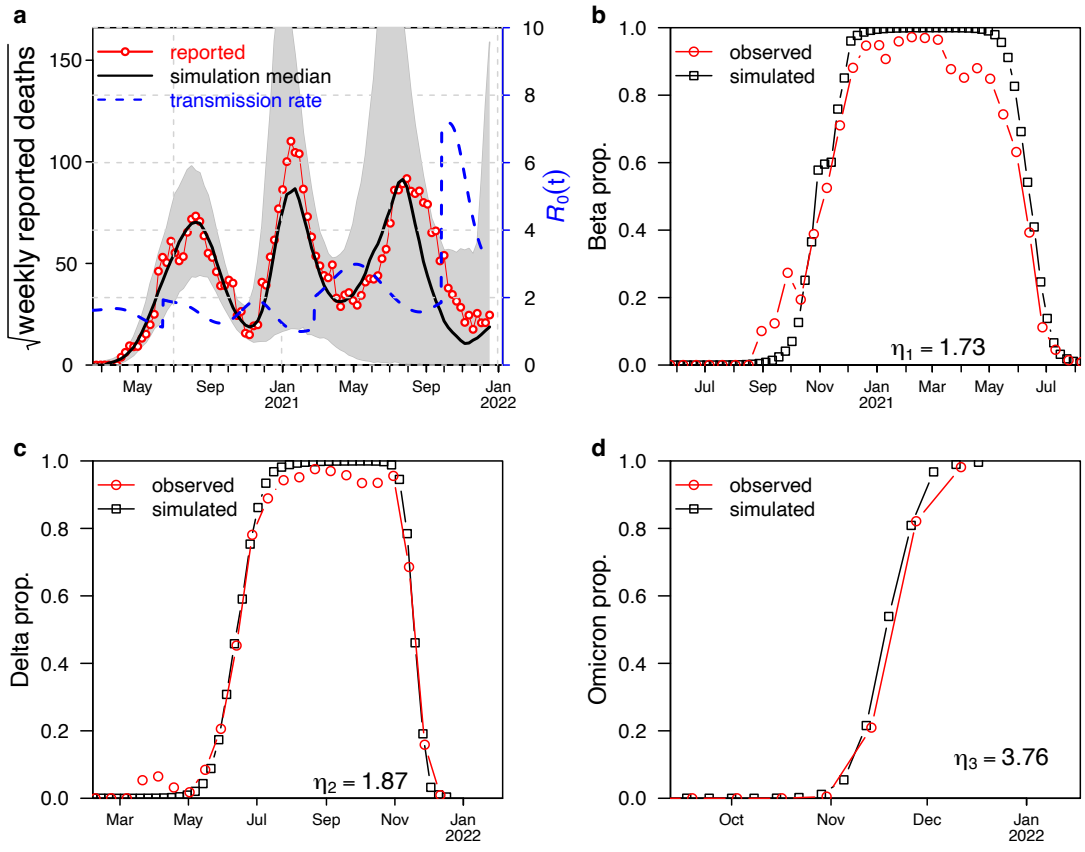
For COVID-19 deaths, a negative binomial measurement model was used to link the simulated model and adjusted reported deaths. For the variant proportions, out of several options, a simple approach was adopted. For the three time intervals, based on the replacement process of the variants, the sum of squared errors between the simulated and reported proportions was calculated. The simulated proportion was defined as either $E_1/(E_1 + E_2)$ or $E_2/(E_1 + E_2)$, which is the ratio of exposed cases of the variant to the exposed cases of both the variant and the preceding strain or variant. The transmissibility of the Beta, Delta, and Omicron variants was denoted as η_1-fold of that of the Wild strain, η_2-fold of that of the Beta strain, and η_3-fold of that of the Delta strain, respectively.

A comparison of the fitting performance of models with different number of nodes in the transmission rate and the second-order Akaike Information Criterion (AICc) revealed that $n_\beta = 13$ yielded the smallest AICc. The emergence times of the three variants were fixed on June 12, 2020, February 23, 2021, and

September 27, 2021. These dates were 1–2 months ahead the first time these variants were reported on August 31, 2020, March 22, 2021, and November 1, 2021. We introduce one exposed case of the dominant variants if there is no exposed case of the dominant variants in a day to mimic the continuous importation of cases.

Aggregated variant proportion data, vaccination data, and stringency index data were obtained from the compiled data source in The Our World in Data (Hannah Ritchie 2020) and original from the GISAID (Shu and McCauley 2017b; Khare et al. 2021; Elbe and Buckland-Merrett 2017; Hale et al. 2021; Mathieu et al. 2021).

2.2.3 Results



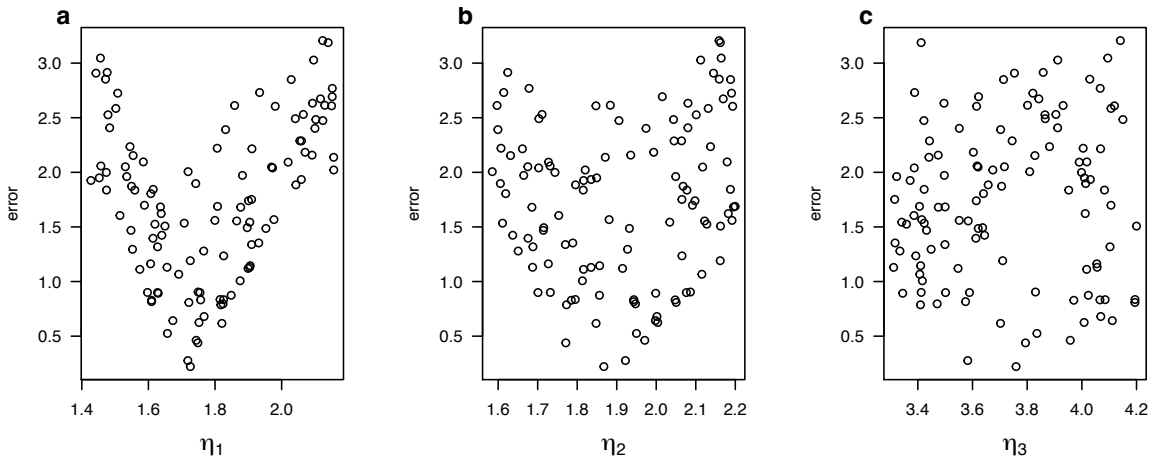
2.6 Results of the model fitting

Figure 2.6 shows the results of the model fitting. Panel 2a shows the model of the simulated deaths and adjusted COVID-19 deaths. Our model simulation largely matched the observed three-fold adjusted COVID-19 deaths in South Africa, with an estimated infection fatality rate of about 1%. Panels b–d show the

simulated proportions of the Beta, Delta, and Omicron variants and the observed proportions. The simulated and observed proportions were closely matched. The estimated relative transmission rate factor is shown in Panels b–d.

(A) Fitting results to three-fold weekly reported deaths (black curve and red circles as the simulated and observed deaths, respectively) and the fitted transmission rate in the units of basic reproduction number (blue dashed curve). n_β=13 nodes were used in the cubic spline (Vetterling et al. 1992) for the transmission rate. The grey region indicates the 95% range of the 1,000 random simulations.

(B-D) The simulated and observed proportions of the Beta, Delta, and Omicron variants among all samples were sequenced.



2.7 The sum of squared errors as functions of η1, η2, and η3 for the model.

These results suggest that our estimates are robust. As indicated, 200 samples were sampled for each of η_1, η_2, and η_3, from a range. They were fixed while fitting the model to adjusted deaths. The sum of squared errors for the three variants was calculated.

Simulated and observed proportions were closely matched. The estimated $\eta_1 = 1.73$, $\eta_2 = 1.87$, and $\eta_3 = 3.76$ are similar to the estimates for the Beta, Delta, and Omicron variants, reported by previous studies at 1.69-fold (Tegally, Wilkinson, Althaus, et al. 2021), 1.65-fold (Campbell and Gustafson 2021), and 3.8-fold of their preceding strain or variant, respectively. In **Figure 2.7**, 200 groups of (η_1, η_2, η_3) values were sampled, and the model was refitted. The sum of squared errors between observed and simulated proportion time series was

calculated. The results in Figure 2.7 suggest that our estimates of η_1, η_2, and η_3 are robust.

In our model, we ignored the effect of re-infection and the effects of vaccination because the coverage of fully vaccinated individuals was relatively low and the invasion of the Omicron variant occurred at the end of the study period. Allowing re-infection would mean a proportion of recovered individuals would become susceptible, leading to an increase in the susceptible pool, i.e., S→S+ΔS. However, concurrently, the transmission rate was increased due to the increased transmissibility of the Omicron variant, i.e., β→β+Δβ. Therefore, the mass action term in our model became βSI→(β+Δβ)(S+ΔS)I = βSI + (ΔβS+ΔSβ+ΔβΔS)I. If possible, disentangling Δβ and ΔS from fitting this type of model to aggregated death data would be difficult, when the time interval covering the Omicron variant was short. The effects of Δβ and ΔS are exchangeable. Therefore, ignoring the immunity evasion-induced ΔS, i.e., assuming ΔS = 0, and synthesis of all effects into Δβ to estimate Δβ was appropriate. Additionally, in interpreting Δβ, Δβ should be emphasized to include both effects from Δβ and ΔS. Furthermore, we sought to estimate how fast the Omicron variant transmits relative to the Delta variant. Other types of study, such as case-control studies, are needed to reveal the underlying mechanism. The sizes of the susceptible pools for a variant, e.g., Omicron, and its preceding variant, e.g., Delta, may be assumed to be the same (i.e., if ΔS = 0, we have $\mathcal{R}_0$(Omicron) = η_3*$\mathcal{R}_0$ (Delta)).

We argue that $\mathcal{R}_0$ (Omicron) could be further categorized into three components: $\mathcal{R}_0$ (reinfection), $\mathcal{R}_0$ (breakthrough), and $\mathcal{R}_0$ (natural). Each of these three components has its own exclusive susceptible pool. $\mathcal{R}_0$ (reinfection) is in the pool of those infected by previous strains/variants. $\mathcal{R}_0$ (breakthrough) is in the pool of the vaccinated population, and $\mathcal{R}_0$ (natural) is in the pool of unvaccinated susceptible individuals.

For the Delta and Omicron variants, the re-infection risks are 15% and 81% (Ferguson et al. 2021). The vaccine breakthrough risk for the Delta variant is 40%. The vaccine efficacy is 60% 3 months after administration of the second vaccine dose, while the vaccine breakthrough risk is high for the Omicron variant. However, the vaccination coverage in South Africa is only 25.96% (fully vaccinated by December 16, 2021). Given the high infection attack rate, the susceptible pool in South Africa is probably 20–30% currently.

Assuming that the susceptible, fully vaccinated, and recovered individuals were 20%, 20%, and 60% of the population in November 2021, in reality, the vaccinated and recovered individuals could overlap. In this study, we assigned the group that overlapped (e.g., infected but vaccinated as well, due to unawareness of infection status) to one of the two groups. Given the above information, the composition of the "susceptible pool" for infection with the Delta and Omicron variants is indicated in **Table 3**.

Table 3. The composition of "susceptible pool" for the Delta and Omicron variants

	Susceptible	Vaccinated	Recovered	Total
Delta	20%	8% (20*40%)	9% (60*15%)	88.6%
Omicron	20%	20% (20*100%)	48.6% (60*81%)	37%

Therefore, the susceptible pool of the Omicron variant was about 2.39-fold that of the Delta variant in November 2021, and the observed transmission advantage was partly due to this difference in the sizes of susceptible pools of the two variants. The natural increase in the transmissibility was 1.57 (1.57*2.39 = 3.76).

2.2.4 Discussion and Conclusion

Several prior studies have examined the relative transmissibility of new SARS-CoV-2 variants compared to their predecessors. Leung et al. conducted an investigation into the early transmissibility of the N501Y mutant strain in the UK from October to November 2020, using bioinformatics and public health data. By analyzing viral genomes carrying the 501Y mutation within spike proteins and incorporating global phylogenetic analysis, the study identified two distinct 501Y variants. Through a competitive transmission model, variant 2 (also known as the Alpha variant) was estimated to have an R_0 value 1.75 times higher than the 501N strain, indicating a 75% increase in transmissibility (Leung et al. 2021). Additionally, Roquebert et al. compared the transmissibility of the Beta variant and Alpha variant in specific regions of France. Using sequencing maps and reverse transcription PCR results to elucidate gene sequence differences and regional distribution, the study employed multinomial log-linear and generalized linear modelling techniques. Their analysis revealed that the Beta variant exhibited a transmission advantage of 15.8% (95% CI: 15.5-16.2%) in Ile-de-France and 17.3% (95% CI: 15.5-16.2%) in Hauts-de-France. For the Omicron variant, Ito et al. investigated its relative instantaneous reproduction number compared to the Delta variant in Denmark. They developed a method to estimate the relative instantaneous reproduction number of one variant relative to another under similar epidemiological conditions. The effective (instantaneous)

reproductive number of the Omicron variant was estimated to be 3.19 times that of the Delta variant (95% CI 2.82-3.61) (Ito, Piantham, and Nishiura 2022).

In terms of mathematical modelling, Chu introduced a fractional order COVID-19 model dynamics with a case study focusing on Saudi Arabia (Chu et al. 2021). The model incorporated classical Caputo-type derivatives of fractional order and considered transmission through the environment. Data from March 02, 2020, to July 31, 2020, were used to estimate parameters, revealing a basic reproduction number (R_0) of 1.2937. Furthermore, Li applied an SEIARD mathematical model to calculate the basic reproduction number using reported cases from March 06, 2021, to April 30, 2021, during the third wave of the pandemic. The calculated R_0 value was 1.2044, with graphical representation of parameter sensitivity and effects on model variables, predicting the peak of infections to occur on May 06, 2021 (Li, Wang, et al. 2021).

We developed a straightforward model consisting of 8 equations to simulate the complex wave patterns observed in COVID-19 deaths and the sequential replacement of variants in South Africa. In tandem, we estimated time-varying transmission rates and calculated three relative transmission rate factors for the Beta, Delta, and Omicron variants. Notably, the relative transmissibility rates of these variants were substantially higher than their predecessors, with increases of 73%, 87%, and 276% for the Beta, Delta, and Omicron variants, respectively. This heightened transmissibility in the Omicron variant stems from two primary sources: an expanded susceptible pool due to enhanced immune evasion capability and an inherent increase in transmission efficiency.

Mathematical models play a pivotal role in pandemic mitigation strategies; however, many existing models are overly complex to yield practical insights. Our approach utilized a concise set of 8 equations to simulate the intricate dynamics of multiple waves and variant replacements, demonstrating that further simplification by removing exposed classes would not significantly alter the outcomes, aligning with prior research findings (Song, Fan, Zhao, et al. 2021).

Nevertheless, our study had certain limitations. We did not account for potentially shortened generation intervals (GI) associated with the Delta and Omicron variants, which could result in an overestimation of the relative transmission rates. Additionally, the model did not incorporate reinfection scenarios, likely leading to an overestimation of transmission rates compared to models that do consider reinfection. Our estimated transmission rates therefore encompass contributions from shortened GIs and the immune evasion properties of the variants.

Furthermore, while reinfection risks were generally low for most variants (e.g., 10% for the Beta variant), cases of reinfection were associated with reduced infectivity and severity. Future models incorporating reinfection should account for these diminished effects. Our estimated transmission rates reflect conditions specific to South Africa and may hold true for other regions with similar infection attack rates, but caution should be exercised in extrapolating these findings to areas with lower infection rates.

Overall, our simplified mathematical framework provides a practical approach to estimating relative transmission rates across multiple variant strains, offering valuable insights for pandemic preparedness and response efforts.

Chapter 3

3 Regional Heterogeneity of in-hospital Mortality of COVID-19 in Brazil

3.1 Introduction

As reported by The World Health Organization (WHO), the global COVID-19 pandemic has led to a staggering total of more than 445 million confirmed cases and approximately 6 million deaths worldwide. Within the context of this global crisis, Brazil, with a population comprising just 2.73% of the world's total, has borne a disproportionately heavy burden, accounting for 8.07% of reported COVID-19 cases and 11.53% of reported deaths globally. This disparity underscores the profound impact of the pandemic on Brazil, highlighting the significant challenges faced by the country's healthcare system and population in grappling with the virus's devastating consequences. The data underscores the urgent need for targeted interventions and support to mitigate the impact of

COVID-19 in Brazil and other heavily affected regions around the world. ('https://www.who.int/emergencies/diseases/novel-coronavirus-2019' 2021)

Variations in COVID-19 mortality rates across different ethnic groups highlight profound social inequalities within societies. Previous research, predominantly conducted in developed nations like the United States, United Kingdom, and various European countries, consistently identified heightened vulnerability among certain ethnic minority groups, such as African Americans and Black British individuals, to SARS-CoV-2 infection (Niedzwiedz et al. 2020). In the context of Brazil, a vast continental nation encompassing diverse ethnicities, deeply ingrained social disparities further compound the impact of the pandemic. Brazil's population is categorized into five racial groups according to the Brazilian Institute of Geography and Statistics: branco (white), pardo (brown or mixed), preto (black), Amarillo (yellow), and caboclo (indigenous). Black and Brown Brazilians, considered disadvantaged, often experience lower educational attainment, diminished income levels, and restricted access to healthcare services compared to their white counterparts (Hone et al. 2017). The epidemic has starkly highlighted these disparities, as marginalized groups struggle within an inadequate welfare system that fails to provide sufficient income support during emergencies and offers limited access to timely information (Ahmed et al. 2020). Furthermore, the economic uncertainty exacerbated by the pandemic significantly impacts the mental well-being of these vulnerable populations, weakening immune resilience and heightening susceptibility to various viral infections. Addressing these systemic inequities is crucial to promoting health equity and ensuring comprehensive, accessible

healthcare for all individuals, irrespective of ethnicity or socioeconomic status (Patel et al. 2020).

In Brazil, the distribution of COVID-19 vaccines encompasses three main types: Coronavac (Sinovac), AZD1222 (AstraZeneca), and BNT162b (Pfizer/BioNTech). These vaccines have played a critical role in the national vaccination effort, aiming to curb the spread of the virus and mitigate its impact on public health. During the administration of the initial dose, data indicates that Coronavac, AZD1222, and BNT162b vaccines were utilized to vaccinate approximately 9.61%, 6.69%, and 0.35% of the Brazilian population, respectively. These figures underscore the diverse distribution of vaccine types and their respective contributions to the nationwide immunization campaign. For the subsequent administration of the second dose, the distribution pattern shows that Coronavac, AZD1222, and BNT162b vaccines were responsible for vaccinating approximately 7.52%, 0.53%, and less than 0.01% of the Brazilian population, respectively. This data highlights the ongoing efforts to ensure complete vaccination coverage among eligible individuals, with varying degrees of uptake observed across different vaccine types. As Brazil continues its vaccination campaign, comprehensive data analysis and strategic distribution efforts remain essential to achieving widespread immunity and combating the COVID-19 pandemic effectively (Boschiero, Palamim, and Marson 2021)

Several prior studies have addressed the disproportionate impact of COVID-19 based on social status and ethnicity in developing nations (Li, Pereira, et al. 2021), This research extends this discussion by investigating detailed epidemiological profiles of patients across diverse ethnic and educational

backgrounds in various geographic regions. Additionally, the study explores how prevalent comorbidities influence in-hospital mortality rates. Furthermore, this paper delves into the effectiveness of vaccines in reducing the Severe Acute Respiratory Infections (SARI) fatality ratio across patients with varying comorbidities and age groups. Specifically, it compares the efficacy of Sinovac and AstraZeneca vaccines in mitigating COVID-19-related mortality outcomes.

3.2 Methods

3.2.1 Data Collection

The patient data utilized in this investigation were sourced from the Influenza Epidemiological Surveillance Information System (SIVEP-Gripe) spanning from February 27, 2020, to March 15, 2022. Demographic data specific to regions were obtained from publicly accessible Census data.

Included in this study were all hospitalized patients with confirmed positive reverse transcription-polymerase chain reaction (RT-PCR) results for SARS-CoV-2 and a documented outcome of either discharge or death within this timeframe. Each patient's dataset contained 154 attributes encompassing demographic details and clinical characteristics, with certain values restricted by government regulations. Key variables selected for analysis included age, gender,

geographical location (state), ethnicity, vaccination status, presence of comorbidities, level of education, and date of hospital admission.

Comorbidity information was absent in approximately 51.4% of cases within the database. In instances where data were unavailable, it was assumed that the respective comorbidities were absent. A total of eleven comorbidities were considered for analysis, including cardiovascular disease, hematologic disorders, liver ailments, neurological conditions, kidney diseases, Down syndrome, asthma, diabetes, immunodeficiency disorders, pulmonary disorders, and obesity.

To account for the dominance of different virus variants during specific periods, distinct time intervals were designated corresponding to the emergence and prevalence of three notable variants subsequent to the original strain. The Ancestor strain was considered dominant from March 13, 2020, to November 23, 2020. This was followed by the Gamma (P.1) variant's dominance from November 23, 2020, to August 16, 2021. The Delta (B.1.617.2) variant then prevailed from August 16, 2021, to December 6, 2021, with the subsequent period from December 16, 2021, onward designated as the Omicron (B.1.1.529) variant dominant period.

3.2.2 Statistical analysis

In this study, the in-hospital mortality rate as the proportion of deaths after hospitalization was selected to be the interest outcome variable. Pearson's Chi-squared test was first performed for categorical variables, and one-way analysis of variance (ANOVA) was examined for continuous variables. Together with the p-value of null hypothesis significance testing, we presented the data showing demographic features comorbidities, and ethnic composition of patients in five macro regions.

To assess the influences of ethnicity, comorbidities, and education on an individual level, we employed mixed-effects Cox proportional hazards regression analysis using the "coxme" package, accounting for geographical variations. This analytical approach mirrors studies examining COVID-19 mortality in intensive care units in the United Kingdom and early-stage mortality in Brazil (Qian et al. 2020). Patient-specific clinical characteristics including age, gender, ethnicity, education, vaccination status, and comorbidities were treated as fixed effects, while region was considered a random effect For categorical variables, White Brazilians, unvaccinated individuals, those with a university degree or higher education, and individuals under the age of 40 were designated as the reference categories for ethnicity, vaccination status, education, and age, respectively. Subsequently, hazard ratios with corresponding 95% confidence

intervals were visualized using a forest plot to illustrate the impact of these factors on mortality risk.

We proceeded to analyze the in-hospital mortality rates across different age groups, focusing on patients with varying features such as comorbidities, ethnic backgrounds, and education levels. Education level was defined based on the highest grade completed by the patient. Furthermore, we investigated the impact of vaccination status on hospitalization outcomes, distinguishing between vaccinated and unvaccinated patients.

In examining vaccine efficacy, we initially compared the Severe Acute Respiratory Infections (SARI) fatality ratio among patients with different numbers of comorbidities, stratified by vaccination status. The SARI fatality ratio was graphically represented by the number of comorbidities per patient. Subsequently, we evaluated the SARI fatality ratio across different age groups, comparing vaccinated and unvaccinated patients, and plotted the results by age.

Moreover, to assess the effectiveness of the two predominant vaccines (Sinovac and AstraZeneca) among hospitalized patients of varying ages and comorbidity profiles, we visualized their respective fatality ratios. This comparative analysis aimed to provide insights into the performance of these vaccines under different patient demographics and health conditions.

3.3 Results

3.3.1 Demographic features of data

We initially presented the demographic characteristics of our dataset concerning patients with outcomes of discharge and death. **Table 4** illustrates the distribution of cases across five age groups, revealing a trend where the proportion of deaths increased with advancing age. The prevalence of deaths was slightly higher among males (34.6%) compared to females (34.4%), among Black and Indigenous Brazilians (39.2% and 38.6%, respectively) compared to other ethnic groups, and among unvaccinated patients (34.7%) compared to vaccinated patients (33.4%). There was a notable predominance of hospitalized deaths in the northeast region, accounting for 42.2% of cases in that area, followed by the northern region (37.2%). The mortality rate was approximately 25% lower in patients with a university degree (23.4%) compared to illiterate patients (48.4%). Nearly all comorbidities examined showed an association with mortality among hospitalized patients.

Subsequently, **Table 5** presents the ethnic composition across different regions. In most regions, the percentages of hospitalizations and deaths among ethnic groups were consistent with their respective population distributions. Notably, in the north, northeast, and central-west regions, White individuals

exhibited a significantly lower proportion of hospitalizations and deaths compared to their representation in the population, whereas Brown individuals had higher rates of hospitalizations and deaths relative to their population size. However, this relative ratio demonstrated contrasting characteristics in the southeast and south regions.

Table 4. Explanatory Variables of Cases

	Death	Discharge	P-value
Age			< 0.001
< 40 years	6550 (13.0%)	43781 (87.0%)	
40 – 49 years	9624 (21.7%)	34669 (78.3%)	
50 – 59 years	16552 (29.9%)	38783 (70.1%)	
60 – 69 years	20641 (41.1%)	29633 (58.9%)	
> 70 years	42457 (54.8%)	35022 (45.2%)	
Gender			0.106
Female	42227 (34.4%)	80576 (65.6%)	
Male	53597 (34.6%)	101312 (65.4%)	
Region			< 0.001
North	6316 (37.2%)	10659 (62.8%)	
Northeast	13775 (42.2%)	18901 (57.8%)	
Central-west	9698 (33.6%)	19171 (66.4%)	
Southeast	42724 (35.1%)	78911 (64.9%)	
South	23311 (30.1%)	54246 (69.9%)	
Ethnic Group			< 0.001
White	55161 (33.2%)	110806 (66.8%)	
Black	4725 (39.2%)	7337 (60.8%)	
Yellow	883 (33.9%)	1722 (66.1%)	
Pardo	34827 (36.1%)	61660 (63.9%)	
Indigenous	228 (38.6%)	363 (61.4%)	
Education Level			< 0.001
Illiterate	4613 (48.4%)	4924 (51.6%)	
Elementary School	16614 (43.8%)	21293 (56.2%)	
Middle School	8722 (34.5%)	16527 (65.5%)	
High School	10878 (26.8%)	29678 (73.2%)	
University	4052 (23.4%)	13291 (76.6%)	
Vaccine			< 0.001
Vaccinated	11645 (33.4%)	23271 (66.6%)	
Unvaccinated	84179 (34.7%)	158617 (65.3%)	
Comorbidities			< 0.001
Cardiovascular	37783 (45.5%)	45241 (54.5%)	
Disease			
Hematology	668 (48.2%)	719 (51.8%)	
Down syndrome	361 (42.9%)	480 (57.1%)	
Liver Disease	1044 (56.8%)	794 (43.2%)	
Asthma	1990 (31.8%)	4276 (68.2%)	
Diabetes	27592 (46.7%)	31484 (53.3%)	
Neurological Disease	4966 (54.4%)	4158 (45.6%)	
Pulmonary	4433 (55.0%)	3634 (45.0%)	
Immunodepression	2834 (54.5%)	2366 (45.5%)	
Renal	4858 (60.1%)	3220 (39.9%)	
Obesity	11203 (42.8%)	14964 (57.2%)	

Table 5. Ethnic Composition in Each Region

	Population	Hospitalization	Death
North			
White	23.2%	1800 (10.6%)	735 (11.6%)
Black	6.5%	604 (3.6%)	219 (3.5%)
Yellow	1.1%	185 (1.1%)	69 (1.1%)
Pardo	67.2%	14127 (83.2%)	5194 (82.2%)
Indigenous	1.9%	259 (1.5%)	99 (1.6%)
Northeast			
White	29.2%	4694 (14.4%)	2175 (15.8%)
Black	9.4%	1521 (4.7%)	683 (4.9%)
Yellow	1.2%	499 (1.5%)	186 (1.4%)
Pardo	59.8%	25912 (79.3%)	10705 (77.7%)
Indigenous	0.4%	50 (0.1%)	26 (0.2%)
Central-west			
White	41.5%	9031 (31.3%)	2990 (30.8%)
Black	6.6%	1228 (4.3%)	472 (4.9%)
Yellow	1.5%	388 (1.3%)	141 (1.5%)
Pardo	49.4%	18076 (62.6%)	6040 (62.3%)
Indigenous	0.9%	146 (0.5%)	55 (0.5%)
Southeast			
White	54.9%	79574 (65.4%)	28150 (65.9%)
Black	7.8%	6934 (5.7%)	2720 (6.4%)
Yellow	1.1%	1176 (1.0%)	387 (0.9%)
Pardo	36%	33896 (27.9%)	11451 (26.8%)
Indigenous	0.1%	55 (0.0%)	16 (0.0%)
South			
White	78.3%	70868 (91.4%)	21111 (90.6%)
Black	4%	1775 (2.3%)	631 (2.7%)
Yellow	0.7%	357 (0.4%)	100 (0.4%)
Pardo	16.7%	4476 (5.8%)	1437 (6.2%)
Indigenous	0.3%	81 (0.1%)	32 (0.1%)

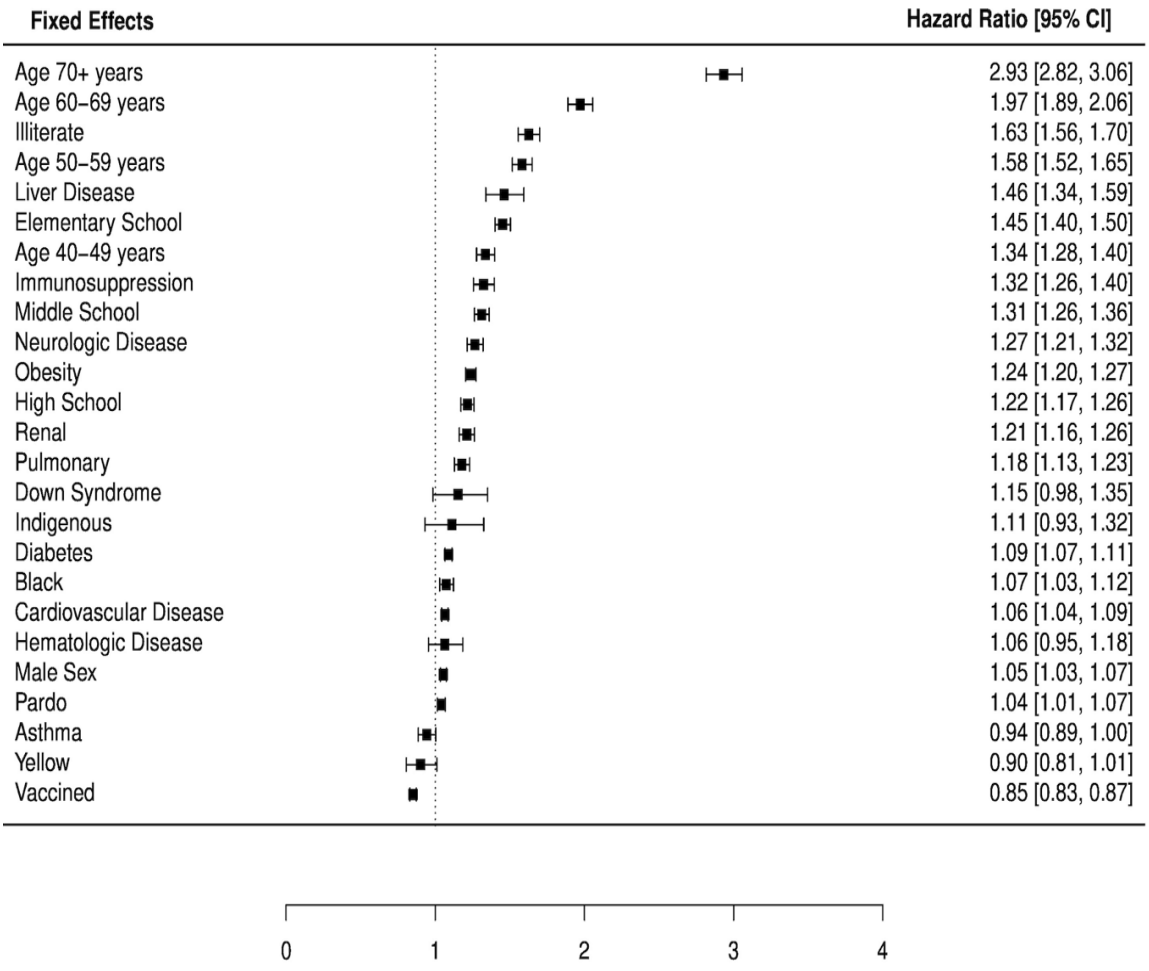
3.3.2 Fitted multivariate mixed-effects Cox model

In the analysis conducted using the multivariate mixed-effects Cox model, **Figure 3.1** displays the estimates of fixed effects, while **Figure 3.2** illustrates the random effects. Notably, patients aged over 70 years exhibited a significantly higher rate of hospital mortality compared to those younger than 40 years (HR: 2.93, 95% CI: 2.82 – 3.06), highlighting age as a critical risk factor for death. Furthermore, patients with lower education levels, particularly illiterate individuals, faced a higher risk of mortality (HR: 1.63, 95% CI: 1.56 – 1.70) compared to those with a university degree or higher.

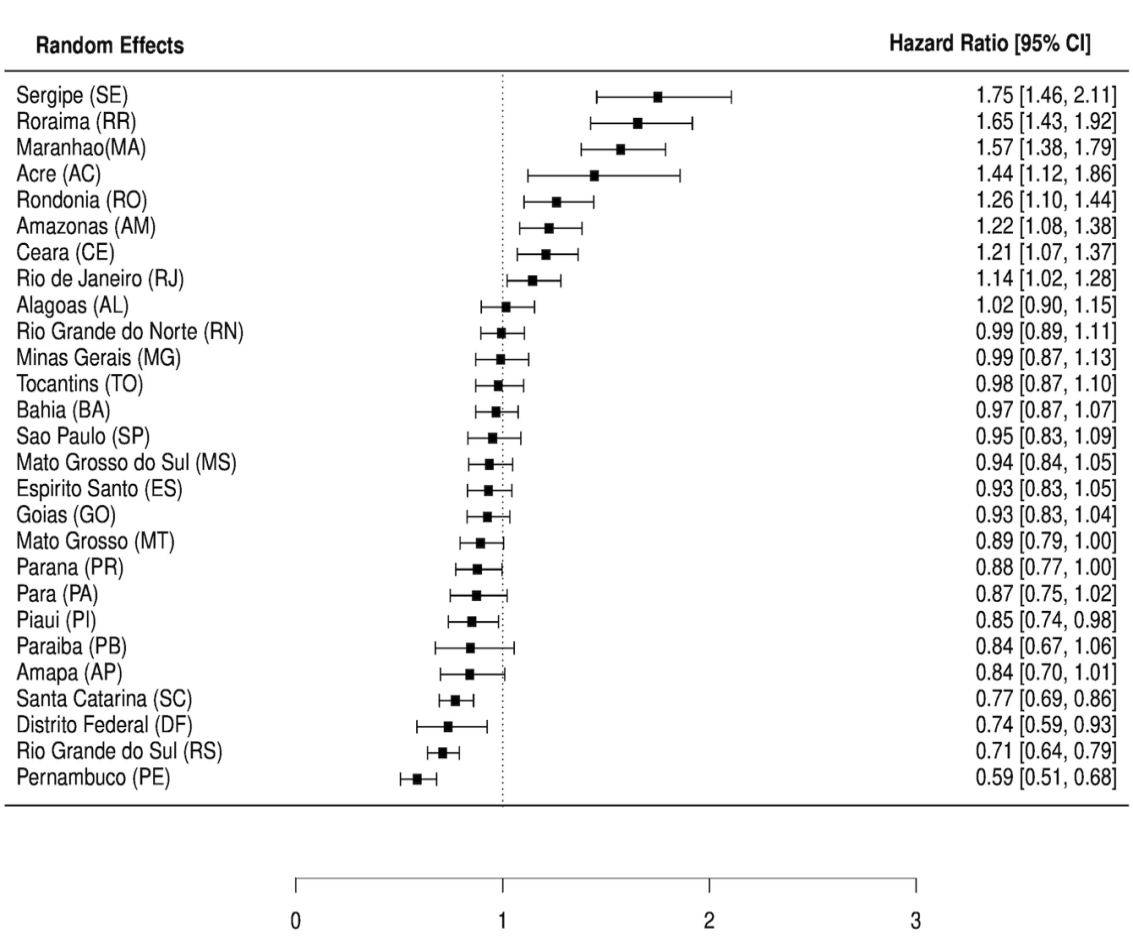
The presence of certain comorbidities also contributed significantly to mortality risk, with conditions such as liver disease (HR: 1.46, 95% CI: 1.34 – 1.59) and immunosuppression (HR: 1.32, 95% CI: 1.26 – 1.40) identified as relatively dangerous factors. Moreover, Black Brazilians experienced a higher risk of death compared to White Brazilians (HR: 1.07, 95% CI: 1.03 – 1.12), underscoring the impact of ethnicity on mortality outcomes.

Significant variations in hazard ratios were observed across different Brazilian states, with states like Sergipe (HR: 1.75, 95% CI: 1.46 – 2.11), Roraima (HR: 1.65, 95% CI: 1.43 – 1.92), Maranhão (HR: 1.57, 95% CI: 1.38 – 1.79), Acre (HR: 1.44, 95% CI: 1.12 – 1.86), and Rondônia (HR: 1.26, 95% CI: 1.10 – 1.44) in the north and northeast regions exhibiting higher hazard ratios

compared to other states. These findings highlight regional disparities in mortality risk within Brazil, emphasizing the need for targeted interventions to address healthcare inequalities across different geographical areas.



3.1 Risk of Death by Clinical Features (Fixed Effects)



3.2 Risk of Death by States (Random Effects)

3.3.3 In-hospital mortality rate

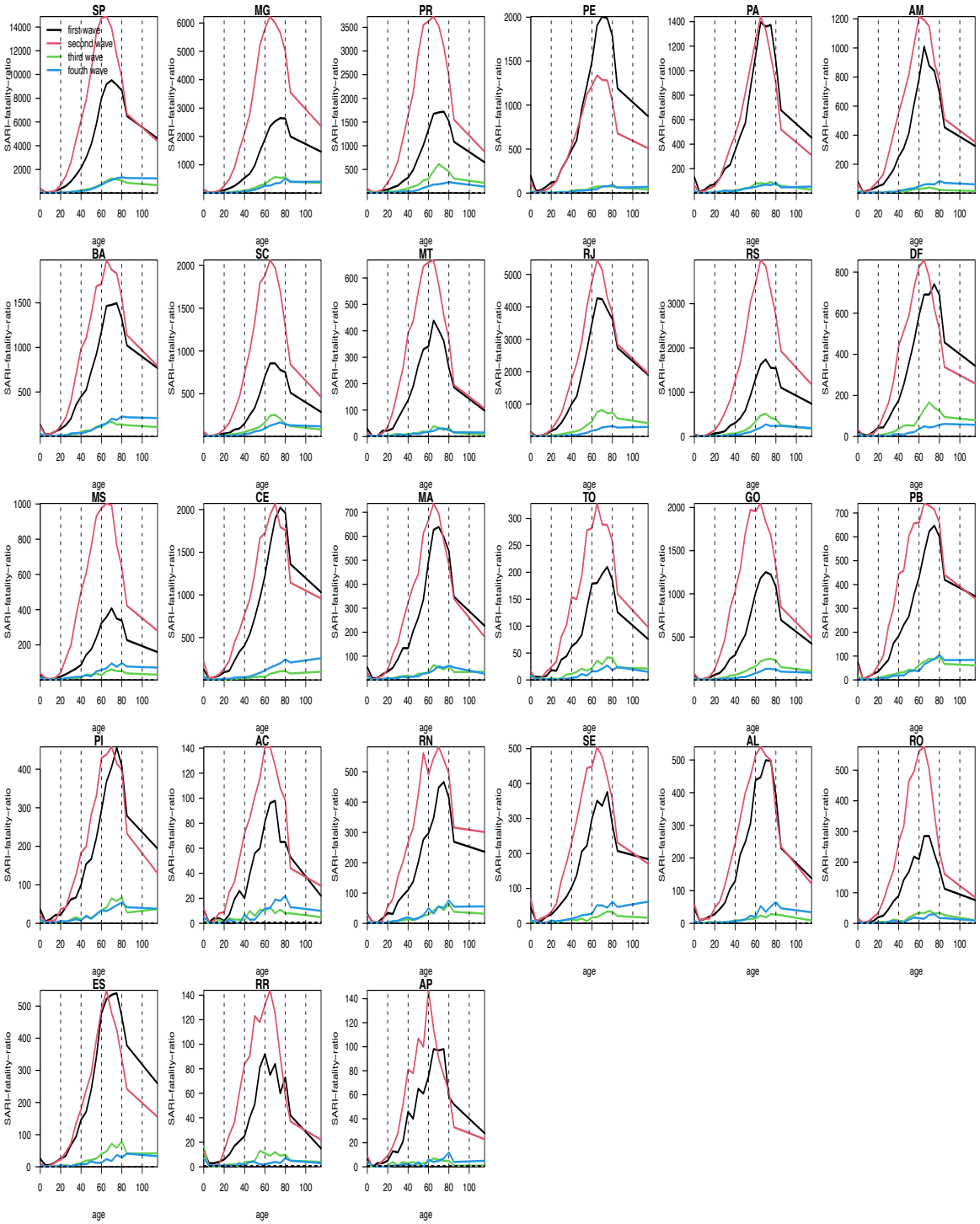
We begin by presenting visualizations of the fatality ratio by age across different dominant variant periods in 27 Brazilian states, providing an overview of respiratory hospitalized mortality trends in Brazil. **Figure 3.3** illustrates the fatality ratio by age during each period of dominant variants across these states. In nearly all states, the peak fatality ratio is observed among patients aged 60 to 70. Furthermore, in most states, the fatality ratio during the initial period of different dominant variants was notably higher compared to other periods with different COVID strains. This indicates that the mortality rate among respiratory inpatients during the dominance of the Gamma (P.1) variant was significantly higher than during other strain dominance periods. Conversely, during the second and third periods dominated by the Delta (B.1.617.2) and Omicron (B.1.1.529) variants, the fatality ratio appears relatively low across all age groups.

Next, we analyze in-hospital fatality rates by age among patients with varying comorbidities, education levels, and ethnic groups. **Figure 3.4** depicts the variation in in-hospital mortality rates among patients with different numbers of comorbidities. Generally, the risk of death for hospitalized patients increases with the number of comorbidities, with the most significant risk gap observed between no comorbidities and one comorbidity (around 20%), which is larger

than the gap between multiple comorbidities (less than 10%). Teenagers with three comorbidities exhibited particularly higher mortality rates.

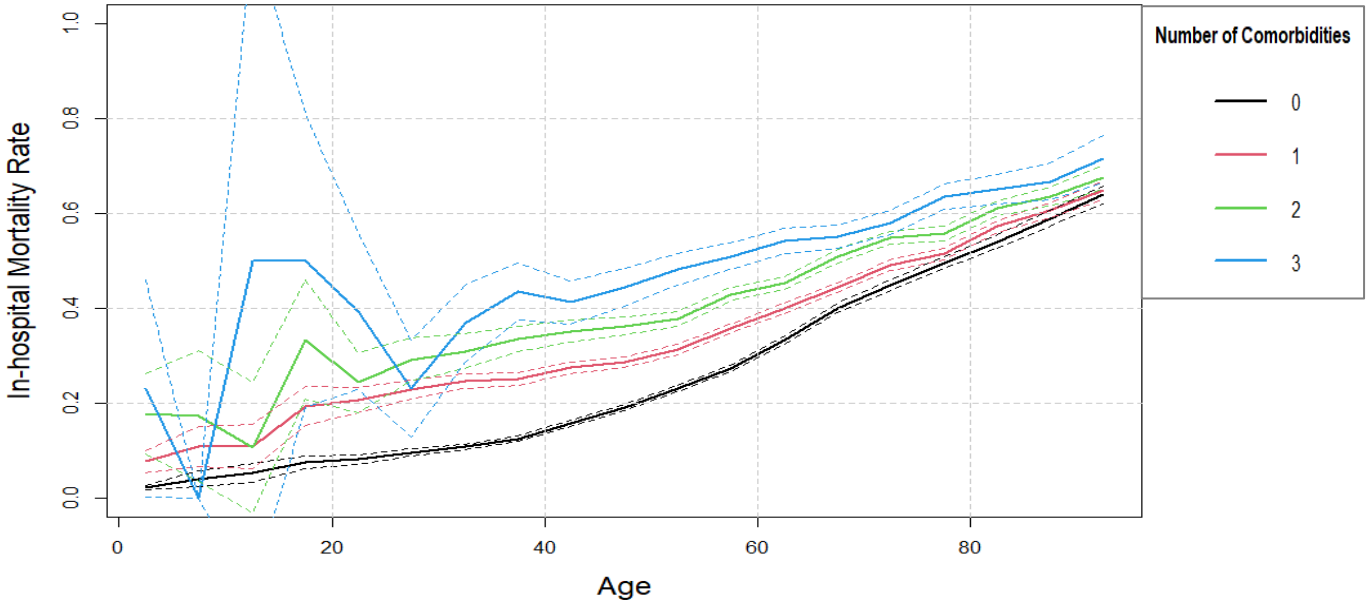
Figure 3.5 presents how mortality rates vary with age across different ethnicities in Brazil. In middle-aged individuals (30 – 60 years old), no significant disparities were observed across races. However, among older individuals (over 70 years old), as well as infants of Indigenous, Yellow, and Black Brazilian descent, there was a higher risk of death. Conversely, Indigenous youth (15 – 24 years old) exhibited relatively lower mortality risk.

In Figure 3.6, we compare in-hospital mortality rates among patients with varying levels of education. For patients under the age of 80, higher education levels were associated with relatively lower risk of mortality. The mortality gap between illiterate individuals and those with some level of education was more pronounced than differences based on specific schooling lengths. Notably, these gaps peaked around age 18 and gradually decreased thereafter, with little variation observed for patients around 80 years old.



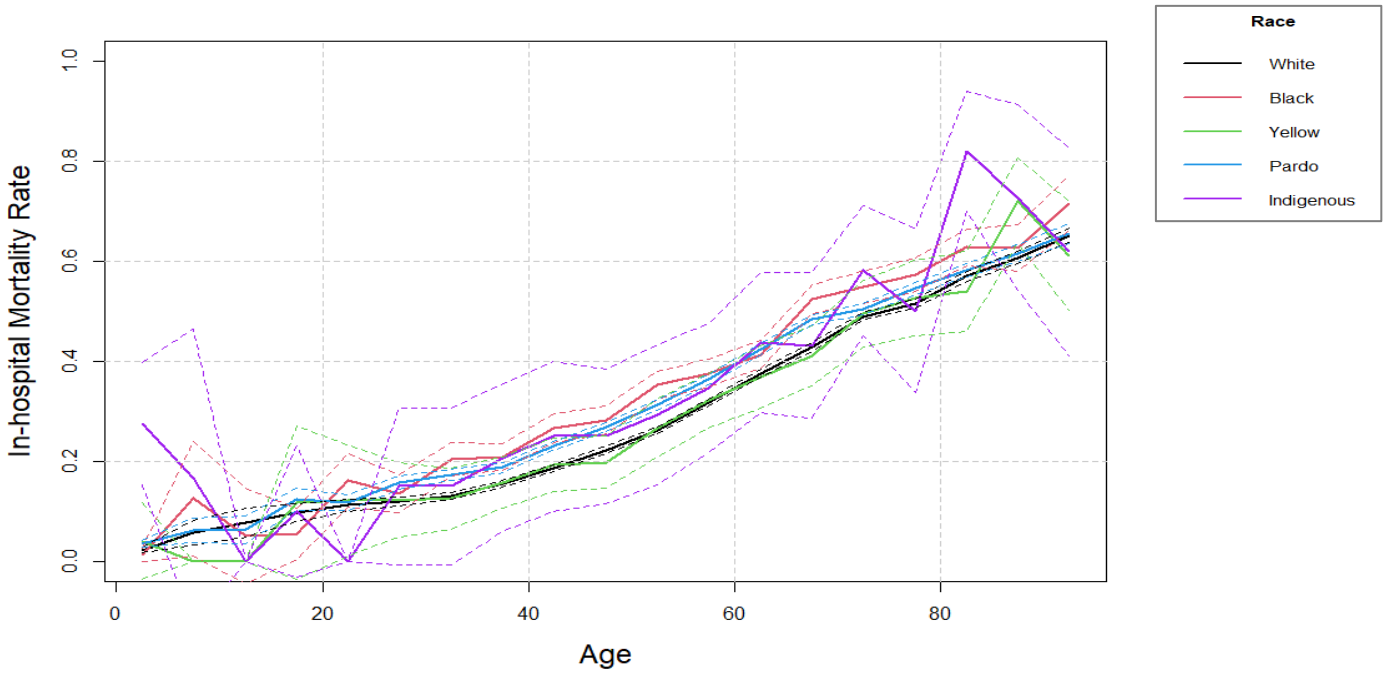
3.3 SARI-fatality-ratio by age in four periods of different dominant variants in Brazilian states

Figure 3.3. Visualizing the fatality ratio by age in four periods of different dominant variants of the pandemic in 27 Brazilian states. Four different colors represent four periods of different dominant variants in Brazil. The black line represents the first period—the Ancestor strain, and the red line means the second period which is dominated by the Gamma (P.1) variant. The green line represents the third period which is the Delta (B.1.617.2) variant-dominated period, and the blue line represents the fourth period which is the Omicron (B.1.1.529) variant-dominated period.



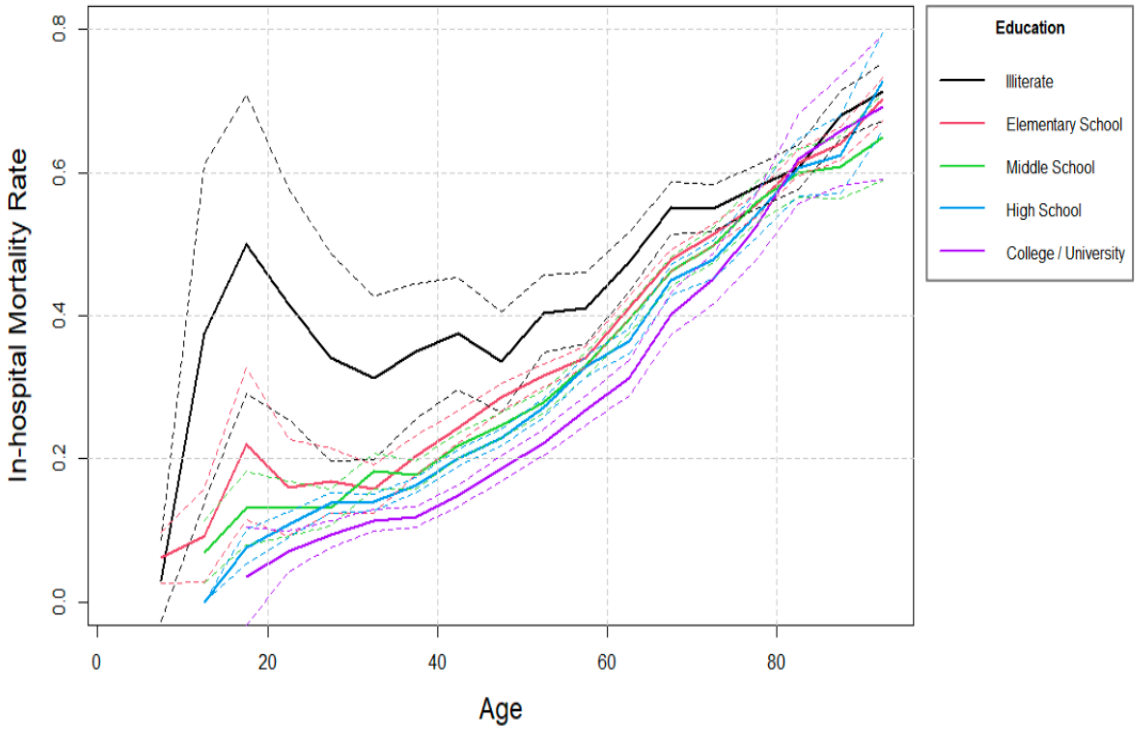
3.4 In-hospital Mortality Rate by Age between Patients with Different Comorbidities

Figure 3.4 compares the in-hospital mortality rate for patients with different numbers of comorbidities at different ages. The black line shows patients with no comorbidities. The red line, the green line, and the blue line represent patients with one, two, and three kinds of comorbidities, respectively.



3.5 In-hospital Mortality Rate by Age between Ethnical Groups

Figure 3.5 shows the in-hospital mortality rate for patients of different ages and ethnic groups. The black line represents white patients, the red line represents black patients, the green line represents yellow patients, the blue line describes pardo patients and the purple line represents indigenous patients.



3.6 In-hospital Mortality Rate by Age between Patients with Different Education Levels

Figure 3.6 compares the in-hospital mortality rate for patients of different ages and levels of education. The black line represents illiterate patients, the red line represents patients with only elementary education, the green line represents patients with only middle school education, the blue line represents patients

whose highest degree was high school, and the purple line represents patients who have been to university or college.

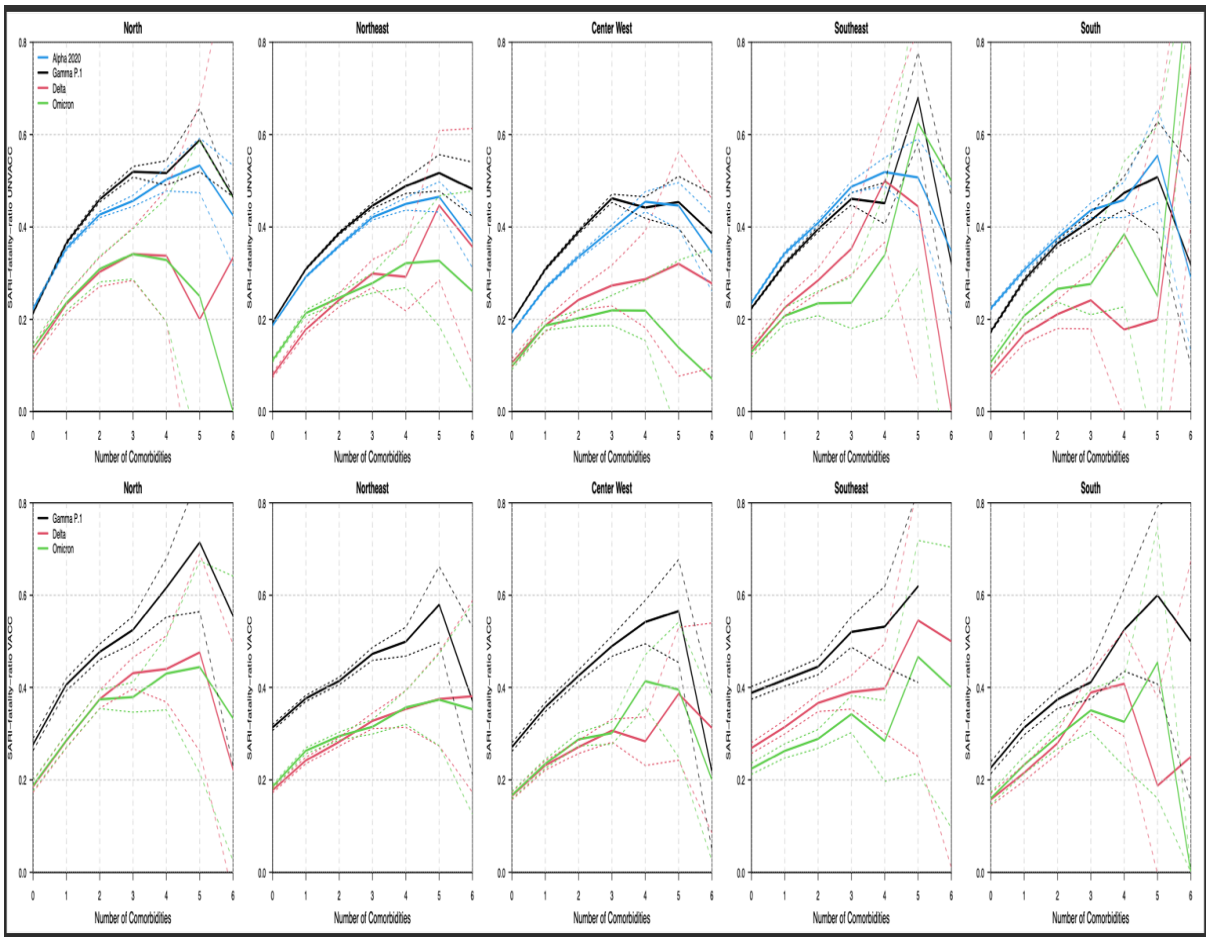
3.3.4 Efficacy of vaccines

In the investigation of vaccination efficacy, **Figure 3.7** illustrates the variation in SARI (Severe Acute Respiratory Infection) fatality ratios between vaccinated and unvaccinated patients categorized by the number of comorbidities and regional disparities. Regions were grouped based on similar income levels and health conditions, highlighting two contrasting areas with the most notable differences. Overall, vaccination did not significantly benefit patients with comorbidities, with fatality rates remaining higher for the Alpha 2020 and Gamma (P.1) variants compared to the Delta (B.1.617.2) and Omicron (B.1.1.529) variants. Across these four variants of coronavirus, fatality ratios increased with the number of comorbidities, peaking around 5 comorbidities and decreasing thereafter.

In **Figure 3.8**, the effectiveness of two vaccine types, Sinovac and AstraZeneca, was evaluated for hospitalized patients with varying numbers of comorbidities across two contrasting regions. During the Gamma (P.1) variant-dominated period, AstraZeneca generally demonstrated better efficacy in preventing deaths among hospitalized patients. However, in the Central-west, Southeast, and South regions, Sinovac showed higher protection for patients with more than four comorbidities. During the Delta (B.1.617.2) variant-dominated

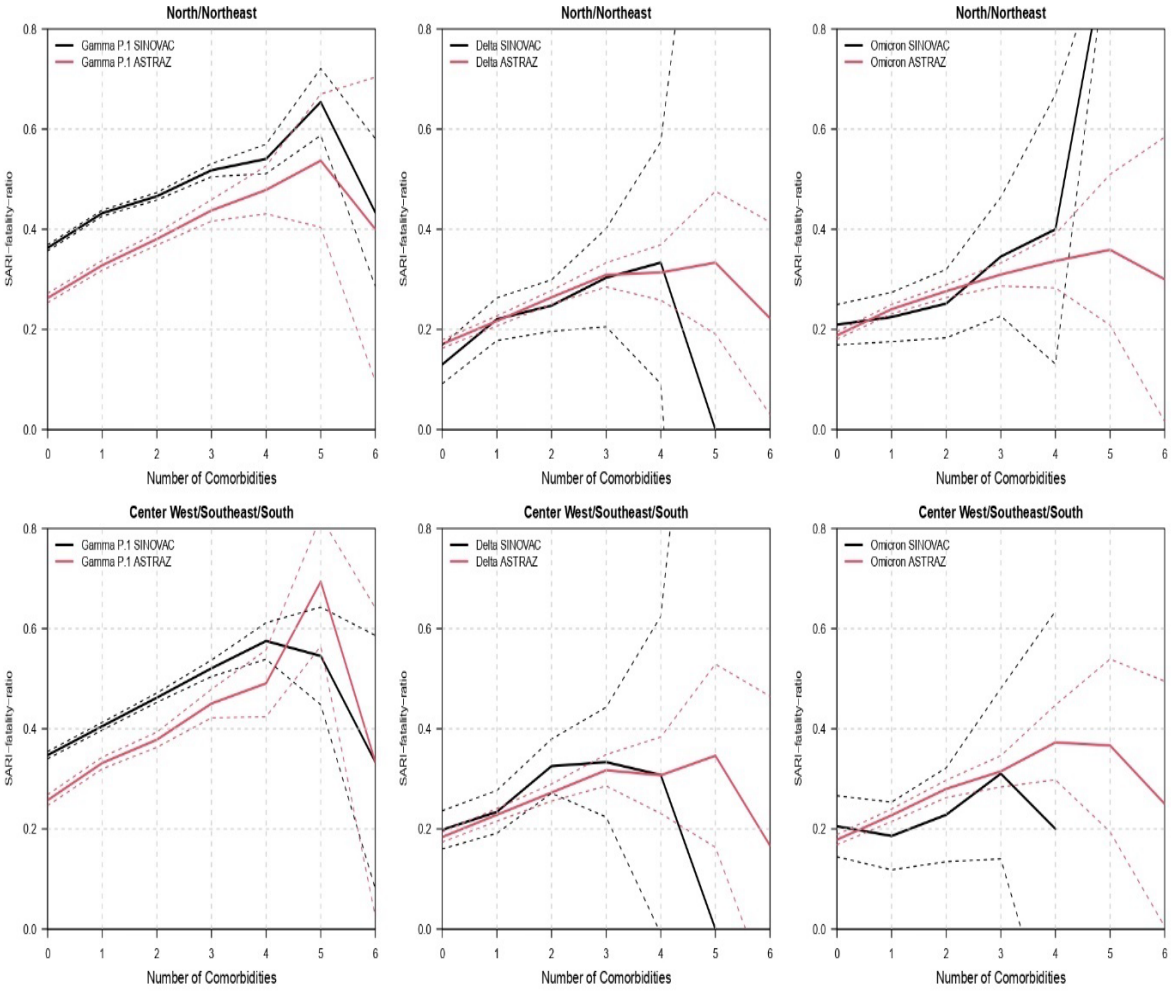
period, both vaccines exhibited similar protective effects for hospitalized patients with no more than four comorbidities. For patients with four or more comorbidities, Sinovac provided greater protection than AstraZeneca. In the Omicron (B.1.1.529) variant-dominated period, AstraZeneca demonstrated significantly greater protection than Sinovac in the North and Northeast regions. Conversely, Sinovac was more protective in the Central-west, Southeast, and South regions, especially for patients with varying numbers of comorbidities.

Figure 3.9 presents the effectiveness of the two vaccines for patients of different ages during different variant periods. Sinovac showed more effectiveness in reducing the SARI fatality ratio among patients under 60 during the Gamma (P.1) variant transmission. Conversely, AstraZeneca exhibited greater protection for patients older than 60. During the Delta (B.1.617.2) variant period, AstraZeneca was more effective in reducing SARI fatality ratios for children and elderly patients in the Central-west, Southeast, and South regions. In the North and Northeast regions, Sinovac provided better protection across all age groups. Similar trends were observed during the Omicron (B.1.1.529) variant epidemic, with Sinovac demonstrating greater protection in the North and Northeast regions and AstraZeneca being more effective for patients aged 40 to 60 in the Central-west, Southeast, and South regions.



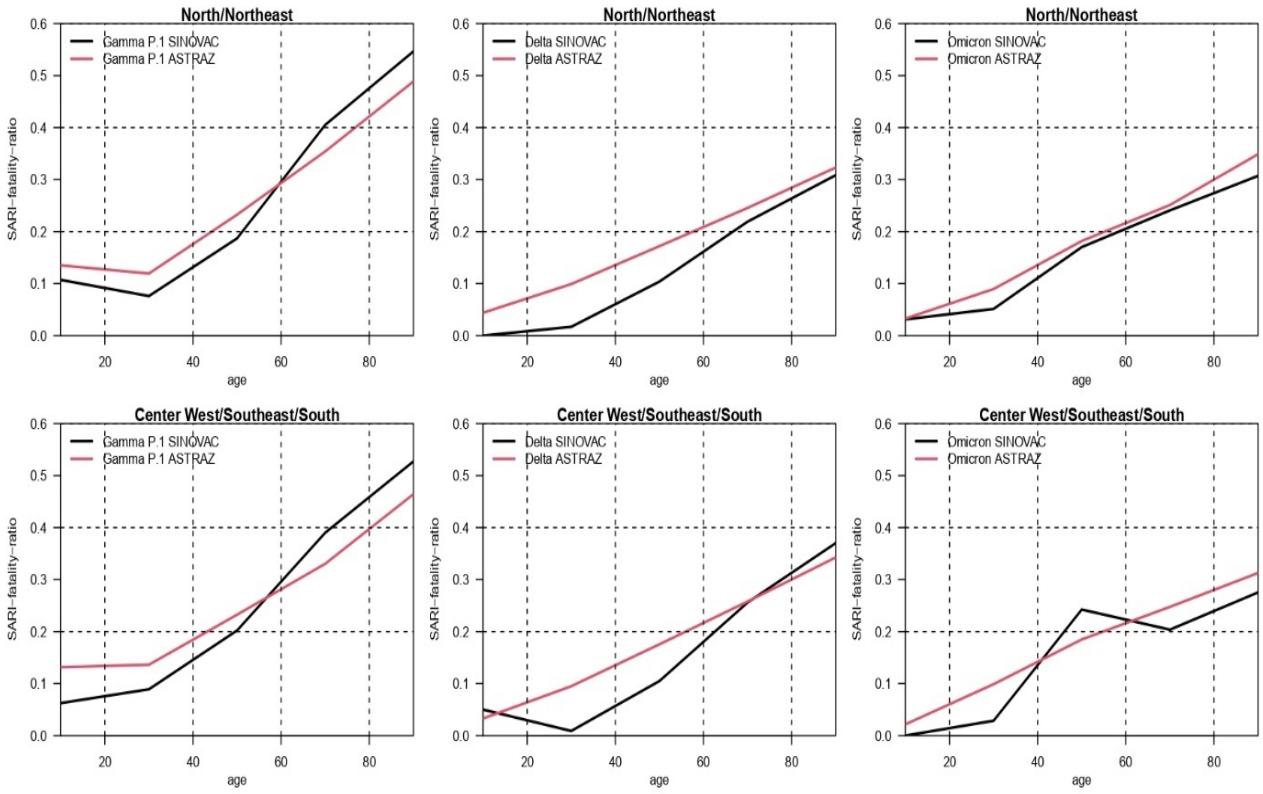
3.7 SARI-fatality ratio in vaccinated and unvaccinated patients by the number of comorbidities and regional differences.

Figure 3.7 The effectiveness of vaccination is assessed by comparing the SARI (Severe Acute Respiratory Infection) fatality ratio in vaccinated and unvaccinated patients with varying numbers of comorbidities. Each period dominated by a specific variant is represented by a different color. The black line corresponds to the Gamma (P.1) variant-dominated period, the red line represents the Delta (B.1.617.2) variant-dominated period, and the green line indicates the Omicron (B.1.1.529) variant-dominated period.



3.8 SINOVAC vs. ASTRAZ by the number of comorbidities

Figure 3.8 The comparison of vaccine effectiveness for patients with varying numbers of comorbidities is depicted in the plot. The red line represents AstraZeneca, while the black line represents Sinovac. The SARI (Severe Acute Respiratory Infection) fatality ratios are shown with a 95% confidence interval displayed by the dashed lines.



3.9 SINOVAC vs. ASTRAZ by age

Figure 3.9 Comparing the effectiveness of two vaccines for patients of different ages. The red line represents the AstraZeneca and the black line represents Sinovac. The efficacy of two vaccines was compared in three different mutation-dominated periods in three regions of Brazil.

3.4 Discussion and Conclusion

In line with global findings, our study corroborates that mortality rates were higher among the elderly (Liu, Chen, et al. 2020), male (Sharma, Volgman, and Michos 2020), and individuals with multiple comorbidities (Halpin et al. 2020). Furthermore, another study published in Nature Medicine analyzed the same dataset, focusing on spatiotemporal variations in COVID-19 mortality rates within Brazilian hospitals (Brizzi et al. 2022). The authors selected 14 state capitals in Brazil during periods dominated by the gamma variant and observed fluctuations in COVID-19 in-hospital mortality rates linked to geographic disparities and healthcare capacity constraints. Using a Bayesian model, they estimated that approximately half of the COVID-19 deaths in hospitals across these 14 cities could have been prevented in the absence of geographic inequalities and healthcare system strain. Our study differs in that we examined trends in geographic disparities of in-hospital mortality across ethnic groups in Brazil and employed models to investigate disparities among these groups.

Upon analyzing the racial composition and mortality rates across various regions, we identified racial disparities in different geographic areas. In comparison to the regional population demographics, hospitalizations and deaths among Pardo Brazilians were notably higher in the north, northeast, and central-west regions, whereas those among White Brazilians were more prevalent in the southeast and south regions. It is noteworthy that the Pardo population is concentrated in the north and northeast regions, whereas the white demographic

is predominant in the south and southeast regions. This observation suggests a correlation between ethnic concentrations and a higher proportion of death cases relative to population size. Additionally, it is observed that White Brazilians tend to have more frequent opportunities for international travel. Considering race as an indicator of intersecting social networks (Neely and Samura 2011), we hypothesize that differential interactions within residential communities may partially contribute to the spread of the epidemic. In the investigation of racial disparities, a study published in the Journal of Public Health utilized an alternative methodology to explore this topic (Peres et al. 2021). The researchers employed the same dataset for their analysis, utilizing logistic regression models to evaluate the relationship between self-reported race and in-hospital mortality while adjusting for clinical characteristics and comorbidities. The median age of the cohort was 61 years, with 57% being male, 35% self-identifying as black/brown, and 35.4% identifying as white. The overall in-hospital mortality rate stood at 37%. Black/brown patients exhibited a higher in-hospital mortality rate compared to white patients (42% vs. 37%) and were less frequently admitted to intensive care units (ICUs) (32% vs. 36%). This study underscores that among hospitalized adults with COVID-19 in Brazil, black/brown patients faced higher mortality rates, utilized fewer hospital resources, and experienced higher rates of illness compared to their white counterparts. These racial disparities in health outcomes and healthcare access underscore the urgent need for implementing strategies to mitigate inequities.

Within the Cox regression model, age emerges as a primary risk factor for mortality. Moreover, liver disease stands out as the most significant

contributor to in-hospital mortality. Among patients with cirrhosis and SARS-CoV-2 infection, notably high rates of mortality and hepatic decompensation are observed. These outcomes likely stem from immune dysfunction associated with cirrhosis (Marjot et al. 2021). This discovery can inform discussions on SARS-CoV-2 vaccination strategies for Brazilian patients with liver cirrhosis or those who have undergone liver transplants. Additionally, obesity has been identified as another significant comorbidity associated with heightened risk. Several mechanisms collectively contribute to this effect (Popkin et al. 2020). A significant concern has been raised regarding the potentially reduced effectiveness of vaccines in obese individuals. Analyzing the relationship between mortality rates and the number of comorbidities reveals a substantial increase in the risk of death when at least one comorbidity is present.

Black individuals face a heightened risk of mortality compared to White individuals. This racial disparity underscores deep-seated social inequalities and highlights the challenges faced by minority populations. Many disadvantaged groups, including Black and Pardo Brazilians, often work in healthcare and nursing roles and reside in environments conducive to infection, leading to disproportionate exposure to risk factors. In the immediate term, policymakers must take swift action to curb the spread of the virus in these vulnerable areas. A paper published in The Lancet Global Health used a similar method (Baqui et al. 2020). The distinction lies in the approach taken by the authors, who utilized the SIVEP-Gripe dataset to conduct cross-sectional research on ethnic and regional disparities in COVID-19 hospital mortality rates in Brazil. This cross-sectional study had a limited data period spanning a few months, during which the authors

examined regional discrepancies in COVID-19 hospital admissions by state and across two socioeconomic sub-regions (five in our study). Furthermore, the authors visualized the prevalence of comorbidities among survivors and non-survivors, as well as the distribution of comorbidities across different racial groups.

In our investigation, survival analysis also revealed that the northern and northeastern regions exhibited higher hazard ratios compared to other states. The northeastern region of Brazil is historically the country's poorest, characterized by a blend of export-oriented plantation agriculture and subsistence farming. Notably, there are extremely impoverished areas within the northern and northeastern regions. Economic disparities are evident among rural households, particularly in the western Amazon, where the rural poverty rate is notably high. This economic context exposes these populations to various social challenges, including limited access to water and inadequate internet connectivity, which may impede their ability to access information about preventive measures against the virus (de León-Martínez et al. 2020). Moreover, climate plays a crucial role in shaping the impact of COVID-19. In tropical regions, favorable climatic conditions can potentially facilitate the spread of the outbreak. The unique rainy season, affecting humidity levels, is likely contributing to the higher mortality rates observed. This scenario may be particularly relevant in areas such as Amazonas (AM), Maranhão (MA), and Ceará (CE) located in the north and northeast regions of Brazil. Conversely, increased hours of solar radiation in relatively drier and sunnier conditions can help mitigate the spread of COVID-19 (Martins et al. 2020).

Education emerged as a significant risk factor for mortality in our study. Our findings highlight a positive correlation between higher education levels and a relatively lower risk of mortality. Individuals with lower levels of education are more susceptible to unemployment issues, and during the COVID-19 pandemic, job loss can be particularly distressing for them. Workers with a high school education or less constitute a disproportionately larger portion of the unemployed population compared to the overall working-age demographic (Daly, Buckman, and Seitelman 2020). The labor market inequality has been exacerbated by the crisis, highlighting unequal access to education as another critical issue. Education level has also been identified as a risk factor in scientific literature (Baqui et al. 2021). Unlike the multivariate mixed-effects Cox model approach, the authors of a scientific report utilized machine learning prediction algorithms to unravel the complex interdependencies among various indicators. The predictive task was framed as a binary classification problem, where '0' denoted death and '1' represented recovery. The analysis leveraged the XGBoost (XGB) algorithm, alongside logistic regression, K-nearest neighbor, neural network, random forest, and support vector machine algorithms. The study revealed that socioeconomic, geographic, and structural factors outweighed individual comorbidities in influencing outcomes in Brazil. Key factors of significance included housing conditions and development indicators, proximity to hospitals (especially in rural or less developed areas), education levels, and the hospital financing model and pressures.

In a study focusing on vaccine effectiveness, a publication in Lancet Regional Health explored the impact of age on the efficacy and longevity of vaccine protection against Vaxzevria and CoronaVac (Cerqueira-Silva et al. 2022). The researchers utilized data not only from SIVEP-Gripe but also e-SUS-Notifica to evaluate vaccine effectiveness using a negative binomial regression model, adjusting for sociodemographic factors. Their findings, akin to our study, demonstrated high vaccine effectiveness (VE) against mortality, with rates of 92.3% for Vaxzevria. They also explored VE against hospitalization and ICU admission, showing rates of 91.4% and 91.1% for Vaxzevria, respectively. For CoronaVac, VE rates against these outcomes were 71.2%, 72.2%, and 73.7%, respectively. VE across all outcomes exhibited a gradual decline with increasing age.

In our investigation, we observed that vaccines did not provide significant protection for hospitalized patients with various comorbidities. The SARI fatality ratio increased with the number of comorbidities, peaking at approximately 5 comorbidities. Comparing Sinovac and AstraZeneca vaccines, AstraZeneca demonstrated greater efficacy than Sinovac in preventing deaths during the Gamma (P.1) variant-dominated period. During the Delta (B.1.617.2) variant period, Sinovac showed superior protection over AstraZeneca for patients with 4 or more comorbidities. In the Omicron (B.1.1.529) variant period, AstraZeneca was more protective than Sinovac in the North and Northeast regions, while Sinovac was more effective in the Central-west, Southeast, and South regions for patients with varying comorbidities.

It's important to note that our analysis was confined to hospitalized patients, and out-of-hospital mortality data were not considered. There could be limitations due to case determination and potential bias from missing information, including delayed hospitalization due to insufficient data. Inequalities in healthcare access may have exacerbated these disparities. Moreover, our study did not account for reinfection, and variant misclassification bias may have occurred due to limited patient sequencing.

Chapter 4

4 Conclusion and Future Work

Our study delved into the notable reduction in the infection fatality rate (IFR) associated with the Omicron (B.1.1.529) variant compared to previous SARS-CoV-2 variants in South Africa. This investigation commenced shortly after the emergence of the Omicron variants, with South Africa being one of the first regions impacted by this novel strain. Prior studies had already highlighted the Omicron variant's heightened immune evasion ability and transmissibility, along with reduced severity. In our research, we developed a sophisticated mathematical model incorporating time-varying transmission rates, vaccination parameters, and immune evasion dynamics. By fitting this model to case and death data up to February 6, 2022, we estimated the transmissibility and infection fatality ratio (IFR) specific to the Omicron variant in South Africa. Our findings revealed a substantial decrease of approximately 78.7% (95% confidence interval: 66.9% to 85.0%) in the infection fatality rate of the Omicron variant compared to its predecessors, highlighting its lower severity despite heightened transmissibility. Thus, our data-driven mathematical modeling provides evidence that the Omicron variant exhibits high transmissibility coupled with significantly reduced infection fatality rates relative to earlier SARS-CoV-2 variants.

In another line of inquiry, we investigated the transmissibility dynamics of all COVID-19 variants circulating in South Africa during the pandemic. South Africa experienced multiple waves of mortality attributed to COVID-19, with three distinct genetic variants of SARS-CoV-2 and their ancestral strain dominating consecutively. Employing an advanced mathematical modeling approach, we estimated the time-varying transmissibility of SARS-CoV-2, specifically examining the relative transmissibility of the Beta, Delta, and Omicron variants. Our analysis revealed substantial increases in transmissibility for these variants, ranging from 73% to 276% higher than their respective predecessors. To our knowledge, our modeling framework represents the first to simulate multiple mortality waves and the replacement dynamics of three variants in South Africa, underscoring the heightened transmissibility of the Omicron variant compared to earlier strains.

Additionally, we explored the regional heterogeneity of in-hospital mortality associated with COVID-19 in Brazil. By fitting a multivariate mixed-effect Cox model to a comprehensive national database of COVID-19 inpatients from February 27, 2020, to March 15, 2022, we discerned critical insights into mortality risks among vaccinated and unvaccinated individuals, accounting for age, ethnicity, education, and comorbidities. Our analysis identified age as the most significant risk factor for death, with notable disparities observed across education levels and geographical regions. In terms of vaccine efficacy, our study highlighted varied protective effects across different vaccine types and regions, emphasizing the intricate interplay between vaccination strategies, social inequality, and regional health outcomes. This investigation underscores the

multifaceted nature of COVID-19 mortality and vaccination dynamics in Brazil, shedding light on critical factors contributing to the pandemic's impact on diverse populations.

As a researcher in mathematical epidemiology field, I believe that preventing the next pandemic should focus on several key scientific approaches:

1. **Early Detection and Monitoring:** Utilize mathematical models and big data to establish early warning systems. By monitoring real-time data like infection rates and geographic spread, potential outbreaks can be identified early, enabling swift intervention.
2. **Transmission Dynamics Modeling:** Understanding disease spread through dynamic models helps predict outbreak trends, evaluate the effectiveness of control measures, and guide decision-making on strategies like social distancing and vaccination.
3. **Optimized Vaccine and Drug Allocation:** Mathematical optimization can determine the most effective distribution of limited vaccines and treatments, especially in resource-constrained situations, to maximize their impact in controlling the outbreak.
4. **Modeling Public Behavior and Social Response:** The spread of a pandemic is influenced by public behavior and societal reactions. Modeling these factors can lead to more accurate predictions and targeted interventions.
5. **International Cooperation and Data Sharing:** Global collaboration and data sharing are crucial. A worldwide data-sharing platform can enhance global early warning and response capabilities.

6. **Continuous Model Updating and Simulation:** Regular updates to models and data, along with scenario simulations, can prepare policymakers to respond effectively to emerging threats.

In summary, mathematical modeling and data analysis are essential for predicting, preventing, and responding to future pandemics, requiring close cooperation between scientists, governments, public health institutions, and the public.

In Appendix part, we observed that the dominance of the Alpha variant was relatively short-lived, lasting only 3 to 4 months, before being swiftly supplanted by the Delta variant in a synchronized manner across numerous countries and regions. The Delta variant initially surfaced in India and neighboring areas before spreading to other parts of the world. In contrast, the Alpha variants, originating in the UK, followed a similar trajectory of regional expansion before becoming globally prevalent. Our visual analysis highlights a remarkable simultaneous replacement of the Alpha variant by the Delta variant in multiple countries and regions. This synchronized substitution contrasts with the earlier replacement of the Alpha variant for the wild strain, which did not exhibit such uniformity. The higher transmissibility of the Delta variant likely contributed to this distinct pattern of viral dissemination during the ongoing COVID-19 pandemic. In addition, we did some preliminary work in MDR-TB field, we hope that the modern machine learning technics based on big data can really predict the MDR-TB and benefit to patients.

For future work, here are some potential methodological improvements for SEIR (Susceptible-Exposed-Infectious-Recovered) infectious disease models:

1. Parameter Estimation Methods:
    - Enhanced Parameter Estimation: Use advanced statistical methods, such as Bayesian inference or machine learning techniques, to more accurately estimate model parameters.
    - Data Integration: Combine data from various sources (e.g., clinical data, demographic data, mobility data) to improve the accuracy of parameter estimation.
2. Model Structure Extensions:
    - Introduce Additional States: For example, incorporate an "isolation" state (SEIQRD model) to better simulate real-world scenarios.
3. Individual Heterogeneity:
    - Individual Differences: Incorporate individual heterogeneity, such as variations in age groups, social behavior, and health status, to enhance the model's adaptability and accuracy.
    - Population Structure: Consider the impact of social networks or group structures, such as households, schools, and workplaces, on transmission patterns.
4. Model Validation and Calibration:
    - Model Validation: Validate the model using multiple data sources and actual observations to ensure its accuracy and reliability.
    - Calibration and Sensitivity Analysis: Conduct systematic model calibration and sensitivity analysis to identify the impact of key parameters on the results.
5. Uncertainty Analysis:

- Incorporate Uncertainty: Perform uncertainty analysis, such as Monte Carlo simulations, to assess the robustness and uncertainty of model predictions.
- Risk Assessment: Evaluate potential extreme scenarios to inform emergency response strategies.

6. Computational Efficiency:
    - Improve Computational Methods: Use more efficient computational methods or optimization algorithms to reduce model run time and computational cost.
    - Parallel Computing: Apply parallel computing techniques to handle large-scale datasets and complex models.

These methodological improvements can make SEIR models more realistic, accurate, and adaptable, thereby better supporting infectious disease research and public health decision-making.

5 Appendices

5.1 Large-scale synchronized replacement of Alpha (B.1.1.7) variant by the Delta (B.1.617.2) variant of SARS-COV-2 in the COVID-19 pandemic

5.1.1 Introduction

The global public health landscape has been profoundly impacted by the coronavirus disease 2019 (COVID-19), stemming from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). By February 2021, over 100 million individuals had received a diagnosis of SARS-CoV-2 infection, with the disease responsible for over 2 million deaths globally (Wang et al. 2021). The virus underwent rapid evolution, leading to the emergence of several variants. By March 2021, the Alpha (B.1.1.7) variant had begun spreading in Cambodia and Thailand. In certain clusters within Thailand, the positivity rate for SARS-CoV-2 testing reached as high as 60%-90% (Chookajorn et al. 2021). According to (Vassallo et al. 2021), in addition to its higher transmissibility compared to the previous wild strain, individuals infected with the Alpha (B.1.1.7) variant faced

an elevated risk of hospitalization relative to those infected with the earlier wild strain, underscoring the heightened virulence of the Alpha (B.1.1.7) variant.

Simultaneously, the emergence of the Gamma (B.1.617.1) and Delta (B.1.617.2) variants in Maharashtra, India led to a resurgence of cases across the country. Unlike the Alpha (B.1.1.7) variant, the Delta (B.1.617.2) lineage is characterized by eight non-synonymous mutations in the S protein. The Delta (B.1.617.2) variant, now widespread in over 200 countries/regions, has been designated as a variant of concern by the CDC (Farinholt et al. 2021). In addition to its heightened transmissibility compared to the Alpha (B.1.1.7) variant, individuals infected with the Delta (B.1.617.2) variant were more than twice as likely to require hospitalization compared to those with the Alpha (B.1.1.7) variant (Liu, Gayle, et al. 2020).

If a virus's genetic code, or genome, undergoes one or more mutations from its original version, it is classified as a variant. Samples collected from patients are sequenced using next-generation sequencing techniques to identify these variants. Comparing the Alpha (B.1.1.7) and Delta (B.1.617.2) variants, the Alpha (B.1.1.7) variant has 19 non-synonymous mutations, including 8 mutations affecting spike proteins. According to Liu et al., their research demonstrated that only the N501Y mutation consistently resulted in increased fitness in experimental models. This mutation, in combination with the eight other Alpha spike mutations, is thought to be the primary driver of enhanced viral transmission observed in the Alpha (B.1.1.7) variant (Liu et al. 2021). The Delta (B.1.617.2) variant exhibits 23 mutations compared to the Alpha (B.1.1.7) strain,

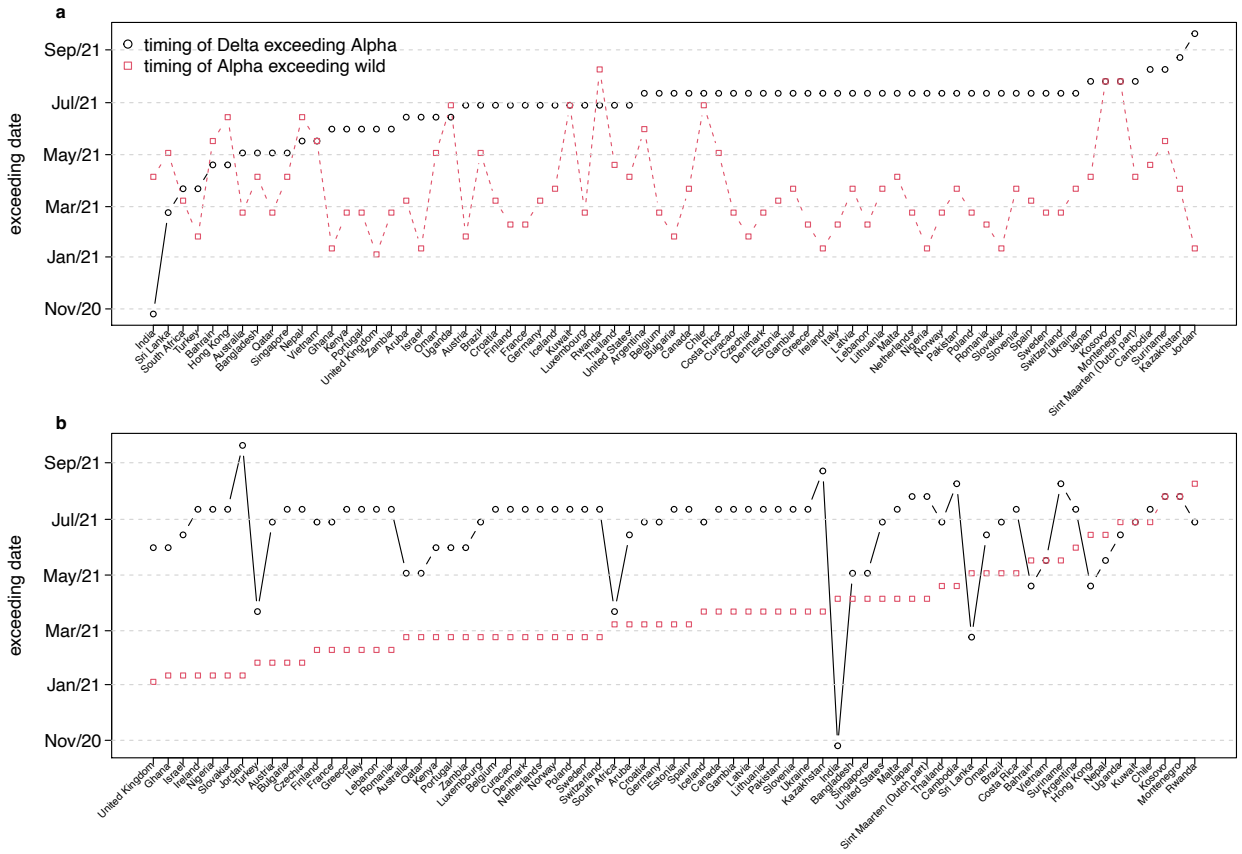
with notable changes in the spike protein believed to contribute to Delta's (B.1.617.2) transmission advantage. Among these mutations are T19R, L452R, T478K, D614G, P681R, and D960N, along with deletions at sites 157 and 158. The most significant spike protein mutations in Delta (B.1.617.2) include L452R and P681R. The L452R mutation replaces leucine at position 452 with arginine, enhancing spike protein affinity for ACE2 receptors and potentially affecting antibody binding that evades vaccine-induced immunity (Mor et al. 2021). The mutation P681R, which replaces proline at position 681 with arginine, enhances the virus's ability to integrate into host cells compared to variants lacking this mutation (Shiehzadegan et al. 2021).

Beginning in June 2020, India initiated 11 phases of unlocking, with the 11th phase announced at the end of March 2021 and set to remain in effect until April 30, 2021. However, starting in mid-April 2021, India experienced a severe surge in the pandemic. By May 10, 2021, more than 388,000 people had been affected (Ghosh et al. 2021). And according to another report by (Shrivastava et al. 2021), distinctive mutations associated with the Delta (B.1.617.2) variant were detected in sequences collected from India in April and May 2021, leading to its emergence as the predominant transmission variant in India during May and June 2021.

5.1.2 Method and result

In this work, we visualize the replacement of the previous strain with the Alpha (B.1.1.7) variant and the replacement of the Alpha (B.1.1.7) variant with Delta (B.1.617.2) variants globally. We find that the Alpha (B.1.1.7) variant only dominated for a short period of 3-4 months and the replacement of Alpha (B.1.1.7) with Delta (B.1.617.2) shows a surprisingly synchronous pattern in a large number of countries/regions.

Proportions of the different variants of concern confirmed over time (https://ourworldindata.org) We downloaded biweekly aggregated variant proportion data from "The Our World in Data" which obtained their data originally from GISAID Initiative(Khare et al. 2021; Elbe and Buckland-Merrett 2017; Shu and McCauley 2017a)

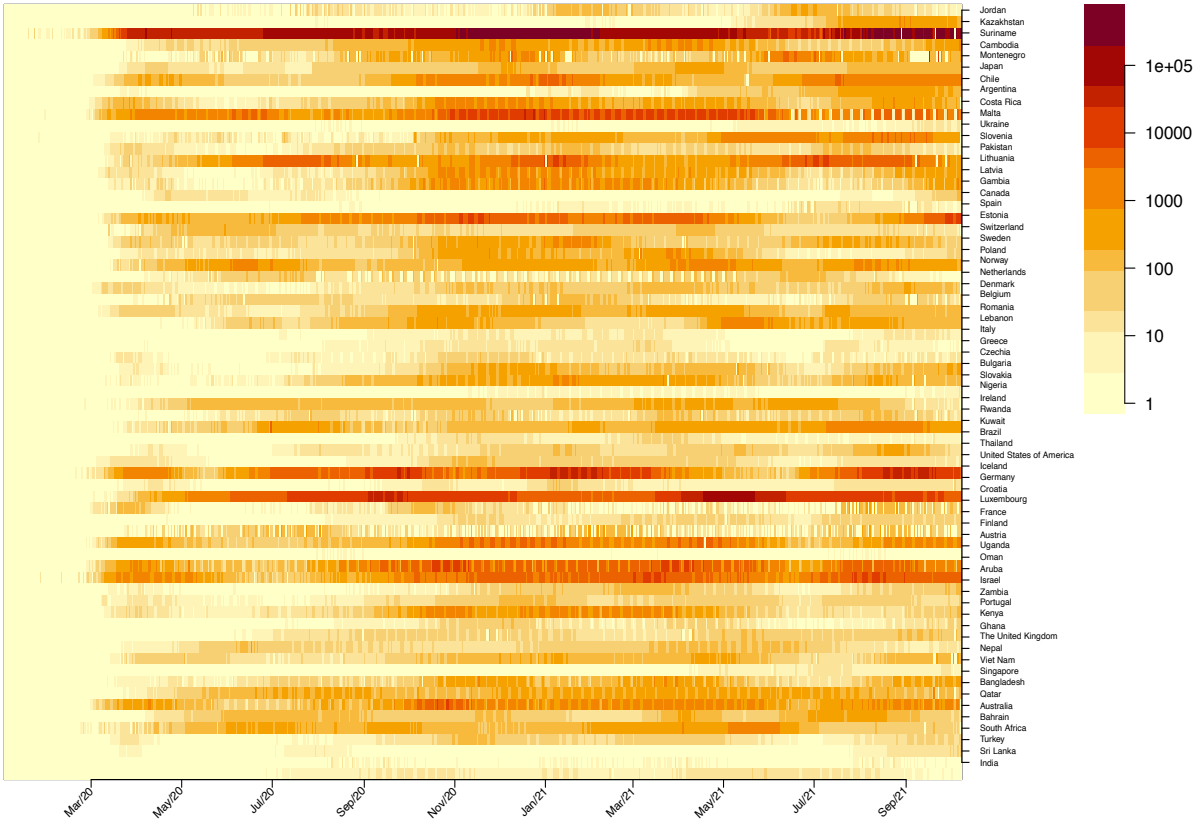


5.1 Timing of the confirmations of Delta (B.1.617.2) variant exceeding Alpha (B.1.1.7) variant.

The first time that the proportion of B.1.617.2 (Delta variant) > the proportion of B.1.1.7 (Alpha variant), denoted as $T_{A\to\Delta}$ in 71 countries/regions, versus timing of the confirmations of Alpha (B.1.1.7) variant exceeding the previous strain $T_{w\to\alpha}$. (a) countries/regions are ordered from left to right according to $T_{A\to\Delta}$. (b) countries/regions are ordered from left to right according to $T_{W\to A}$.

From **Figure 5.1**, we can see that the Delta (B.1.617.2) variant was first found in India and spread first in neighboring countries/regions, gradually to other countries/regions. Alpha (B.1.1.7) variant was first found in the United Kingdom and spread first in neighboring countries/regions, gradually to the rest of the world. $T_{A\to\Delta}$ (Black circle) has a surprising synchrony pattern across a large number of countries/regions (ie, a horizontal line across a large number of countries/regions in July 2021). This synchrony pattern is less evident in the $T_{W\to A}$ (red squares). In other words, the timing of the Delta (B.1.617.2) variant replacing the Alpha (B.1.1.7) variant occurred simultaneously in many countries/regions. In contrast, the timing of the Alpha (B.1.1.7) variant replacing the previous wild strain did not show a strong synchronous substitution trend. This could be related to a higher transmissibility of Delta (B.1.617.2) variant compared to other previous strains. Delta (B.1.617.2) variant also possesses a shortened incubation period and increased viral load. In particular, the viral load is about 1000 times higher in patients who were infected with the Delta (B.1.617.2) variant than patients who were infected with the original strain. The first detectable time for the Delta (B.1.617.2) variant is 4 days after infection which is longer than the average detectable time of the original strain (6 days) (Reardon 2021)

Also, the invasion time of the Delta (B.1.617.2) variant happened at the tail of the wave of Alpha (B.1.1.7) variant and coincided with a relaxation of social distancing in many countries (see **Figure 5.2**).



5.2 The population standardized daily reported COVID-19 deaths

This figure report the population standardized daily reported COVID-19 deaths in these countries/regions listed. The countries/regions skipped by the Alpha (B.1.1.7) variant (e.g., India) had a mild 2020 year. The synchronized $T_{\alpha\to\delta}$ coincided by a trough of deaths in European countries. Here we show population standardized daily data (daily reported COVID-19 per 1 million population). A bright band (low deaths) can be seen in June-July 2021 when Delta (B.1.617.2) invaded. Data are obtained from https://covid19.who.int/info/.

5.1.3 Conclusion and Discussion

The dominance time of the Alpha (B.1.1.7) variant is between $T_{W\to A}$, and $T_{A\to \Delta}$. In South and Southeast Asia, e.g. India, the Alpha (B.1.1.7) variant failed to dominate, which could be associated with a mild first wave (with low deaths per capita in India in 2020).

In Scotland, BNT162b2 and AstraZeneca vaccines were 79% and 60% effective at preventing SARS-COV-2 Delta (B.1.617.2) variant infection after two doses, respectively. However, the Pfizer-Biontech and AstraZeneca vaccines maintain a high degree of protection against any infection similar to the Alpha (B.1.1.7) variant (Cevik et al. 2021).

In summary, we reported a large-scale synchronized replacement of the Alpha (B.1.1.7) variant by the Delta (B.1.617.2) variant which could be due to the invasion timing of Delta (B.1.617.2) variant and its relatively high transmissibility. Also, we note that these countries/regions in South and Southeast Asia experienced a mild 2020 year that was largely skipped by the Alpha (B.1.1.7) variant. Modeling simulation can be done using a model framework similar to those in previous works(Rohani, Earn, and Grenfell 1999; Rohani et al. 2003).

5.2 Preliminary work about MDR-TB

For future research work, Tuberculosis (TB) -related work especially the multidrug-resistant tuberculosis (MDR-TB). Tuberculosis is a bacterial disease that infects about one-third of the population worldwide. About 10% of those infected will develop active TB disease. The incubation period of TB ranges from a few months to decades, and the treatment period is approximately 6-9 months(Liu and Sun 2010). Tuberculosis remains one of the leading causes of morbidity and mortality worldwide. According to WHO ('World Health Organization. Global Tuberculosis Report' 2021), in 2020 an estimated 10 million people became infected with TB and 1.5 million died from TB, which is equivalent to 4100 deaths per day.

As TB is gradually brought under control with the incidence falling at about 2% per year globally, the spread of multidrug-resistant tuberculosis (MDR-TB, i.e. tuberculosis with resistance to isoniazid and rifampicin that does not respond to treatment or for which treatment is discontinued because of side effects) has a negative impact on TB control worldwide(Gandhi et al. 2010). The more severe format of MDR-TB is extensively drug-resistant tuberculosis (XDR-TB). XDR- TB is defined as resistance to isoniazid, rifampicin, any fluoroquinolone and at least one injectable drugs (capreomycin, kanamycin and amikacin) (Migliori et al. 2007; Velayati et al. 2009; Gandhi et al. 2010). The emergence of drug resistance is a major threat to TB control because it raises the prospect of returning the disease to a time when there was no cure (Raviglione 2006). The emergence of drug resistance in TB patients is due to spontaneous

mutations in the M tuberculosis genome that occur at a predictable rate (David 1970; Supply et al. 2003). And through artificial drug selection, these resistant bacteria gradually become the dominant strain, so to some extent, drug-resistant TB can also be considered as man-made event(Kaplan et al. 2003; Post et al. 2004). Drug resistance is a growing threat to world public health, in 2017, MDR-TB already killed 14% of tuberculosis patients all over the world (O'Neill 2016). Early studies explored mechanism of drug-resistance TB (Seung et al. 2004) (Borrell and Gagneux 2009).Antibiotic resistance first occurs when drug-sensitive TB patients receive inadequate or inappropriate treatment. Once drug-resistant bacteria are present in infected patients, it is possible that these bacteria can spread the resistant bacteria to others through respiratory modes of transmission, such as droplet transmission. And there is evidence that appear in the selection of resistant bacteria which is easy to spread. However, due to the complexity of the anti-drug mechanism of TB, no consensus is reached yet.

Mathematical modelling is a very important tool to study disease epidemics. For the prediction of MDR-TB, machine learning model is an effective method for early identification and detection of MDR-TB. A variety of machine learning models have been used to predict multidrug-resistant TB and feature selection(Ali et al. 2021; Evora, Seixas, and Kritski 2017; Chen et al. 2018; Solari et al. 2008).

In Pakistan, Mian et al (Ali et al. 2021) focused on the machine learning feature selection algorithm in order to identify the multidrug-resistant tuberculosis. They exploited decision tree, random forest, k-nearest neighbors, support vector machine (SVM), logistic regression, least absolute shrinkage and selection operator (LASSO), and artificial neural networks (ANNs) to analyze a

case-control dataset. The study demonstrated that close contact with MDR-TB patients, tobacco using, previous tuberculosis history, depression and improper treatment and the interruption of treatment may contribute to the MDR-TB, and SVM and Random Forest are the models with best performance to classify patients in this case.

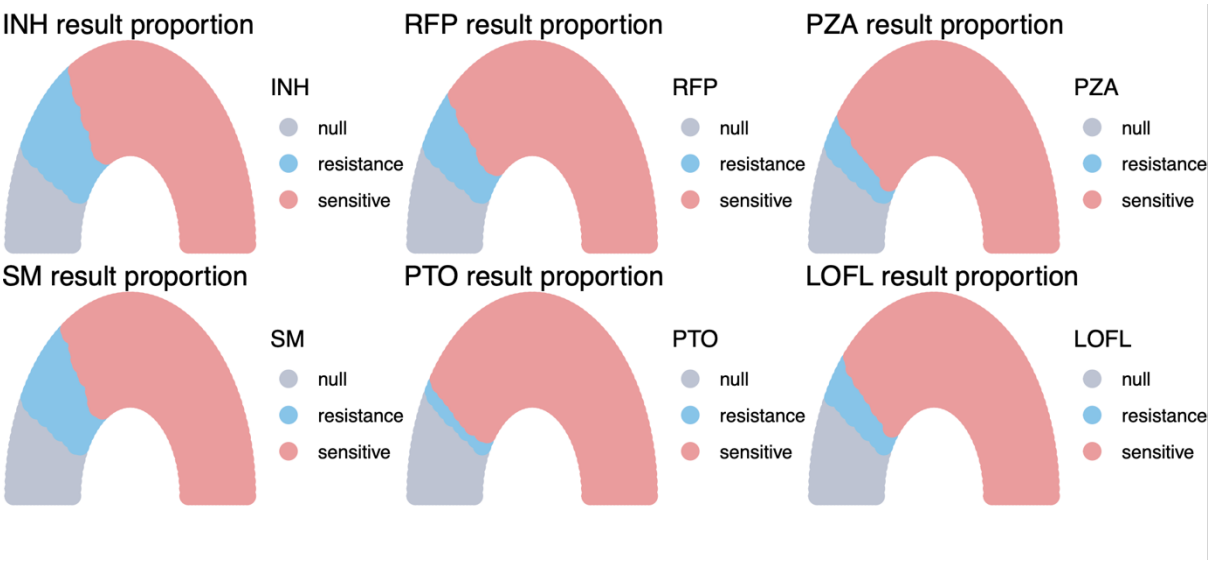
Évora et al (Evora, Seixas, and Kritski 2017) used machine learning methods to identify MDR-TB patients in Rio de Janeiro. Their data included clinical and demographic information. They used classification and regression trees as well as artificial neural networks to classified the MDR-TB patients, where the ANN algorithm achieved ideal accuracy and concluded that history of tuberculosis, close exposure to drug-resistant tuberculosis, high temperature, smoking, and hemoptysis were important factors.

Chen and Michael (Chen et al. 2018) obtained data of over 3000 tuberculosis patients, of which 1228 were MDR-TB cases. They attempted to predict drug-resistance tuberculosis by machine learning algorithms including random forest, logistic regression, and deep neural network (DNN). The performance of the three models were compared by determining the specificity, sensitivity and accuracy. The random forest demonstrated better performance than others with specificity of 92.7%, sensitivity of 93.7% and accuracy of 97.9%. The bloody cough, close contact with drug-resistant patients, history of tuberculosis, drug and alcohol abuse, poor medication management and high temperature were significant predictors.

This research will use the data from Shandong which includes the patient's own living habits, age, gender, medical condition and other variables to fit different machine learning models, e.g. decision tree, random forest, logistic

regression, support vector machine, to classify the MDR-TB patients and do the feature selection to obtain the important factors that may cause the emergence and transmission of MDR-TB. The best machine learning model for predicting MDR-TB cases in the future will be selected by parameter adjustment.

In the preliminary work, we utilized drug-susceptibility data in 2021 from Shandong province to do the exploratory analysis, including the age distribution of TB patients by gender, the proportions of TB cases with resistance to some common drugs and drug sensitivity tests. Further, for drugs with a higher proportion of reported resistance, the age distribution of drug-resistant cases by gender were analysed. By exploring the 2021 data, we envision that for the future research, we may have to quantify the heterogeneities of tuberculosis and multidrug-resistant tuberculosis incidence by gender, age group, resident status, occupation and other features, to predict the future trend of multidrug-resistant tuberculosis.

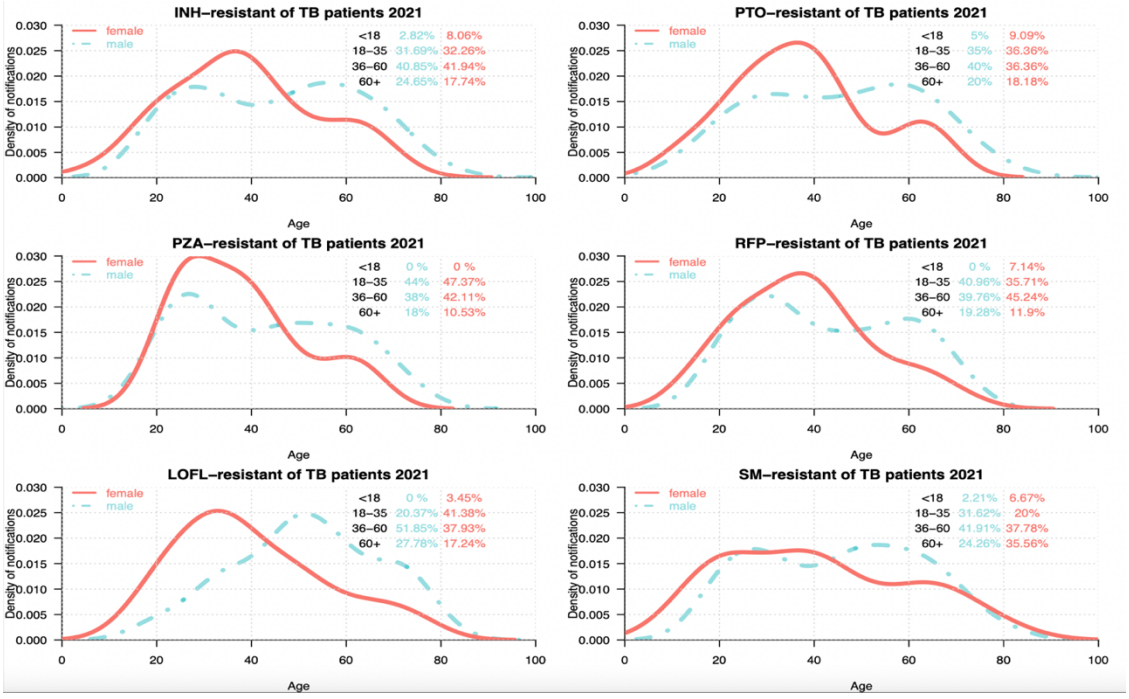


5.3 Drug sensitivity test of some common tuberculosis drugs

.

The red color represents patients with drug sensitivity, the blue color represents patients with drug resistance, the grey one represents missing values.

Figure 5.3 demonstrates drug sensitivity test results and the proportion of sensitive and resistance are shown. Due to the variety of drugs used by different patients, there are significant proportion of missing values in the drug sensitivity test results. By checking the commonly used tuberculosis drugs, about 18% and 11% of the tuberculosis patients resistant to the two most common TB drugs Isoniazid (INH) and Rifampicin (RFP) respectively.



5.4 Case densities and age distributions of TB cases by male and female with resistance to different drugs.

The red solid line represents density of female patients, and the green dashed-and-dotted line represents density of male patients.

Figure 5.4 shows the case densities and age distributions of TB cases by male and female with resistance to different drugs. Six drugs with higher proportion of reported resistance were selected. For each drug, the ages of the cases were grouped into four, i.e., <18, 18-35, 36-60 and 60+. For Isoniazid (INH), to which 18.1% of the cases were resistant, in both female and male, over 40% were in age group 36-60. In general, most patients with drug resistance were under 60 years old. And the proportion of elderly patients with drug resistance was not high, except for Streptomycin (SM), to which 36% of female patients aged 60 and above were resistant.

6 Bibliography

Abdool Karim, Salim S, and Cheryl Baxter. 2021. 'Impact of SARS-CoV-2 variants of concern on Covid-19 epidemic in South Africa', *Transactions of the Royal Society of South Africa*: 1-4.

Abdullah, F, J Myers, D Basu, G Tintinger, V Ueckermann, M Mathebula, R Ramlall, S Spoor, T de Villiers, and Z Van der Walt. 2021. 'Decreased severity of disease during the first global omicron variant covid-19 outbreak in a large hospital in tshwane, south africa', *International Journal of Infectious Diseases*.

Ahmed, Faheem, Na'eem Ahmed, Christopher Pissarides, and Joseph Stiglitz. 2020. 'Why inequality could spread COVID-19', *The Lancet Public Health*, 5: e240.

Ali, Mian Haider, Dost Muhammad Khan, Khalid Jamal, Zubair Ahmad, Sadaf Manzoor, and Zardad Khan. 2021. 'Prediction of multidrug-resistant tuberculosis using machine learning algorithms in swat, Pakistan', *Journal of healthcare engineering*, 2021.

Bai, Yuan, Zhanwei Du, Mingda Xu, Lin Wang, Peng Wu, Eric HY Lau, Benjamin J Cowling, and Lauren Ancel Meyers. 2021. 'International risk of SARS-CoV-2 Omicron variant importations originating in South Africa', *medRxiv*.

Baqui, Pedro, Ioana Bica, Valerio Marra, Ari Ercole, and Mihaela van Der Schaar. 2020. 'Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study', *The Lancet Global Health*, 8: e1018-e26.

Baqui, Pedro, Valerio Marra, Ahmed M Alaa, Ioana Bica, Ari Ercole, and Mihaela van Der Schaar. 2021. 'Comparing COVID-19 risk factors in Brazil using machine learning: the importance of socioeconomic, demographic and structural factors', *Scientific reports*, 11: 1-10.

Bernoulli, D. 1766. 'Essai d'une nouvelle analyse de la mortalité causé e par la petite vé role, et des avantages de l'inoculation pour la prévenir. Mem Math Phy Acad Roy Sci Paris 1–45', *Trans. Bradley L (1971):"An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it." Adult Education Department: Nottingham, UK*.

Bernoulli, Daniel, and Sally Blower. 2004. 'An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it', *Reviews in medical virology*, 14: 275.

Bishai, David, Benjamin Johns, Divya Nair, Juliet Nabyonga-Orem, Braka Fiona-Makmot, Emily Simons, and Alya Dabbagh. 2011. 'The cost-effectiveness of supplementary immunization activities for measles: a stochastic model for Uganda', *The Journal of infectious diseases*, 204: S107-S15.

Borrell, S, and S Gagneux. 2009. 'Infectiousness, reproductive fitness and evolution of drug-resistant Mycobacterium tuberculosis [State of the art]', *The International Journal of Tuberculosis and Lung Disease*, 13: 1456-66.

Boschiero, Matheus Negri, Camila Vantini Capasso Palamim, and Fernando Augusto Lima Marson. 2021. 'The hindrances to perform the COVID-19 vaccination in Brazil', *Human Vaccines & Immunotherapeutics*, 17: 3989-4004.

Bourouiba, Lydia. 2021. 'The fluid dynamics of disease transmission', *Annual Review of Fluid Mechanics*, 53: 473-508.

Brizzi, Andrea, Charles Whittaker, Luciana Servo, Iwona Hawryluk, Carlos A Prete, William M de Souza, Renato S Aguiar, Leonardo JT Araujo, Leonardo S Bastos, and Alexandra Blenkinsop. 2022. 'Spatial and temporal fluctuations in COVID-19 fatality rates in Brazilian hospitals', *Nature medicine*: 1-10.

Campbell, Finlay, Brett Archer, Henry Laurenson-Schafer, Yuka Jinnai, Franck Konings, Neale Batra, Boris Pavlin, Katelijn Vandemaele, Maria D Van Kerkhove, and Thibaut Jombart. 2021. 'Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021', *Eurosurveillance*, 26: 2100509.

Campbell, Harlan, and Paul Gustafson. 2021. 'Inferring the COVID-19 IFR with a simple Bayesian evidence synthesis of seroprevalence study data and imprecise mortality data', *medRxiv*: 21256975.

Cao, Yunlong, Jing Wang, Fanchong Jian, Tianhe Xiao, Weiliang Song, Ayijiang Yisimayi, Weijin Huang, Qianqian Li, Peng Wang, and Ran An. 2021. 'B. 1.1. 529 escapes the majority of SARS-CoV-2 neutralizing antibodies of diverse epitopes', *BioRxiv*: 2021.12. 07.470392.

Cele, Sandile, Inbal Gazy, Laurelle Jackson, Shi-Hsia Hwa, Houriiyah Tegally, Gila Lustig, Jennifer Giandhari, Sureshnee Pillay, Eduan Wilkinson, and Yeshnee Naidoo. 2021. 'Escape of SARS-CoV-2 501Y. V2 from neutralization by convalescent plasma', *Nature*, 593: 142-46.

Cerqueira-Silva, Thiago, Vinicius de Araújo Oliveira, Viviane S Boaventura, Julia M Pescarini, Juracy Bertoldo Júnior, Tales Mota Machado, Renzo Flores-Ortiz, Gerson O Penna, Maria Yury Ichihara, and Jacson Venâncio de Barros. 2022. 'Influence of age on the effectiveness and duration of protection of Vaxzevria and CoronaVac vaccines: a population-based study', *The Lancet Regional Health-Americas*, 6: 100154.

Cevik, Muge, Nathan D Grubaugh, Akiko Iwasaki, and Peter Openshaw. 2021. 'COVID-19 vaccines: Keeping pace with SARS-CoV-2 variants', *Cell*.

Chan, Paul KS. 2002. 'Outbreak of avian influenza A (H5N1) virus infection in Hong Kong in 1997', *Clinical infectious diseases*, 34: S58-S64.

Chemaitelly, Hiam, Roberto Bertollini, and Laith J Abu-Raddad. 2021. 'Efficacy of Natural Immunity against SARS-CoV-2 Reinfection with the Beta Variant', *New England Journal of Medicine*, 385: 2585-86.

Chen, J, R Wang, NB Gilby, and GW Wei. 'Omicron (B. 1.1. 529): Infectivity, vaccine breakthrough, and antibody resistance. arXiv 2021', *arXiv preprint arXiv:2112.01318*.

Chen, Michael L, Akshith Doddi, Jimmy Royer, Luca Freschi, Marco Schito, Matthew Ezewudo, Isaac S Kohane, Andrew Beam, and Maha Farhat. 2018. 'Deep learning predicts tuberculosis drug resistance status from genome sequencing data', *BioRxiv*: 275628.

Chookajorn, Thanat, Theerarat Kochakarn, Chaiwat Wilasang, Namfon Kotanan, and Charin Modchang. 2021. 'Southeast Asia is an emerging hotspot for COVID-19', *Nature medicine*: 1-2.

Chu, Yu-Ming, Aatif Ali, Muhammad Altaf Khan, Saeed Islam, and Saif Ullah. 2021. 'Dynamics of fractional order COVID-19 model with a case study of Saudi Arabia', *Results in Physics*, 21: 103787.

Correction Naghavi, Mohsen, Haidong Wang, Rafael Lozano, Adrian Davis, Xiaofeng Liang, Maigeng Zhou, Stein Emil Vollset, Ayse Abbasoglu Ozgoren, Safa Abdalla, and Foad Abd Allah. 2015. 'Global, regional, and national age-sex

specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013', *The Lancet*, 385: 117-71.

Daly, Mary C, Shelby R Buckman, and Lily M Seitelman. 2020. 'The unequal impact of COVID-19: why education matters', *FRBSF Economic Letter*, 17: 1-5.

David, Hugo L. 1970. 'Probability distribution of drug-resistant mutants in unselected populations of Mycobacterium tuberculosis', *Applied microbiology*, 20: 810-14.

de León-Martínez, Lorena Díaz, Andrés Palacios-Ramírez, Maribel Rodriguez-Aguilar, and Rogelio Flores-Ramírez. 2020. 'Critical review of social, environmental and health risk factors in the Mexican indigenous population and their capacity to respond to the COVID-19', *Science of The Total Environment*, 733: 139357.

Deeks, Steven G, Julie Overbaugh, Andrew Phillips, and Susan Buchbinder. 2015. 'HIV infection', *Nature reviews Disease primers*, 1: 1-22.

Dejnirattisai, Wanwisa, Robert H Shaw, Piyada Supasa, Chang Liu, Arabella SV Stuart, Andrew J Pollard, Xinxue Liu, Teresa Lambe, Derrick Crook, and Dave I Stuart. 2022. 'Reduced neutralisation of SARS-CoV-2 omicron B. 1.1. 529 variant by post-immunisation serum', *The Lancet*, 399: 234-36.

Del Rio, Carlos, Preeti N Malani, and Saad B Omer. 2021. 'Confronting the delta variant of SARS-CoV-2, summer 2021', *Jama*, 326: 1001-02.

Demmler, Gail J, and B Lee Ligon. 2003. "Severe acute respiratory syndrome (SARS): a review of the history, epidemiology, prevention, and concerns for the future." In *Seminars in pediatric infectious diseases*, 240-44. Elsevier.

Dyer, Owen. 2021. 'Covid-19: Peru's official death toll triples to become world's highest', *BMJ (Clinical research ed.)*, 373: n1442.

ECDC. 2022. 'SARS-CoV-2 Variants of Concern as of 11 March 2022'.

Elbe, Stefan, and Gemma Buckland-Merrett. 2017. 'Data, disease and diplomacy: GISAID's innovative contribution to global health', *Global challenges*, 1: 33-46.

Evora, LHRA, JM Seixas, and Afrânio Lineu Kritski. 2017. 'Neural network models for supporting drug and multidrug resistant tuberculosis screening diagnosis', *Neurocomputing*, 265: 116-26.

Faria, Nuno R, Andrew Rambaut, Marc A Suchard, Guy Baele, Trevor Bedford, Melissa J Ward, Andrew J Tatem, João D Sousa, Nimalan Arinaminpathy, and Jacques Pépin. 2014. 'The early spread and epidemic ignition of HIV-1 in human populations', *Science*, 346: 56-61.

Farinholt, Timothy, Harsha Doddapaneni, Xiang Qin, Vipin Menon, Qingchang Meng, Ginger Metcalf, Hsu Chao, Marie-Claude Gingras, Vasanthi Avadhanula, and Paige Farinholt. 2021. 'Transmission event of SARS-CoV-2 Delta variant reveals multiple vaccine breakthrough infections', *BMC medicine*, 19: 1-6.

Fauci, Anthony S. 2001. 'Infectious diseases: considerations for the 21st century', *Clinical infectious diseases*, 32: 675-85.

Ferguson, Neil, Azra Ghani, Anne Cori, Alexandra Hogan, Wes Hinsley, and E Volz. 2021. 'Report 49: Growth, population distribution and immune escape of Omicron in England', *Imperial College London (16-12-2021), doi: https://doiorg/1025561*, 93038.

Freitas, André Ricardo Ribas, Otto Albuquerque Beckedorff, Luciano Pamplona de Góes Cavalcanti, Andre M Siqueira, Daniel Barros de Castro, Cristiano Fernandes da Costa, Daniele Rocha Queiróz Lemos, and Eliana NC Barros. 2021. 'The emergence of novel SARS-CoV-2 variant P. 1 in Amazonas (Brazil)

was temporally associated with a change in the age and sex profile of COVID-19 mortality: A population based ecological study', *The Lancet Regional Health–Americas*, 1.

Gandhi, Neel R, Paul Nunn, Keertan Dheda, H Simon Schaaf, Matteo Zignol, Dick Van Soolingen, Paul Jensen, and Jaime Bayona. 2010. 'Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis', *The Lancet*, 375: 1830-43.

Gardner, Billy J, and A Marm Kilpatrick. 2021. 'Estimates of reduced vaccine effectiveness against hospitalization, infection, transmission and symptomatic disease of a new SARS-CoV-2 variant, Omicron (B. 1.1. 529), using neutralizing antibody titers', *medRxiv*.

Gaudart, Jean, Mohamad Ghassani, Julie Mintsa, Jules Waku, Mustapha Rachdi, Ogobara K Doumbo, and Jacques Demongeot. 2010. "Demographic and spatial factors as causes of an epidemic spread, the copule approach: application to the retro-prediction of the black death epidemy of 1346." In *2010 IEEE 24th International conference on advanced information networking and applications workshops*, 751-58. IEEE.

Gaudart, Jean, Ousmane Touré, Nadine Dessay, A Lassane Dicko, Stéphane Ranque, Loic Forest, Jacques Demongeot, and Ogobara K Doumbo. 2009. 'Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali', *Malaria journal*, 8: 1-12.

Ghosh, Nimisha, Indrajit Saha, Jnanendra Prasad Sarkar, and Ujjwal Maulik. 2021. 'Strategies for COVID-19 Epidemiological Surveillance in India: Overall Policies Till June 2021', *Frontiers in Public Health*, 9.

Gottlieb, Michael S, Robert Schroff, Howard M Schanker, Joel D Weisman, Peng Thim Fan, Robert A Wolf, and Andrew Saxon. 1981. 'Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency', *New England journal of medicine*, 305: 1425-31.

Gu, Haogao, Pavithra Krishnan, Daisy YM Ng, Lydia DJ Chang, Gigi YZ Liu, Samuel SM Cheng, Mani MY Hui, Mathew CY Fan, Jacob HL Wan, and Leo HK Lau. 2022. 'Probable transmission of SARS-CoV-2 omicron variant in quarantine hotel, Hong Kong, China, November 2021', *Emerging infectious diseases*, 28: 460.

Guldberg, CM, P Waage, and EW Lund. 1864. "Studies concerning affinity. CM Forhandlinger: Videnskabs-Selskabet i Christiana, 35." In, 548-50.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, and Saptarshi Majumdar. 2021. 'A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)', *Nature Human Behaviour*, 5: 529-38.

Halpin, David MG, Rosa Faner, Oriol Sibila, Joan Ramon Badia, and Alvar Agusti. 2020. 'Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection?', *The Lancet Respiratory Medicine*, 8: 436-38.

Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, Max Roser. 2020. 'Coronavirus Pandemic (COVID-19)', Accessed Feb 28. https://ourworldindata.org/coronavirus.

He, Daihai, Edward L Ionides, and Aaron A King. 2010a. 'Plug-and-play inference for disease dynamics: measles in large and small populations as a case study', *J Royal Society Interface*, 7: 271-83.

———. 2010b. 'Plug-and-play inference for disease dynamics: measles in large and small populations as a case study', *Journal of the Royal Society Interface*, 7: 271-83.
He, Xuemei, Weiqi Hong, Xiangyu Pan, Guangwen Lu, and Xiawei Wei. 2021. 'SARS-CoV-2 Omicron variant: characteristics and prevention', *MedComm*.
Hodcroft, Emma B. 2021. 'CoVariants: SARS-CoV-2 Mutations and Variants of Interest.'. https://covariants.org.
Hone, Thomas, Davide Rasella, Mauricio L Barreto, Azeem Majeed, and Christopher Millett. 2017. 'Association between expansion of primary healthcare and racial inequalities in mortality amenable to primary care in Brazil: a national longitudinal analysis', *PLoS medicine*, 14: e1002306.
'https://www.who.int/emergencies/diseases/novel-coronavirus-2019'. 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
Huang, Chaolin, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, and Xiaoying Gu. 2020. 'Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China', *The Lancet*, 395: 497-506.
Hui, David SC, and Paul KS Chan. 2010. 'Severe acute respiratory syndrome and coronavirus', *Infectious Disease Clinics*, 24: 619-38.
Ionides, Edward L, Anindya Bhadra, Yves Atchadé, and Aaron King. 2011. 'Iterated filtering'.
Ionides, Edward L, Carles Bretó, and Aaron A King. 2006. 'Inference for nonlinear dynamical systems', *Proceedings of the National Academy of Sciences*, 103: 18438-43.
Ito, Kimihito, Chayada Piantham, and Hiroshi Nishiura. 2021. 'Predicted dominance of variant Delta of SARS-CoV-2 before Tokyo Olympic Games, Japan, July 2021', *Eurosurveillance*, 26: 2100570.
———. 2022. 'Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark', *Journal of medical virology*, 94: 2265-68.
'Johns Hopkins University CSSE.'. 2022. https://github.com/CSSEGISandData/COVID-19.
Kaplan, Gilla, Frank A Post, Andre L Moreira, Helen Wainwright, Barry N Kreiswirth, Melike Tanverdi, Barun Mathema, Srinivas V Ramaswamy, Gabi Walther, and Lafras M Steyn. 2003. 'Mycobacterium tuberculosis Growth at theCavity Surface: a Microenvironment with FailedImmunity', *Infection and immunity*, 71: 7099-108.
Karim, Salim S Abdool, and Quarraisha Abdool Karim. 2021. 'Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic', *The Lancet*, 398: 2126-28.
Keele, Brandon F, Fran Van Heuverswyn, Yingying Li, Elizabeth Bailes, Jun Takehisa, Mario L Santiago, Frederic Bibollet-Ruche, Yalu Chen, Louise V Wain, and Florian Liegeois. 2006. 'Chimpanzee reservoirs of pandemic and nonpandemic HIV-1', *Science*, 313: 523-26.
Kermack, William Ogilvy, and Anderson G McKendrick. 1927. 'A contribution to the mathematical theory of epidemics', *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115: 700-21.
———. 1932. 'Contributions to the mathematical theory of epidemics. II.—The problem of endemicity', *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 138: 55-83.

———. 1933. 'Contributions to the mathematical theory of epidemics. III.—Further studies of the problem of endemicity', *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 141: 94-122.

Kevin Baird, J. 2017. 'Malaria control by commodities without practical malariology', *BMC public health*, 17: 590.

Khare, Shruti, Céline Gurry, Lucas Freitas, Mark B Schultz, Gunter Bach, Amadou Diallo, Nancy Akite, Joses Ho, Raphael TC Lee, and Winston Yeo. 2021. 'GISAID's Role in Pandemic Response', *China CDC Weekly*, 3: 1049.

Khoury, David S, Megan Steain, James Triccas, Alex Sigal, Miles Philip Davenport, and Deborah Cromer. 2021. 'Analysis: A meta-analysis of Early Results to predict Vaccine efficacy against Omicron', *medRxiv*.

King, Aaron A, Dao Nguyen, and Edward L Ionides. 2015. 'Statistical inference for partially observed Markov processes via the R package pomp', *arXiv preprint arXiv:1509.00503*.

Korber, Bette, Will M Fischer, Sandrasegaram Gnanakaran, Hyejin Yoon, James Theiler, Werner Abfalterer, Nick Hengartner, Elena E Giorgi, Tanmoy Bhattacharya, and Brian Foley. 2020. 'Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus', *Cell*, 182: 812-27. e19.

Ksiazek, Thomas G, Dean Erdman, Cynthia S Goldsmith, Sherif R Zaki, Teresa Peret, Shannon Emery, Suxiang Tong, Carlo Urbani, James A Comer, and Wilina Lim. 2003. 'A novel coronavirus associated with severe acute respiratory syndrome', *New England journal of medicine*, 348: 1953-66.

Kuhlmann, Constanze, Carla Konstanze Mayer, Mathilda Claassen, Tongai G Maponga, Andrew D Sutherland, Tasnim Suliman, Megan Shaw, and Wolfgang Preiser. 2021. 'Breakthrough infections with SARS-CoV-2 Omicron variant despite booster dose of mRNA vaccine', *Available at SSRN 3981711*.

Kumar, Ashutosh, Adil Asghar, Khursheed Raza, Ravi K Narayan, Rakesh K Jha, Abhigyan Satyam, Gopichand Kumar, Prakhar Dwivedi, Chetan Sahni, and Chiman Kumari. 2021. 'Demographic characteristics of SARS-CoV-2 B. 1.617. 2 (Delta) variant infections in Indian population', *medRxiv*: 2021.09. 23.21263948.

Kumar, Suresh, Thiviya S Thambiraja, Kalimuthu Karuppanan, and Gunasekaran Subramaniam. 2021. 'Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein', *Journal of medical virology*.

Lambert, JH. 1772. 'Die Tödlichkeit der Kinderblattern', *Beyträage zum Gebrauche der Mathematik und deren Anwendung*, 3: 568.

Lan, Jun, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, and Linqi Zhang. 2020. 'Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor', *Nature*, 581: 215-20.

'Latest situation of COVID-19 (as of 4 April 2022)'. 2022. Centre for Health Protection (CHP) of the Department of Health (DH) Hong Kong, Accessed April 4. https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf.

Le Rutte, Epke A, Andrew J Shattock, Nakul Chitnis, Sherrie L Kelly, and Melissa A Penny. 2021. 'Assessing impact of Omicron on SARS-CoV-2 dynamics and public health burden', *medRxiv*.

Lekone, Phenyo E, and Bärbel F Finkenstädt. 2006. 'Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study', *Biometrics*, 62: 1170-77.

Leung, Kathy, Marcus HH Shum, Gabriel M Leung, Tommy TY Lam, and Joseph T Wu. 2021. 'Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020', *Eurosurveillance*, 26: 2002106.

Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, and Jessica Y Wong. 2020. 'Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia', *New England journal of medicine*, 382: 1199-207.

Li, Sabrina L, Rafael HM Pereira, Carlos A Prete Jr, Alexander E Zarebski, Lucas Emanuel, Pedro JH Alves, Pedro S Peixoto, Carlos KV Braga, Andreza A de S Santos, and William M de Souza. 2021. 'Social and racial inequalities as contributing factors to COVID-19 vulnerability in São Paulo, Brazil'.

Li, Xiao-Ping, Ye Wang, Muhammad Altaf Khan, Mohammad Y Alshahrani, and Taseer Muhammad. 2021. 'A dynamical study of SARS-COV-2: A study of third wave', *Results in Physics*, 29: 104705.

Lin, Lixin, Boqiang Chen, Yanji Zhao, Weiming Wang, and Daihai He. 2021. 'Two Waves of COVID-19 in Brazilian Cities and Vaccination Impact', *Available at SSRN 3977464*.

Lin, Lixin, Yanji Zhao, Boqiang Chen, and Daihai He. 2021. 'Model analysis of vaccination effectiveness by state in the United States', *Available at SSRN*.

Lin, Qianying, Alice PY Chiu, Shi Zhao, and Daihai He. 2018. 'Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia', *Statistical methods in medical research*, 27: 1968-78.

Liu, Kai, Ying Chen, Ruzheng Lin, and Kunyuan Han. 2020. 'Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients', *Journal of Infection*, 80: e14-e18.

Liu, Yang, Jianying Liu, Kenneth S Plante, Jessica A Plante, Xuping Xie, Xianwen Zhang, Zhiqiang Ku, Zhiqiang An, Dionna Scharton, and Craig Schindewolf. 2021. 'The N501Y spike substitution enhances SARS-CoV-2 infection and transmission', *Nature*: 1-9.

Liu, Ying, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. 2020. 'The reproductive number of COVID-19 is higher compared to SARS coronavirus', *Journal of travel medicine*.

Liu, Ying, and Joacim Rocklöv. 2021. 'The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus', *Journal of Travel Medicine*.

Liu, Yongqi, and Zhendong Sun. 2010. "A new model for MDR-TB infection with undetected TB cases." In *2010 Chinese Control and Decision Conference*, 792-97. IEEE.

Luna-Muschi, Alessandra, Igor C Borges, Elizabeth de Faria, Antonio S Barboza, Fernando L Maia, Mariana D Leme, Ana Rubia Guedes, Maria Cassia Mendes-Correa, Esper G Kallas, and Aluísio C Segurado. 2022. 'Clinical features of COVID-19 by SARS-CoV-2 Gamma variant: A prospective cohort study of vaccinated and unvaccinated healthcare workers', *Journal of Infection*, 84: 248-88.

Madhi, Shabir A, Vicky Baillie, Clare L Cutland, Merryn Voysey, Anthonet L Koen, Lee Fairlie, Sherman D Padayachee, Keertan Dheda, Shaun L Barnabas, and Qasim

E Bhorat. 2021. 'Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B. 1.351 variant', *New England Journal of Medicine*, 384: 1885-98.
Marjot, Thomas, Gwilym J Webb, Alfred S Barritt, Andrew M Moon, Zania Stamataki, Vincent W Wong, and Eleanor Barnes. 2021. 'COVID-19 and liver disease: mechanistic and clinical perspectives', *Nature reviews Gastroenterology & hepatology*, 18: 348-64.
Martínez-García, Laura, Marco Antonio Espinel, Melanie Abreu, José María González-Alba, Desirèe Gijón, Amaranta McGee, Rafael Cantón, Juan Carlos Galán, and Jesús Aranaz. 2021. 'Emergence and spread of B. 1.1. 7 lineage in primary care and clinical impact in the morbi-mortality among hospitalized patients in Madrid, Spain', *Microorganisms*, 9: 1517.
Martins, Leila Droprinchinski, Iara da Silva, Wellington Vinicius Batista, Maria de Fátima Andrade, Edmilson Dias de Freitas, and Jorge Alberto Martins. 2020. 'How socio-economic and atmospheric variables impact COVID-19 and influenza outbreaks in tropical and subtropical regions of Brazil', *Environmental research*, 191: 110184.
Mathieu, Edouard, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. 2021. 'A global database of COVID-19 vaccinations', *Nature human behaviour*, 5: 947-53.
Migliori, GB, G De Iaco, G Besozzi, R Centis, and DM Cirillo. 2007. 'First tuberculosis cases in Italy resistant to all tested drugs', *Weekly releases (1997–2007)*, 12: 3194.
Mlcochova, Petra, Steven A Kemp, Mahesh Shanker Dhar, Guido Papa, Bo Meng, Isabella ATM Ferreira, Rawlings Datir, Dami A Collier, Anna Albecka, and Sujeet Singh. 2021. 'SARS-CoV-2 B. 1.617. 2 Delta variant replication and immune evasion', *Nature*, 599: 114-19.
Mor, Orna, Michal Mandelboim, Shay Fleishon, Efrat Bucris, Dana Bar-Ilan, Michal Linial, Ital Nemet, Limor Kliker, Yaniv Lustig, and Ella S Mendelson. 2021. 'The rise and fall of a local SARS-CoV-2 variant with the spike protein mutation L452R', *Vaccines*, 9: 937.
Musa, Salihu Sabiu, Xueying Wang, Shi Zhao, Shudong Li, Nafiu Hussaini, Weiming Wang, and Daihai He. 2021. 'Heterogeneous Severity of COVID-19 in African Countries: A Modeling Approach'.
———. 2022. 'The heterogeneous severity of COVID-19 in African countries: a modeling approach', *Bulletin of mathematical biology*, 84: 1-16.
Neely, Brooke, and Michelle Samura. 2011. 'Social geographies of race: Connecting race and space', *Ethnic and racial studies*, 34: 1933-52.
Niedzwiedz, Claire L, Catherine A O'Donnell, Bhautesh Dinesh Jani, Evangelia Demou, Frederick K Ho, Carlos Celis-Morales, Barbara I Nicholl, Frances S Mair, Paul Welsh, and Naveed Sattar. 2020. 'Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank', *BMC medicine*, 18: 1-14.
Nishiura, Hiroshi, Kimihito Ito, Asami Anzai, Tetsuro Kobayashi, Chayada Piantham, and Alfonso J Rodríguez-Morales. 2021. "Relative reproduction number of SARS-CoV-2 Omicron (B. 1.1. 529) compared with Delta variant in South Africa." In, 30. MDPI.
Nyberg, Tommy, Neil M Ferguson, Sophie G Nash, Harriet H Webster, Seth Flaxman, Nick Andrews, Wes Hinsley, Jamie Lopez Bernal, Meaghan Kall, and Samir Bhatt. 2022. 'Comparative analysis of the risks of hospitalisation and death

associated with SARS-CoV-2 omicron (B. 1.1. 529) and delta (B. 1.617. 2) variants in England: a cohort study', *The Lancet*, 399: 1303-12.
O'Neill, Jim. 2016. 'Tackling drug-resistant infections globally: final report and recommendations'.
Patel, JA, FBH Nielsen, AA Badiani, S Assi, VA Unadkat, B Patel, R Ravindrane, and H Wardle. 2020. 'Poverty, inequality and COVID-19: the forgotten vulnerable', *Public health*, 183: 110.
Peres, Igor Tona, LdSL Bastos, JG Mayrinck Gelli, JF Marchesi, LF Dantas, BBP Antunes, PM Maçaira, FA Baião, S Hamacher, and Fernando A Bozza. 2021. 'Sociodemographic factors associated with COVID-19 in-hospital mortality in Brazil', *Public health*, 192: 15-20.
Popkin, Barry M, Shufa Du, William D Green, Melinda A Beck, Taghred Algaith, Christopher H Herbst, Reem F Alsukait, Mohammed Alluhidan, Nahar Alazemi, and Meera Shekar. 2020. 'Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships', *Obesity Reviews*, 21: e13128.
Post, Frank A, Paul A Willcox, Barun Mathema, Lafras M Steyn, Karen Shean, Srinivas V Ramaswamy, Edward A Graviss, Elena Shashkina, Barry N Kreiswirth, and Gilla Kaplan. 2004. 'Genetic polymorphism in Mycobacterium tuberculosis isolates from patients with chronic multidrug-resistant tuberculosis', *The Journal of infectious diseases*, 190: 99-106.
Poudel, Sujan, Angela Ishak, Javier Perez-Fernandez, Efrain Garcia, Darwin A León-Figueroa, Luccio Romaní, D Katterine Bonilla-Aldana, and Alfonso J Rodriguez-Morales. 2022. 'Highly mutated SARS-CoV-2 Omicron variant sparks significant concern among global experts–What is known so far?', *Travel medicine and infectious disease*, 45: 102234.
Pulliam, Juliet RC, Cari van Schalkwyk, Nevashan Govender, Anne von Gottberg, Cheryl Cohen, Michelle J Groome, Jonathan Dushoff, Koleka Mlisana, and Harry Moultrie. 2021. 'Increased risk of SARS-CoV-2 reinfection associated with emergence of the Omicron variant in South Africa', *medRxiv*.
———. 2022. 'Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa', *Science*, 376: eabn4947.
Qian, Zhaozhi, Ahmed M Alaa, Mihaela van der Schaar, and Ari Ercole. 2020. 'Between-centre differences for COVID-19 ICU mortality from early data in England', *Intensive care medicine*, 46: 1779-80.
Raviglione, Mario. 2006. 'XDR-TB: entering the post-antibiotic era?', *The International Journal of Tuberculosis and Lung Disease*, 10: 1185-87.
Reardon, Sara. 2021. 'How the Delta variant achieves its ultrafast spread', *Nature*, 21.
Rohani, P, CJ Green, NB Mantilla-Beniers, and BT Grenfell. 2003. 'Ecological interference between fatal diseases', *Nature*, 422: 885-88.
Rohani, Pejman, David JD Earn, and Bryan T Grenfell. 1999. 'Opposite patterns of synchrony in sympatric disease metapopulations', *Science*, 286: 968-71.
'Science Brief: Omicron (B.1.1.529) Variant'. 2021. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-omicron-variant.html.
Seung, Kwonjune J, Irina E Gelmanova, Gennadiy G Peremitin, Vera T Golubchikova, Vera E Pavlova, Olga B Sirotkina, Galina V Yanova, and Aivar K Strelis. 2004. 'The effect of initial drug resistance on treatment response and acquired drug

resistance during standardized short-course chemotherapy for tuberculosis', *Clinical infectious diseases*, 39: 1321-28.
Sharma, Garima, Annabelle Santos Volgman, and Erin D Michos. 2020. 'Sex differences in mortality from COVID-19 pandemic: are men vulnerable and women protected?', *Case Reports*, 2: 1407-10.
Sheikh, Aziz, Steven Kerr, Mark Woolhouse, Jim McMenamin, Chris Robertson, Colin Richard Simpson, Tristan Millington, Ting Shi, Utkarsh Agrawal, and Safraj Shahul Hameed. 2022. 'Severity of omicron variant of concern and effectiveness of vaccine boosters against symptomatic disease in Scotland (EAVE II): a national cohort study with nested test-negative design', *The Lancet Infectious diseases*, 22: 959-66.
Sheikh, Aziz, Jim McMenamin, Bob Taylor, and Chris Robertson. 2021. 'SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness', *The Lancet*, 397: 2461-62.
Shi, Heshui, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, and Chuansheng Zheng. 2020. 'Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study', *The Lancet Infectious diseases*, 20: 425-34.
Shiehzadegan, Shayan, Nazanin Alaghemand, Michael Fox, and Vishwanath Venketaraman. 2021. 'Analysis of the delta variant B. 1.617. 2 COVID-19', *Clinics and Practice*, 11: 778-84.
Shrivastava, Shubham, Suhas T Mhaske, Meera S Modak, Rashmi G Virkar, Shamburaje S Pisal, Akhilesh Chandra Mishra, and Vidya A Arankalle. 2021. 'Emergence of Two Distinct Variants of SARS-CoV-2 and Explosive Second Wave of COVID-19: An Experience From A Tertiary Care Hospital, Pune, India'.
Shu, Y, and J McCauley. 2017a. "Eurosurveillance 2017, DOI: 10.2807/1560-7917." In.: ES.
Shu, Yuelong, and John McCauley. 2017b. 'GISAID: Global initiative on sharing all influenza data–from vision to reality', *Eurosurveillance*, 22: 30494.
Sinha, Siddharth, Benjamin Tam, and San Ming Wang. 2021. 'Altered interaction between RBD and ACE2 receptor contributes towards the increased transmissibility of SARS CoV-2 delta, kappa, beta, and gamma strains with RBD double mutations', *BioRxiv*: 2021.08. 30.458303.
Smith, David L, Katherine E Battle, Simon I Hay, Christopher M Barker, Thomas W Scott, and F Ellis McKenzie. 2012. 'Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens', *PLoS pathogens*, 8: e1002588.
Solari, L, C Acuna-Villaorduna, A Soto, J Agapito, F Perez, F Samalvides, J Zegarra, J Diaz, E Gotuzzo, and Patrick Van Der Stuyft. 2008. 'A clinical prediction rule for pulmonary tuberculosis in emergency departments', *The International Journal of Tuberculosis and Lung Disease*, 12: 619-24.
Song, Haitao, Guihong Fan, Yuan Liu, Xueying Wang, and Daihai He. 2021. 'The Second Wave of COVID-19 in South and Southeast Asia and the Effects of Vaccination', *Frontiers in Medicine*, 8.
Song, Haitao, Guihong Fan, Shi Zhao, Huaichen Li, Qihua Huang, and Daihai He. 2021. 'Forecast of the COVID-19 trend in India: A simple modelling approach', *Mathematical Biosciences and Engineering*, 18: 9775-86.
Starr, Tyler N, Allison J Greaney, Sarah K Hilton, Daniel Ellis, Katharine HD Crawford, Adam S Dingens, Mary Jane Navarro, John E Bowen, M Alejandra Tortorici, and Alexandra C Walls. 2020. 'Deep mutational scanning of SARS-CoV-2

receptor binding domain reveals constraints on folding and ACE2 binding', *Cell*, 182: 1295-310. e20.

Supply, Philip, Robin M Warren, Anne-Laure Bañuls, Sarah Lesjean, Gian D Van Der Spuy, Lee-Anne Lewis, Michel Tibayrenc, Paul D Van Helden, and Camille Locht. 2003. 'Linkage disequilibrium between minisatellite loci supports clonal evolution of Mycobacterium tuberculosis in a high tuberculosis incidence area', *Molecular microbiology*, 47: 529-38.

Tang, Xiujuan, Salihu S Musa, Shi Zhao, Shujiang Mei, and Daihai He. 2021. 'Using Proper Mean Generation Intervals in Modeling of COVID-19', *Frontiers in public health*, 9.

Tegally, Houriiyah, Eduan Wilkinson, Christian L Althaus, Marta Giovanetti, James Emmanuel San, Jennifer Giandhari, Sureshnee Pillay, Yeshnee Naidoo, Upasana Ramphal, and Nokukhanya Msomi. 2021. 'Rapid replacement of the Beta variant by the Delta variant in South Africa', *medRxiv*.

Tegally, Houriiyah, Eduan Wilkinson, Marta Giovanetti, Arash Iranzadeh, Vagner Fonseca, Jennifer Giandhari, Deelan Doolabh, Sureshnee Pillay, Emmanuel James San, and Nokukhanya Msomi. 2021. 'Detection of a SARS-CoV-2 variant of concern in South Africa', *Nature*, 592: 438-43.

Torjesen, Ingrid. 2021a. "Covid-19: Delta variant is now UK's most dominant strain and spreading through schools." In.: British Medical Journal Publishing Group.

———. 2021b. "Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear." In.: British Medical Journal Publishing Group.

Vassallo, Matteo, Sabrina Manni, Camille Klotz, Roxane Fabre, Paola Pini, Elea Blanchouin, Audrey Sindt, Laurene Lotte, Jean Marc Dubertrand, Stephane Liguori, Nathalie Berkane, Yannick Duval, Fabien Rolland, and Christian Pradier. 2021. 'Patients admitted for variant alpha COVID-19 have poorer outcomes than those infected with the old strain', *Journal of clinical medicine*, 10: 3550.

Vaughan, Adam. 2021. "Omicron emerges." In.: Elsevier.

Velayati, Ali Akbar, Mohammad Reza Masjedi, Parissa Farnia, Payam Tabarsi, Jalladein Ghanavi, Abol Hassan ZiaZarifi, and Sven Eric Hoffner. 2009. 'Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran', *Chest*, 136: 420-25.

Vetterling, William T, William H Press, Saul A Teukolsky, and Brian P Flannery. 1992. *Numerical recipes: example book C (The Art of Scientific Computing)* (Press Syndicate of the University of Cambridge).

Wang, Chengdi, Zhoufeng Wang, Guangyu Wang, Johnson Yiu-Nam Lau, Kang Zhang, and Weimin Li. 2021. 'COVID-19 in early 2021: current status and looking forward', *Signal Transduction and Targeted Therapy*, 6: 1-14.

Wang, Rong-Hua, Zhen Jin, Quan-Xing Liu, Johan van de Koppel, and David Alonso. 2012. 'A simple stochastic model with environmental transmission explains multi-year periodicity in outbreaks of avian flu', *PloS one*, 7: e28873.

Watkins, Kevin. 2018. 'Emerging infectious diseases: a review', *Current Emergency and Hospital Medicine Reports*, 6: 86-93.

Webster, Robert G. 1998. 'Influenza: an emerging disease', *Emerging Infectious Diseases*, 4: 436.

Wilhelm, Alexander, Marek Widera, Katharina Grikscheit, Tuna Toptan, Barbara Schenk, Christiane Pallas, Melinda Metzler, Niko Kohmer, Sebastian Hoehl,

and Fabian A Helfritz. 2021. 'Reduced neutralization of SARS-CoV-2 omicron variant by vaccine sera and monoclonal antibodies', *medRxiv*.

Wolter, Nicole, Waasila Jassat, Sibongile Walaza, Richard Welch, Harry Moultrie, Michelle Groome, Daniel Gyamfi Amoako, Josie Everatt, Jinal N Bhiman, and Cathrine Scheepers. 2022. 'Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study', *The Lancet*, 399: 437-46.

'World Health Organization. Global Tuberculosis Report'. 2021.

Young, Megan, Harry Crook, Janet Scott, and Paul Edison. 2022. 'Covid-19: virology, variants, and vaccines', *BMJ medicine*, 1.

Yu, Yangyang, Yuan Liu, Shi Zhao, and Daihai He. 2021. 'A Simple Model to Estimate the Transmissibility of SARS-COV-2 Beta, Delta and Omicron Variants in South Africa', *Delta and Omicron Variants in South Africa (December 20, 2021)*.

Yuki, Koichi, Miho Fujiogi, and Sophia Koutsogiannaki. 2020. 'COVID-19 pathophysiology: A review', *Clinical immunology*, 215: 108427.

Zhang, Li, Qianqian Li, Ziteng Liang, Tao Li, Shuo Liu, Qianqian Cui, Jianhui Nie, Qian Wu, Xiaowang Qu, and Weijin Huang. 2022. 'The significant immune escape of pseudotyped SARS-CoV-2 Variant Omicron', *Emerging microbes & infections*, 11: 1-5.

Zhang, Xiantao, Shijian Wu, Bolin Wu, Qirui Yang, Achun Chen, Yuzhuang Li, Yiwen Zhang, Ting Pan, Hui Zhang, and Xin He. 2021. 'SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance', *Signal Transduction and Targeted Therapy*, 6: 1-3.

Zhao, Shi, Lewi Stone, Daozhou Gao, and Daihai He. 2018. 'Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination', *PLoS Negl Trop Dis*, 12: e0006158.