frontiers | Research Topics

Modeling of COVID-19 and other infectious diseases: Mathematical, statistical and biophysical analysis of spread patterns

Edited by
Omar El Deeb, Gopi Battineni, Samer A. Kharroubi and Khalid Hattaf

Published in
Frontiers in Public Health
Frontiers in Applied Mathematics and Statistics
Frontiers in Medicine

ISSN 1664-8714
ISBN 978-2-83252-138-0
DOI 10.3389/978-2-83252-138-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Modeling of COVID-19 and other infectious diseases: Mathematical, statistical and biophysical analysis of spread patterns

Topic editors

Omar El Deeb — Lebanese American University, Lebanon
Gopi Battineni — University of Camerino, Italy
Samer A. Kharroubi — American University of Beirut, Lebanon
Khalid Hattaf — Centre Régional des Métiers de l'Education et de la Formation (CRMEF), Morocco

Citation

El Deeb, O., Battineni, G., Kharroubi, S. A., Hattaf, K., eds. (2023). *Modeling of COVID-19 and other infectious diseases: Mathematical, statistical and biophysical analysis of spread patterns*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83252-138-0

Table of contents

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Editorial
PUBLISHED 27 March 2023
DOI 10.3389/fams.2023.1178479

OPEN ACCESS

EDITED AND REVIEWED BY
Axel Hutt,
Inria Nancy - Grand-Est Research Centre, France

*CORRESPONDENCE
Omar El Deeb
✉ omar.eldeeb@lau.edu.lb

SPECIALTY SECTION
This article was submitted to Dynamical Systems, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 02 March 2023
ACCEPTED 13 March 2023
PUBLISHED 27 March 2023

CITATION
El Deeb O, Hattaf K and Kharroubi SA (2023) Editorial: Modeling of COVID-19 and other infectious diseases: Mathematical, statistical and biophysical analysis of spread patterns. *Front. Appl. Math. Stat.* 9:1178479. doi: 10.3389/fams.2023.1178479

Editorial: Modeling of COVID-19 and other infectious diseases: Mathematical, statistical and biophysical analysis of spread patterns

Omar El Deeb[1]*, Khalid Hattaf[2] and Samer A. Kharroubi[3]

[1]Department of Natural Sciences, Lebanese American University, Beirut, Lebanon, [2]Centre Régional des Métiers de l'Education et de la Formation, Casablanca, Morocco, [3]Department of Nutrition and Food Sciences, American University of Beirut, Beirut, Lebanon

KEYWORDS

infectious diseases, mathematical modeling, COVID-19, coronavirus, spread patterns

Editorial on the Research Topic
Modeling of COVID-19 and other infectious diseases: Mathematical, statistical and biophysical analysis of spread patterns

Viral pandemics and infectious diseases have always constituted an imminent threat to humanity. Throughout history, humanity has struggled with viral outbreaks and pandemics. Some of the most significant examples include the Spanish Flu in 1918-1919, which infected an estimated 500 million people worldwide and caused 50 million deaths. Efforts to control the spread of infectious diseases have led to the development of vaccines, antibiotics, and public health measures, but these diseases continue to pose a significant threat to global health and wellbeing. Scientific advancements in science, public health, medications, and vaccines are pivotal in the containment and even the elimination of many of those threats. However, these efforts can be challenged by factors such as limited resources, anti-vaccine sentiment, and difficulty in controlling the spread in communities with limited access to healthcare.

The coronavirus disease 2019 (COVID-19) continues as the main cause of hospitalization and death and as a main public health risk since the first case was registered in December 2019 in China. It was declared a global pandemic by the World Health Organization (WHO) on March 11, 2020. The recent emergence of variants creates a major cause of concern since they can lead to an epidemic rebound especially with the possibility of the emergence of vaccine resisting, deadlier or more transmissible future variants. The COVID-19 infection reached at least 680 million people globally and caused over 6.80 million deaths [1]. Mathematical models and simulations were used to describe the spread of infectious diseases by using mathematical equations to represent the spread of an infectious agent in a population. The models consider factors such as the infectiousness of the disease, the number of susceptible and infected individuals, and the rate at which people recover or die. The output of these models can provide important insights into the dynamics of disease spread, and can be used to predict the potential spread of an outbreak, inform public health policy, and evaluate the efficacy of interventions. Simulations of these models can also be used to visualize the spread of the disease over time, and to test the impact

of different control measures. Such studies include: Non-linear dynamics, non-equilibrium processes, and self-organization modeling of infectious diseases, statistical, spatiotemporal and big data analytics of COVID-19, applications from the social sciences, public health, economics, engineering in relation to COVID-19 pandemics, and modeling, simulations and forecasting of spread patterns, vaccine efficiency, treatment, behavioral aspects and public policies [2–10].

Another import field of study is that of vaccines, their efficacies and their implementation schemes. Studies on the efficacy of vaccines in the fight against infectious diseases like COVID-19 involve evaluating the effectiveness of the vaccine in reducing the incidence of the disease, severity of symptoms, and the transmission of the virus. These studies typically use randomized controlled trials (RCTs), where individuals are randomly assigned to receive either the vaccine or a placebo, and are followed over time to determine the incidence of the disease in each group. The efficacy of the vaccine is then determined by comparing the incidence of the disease in the vaccinated group to that of the unvaccinated group. It is important to note that vaccine efficacy can vary based on various factors such as the population being studied, the duration of follow-up, and the level of circulating virus in the population [11, 12]. Ongoing monitoring and analysis of vaccine efficacy is crucial to ensure the continued safety and effectiveness of vaccines as the pandemic continues to evolve.

This Research Topic includes articles that study spread patterns of infectious diseases using several mathematical, statistical, computational, and biophysical methods covering compartmental models, agent-based models (ABM), spatiotemporal analysis, data-driven analysis, artificial intelligence, and analytic methods. Sherwani et al. analyzed the seroprevalence of anti-S1-RBD antibodies in pre-pandemic and pandemic subjects from Saudi Arabia and found out that antibody levels increased in samples collected during the pandemic, even though these subjects were not clinically COVID-19 positive. Zhao and Liu reviewed the distribution characteristics of COVID-19 in America based on space-time scan statistic. Their empirical results reveal the relative risk of the first-level and the second-level clustering area of the epidemic across several states. The influence of co-morbidities during SARS-CoV-2 infection in the Indian Population was analyzed by Matysek et al.. They found out that the highest correlation coefficient were age, random serum glucose, serum urea, gender and serum cholesterol, whereas the highest inverse correlation coefficient was assessed for alanine transaminase, red blood cells count and serum creatinine. Wu et al. worked on modeling the small scale outbreak of COVID-19, especially in China and proposed a new version of cellular automata with a time matrix, to simulate outbreaks.

The policy choices for Shanghai in response to challenges of Omicron were inspected by Qian et al. who showed that effective policies for Omicron include high level of testing capacity to identify and quarantine the infected cases, especially the asymptomatic cases in addition to immediate home-isolation and fast transfer to centralized quarantine location. Harris showed how controls on access alone through concentric regulatory zones, without restrictions on movement, were inadequate to halt an advancing COVID-19 outbreak in the highly populous area of South Brooklyn, New York. Kim et al. introduced a mathematical modeling approach to study the economic impact of COVID-19 interventions and their study asserts the importance of the rapidity of vaccine rollout to the cost effective control of the number of infections and deaths. Yu et al. employed machine learning models based on blood inflammatory cytokines to identify the hospitalized mortality of patients with COVID-19. Their study constructs predictive models to assess patients who may have poor prognoses early and accurately. Liu et al. modeled the effects of vaccination, nucleic acid testing, and face mask wearing interventions against COVID-19 in large sports events using an SEIR model and found out that the combined use of these measures could largely decrease the number of infections. Future possible studies in mathematical modeling of infectious diseases include the incorporation of spatial dynamics, social dynamics, genomic data, more accurate models of transmission, impact of vaccination and developing real-time models for outbreak responses.

Author contributions

OE prepared the draft of the manuscript. SK and KH revised and improved the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. *COVID-19 Coronavirus Pandemic, Worldometer*. Available online at: https://www.worldometers.info/coronavirus/ (accessed on March 2, 2023).

2. Chintalapudi N, Battineni G, Amenta F. Sentimental analysis of COVID-19 tweets using deep learning models. *Infect Dis Rep*. (2021) 13:329–39. doi: 10.3390/idr13020032

3. Kharroubi S, Saleh F. Are lockdown measures effective against COVID-19?. *Front Public Health*. (2020) 8:549692. doi: 10.3389/fpubh.2020.549692

4. Kharroubi S, Diab-El-Harake M. Sex-differences in COVID-19 diagnosis, risk factors and disease comorbidities: a large US-based cohort study. *Front Public Health*. (2020) 10:1029190. doi: 10.3389/fpubh.2022.1029190

5. El Deeb O, Jalloul M. Forecasting the outbreak of COVID-19 in Lebanon. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.09.03.20187880

6. Kharroubi S. Modeling the spread of COVID-19 in Lebanon: a Bayesian perspective. *Front Appl Math Stat.* (2020) 6:40. doi: 10.3389/fams.2020.00040

7. Hattaf K, Mohsen A. Dynamics of a generalized fractional epidemic model of COVID-19 with carrier effect. *Adv Syst Sci Appl.* (2022) 22:126–38. doi: 10.25728/assa.2022.22.3.1172

8. El Deeb O, Jalloul M. The dynamics of COVID-19 spread: evidence from Lebanon. *Math Biosci Eng.* (2020) 17:5618–32. doi: 10.3934/mbe.2020302

9. El Deeb O. Spatial autocorrelation and the dynamics of the mean center of COVID-19 infections in Lebanon. *Front Appl Math Stat.* (2021) 6:620064. doi: 10.3389/fams.2020.620064

10. Hattaf K, El Karimi MI, Mohsen AA, Hajhouji Z, El Younoussi M, Yousfi N. Mathematical modeling and analysis of the dynamics of RNA viruses in presence of immunity and treatment: a case study of SARS-CoV-2. *Vaccines*. (2023) 11:201. doi: 10.3390/vaccines11020201

11. Semlali M, Hattaf K, El Kettani M. Modeling and analysis of the dynamics of COVID-19 transmission in presence of immigration and vaccination. *Commun Math Biol Neurosci.* (2022) 36. doi: 10.28919/cmbn/7270

12. El Deeb O, Jalloul M. Efficacy versus abundancy: Comparing vaccination schemes. *PLoS ONE*. (2022) 17:e0267840. doi: 10.1371/journal.pone.0267840

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fpubh.2022.874741

Seroprevalence of Anti-S1-RBD Antibodies in Pre-pandemic and Pandemic Subjects From Hail Region, KSA

Subuhi Sherwani[1], Mohd Wajid Ali Khan[2], Arshi Mallik[3], Mahvish Khan[1], Mohd Saleem[4], Mohamed Raafat[5], Ayed A. Shati[6] and Noor Alam[7]*

[1] Department of Biology, College of Sciences, University of Ha'il, Ha'il, Saudi Arabia, [2] Department of Chemistry, College of Sciences, University of Ha'il, Ha'il, Saudi Arabia, [3] Department of Clinical Biochemistry, College of Medicine, King Khalid University, Abha, Saudi Arabia, [4] Department of Pathology, Sub-division of Medical Microbiology, College of Medicine, University of Ha'il, Ha'il, Saudi Arabia, [5] Department of Physiotherapy, College of Applied Medical Sciences, University of Ha'il, Ha'il, Saudi Arabia, [6] Department of Child Health, College of Medicine, King Khalid University, Abha, Saudi Arabia, [7] Department of Basic Sciences, Deanship of Preparatory Year, University of Ha'il, Ha'il, Saudi Arabia

OPEN ACCESS

Edited by:
Omar El Deeb, Lebanese American University, Lebanon

Reviewed by:
Sardar Sindhu, Dasman Diabetes Institute, Kuwait
Waleed Mahallawi, Taibah University, Saudi Arabia

****Correspondence:***
Subuhi Sherwani
susherwani@gmail.com;
s.sherwani@uoh.edu.sa

Specialty section:
This article was submitted to Infectious Diseases – Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

Received: *12 February 2022*
Accepted: *29 April 2022*
Published: *09 June 2022*

Citation:
Sherwani S, Khan MWA, Mallik A, Khan M, Saleem M, Raafat M, Shati AA and Alam N (2022) Seroprevalence of Anti-S1-RBD Antibodies in Pre-pandemic and Pandemic Subjects From Hail Region, KSA. Front. Public Health 10:874741. doi: 10.3389/fpubh.2022.874741

Background: Two years into the pandemic, yet the threat of new SARS-CoV-2 variants continues to loom large. Sustained efforts are required to fully understand the infection in asymptomatic individuals and those with complications. Identification, containment, care, and preventative strategies rely on understanding the varied humoral immune responses.

Methods: An in-house ELISA was developed and standardized to screen for serum IgG antibodies against the SARS-CoV-2 S1-RBD protein as an antigen. This study aims to investigate the seroprevalence of serum antibodies against S1-RBD antigen in pre-pandemic ($n = 120$) and during the early pandemic period ($n = 120$) in subjects from the Hail region, KSA and to correlate it with clinical and demographic factors.

Results: Samples collected from both male ($n = 60$) and female ($n = 60$) subjects during the pandemic in the age groups of 20–40 (0.31 ± 0.029 and 0.29 ± 0.024, respectively) and 41–60 years (0.35 ± 0.026 and 0.30 ± 0.025, respectively) showed significantly higher levels of serum antibodies against S-RBD antigen than the age-matched pre-pandemic samples [male ($n = 60$) and female ($n = 60$)]. Pandemic subjects exhibited significantly ($p < 0.01$) higher inhibition (80–88%) than age-matched pre-pandemic subjects (32–39%). Antibodies against S1-RBD antigen were detected in approximately 10% of the total pre-pandemic population (males and females). However, subjects > 60 years did not show antibodies.

Conclusion: Antibody levels increased in samples collected during the pandemic, even though these subjects were not clinically COVID-19 positive. A small number of pre-pandemic subjects showed serum antibodies, suggesting prior exposure to other coronaviruses in the region. With dwindling neutralizing antibody levels and reduced vaccine efficacy against newer variants, it remains crucial to develop better assays for surveillance, management, and future research.

Keywords: SARS-CoV-2, S1-RBD, COVID-19, ELISA, antibodies, seroprevalence

INTRODUCTION

The end of 2019 witnessed the emergence, rise, and rapid spread of a highly contagious novel coronavirus known as Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), the causative pathogen of the highly contagious Corona Virus Disease 2019 or COVID-19, to almost every corner of the world (1). COVID-19 continues to be a threat, with the possible emergence of new variants with the ability to spread more rapidly and target children. Factors such as gender, age, and comorbid conditions contribute to disease severity and complications (2). The repercussions of this health crisis will be felt for many years to come.

The phylogenetically similar coronaviruses—SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV)—are beta coronaviruses, emergent from animal reservoirs, capable of rapid transmission and serious infectious outcomes in humans (3). The primary mode of COVID-19 virus transmission responsible for the pandemic is human-to-human, *via* aerosols and droplets, from infected individuals through talking, coughing, or sneezing (4). COVID-19 has a probable asymptomatic incubation period between 2 and 14 days, with newer variants displaying even lower incubation periods (5).

Those infected with the virus can broadly be classified according to their level of infection and the severity of the disease. Some infected individuals remain asymptomatic, whereas others experience mild, transient symptoms. A substantial number of infected individuals with advanced age and medical comorbidities such as diabetes, hypertension, or immunocompromised states are hospitalized due to complications (2). Depending on their immunological condition, individuals infected with COVID-19 experience mild, moderate, or severe symptoms. Common symptoms include dry cough, fever, fatigue, loss of taste or smell, and diarrhea. Severe symptoms include dyspnea and chest pains (6). Severe pathological manifestations of the disease in the infected population with comorbidities include acute respiratory distress syndrome (ARDS) and respiratory failure (7). Thus, age, pre-comorbidities, an increased viral load, low SARS-CoV-2 antibody response, or an excessive systemic inflammatory response known as a cytokine storm are contributory risk factors to adverse patient outcomes (2, 8).

SARS-CoV-2 is an enveloped virus with a linear, unsegmented positive-sense RNA genome. The nucleocapsid of the virion consists of N-phosphoprotein (NP) and RNA, surrounded by lipid bilayers (9). The (S1) spike glycoprotein peplomer mediates viral attachment, followed by membrane fusion. This glycoprotein is immunogenic and hence the target of IgM and IgG humoral circulating antibodies (Abs) (10, 11). SARS-CoV-2 infected patients produce antibodies 4–8 days post-onset (12). Recent studies suggest the role of serum antibodies, memory B cells, and cross-reactive T-cells in conferring immune protection against the virus (13). However, more region-specific studies are needed to ascertain host vulnerability, the nature of immune responses in individuals, and the extent and duration of protection.

In spite of a plethora of primary studies conducted during the pandemic about serum immunoglobulin G (IgG) generated against S1-RBD in COVID-19, details about the prevalence, durability, response, and degree of the conference of immunity from previous infections remain understudied. In a study investigating anti-S1RBD IgG in COVID-19 hospital patients during the early pandemic with a commercial ELISA, it was found that median OD values were to be higher in patients with the severe disease than those with the mild, moderate and critical disease. However, the same pattern was not observed with respect to anti-NP IgG (14). Also, the same study found that anti-S1RBD IgG levels remained stably above positive threshold values in patients with severe infections but were lower in patients with mild or moderate infections (14). In a separate cross-sectional study of unvaccinated U.S. adults, anti-S1RBD antibodies were detected in 99% of individuals who reported a positive COVID-19 test, 55% of individuals who believed they had COVID-19 but were not tested, and 11% of individuals who believed that they never had a COVID-19 infection. Also, in individuals with a positive COVID-19 result, anti-RBD levels were detectable for up to 20 months (15).

The current study aims to investigate the seroprevalence of serum antibodies against SARS-CoV-2 in the general population before the emergence of the virus and during the early phase of the COVID-19 pandemic, using an ELISA designed to screen IgG antibodies directed against viral S1-glycoprotein receptor-binding domain (S1-RBD) protein antigen. It is a crucial first step in determining the humoral immune response of asymptomatic and subclinical infections in individuals and their associated implications. Such information is vital for both researchers and policymakers in developing successful surveillance and management strategies for vaccine delivery, care of unvaccinated and vaccinated infected individuals, and effective age-related outcomes.

MATERIALS AND METHODS

Study Subjects–Sera Collection Pre-pandemic and During COVID-19 Pandemic

A total of 240 sera samples were collected from healthy individuals before and during the early months of the COVID-19 pandemic from the Hail region, Saudi Arabia, with their prior consent. The research study was carried out per the Declaration of Helsinki (1964). Of the samples collected, 120 sera samples were from individuals who were not diagnosed with any disease. Furthermore, individuals with immune disorders, immunodeficiencies, allergies, cancer, pregnant women, and those with serious lung, heart, kidney, or liver disease were excluded from the study. An equal number of sera samples were collected from individuals during the COVID-19 pandemic with no history of COVID-19 infection and no administration of any COVID-19 vaccine. Samples were collected under the Research Ethics Committee; the University of Hail approved the study protocol H-2021-122. Subjects with any previous

TABLE 1 | Group characteristics of study population.

Groups (Age in Years) *n* = subjects	Fasting blood glucose (mg/dl)	HbA1c (%)	BMR (cal/sq.m/hr)	Number of smokers (duration; years ± SD)	Number of subjects with			
					Fever	Fatigue	Cough	Myalgia
M-Pre-P (20-40) $n = 20$	85.2 ± 5.4	5.5 ± 0.3	37.6 ± 2.3	8 (9.4 ± 5.1)	1	2	—	—
M-Pan (20-40) $n = 20$	83.0 ± 7.1	5.4 ± 0.3	38.7 ± 2.1	7 (8.2 ± 4.2)	4	3	3	2
F-Pre-P (20-40) $n = 20$	84.3 ± 8.3	5.5 ± 0.3	33.6 ± 3.3	−3 (4.1 ± 4.2)	1	2	1	—
F-Pan (20-40) $n = 20$	84.6 ± 5.3	5.6 ± 0.4	35.0 ± 2.9	−3 (6.3 ± 2.1)	1	2	2	—
M-Pre-P (41-60) $n = 20$	89.1 ± 7.1	5.5 ± 0.4	35.7 ± 2.6	−9 (12.6 ± 6.3)	1	2	1	—
M-Pan (41-60) $n = 20$	89.6 ± 8.5	5.6 ± 0.4	41.8 ± 3.2	−7 (15.4 ± 6.4)	4	3	4	2
F-Pre-P (41-60) $n = 20$	88.9 ± 7.3	5.4 ± 0.3	31.2 ± 3.1	−3 (11.7 ± 3.4)	—	2	1	—
F-Pan (41-60) $n = 20$	90.1 ± 7.9	5.6 ± 0.3	33.0 ± 2.7	−4 (7.4 ± 2.6)	2	2	2	—
M-Pre-P (>60) $n = 20$	97.0 ± 11.3	5.8 ± 0.5	33.9 ± 2.4	7 (22.3 ± 5.8)	—	2	2	—
M-Pan (>60) $n = 20$	98.3 ± 9.5	5.7 ± 0.3	39.9 ± 2.8	7 (26.4 ± 4.8)	6	6	6	3
F-Pre-P (>60) $n = 20$	96.8 ± 8.8	5.7 ± 0.4	31.0 ± 2.6	4 (13.1 ± 3.9)	1	2	1	—
F-Pan (>60) $n = 20$	97.2 ± 8.8	5.8 ± 0.5	36.7 ± 3.4	4 (16.1 ± 3.9)	3	3	2	1

M-Pre-P, M-Pan, F-Pre-P and F-Pan represents Male subjects' pre-pandemic, Male subjects during pandemic, Female subjects' pre-pandemic and female subjects during pandemic, respectively. Normal ranges for FBG are 70-99 mg/dl. Normal ranges for BMR adult men and women are 35–38 and 32–35 cal/sq.m/hr, respectively.

history of disease or associated complications, including COVID-19, were excluded from this study. Serum samples were kept in temperature-controlled environments (-20 to−80°C). Demographic data collected for the sera samples included age, gender, fasting blood glucose (FBG), basal metabolic rate (BMR), and smoking history. FBG, HbA1c, and BMR were assessed using well-known methods prescribed regularly in the clinics. Participants were asked to report symptoms such as fever, fatigue, cough, or myalgia in the 14 days prior to sample collection, as these may assist in the interpretation of antibody results.

The participants in this study were divided into groups, each comprising 20 volunteers ($n = 20$), assorted in both gender and age. The demographic data of the groups are represented in **Table 1**. The distribution is as follows: serum samples collected from men pre-pandemic and aged 20–40 years old (M-Pre-P, 20–40 years); serum samples collected from men during the pandemic, aged 20–40 years old (M-Pan, 20–40 years); serum samples collected from women pre-pandemic aged 20–40 years old (F-Pre-P, 20–40 years); serum samples collected from women during the pandemic aged 20–40 years old (F-Pan, 20–40 years); serum samples collected from men pre-pandemic who were 41–60 years old (M-Pre, 41–60 years); serum samples collected from men during the pandemic who were 41–60 years old (M-Pan (41–60 years); serum samples collected from women pre-pandemic who were 41–60 years old (F-Pre-P, 41–60 years); serum samples collected from women during the pandemic who were aged 41–60 years (F-Pan, 41–60 years); serum samples collected from men pre-pandemic who were more than 60 years old (M-Pre-P, >61 years); serum samples collected from men during the pandemic who were more than 60 years old (M-Pan, >61 years); serum samples collected from women pre-pandemic who were more than 60 years old (F-Pre-P > 61 years) and serum samples collected from women during the pandemic who were more than 60 years old (F-Pan, > 61 years).

Estimation of Inflammatory Cytokines IL-6 and TNF-α

Cytokines IL-6 and TNF-α levels were analyzed in serum samples from all the cohorts using commercially available quantitative sandwich immunoassay kits (R&D System, Minneapolis, MN, USA). The sensitivities of the ELISA kits were <0.5 pg/mL. Samples were assayed in triplicate.

Optimization of Antigen Concentration for Indirect Binding ELISA

The recombinant S1-RBD-protein antigen (MyBioSource, USA) coating concentration was optimized as described previously with slight modifications (16, 17); varying concentrations 0.1, 1, 2, 4, 8, and 10 μg/ml) of S1-RBD protein in coating buffer (0.05 M carbonate-bicarbonate buffer, pH 9.6) was coated on the ELISA plate. The plate was incubated for 2 h, and unbound antigens were removed by washing using phosphate buffer saline (PBS).

Unbound spaces were blocked with 2.5% BSA and incubated for 1 h at 37°C. The ELISA plate was washed three to five times with PBS-Tween20 (PBS-T). Test samples [anti-R-C19-S1-RBD IgG (MyBioSource, California, USA)] and serum samples from three COVID-19 convalescent patients were diluted (1:100) in dilution buffer [phosphate buffer saline (PBS)] and were added to each well (100 µl/well). The plates were incubated at 37°C for 2 h, and after incubation, the ELISA plate was washed three to five times with PBS-T. The secondary antibody against anti-R-C19-S1-RBD IgG (MyBioSource, California, USA) was diluted as per the manufacturer's instruction to 1:50,000 and added to the ELISA plate (100 µl/well). For serum samples, secondary anti-human IgG diluted 1:2,000 in dilution buffer was added to each well (100 µl/well). Secondary antibodies were conjugated to horseradish peroxidase (GE Healthcare). After incubation at 37°C for 2 h, the plate was washed three to five times with PBS-T. 3,3',5,5'-Tetramethylbenzidine (TMB). Stabilized substrate (Promega) was added to each well (100 µl) for 20 min. One molar of H_2SO_4 (100 µl) was used to stop the enzyme reaction after 20 min incubation at room temperature. The results were expressed as optical density (OD) (OD = mean of triplicate wells minus mean of the blank wells). The OD of the reaction product was read at 450 nm on an ELISA plate reader.

Optimization of Serum Dilution for Indirect Binding ELISA

Serum dilution from three COVID-19 convalescent patients was optimized for the ELISA assay. The ELISA plate was coated with 2 µg/ml of S1-RBD antigen in coating buffer and incubated for 2 h. Unbound antigens were removed by washing with PBS. Unbound spaces in the ELISA plate were blocked with 2.5% BSA and incubated for 1 h at 37°C. The plate was washed three to five times with PBS-T. Serum samples ($n = 3$) diluted serially (1:50, 1:100, 1:200, 1:400, 1:800, and 1:1,600) in PBS were added to each well (100 µl/well). The plates were incubated at 37°C for 2 h and, after incubation, washed three to five times with PBS-T. Secondary anti-human IgG conjugated to horseradish peroxidase was diluted at 1:2,000 in a dilution buffer and added to each well (100 µl/well). The remaining steps were the same as those given above.

Determination of the Threshold Value by Indirect ELISA

The OD450 nm value of 20 sera samples of normal individuals from our laboratory, obtained before the COVID-19 outbreak, was detected with the optimum concentration of protein and antibody by indirect ELISA. The results were statistically analyzed to determine the cut-off value. The mean (X) and standard deviation (SD) of the 20 samples were calculated. The cut-off value was X + 3 SD, which was positive when the IgG OD450 nm value of the samples to be tested was ≥X + 3 SD and negative when the IgG OD450 nm value was <X + 3 SD.

ELISA Plate Description

To test the sera samples, the following design of a 96-well ELISA plate was used: twenty serum samples from each group were tested in triplicate on each plate along with antibody specific

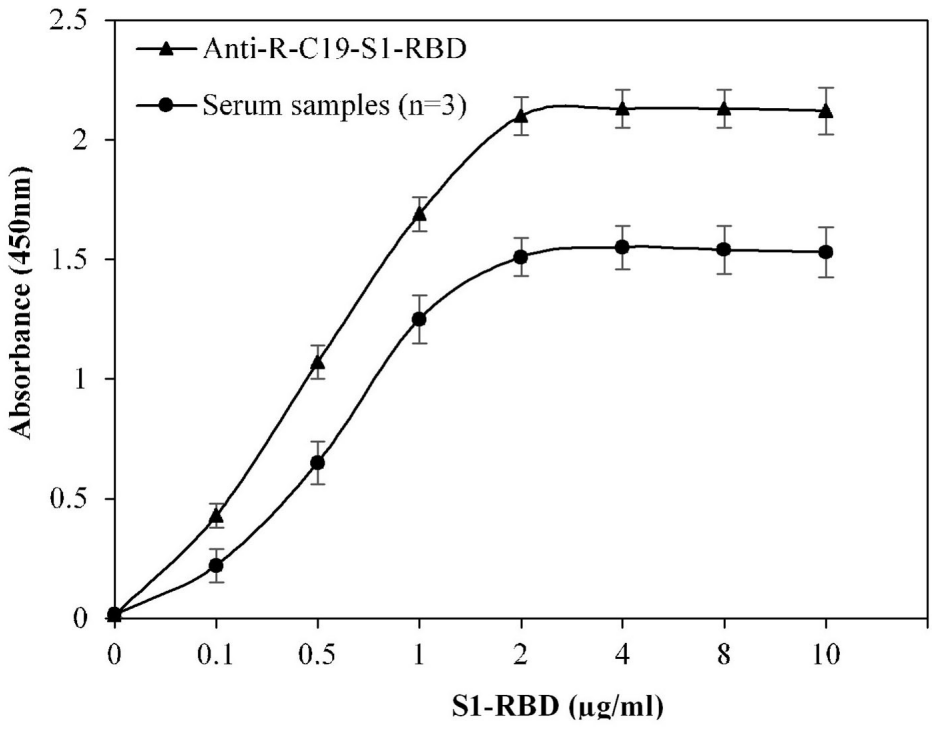


FIGURE 1 | Optimization of antigen (S1-RBD) concentration for ELISA. Serum samples from COVID-19 convalescent individuals ($n = 3$) were used in the assay. Anti-R-C19-S1-RBD IgG was used a positive control. Each sample was run in triplicate under similar conditions.

(anti-R-C19-S1-RBD IgG) for S1-RBD antigen, which served as a positive control (three wells). Six wells included two different pre-pandemic serum samples that showed <0.1 OD in indirect binding ELISA and were considered negative controls. Blanks were also included in three wells. The results were expressed as OD (OD = mean of triplicate wells minus the mean of the blank wells).

Isolation of Serum IgG

Affinity chromatography was applied to isolate IgG from serum samples using a Protein A-Agarose column (Sigma-Aldrich, USA). The Protein A-Agarose column was washed 2–3 times using PBS buffer (pH 7.4) prior to the addition of the sample. A volume of 0.5 ml of serum sample was diluted with an equal volume of PBS (pH 7.4) and run through the column. Samples were re-eluted 2–3 times for efficient binding of IgG. Unbound IgG was removed by extensive washing with the same washing buffer. Serum IgG, which was bound to the column, was eluted with elution buffer (acetic acid (0.58%) in sodium chloride (0.85%) and neutralized with 1 M Tris-HCl (pH 8.5). About 2–3 ml of fractions were collected in serum tubes, and each tube was read at 251 and 278 nm. The concentration of IgG was estimated as 1 mg/ml at 1.4 OD.

Specificity and Reproducibility of Indirect ELISA

The specificity of the indirect binding ELISA was assessed by evaluating the presence of SARS-CoV-2 specific IgG against S1-RBD in the sera samples of COVID-19 convalescent patients ($n = 3$), hepatitis C virus (HCV) ($n = 3$), tuberculosis (TB) ($n = 3$), and rheumatoid arthritis (RA) ($n = 3$) patients. The ELISA plates were coated with 2 g/ml of S1RBD antigen, and the specificity of the method was evaluated with the established indirect ELISA method (15, 16). Anti-R-C19-S1-RBD IgG served as a positive control, and IgG from pre-pandemic subjects who showed <0.2 OD served as a negative control.

Plate-to-plate variation was monitored by comparing the control panel results between the different wells of the same plate; the same sera samples were run on different plates on the same day as well as on different days.

Indirect Binding ELISA

The binding activity of serum antibodies to S1-RBD antigen was detected by indirect binding ELISA as described above with slight modifications (16–19).

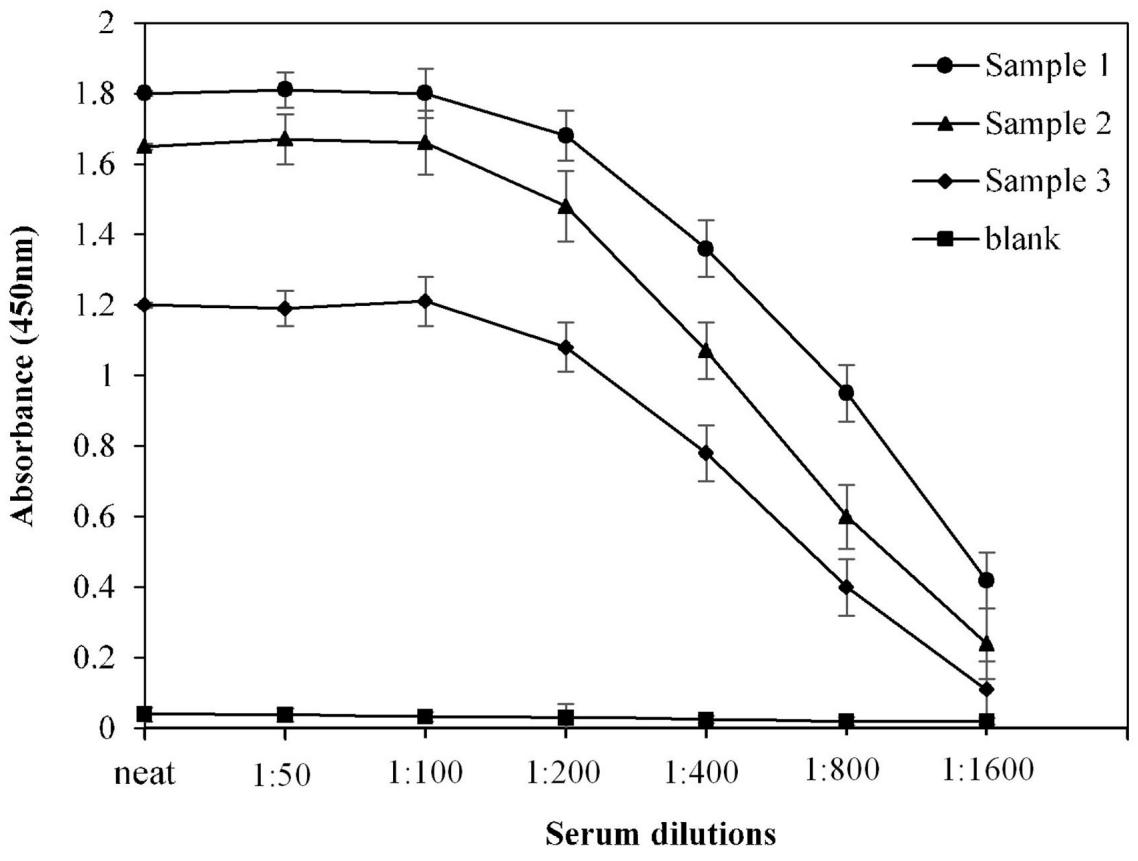


FIGURE 2 | Optimization of serum dilution for ELISA. Serum samples from COVID-19 convalescent individuals ($n = 3$) were used in the assay. Each sample was run in triplicate under similar conditions.

Inhibition ELISA

The specificities of S1-RBD antigen and serum IgG were estimated by competition ELISA (18–20). Increasing concentrations of S1-RBD antigen (0–10 μg/ml) were allowed to interact with a constant amount of serum autoantibodies from individuals of different groups for 2 h at room temperature and overnight at 4°C. After incubation, the immune complex formed was incubated in the microplate wells (instead of the serum taken in indirect binding ELISA), and the bound antibody levels were detected as in indirect binding ELISA. The percent inhibition was calculated using the formula:

$$\text{Percent inhibition} = [1 - (A_{\text{inhibited}}/A_{\text{uninhibited}})] \times 100,$$

where $A_{\text{inhibited}}$ is the absorbance at 10 μg/mL of inhibitor concentration and $A_{\text{uninhibited}}$ is the absorbance at zero inhibitor concentration.

Statistical Analysis

Statistical analyses were carried out using OriginPro v6.1. One-way or two-way analysis of variance (ANOVA) was applied to test for statistical significance. Only *p*-values of 0.05 or lower were considered statistically significant [$p > 0.05$ (ns, not significant), $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***)].

RESULTS

Optimization of Antigen for ELISA

Multiple steps were included in the ELISA method. To develop an efficient ELISA assay, it is essential to standardize all steps. The concentration of S1-RBD protein antigen used to coat the microplate was optimized, which effectively covered the bottom of the microplate wells. **Figure 1** shows that at a concentration of 2 μg/ml, recombinant S1-RBD protein antigen exhibited maximum absorbance, which was recorded for both anti-R-C19-S1-RBD IgG and serum samples. Hence, 2 μg/ml of recombinant S1-RBD protein antigen was used for all further ELISA assays.

Optimization of Serum Dilution for ELISA

For optimization of serum dilution used in indirect binding ELISA, serum samples from three COVID-19 convalescent patients were diluted with varying ratios. Maximum absorbance

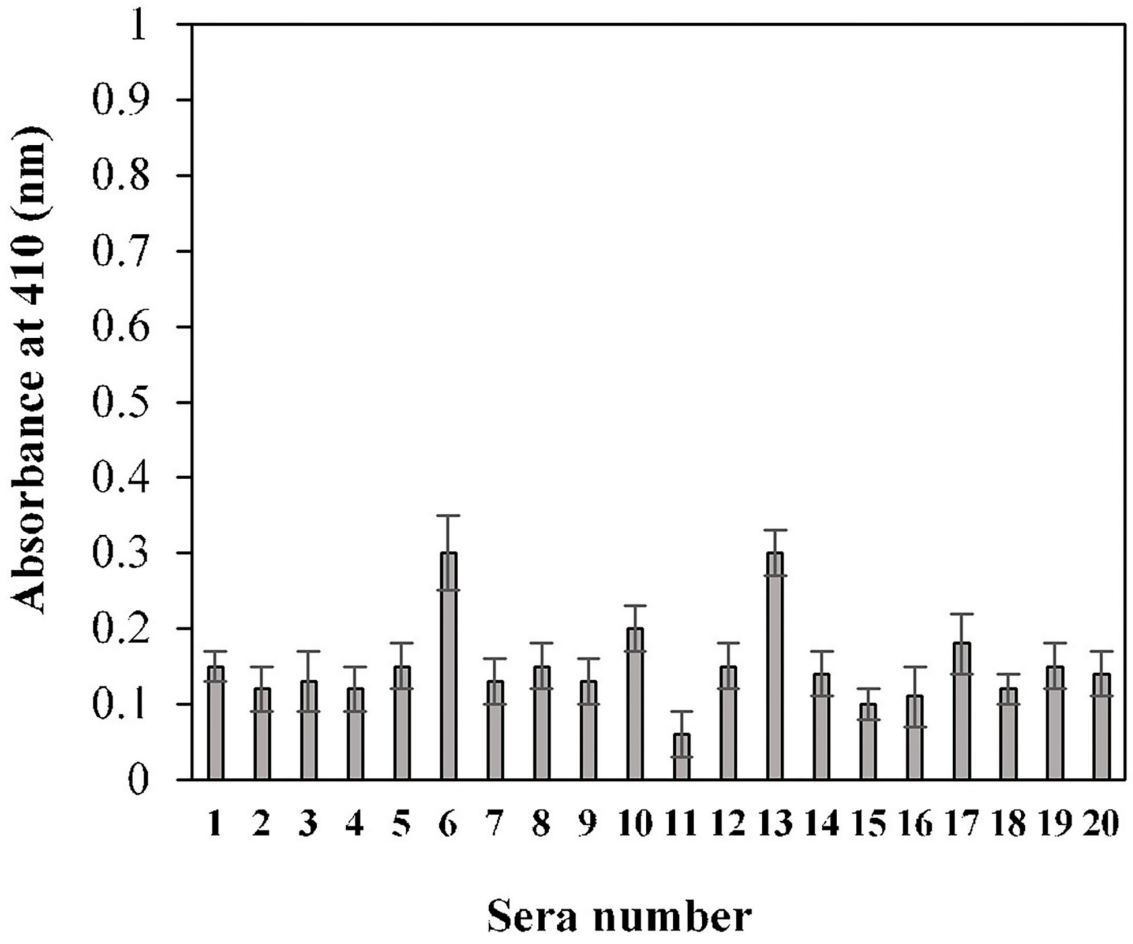


FIGURE 3 | Estimation of cut-off value for indirect binding ELISA. Pre-pandemic serum samples ($n = 20$) were used in the ELISA (1:100 dilution) assay against the S1-RBD antigen (2 μg/ml). Each sample was run in triplicate.

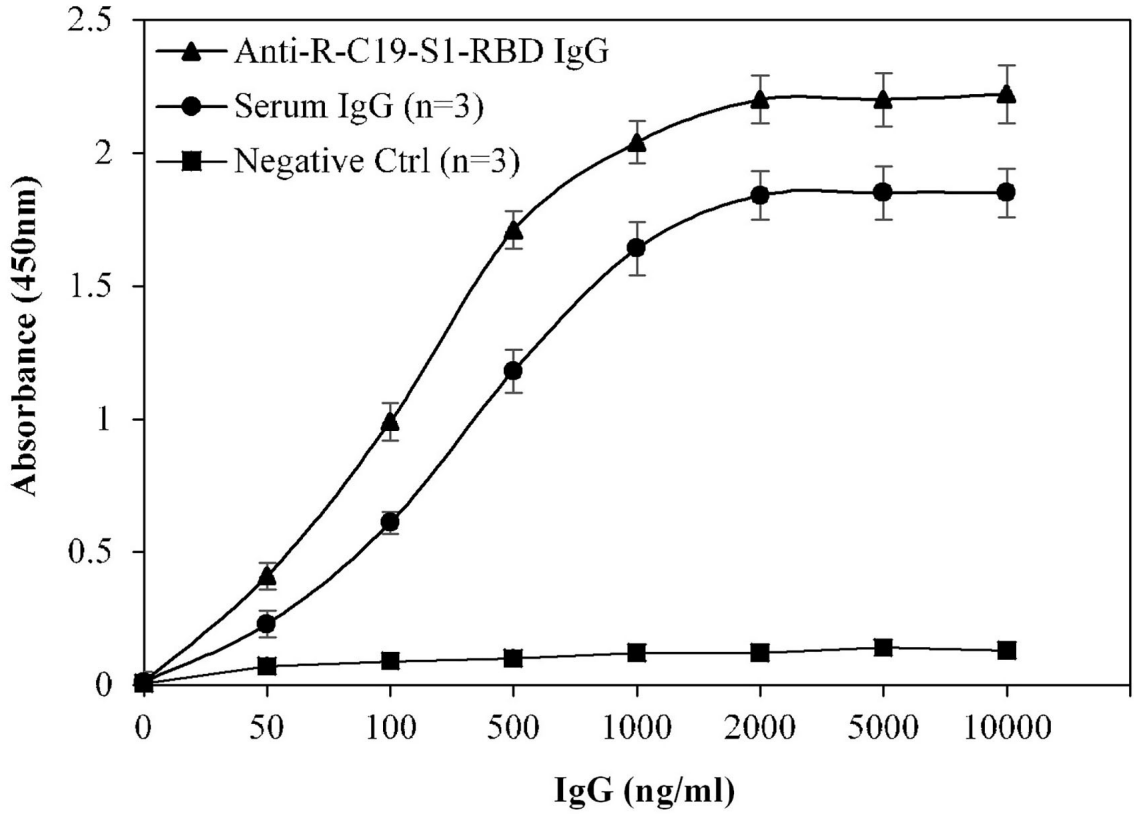


FIGURE 4 | Titration curves for the optimization of serum IgG (COVID-19 patients; $n = 3$). Commercially available anti-R-C19-S1-RBD IgG was used as a positive control. Three nonreactive serum samples from pre-pandemic subjects served as a negative control. ELISA plates were coated with an antigen (S1-RBD) concentration of 2 μg/ml. Each sample was run in triplicate under similar conditions.

against commercially available S1-RBD protein antigen was observed at neat as well as at 1:50 and 1:100 dilutions (**Figure 2**). Therefore, for further ELISA assays, a serum dilution of 1:100 was used.

Determination of the Threshold Value by Indirect ELISA

The calculated cut-off value, using the given method for the 20 randomly selected pre-pandemic normal sera samples, was found to be 0.2 (OD) (**Figure 3**). An in-house threshold or cut-off ratio value, which best distinguished elevated anti-S1-RBD antibody levels from healthy control individuals, was established to be 0.2.

Specificity and Reproducibility of Indirect ELISA

Specificity of the assay was investigated using a titration assay of R-C19-S1-RBD IgG, purified IgGs from COVID-19 convalescent patients ($n = 3$), and pre-pandemic serum ($n = 3$) (**Figure 4**). At a concentration of 2,000 ng/ml, COVID-19 convalescent patients' IgG exhibited higher specificity (1.84 ± 0.09; $p < 0.0001$) than pre-pandemic subjects' IgG (0.12 ± 0.10). R-C19-S1-RBD IgG served as a positive control (2.20 ± 0.09).

Furthermore, the cross-reactivity of serum IgG was investigated using isolated serum IgGs ($n = 3$) from patients with HCV, TB, and RA, showing negligible binding (0.12 ± 0.054, 0.10 ± 0.05 and 0.11 ± 0.045, respectively) (**Figure 5**). However, serum IgG from COVID-19 convalescent patients (OD: 1.87 ± 0.18) exhibited significantly higher ($p < 0.0001$) binding compared to HCV, TB, RA, and blank (**Figure 5**). Anti-R-C19-S1-RBD IgG (2.18 ± 0.07) and IgG from pre-pandemic subjects (0.13 ± 0.05) served as positive and negative controls, respectively.

Inflammatory Cytokine Levels

Inflammatory cytokines such as IL-6 and TNF-α were estimated in serum samples of all the subjects from different groups (**Figure 6**). Post-pandemic subjects exhibited slightly elevated levels of IL-6. However, these differences were non-significant. No remarkable changes were observed in TNF levels in all groups.

Clinical and Epidemiological Characterizations

Clinical and epidemiological data for 240 sera samples from different groups assorted by age and gender are presented in **Table 1**. Volunteers included equal numbers of men and women

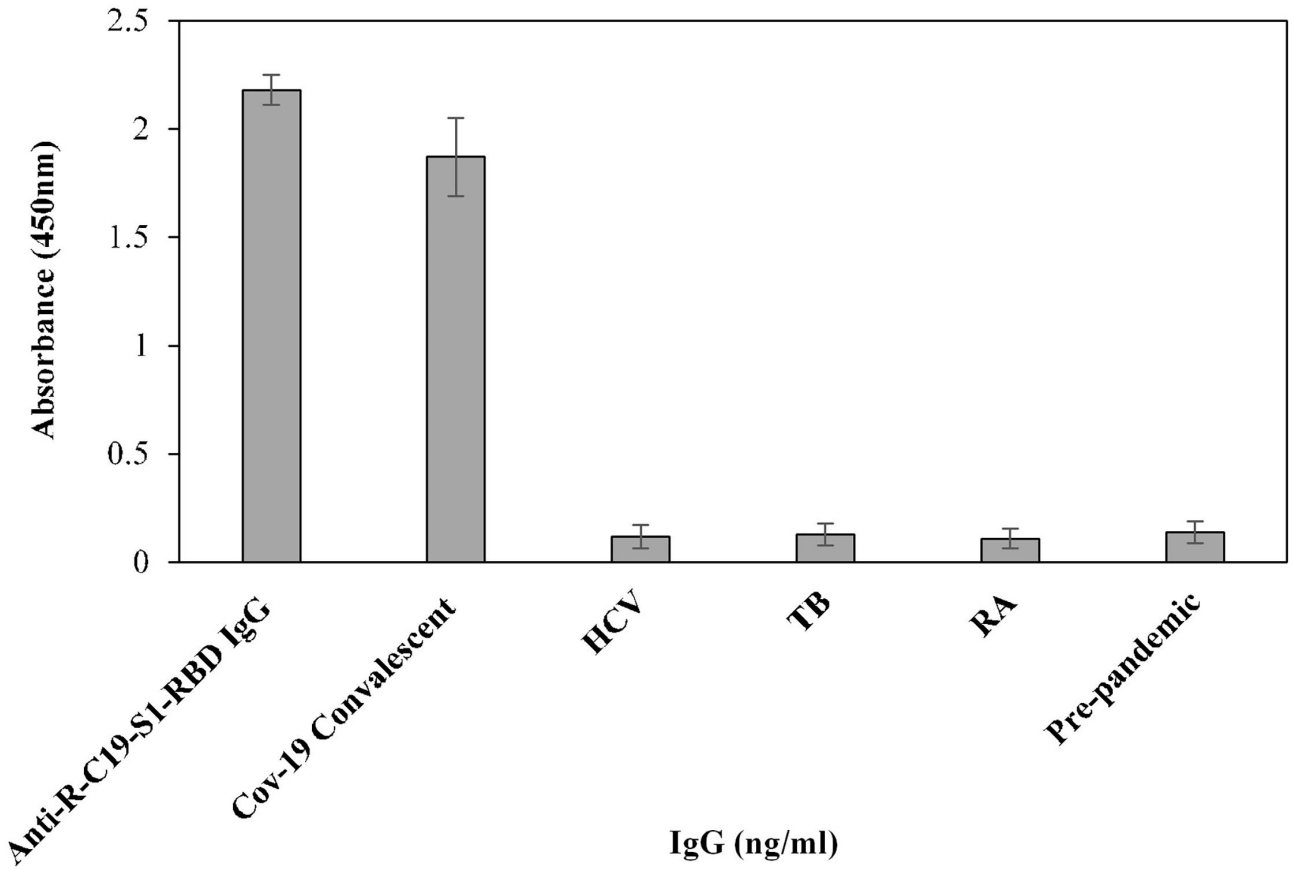


FIGURE 5 | The specificity of the indirect binding ELISA was established by screening three serum samples each from COVID-19 convalescent, HCV, TB and RA patients. Anti-S1-RBD-IgG and pre-pandemic subjects ($n = 3$) were used as positive and negative controls, respectively.

in each group. FBG levels were within the normal range (70–98 mg/dl) for all groups. However, increased FBG levels were observed in older age groups (>60 years) as compared to other age groups (<60 years). HbA1c was found to be within the normal range; however, slightly elevated levels (non-significant) were observed in subjects aged more than 60 years. BMR for pandemic groups showed slightly increased values compared to pre-pandemic groups. In women, the BMR was lower than in men of corresponding age groups. Significantly high BMR was found in groups [M-Pan (20–40, $p < 0.05$), M-Pan (41–60, $p < 0.01$), M-Pan (>60, $p < 0.01$), and F-Pan (>60, $p < 0.05$)] in which subject(s) showed symptoms of fever, cough, and myalgia altogether (**Table 1**). Additionally, this trend was observed only in groups with a higher number of smokers and with an increased smoking duration.

Indirect Binding ELISA

The binding efficiency of serum antibodies and S1-RBD protein antigen was evaluated for all age and gender assorted groups. Serum samples were tested at a dilution of 1:100 in an indirect binding ELISA against the S1-RBD antigen (2 μg/ml).

The binding specificities of serum antibodies against the S1-RBD antigen in samples collected before and during the COVID-19 pandemic were found to vary among the 20–40-year-old age group subjects. Samples collected from both male and female subjects pre-COVID showed low binding toward antigen, i.e., 0.17 ± 0.016 and 0.16 ± 0.018, respectively. However, significantly ($p < 0.05$) higher binding was observed in sera samples collected during the pandemic from subjects of both genders (male and female) corresponding to the same age groups, 0.31 ± 0.029 and 0.29 ± 0.024, respectively (**Figure 7**). Only two pre-pandemic serum samples, each from men and women aged 20–40 years, were found to be positive (average values; 0.33 ± 0.024 and 0.34 ± 0.027, respectively). However, for the same age and gender-matched samples collected during the pandemic, seven samples were found to be positive for men and women (average values; 0.59 ± 0.039 and 0.55 ± 0.035, respectively) each (**Figure 7**).

Serum antibody binding patterns against the S1-RBD antigen were evaluated for both male and female subjects aged 41–60 years (**Figure 8**). Low levels of binding were observed in both male (0.19 ± 0.018) and female (0.18 ± 0.017) subjects from serum samples collected before the start of the pandemic. Samples collected during the pandemic from both male and female individuals exhibited significantly ($p < 0.05$) higher binding (0.35 ± 0.026 and 0.30 ± 0.025, respectively) against the

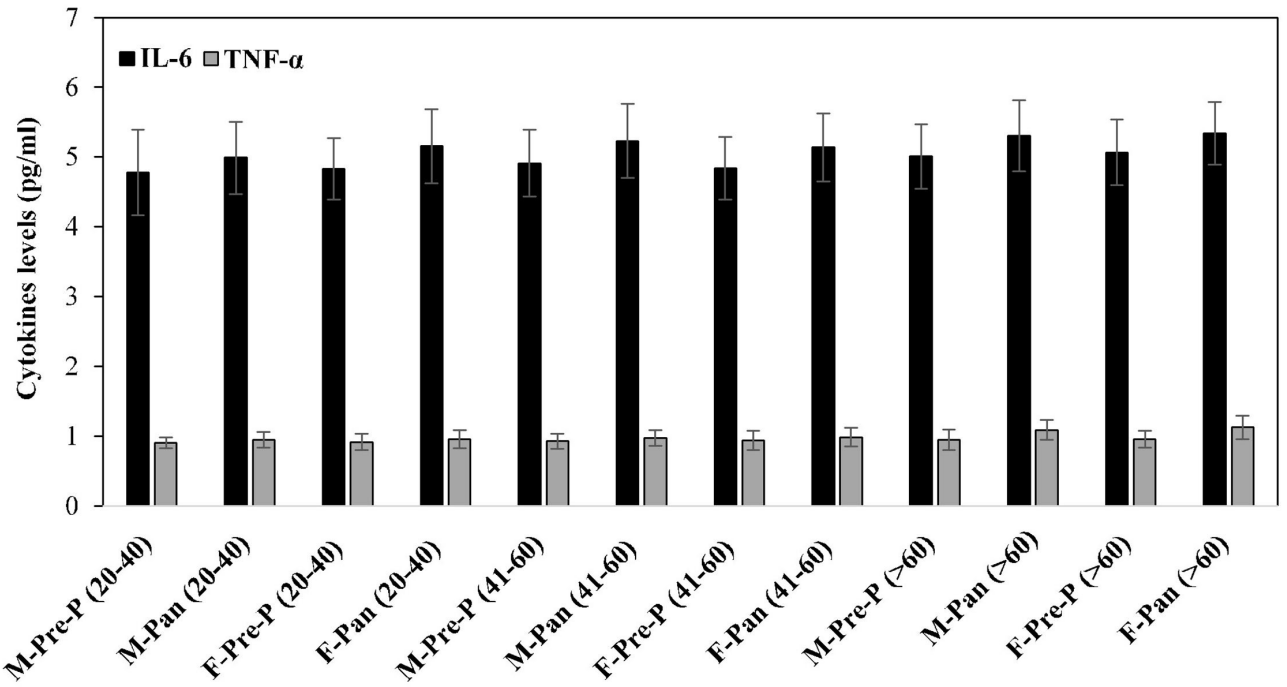


FIGURE 6 | Serum inflammatory cytokines TNF-α and IL-6 levels (pg/ml) were estimated from all the studied groups. All samples were in triplicates, and values are given as mean ± SD.

antigen as compared to the age-matched pre-pandemic subjects. From samples collected during the pandemic in the age group of 41–60 years, eight samples from men (0.63 ± 0.041) and six samples from women (0.59 ± 0.043) showed positive binding with high reactivity (**Figure 8**). Comparatively, a much smaller number of pre-pandemic samples showed positive binding and reactivity; two samples were from men (0.41 ± 0.033) and one sample (0.43 ± 0.032) from women.

In the sample group > 60 years of age, no binding activity was detected among both male (0.17 ± 0.017) and female (0.16 ± 0.017) subjects in samples collected prior to the pandemic (**Figure 9**). Samples for both men (0.20 ± 0.022) and women (0.23 ± 0.021) collected during the pandemic showed less reactivity toward the antigen. However, in the same age group, seropositivity was detected in four men (0.285 ± 0.027) and six women (0.38 ± 0.039) samples collected during the pandemic. Moreover, the positive samples from > 60 years olds showed a low level of reactivity when compared to positive samples from other groups (20–40 and 41–60 years) collected during the pandemic.

Correlation Analysis

Correlation analysis was performed for all pre-pandemic samples as well as samples collected during the pandemic for antibodies against S1-RBD in different age groups and various parameters (FBG, BMR, smoking, fever, fatigue, cough, and myalgia) (see **Table 2**). This analysis showed that the data for fever, fatigue, cough, and myalgia significantly correlated with antibodies against S1-RBD in samples collected during the pandemic in higher age groups (41–60 and >60) for both male and female subjects. However, for the age group 41–60 years, more parameters showed a correlation for male subjects [M-Pan (41–60)] as compared to female subjects [F-Pan (41–60)]. Parameters such as fever, fatigue, and cough consistently correlated with samples collected during the pandemic for all groups. FBG levels did not exhibit a correlation with any of the groups. Correlation analysis showed that myalgia could be strongly correlated with circulating IgG against COVID-19 infection, even though subjects had not been diagnosed with the disease.

Inhibition ELISA of Serum Antibodies Against S1-RBD

The binding specificity of the circulating antibodies to the S1-RBD antigen was further ascertained by inhibition ELISA using S1-RBD as an inhibitor, as given in **Table 3**. As shown in indirect binding ELISA results, inhibition ELISA of subjects from samples collected during the pandemic showed a significantly ($p < 0.001$) higher maximum percent inhibition than the age-matched subjects from the pre-pandemic group. The highest mean percent inhibition was detected in subjects from groups M-Pan (41–60), followed by M-Pan (20-40), F-Pan (20-40), M-Pan (41-60), F-Pan (>60), and M-Pan (>60) (see **Table 3**). In contrast, very low mean percent inhibitions (39 ± 3.4 – 32 ± 3.0) were observed in most of the groups of the pre-pandemic subjects. M-Pre (>60) and F-Pre (>60) did not have any positive samples. However, two samples were randomly

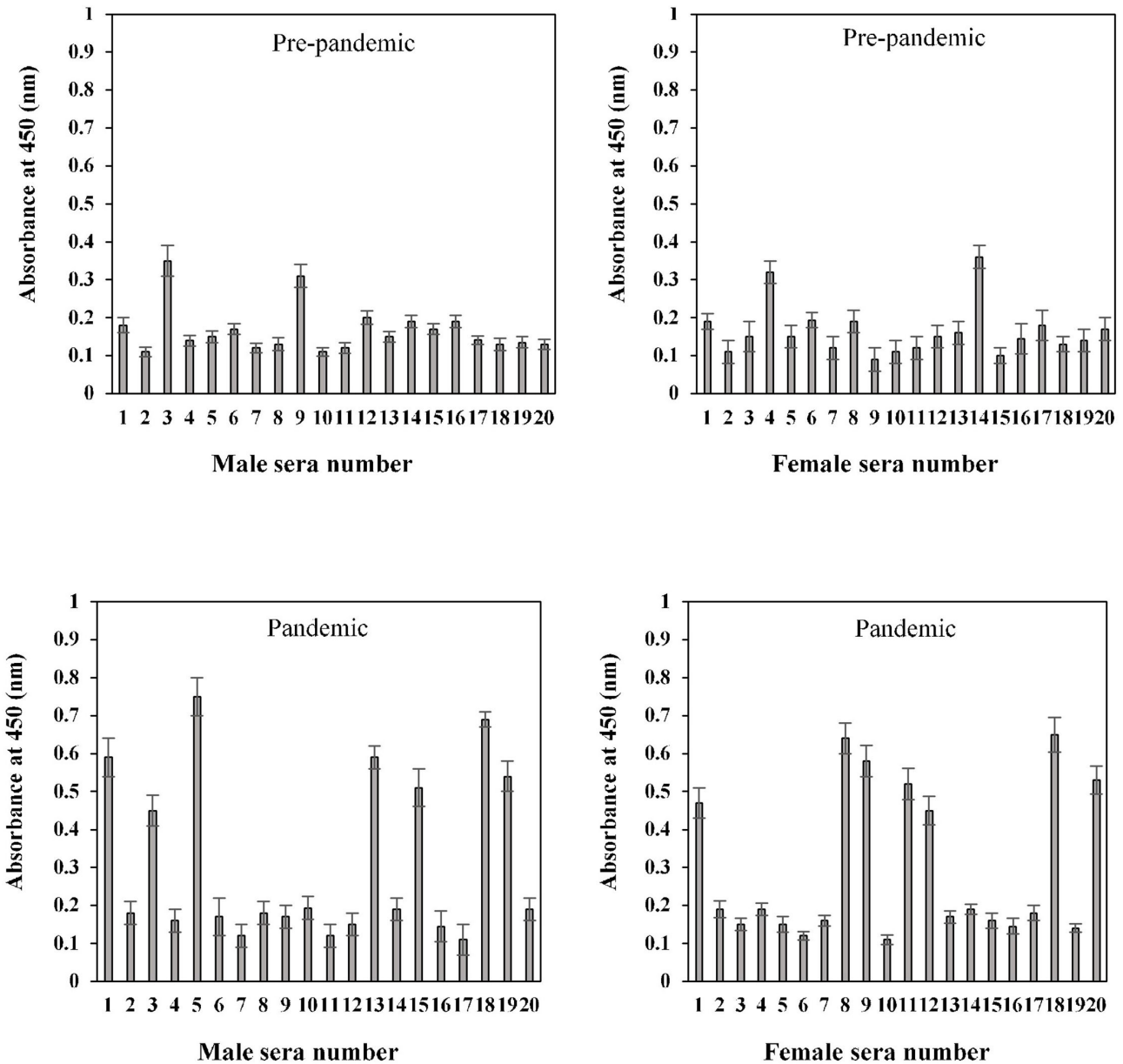


FIGURE 7 | Indirect binding ELISA of pre-pandemic samples and samples collected during the pandemic from men and women aged 20–40 years. ELISA plates were coated with an antigen concentration of 2 μg/ml. Each sample was run in triplicate under similar conditions.

selected from these groups and tested for inhibition ELISA (see **Table 3**).

This data showed that pre-pandemic subjects older than 60 years did not have antibodies when compared to subjects aged 40 years or less. This may contribute to the higher death rate in older age groups (>60 years) in COVID-19 patients. Even circulating antibodies in positive subjects belonging to older age groups (>60 years) collected during the pandemic showed less specificity for the S1-RBD antigen than in subjects aged <40 years.

DISCUSSION

This study highlights the potential contribution of serology in understanding the difference in the immune status of individuals before the onset of and during the early phase of the COVID-19 pandemic. We report an in-house ELISA assay, developed and optimized to detect antibodies against the S1 domain of SARS-CoV-2 spike protein characteristics of SARS-CoV2 anti-S1-RBD IgG antibodies in sera samples of individuals collected pre-pandemic and during the COVID-19 pandemic.

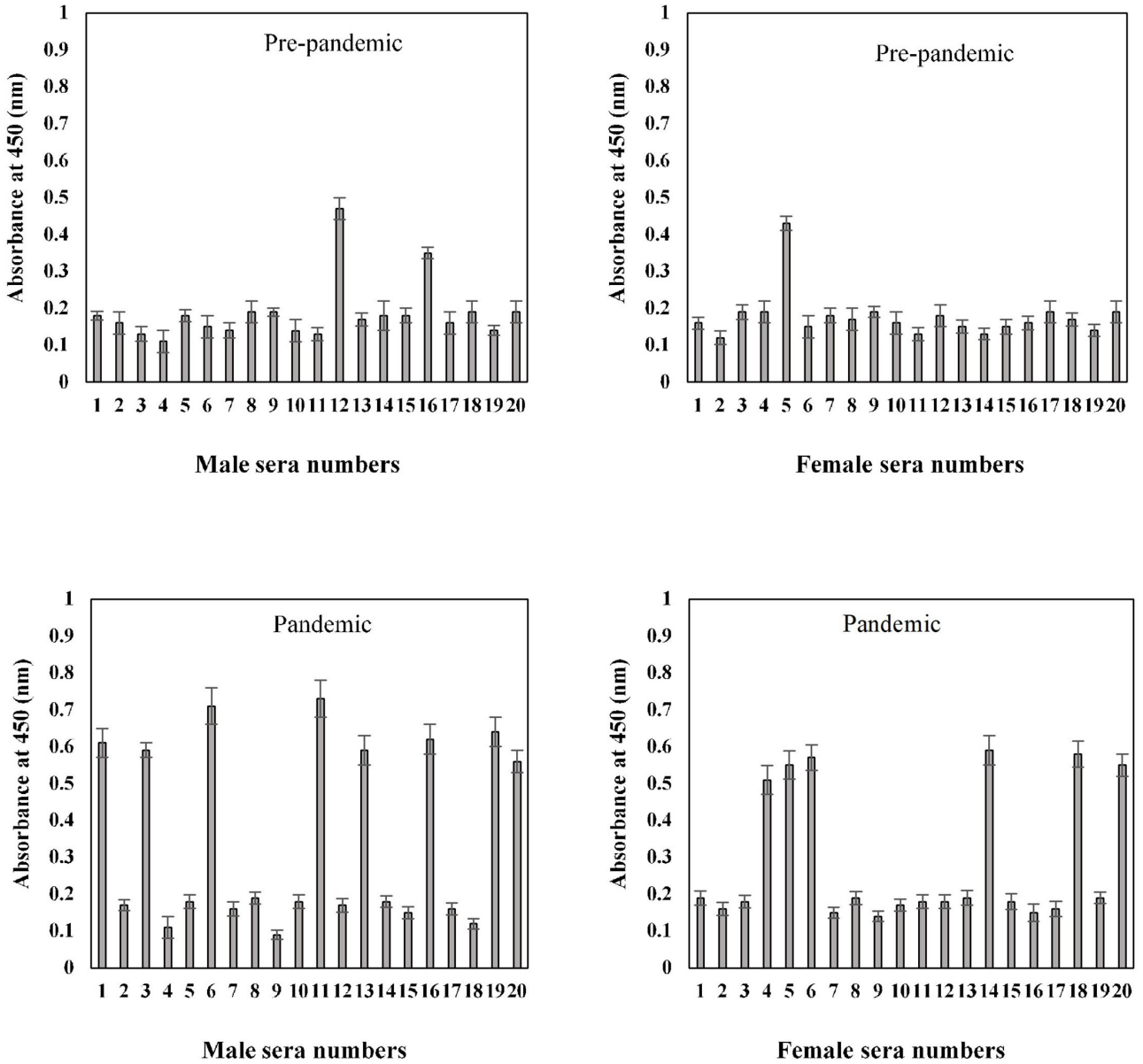


FIGURE 8 | Indirect binding ELISA from samples collected pre-pandemic and during the pandemic from men and women aged 41–60 years. ELISA plates were coated with an antigen concentration of 2 μg/ml. Each sample was run in triplicate under similar conditions.

The preferred method for diagnosing SARS-CoV-2 infection has been through viral nucleic acid or RT-qPCR tests, which require pharynx swab samples. This method, although highly sensitive, is subject to sampling techniques. The ELISA assay is an alternate assay that is both highly specific, sensitive, and cost-effective. Not only is this method suitable for large-scale sample testing and diagnosis but it also provides valuable information about the humoral state of the subject (21, 22). This study focuses on the development of an easy-to-use and high-throughput serological ELISA method with a low threshold value, which is specific, sensitive, and reproducible. This detection method can potentially identify asymptomatic, subclinical, or prior infections. Our optimized protocol can be implemented to accommodate large-scale automated testing of COVID-19 antibodies against the S1-RBD protein antigen.

The study aimed to investigate differences in anti-S1-RBD antibody profiles in sera samples collected from equal numbers of gender and age-matched individuals pre-pandemic and during the pandemic in 2020 to discern differences in trends. Along with demographic data, it is essential to study various clinical characteristics such as fever, cough, myalgia, fatigue, FBG, and BMR levels that may affect disease prognosis (23–26). High FBG and HbA1c levels were detected in older age groups (>60 years), which is a normal pattern in the elderly (17). BMR was also

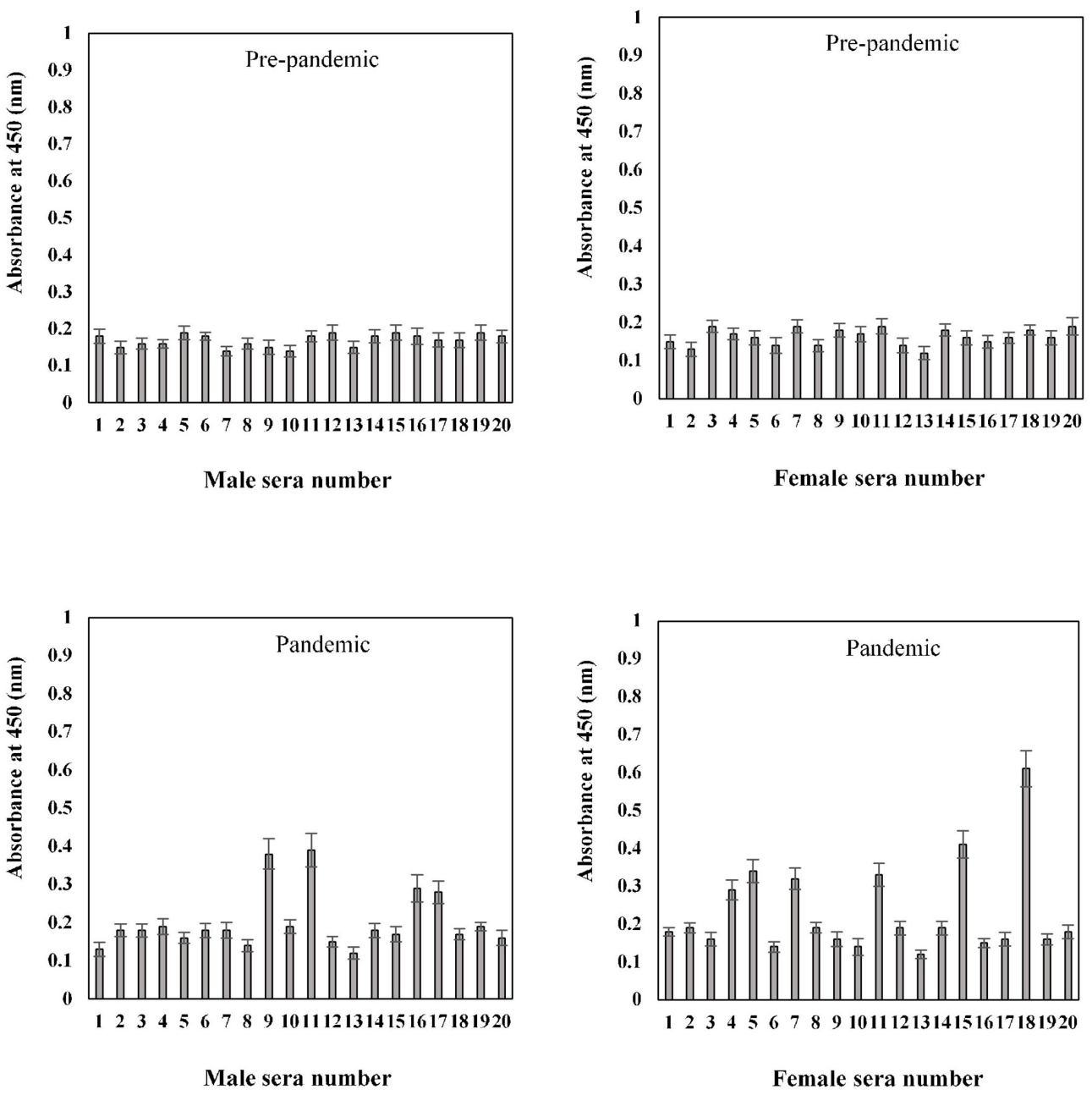


FIGURE 9 | Indirect binding ELISA from samples collected pre-pandemic and during the pandemic from men and women aged > 60 years. ELISA plates were coated with an antigen concentration of 2 μg/ml. Each sample was run in triplicate under similar conditions.

found to decrease linearly with age. Women exhibited lower BMR than men of corresponding age groups. However, participants reporting symptoms of moderate fever, fatigue, cough, and myalgia had a significantly ($p < 0.05$) higher BMR, with the trend most apparent in men with a history of smoking. They increased smoking duration (23, 24).

The binding specificity profiles of serum antibodies against the S1-RBD antigen were ascertained by indirect binding ELISA for all age and gender assorted groups, collected pre-pandemic and during the pandemic. A smaller number of subjects (10%) with low levels of circulating antibodies against S1-RBD were identified in pre-pandemic groups (>60 years). The prior exposure of these subjects to other coronaviruses like MERS-CoV cannot be ruled out. However, a higher number of non-positive pre-pandemic subjects might not have been exposed to any coronaviruses and thus did not display antibodies, which appear to be rare (27). No positive sera samples were detected in pre-pandemic samples of both men and women aged > 60

TABLE 2 | Correlation between anti-S1-RBD antibody levels and study parameters in various age groups.

		M-Pre-P (20-40)	M-Pan (20-40)	F-Pre-P (20-40)	F-Pan (20-40)	M-Pre-P (41-60)	M-Pan (41-60)	F-Pre-P (41-60)	F-Pan (41-60)	M-Pre-P (>60)	M-Pan (>60)	F-Pre-P (>60)	F-Pan (>60)
		r	r	r	r	R	r	r	r	r	r	r	r
Antibodies Against S1-RBD	FBG	0.05	0.32	0.37	0.1	0.35	0.05	0.07	0.04	0.15	0.39	0.23	0.1
	BMR	0.08	0.61*	0.44	0.25	0.13	0.77**	0.28	0.38	0.2	0.71**	0.16	0.64*
	Smoking	0.13	0.34	0.05	0.43	0.23	0.91***	0.67*	0.21	0.4	0.63*	0.3	0.76**
	Fever	0.03	0.6*	0.19	0.42	0.04	0.68*	NA	0.6*	NA	0.87***	0.03	0.81***
	Fatigue	0.09	0.67*	0.3	0.61*	0.02	0.60*	0.11	0.44*	0.1	0.71**	0.12	0.74**
	Cough	NA	0.67*	0.26	0.61*	0.04	0.69*	0.17	0.6*	0.3	0.68*	0.19	0.68*
	Myalgia	NA	0.61*	NA	NA	NA	0.63*	NA	NA	NA	0.86***	NA	0.74**
	IL-6	0.06	0.08	0.07	0.08	0.06	0.09	0.05	0.09	0.07	0.09	0.06	0.08
	TNF-α	0.03	0.04	0.03	0.03	0.04	0.03	0.04	0.04	0.03	0.05	0.04	0.06

*Correlation analyses were performed using SPSS18. Pearson correlation tests were adopted to analyze the correlations between parameters. Significance is defined as *P < 0.05, **P < 0.001, ***P < 0.0001. NA represents not applicable.*

TABLE 3 | Pre- and intra-pandemic comparison of serum IgG inhibition in various age group.

Subject groups (years)	Mean maximum percent inhibition	
	Pre-pandemic	Pandemic
Male (20–40)	37 ± 3.1 (2)*	83 ± 4.8 (7)*
Female (20–40)	39 ± 3.4 (2)*	82 ± 4.3 (7)*
Male (41–60)	34 ± 3.8 (2)*	88 ± 5.5 (8)*
Female (41–60)	32 ± 3.0 (2)*	80 ± 5.1 (7)*
Male (≥61)	4.2 ± 1.6 (2)‡	43 ± 4.9 (4)*
Female (≥61)	3.9 ± 1.4 (2)‡	48 ± 4.6 (6)*

**Values in parenthesis showed positive sera samples against S1-RBD antigen. ‡Randomly selected non-positive samples against S1-RBD antigen. The ELISA plates were coated with antigen (2 μg/ml). Antigen was used as an inhibitor.*

years. Age is a significant risk factor for COVID-19 infection that has been explored in many studies (28–30). Our study is in agreement with the findings of another age-structured study which showed that in individuals aged younger than 20 years of age, susceptibility to COVID infection was found to be approximately half that of adults aged over 20 years (31). Also, the incidence of clinical symptoms exhibited in infected individuals increased with increasing age. The implications of age-related susceptibility to infection and immune outcomes are essential factors in consideration of the burden of disease. These can further be used for age-structured correlation studies for a population and projections of subclinical and clinal infections.

A recent report suggested the presence of pre-existing cross-reactive antibodies to the SARSCoV2 spike in young people, including children, mainly against the S2 domain (31, 32). However, the complete etiology of the presence of these antibodies is unknown, and hence, further investigations would be necessary for protection against future SARS-related infections.

The binding specificities were significantly higher for sera collected during the pandemic from age and gender-matched subjects, indicating higher chances of exposure to the S1-RBD antigen of SARS-CoV2. A total of 38 sera samples were strongly positive, with more male sera samples found to be positive in the age groups of 20–40 and 41–60 years, perhaps due to the higher probability of exposure, the role of sex hormones in immune activation and increased age (33). These findings can also be explained by previous studies highlighting inadequate compliance with recommended hygiene measures and contact restrictions in younger age groups and men (34).

Conversely, fewer female samples showed strong positivity in the 20–40 and 41–60-year age groups compared to male samples. This finding is in agreement with previous studies, which indicate sexual dimorphism in immune response and lower neutralizing antibody titers are significantly associated with the female sex

(33, 35). However, a slightly higher number of female samples were found to be positive in the age group of >60 years, indicating a more robust immune response than men. However, the reason behind this difference is not apparent and warrants further investigation. This shows that gender is an essential factor in subjects of older age groups for the presence of antibodies against the S1-RBD antigen.

Antibodies with neutralizing activity are considered necessary for protection against SARS-CoV-2 infection. Many studies have demonstrated a close correlation between anti-SARS-CoV-2 spike IgG antibody levels and neutralizing activity (27, 36), which was also shown in this study, suggesting a critical role of anti-spike antibodies in virus neutralization. The duration of persistence of these antibodies is key to devising strategies to combat newly emergent highly transmissible variants.

Strong correlations were observed for fever, fatigue, cough, and myalgia with antibodies against S1-RBD in samples collected during the pandemic (30). However, FBG, HbA1c, and BMR either did not correlate or inconsistently correlated with a few groups. Inflammatory cytokines IL-6 and TNF-α also did not show any correlation with serum IgGs against the S1-RBD antigen. Hence, these factors cannot be ignored when assessing the disease diagnosis and the level of infection. To the best of our knowledge, this is the first seroprevalence study of IgG specific for the COVID-19 virus antigen "S1-RBD" from the Hail region in KSA, providing valuable information about IgG levels in different age groups as well as genders. Interestingly, approximately <10% of subjects exhibited the presence of these IgGs in serum samples obtained prior to the COVID-19 pandemic. The reason for this remains uninvestigated, although prior exposure of the population to another SARS virus may be a possibility. The limitations of our study include the limited numbers of samples and clinical data collected for this cross-sectional study conducted during the early phase of the pandemic due to the strict health and safety policies and restrictions. Future longitudinal studies with a larger sample size would be valuable for a comprehensive comparison of data for such samples. They may provide a better understanding of social determinants and the overall humoral immune response.

RBD-specific antibodies detected in the plasma of infected patients showed potent antiviral activity in all infected individuals, suggesting a broader role of neutralizing antibodies in COVID-19 infection, which may contribute to overall vaccine design and efficacy (37). Inhibition ELISA performed for all groups exhibited low antigen-antibody binding specificities in older age group subjects than younger individuals. Seroprevalence and surveillance studies can help identify asymptomatic or subclinical infections in a population. Such studies can offer insight into the subtle differences/variations in the underlying immunological mechanisms. Knowledge of the contribution of any pre-existing immunity to SARS-CoV-2 and the role of the humoral immune response in asymptomatic and subclinical infections is vital in devising strategies for surveillance and containment. Even with the increased availability of a range of vaccines against SARS-CoV-2, the possibility of reinfection still looms large with the threat of newer and more transmissible variants and dual viral infections, partially vaccinated or unvaccinated populations, and waning immunity levels. Understanding the extent and duration of protective immunity in individuals of a population is essential for the protection of vulnerable groups and facilitating the return of society to a state of normalcy.

CONCLUSION

The current study encompasses a seroepidemiological study of anti-S1-RBD antibodies in a population in the Hail region, KSA, before the start of and during the early phase of the COVID-19 pandemic. Pre-pandemic subjects >60 years exhibited about 10 percent circulating antibodies against S1-RBD antigen, which is indicative of earlier exposure to other coronaviruses. In early pandemic subjects, the percentage of anti-S1-RBD antibodies significantly increased to 35 percent. These serum antibodies showed a strong correlation with symptoms of fever, fatigue, cough, and myalgia. Higher antibody titers are significantly associated with the male sex. However, these antibodies decreased in the elderly. This ELISA assay is an important and valuable tool for screening large numbers of samples from different age groups and assessing immune status. Age-specific antibody profiles indicate the need for targeted monitoring strategies for prevention, disease management, and vaccine effectiveness. The rise of newer variants, waning antibody levels, and reduced vaccine efficacy raise concerns about the durability of responses in clinical protection.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee (REC) at the University of Hail. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SS, MWAK, AM, NA, and MR: formal analysis. SS and MWAK: investigation, conceptualization, data curation, and writing the original draft preparation. SS, MWAK, and MS: resources. AS, AM, MK, NA, and MR: writing-review and editing. SS: supervision and funding acquisition. SS, MWAK, AM, and MS: methodology. All authors read and approved the final manuscript.

FUNDING

This research has been funded by the Scientific Research Deanship at the University of Ha'il – Saudi Arabia through project number RG-20 077.

REFERENCES

1. Eastin C, Eastin T. Clinical characteristics of coronavirus disease 2019 in China: Guan W, Ni Z, Hu Y, et al. *N Engl J Med*. (2020) [Online ahead of print]. *J Emerg Med.* (2020) 58:711–2. doi: 10.1016/j.jemermed.2020.04.004
2. Biswas M, Rahaman S, Biswas TK, Haque Z, Ibrahim B. Association of sex, age, and comorbidities with mortality in COVID-19 patients: a systematic review and meta-analysis. *Intervirology*. (2021) 64:36–47. doi: 10.1159/000512592
3. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. COVID-19, SARS and MERS: are they closely related? *Clin Microbiol Infect.* (2020) 26:729–34. doi: 10.1016/j.cmi.2020.03.026
4. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* (2020) 579:265–9. doi: 10.1038/s41586-020-2008-3
5. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann of Intern Med*. (2020) 172:577–5 doi: 10.7326/M20-0504
6. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med.* (2020) 26:1037–40. doi: 10.1038/s41591-020-0916-2
7. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med.* (2020) 180:934–43. doi: 10.1001/jamainternmed.2020.0994
8. Sherwani S, Khan MWA. Cytokine response in SARS-CoV-2 infection in the elderly. *J Inflamm Res.* (2020) 13:737–47. doi: 10.2147/JIR.S276091
9. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* (2016) 24:490–502. doi: 10.1016/j.tim.2016.03.003
10. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* (2020) 181:281–92. doi: 10.1016/j.cell.2020.02.058
11. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.* (2015) 1282:1–23. doi: 10.1007/978-1-4939-2438-7_1
12. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis.* (2020) 21:1–8. doi: 10.1093/cid/ciaa310
13. Huang AT, Garcia-Carreras B, Hitchings MDT, Yang B, Katzelnick LC, Rattigan SM, et al. A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. *Nat Commun.* (2020) 11:4704. doi: 10.1038/s41467-020-18450-4
14. Bavaro DF, Laghetti P, Milano E, Brindicci G, Volpe A, Lagioia LA, et al. Anti-spike S1 receptor-binding domain antibodies against SARS-CoV-2 persist several months after infection regardless of disease severity. *J Med Virol.* (2021) 93:3158–64. doi: 10.1002/jmv.26878
15. Alejo JL, Mitchell J, Chang A, Chiang TPY, Massie AB, Segev DL, et al. Prevalence and durability of SARS-CoV-2 antibodies among unvaccinated us adults by history of COVID-19. *JAMA*. (2022) 327:1085–7. doi: 10.1001/jama.2022.1393
16. Sherwani S, Chowdhury M, Bugert JJ. ELISA for molluscum contagiosum virus. *Curr Protoc Microbiol*. (2017) 47:14A.6.1–9. doi: 10.1002/cpmc.42
17. Sherwani S, Farleigh L, Agarwal N, Loveless S, Robertson N, Hadaschik E, et al. Seroprevalence of Molluscum contagiosum virus in German and UK populations. *PLoS ONE.* (2014) 9:e88734. doi: 10.1371/journal.pone.0088734
18. Khan MW, Banga K, Mashal SN, Khan WA. Detection of autoantibodies against reactive oxygen species modified glutamic acid decarboxylase-65 in type 1 diabetes associated complications. *BMC Immunol.* (2011) 12:19. doi: 10.1186/1471-2172-12-19
19. Khan MW, Qadrie ZL, Khan WA. Antibodies against gluco-oxidatively modified human serum albumin detected in diabetes-associated complications. *Int Arch Allergy Immunol.* (2010) 153:207–14. doi: 10.1159/000312639
20. Khan MW, Sherwani S, Khan WA, Ali R. Characterization of hydroxyl radical modified GAD65: a potential autoantigen in type 1 diabetes. *Autoimmunity*. (2009) 42:150–8. doi: 10.1080/08916930802468276
21. Galipeau Y, Greig M, Liu G, Driedger M, Langlois MA. Humoral responses and serological assays in SARS-CoV-2 infections. *Front Immunol.* (2020) 11:610688. doi: 10.3389/fimmu.2020.610688
22. Zhou C, Bu G, Sun Y, Ren C, Qu M, Gao Y, et al. Evaluation of serum IgM and IgG antibodies in COVID-19 patients by enzyme linked immunosorbent assay. *J Med Virol.* (2021) 93:2857–66. doi: 10.1002/jmv.26741
23. Metsios GS, Stavropoulos-Kalinoglou A, Nevill AM, Douglas KM, Koutedakis Y, Kitas GD. Cigarette smoking significantly increases basal metabolic rate in patients with rheumatoid arthritis. *Ann Rheum Dis.* (2008) 67:70–3. doi: 10.1136/ard.2006.068403
24. Audrain-McGovern J, Benowitz NL. Cigarette smoking, nicotine, and body weight. *Clin Pharmacol Ther.* (2011) 90:164–8. doi: 10.1038/clpt.2011.105
25. Zhang C, Lin L, Tang D, Liu F, Li M, Li Q, et al. Antibody responses to SARS-CoV-2 in healthy individuals returning to Shenzhen. *J Med Virol.* (2021) 93:1154–7. doi: 10.1002/jmv.26355
26. He B, Wang J, Wang Y, Zhao J, Huang J, Tian Y, et al. The metabolic changes and immune profiles in patients with COVID-19. *Front Immunol.* (2020) 11:2075. doi: 10.3389/fimmu.2020.02075
27. Lv H, Wu NC, Tsang OT, Yuan M, Perera RAPM, Leung WS, et al. Cross-reactive antibody response between SARS-CoV-2 and SARS-CoV infections. *Cell Rep.* (2020) 31:107725. doi: 10.1016/j.celrep.2020.107725
28. Bates TA, Leier HC, Lyski ZL, Goodman JR, Curlin ME, Messer WB, et al. Age-dependent neutralization of SARS-CoV-2 and P. 1 variant by vaccine immune serum samples. *JAMA*. (2021) 326:868–9. doi: 10.1001/jama.2021.11656
29. Schlickeiser S, Schwarz T, Steiner S, Wittke K, Al Besher N, Meyer O, et al. Disease severity, fever, age, and sex correlate with SARS-CoV-2 neutralizing antibody responses. *Front Immunol.* (2021) 11:628971. doi: 10.3389/fimmu.2020.628971
30. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID Covid-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med.* (2020) 26:1205–11. doi: 10.1038/s41591-020-0962-9
31. Ng KW, Faulkner N, Cornish GH, Rosa A, Harvey R, Hussain S, et al. Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science.* (2020) 370:1339–43. doi: 10.1126/science.abe1107
32. Jian-Min J, Peng B, Wei H, Fei W, Xiao-Fang L, De-Min H, et al. Gender differences in patients with COVID-19: Focus on severity and mortality. *Front Public Health.* (2020) 8:152. doi: 10.3389/fpubh.2020.00152
33. Markmann AJ, Giallourou N, Bhowmik DR, Hou YJ, Lerner A, Martinez DR, et al. Sex disparities and neutralizing antibody durability to SARS-CoV-2 infection in convalescent individuals. *medRxiv*. (2021) 6:e0027521. doi: 10.1128/mSphere.00736-21
34. Bazaid AS, Aldarhami A, Binsaleh NK, Sherwani S, Althomali OW. Knowledge and practice of personal protective measures during the COVID-19 pandemic: a cross-sectional study in Saudi Arabia. *PLoS ONE.* (2020) 15:e0243695. doi: 10.1371/journal.pone.0243695
35. Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature.* (2020) 588:315–20. doi: 10.1038/s41586-020-2700-3
36. Shrwani K, Sharma R, Krishnan M, Jones T, Mayora-Neto M, Cantoni D, et al. Detection of serum cross-reactive antibodies and memory response to SARS-CoV-2 in prepandemic and post-COVID-19 convalescent samples. *J Infect Dis.* (2021) 224:1305–15. doi: 10.1093/infdis/jiab333

37. Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature.* (2020) 584:437–42. doi: 10.1038/s41586-020-2456-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fpubh.2022.907814

Modeling of the Small-Scale Outbreak of COVID-19

Ze-Yang Wu[1], *Hong-Bo Zhang*[1,2*] *and Hong-Fei Zhao*[3]

[1] *Department of Digital Media Technology, College of Computer Science and Technology, Huaqiao University, Xiamen, China,* [2] *Fujian Key Laboratory of Big Data Intelligence and Security, Huaqiao University, Xiamen, China,* [3] *Faculty of Science, National University of Singapore, Singapore, Singapore*

With the improvement of treatment and prevention methods, many countries have the pandemic under control. Different from the globally large-scale outbreak of COVID-19 in 2020, now the outbreak in these countries shows new characteristics, which calls for an effective epidemic model to describe the transmission dynamics. Meeting this need, first, we extensively investigate the small-scale outbreaks in different provinces of China and use classic compartmental models, which have been widely used in predictions, to forecast the outbreaks. Additionally, we further propose a new version of cellular automata with a time matrix, to simulate outbreaks. Finally, the experimental results show that the proposed cellular automata could effectively simulate the small-scale outbreak of COVID-19, which provides insights into the transmission dynamics of COVID-19 in China and help countries with small-scale outbreaks to determine and implement effective intervention measures. The countries with relatively small populations will also get useful information about the epidemic from our research.

Keywords: COVID-19, small-scale outbreak, cellular automata, time matrix, simulation

OPEN ACCESS

Edited by:
Gopi Battineni,
University of Camerino, Italy

Reviewed by:
Muhammad Imran Khan,
The University of Haripur, Pakistan
Lin Wang,
University of Cambridge, United Kingdom

***Correspondence:**
Hong-Bo Zhang
zhanghongbo@hqu.edu.cn

Specialty section:
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

Received: *30 March 2022*
Accepted: *06 June 2022*
Published: *01 July 2022*

Citation:
Wu Z-Y, Zhang H-B and Zhao H-F (2022) Modeling of the Small-Scale Outbreak of COVID-19. Front. Public Health 10:907814. doi: 10.3389/fpubh.2022.907814

1. INTRODUCTION

In December 2019, the high-speed expansion of COVID-19 managed itself into a global pandemic in a minute, which ended up as a global crisis. China launches a resolute battle to prevent and control the spread of COVID-19, within 4 months the transmission of the virus has been successfully cut off. The daily confirmed cases in China mainland dropped below 100 and further declined to a single digit. Hard work of China had gained remarkable achievement. Unfortunately, the virus has mutated in a way that might spread easier, which poses a great challenge to epidemic prevention.

According to the data from the National Health Commission of China, we can obtain some general principles underlying the spread of the virus, as shown in **Figure 1**. During outbreaks, the daily confirmed cases are less than one hundred. The outbreaks will last for around 30 days. It is noticeable that the trend of daily confirmed cases reached its peak around 15 days after the outbreak and the daily new recovered cases peak at 20 days later.

With the experience of fighting against COVID-19 in Wuhan, the Chinese government has had science-based measures for COVID-19 prevention and control. In this article, we set our attention on the small-scale outbreak and transmission of COVID-19 in various provinces of China, and try to reveal the general principles underlying the spread of the virus to provide theoretical support for epidemic prevention. With appropriate parameter setting and transmission rules, this study can also be used for epidemic analysis in many other countries. The stability of the proposed model is tested with COVID-19 data in Potter County, Texas US. This study can provide important information for making appropriate decisions in countries that lack medical resources.

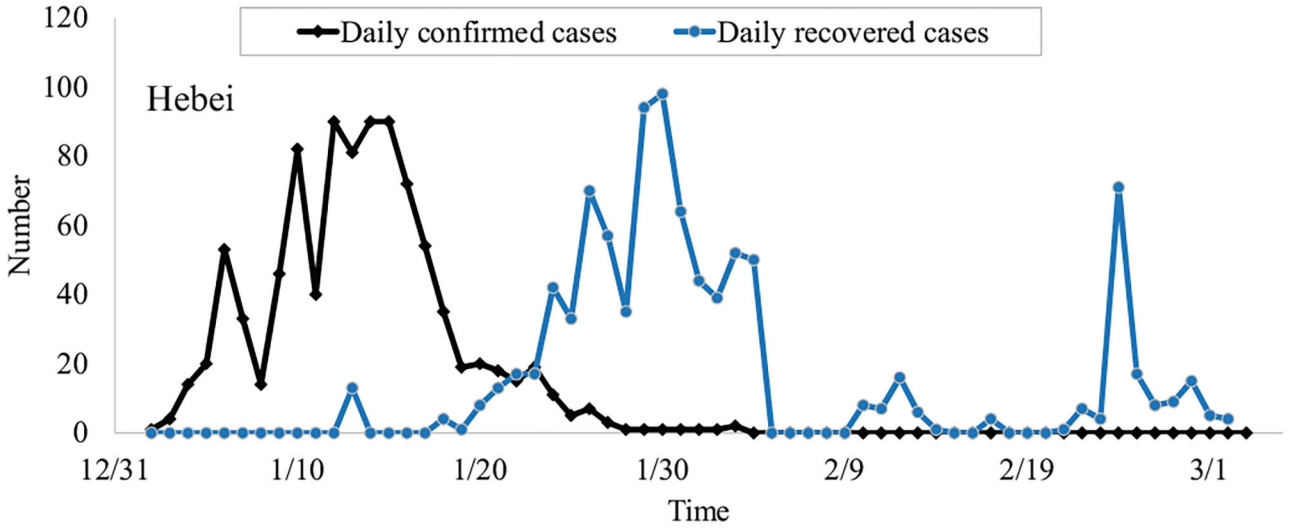


FIGURE 1 | The daily confirmed cases and recovered cases of COVID-19 in Hebei province on January 2021.

Many models can be applied to simulate the spread of the epidemic. The traditional SIR model has already been widely used to simulate the epidemic at a country level, generally with numerous infected cases a huge population (1–4). Nowadays, many countries already have the COVID-19 under control, and the infected cases in a city or county with a relatively small population, are marginal compared with the total infected cases at the country level. Most of the previous studies focused their attention on outbreaks with sufficient infected cases, like the outbreak in Wuhan (4, 5). We believe that COVID-19 will show a trend of small outbreaks within a certain range. It is significant to explore the regularity of small-scale outbreaks of COVID-19. In this study, we focus our attention on small-scale outbreaks with a limited number of cases. First, we applied traditional SIR model and $SEI_UI_DR_UR_D$ model simplified from SEI_DI_UQHRD (6). Then a new version of CA is proposed to carry out the experiment. In the improved CA model, the parameters and transmission rules are set according to the data from the local health department. During the experiment, we simulate the outbreaks in two provinces of China, Heilongjiang and Hebei, and Potter County, Texas US. The results show that our improved CA has a better performance compared with the two compartments models mentioned above.

The contribution of this study can be summarized as follows. To the best of our knowledge, this is the first study that tries to simulate the small-scale outbreak of COVID-19. Additionally, a new version of CA has been proposed with a time matrix that can simulate the outbreak well by setting transmission rules. Utilizing this model, epidemic trends in small-scale outbreaks can be used to help health officials make decisions on public health policies.

2. RELATED STUDY

In 2002, the SARS (Severe Acute Respiratory Syndrome) virus was first found in Guangdong China. Classical compartmental models SIR have been used in simulation and prediction (7–9). In their studies, the number of susceptible, infected, and recovered from Beijing have been calculated using the SIR model, and all parameters with epidemiological meaning including transmission rate, removal rate, and basic reproduction number have been estimated. The same methods have been used in investigating the transmission rules of SARS in Guangdong province (10). The studies showed that transmission dynamic models, in the form of differential equations, could simulate the process of SARS transmission with reasonable parameters and reflect the dynamic of SARS transmission.

Since December 2019, the COVID-19 started its transmission, and classical compartmental models have been widely used in predictions. But the rate of transmission and many other parameters in classical models are constants. For better simulation, numerous researchers have proposed many methods of predicting the parameters dynamically (7, 11, 12). The improved SEIR model has been used in forecasting the outbreak and combined with a series of interventions formulated by the government. With the development of machine learning, a dynamic prediction method of the infection rate was derived based on long short-term memory (LSTM) and has a better performance compared with that of the traditional SEIR model (13, 14). These models assumed that populations are completely mixed and ignore spatial effects of spread epidemics; also interaction between individuals is neglected since they model populations as continuous entities (15).

Cellular automata are dynamic systems with discrete time, space, and state. It discusses the overall properties on the premise of synchronous updating based on local principles, which is expected to simulate the real epidemic situation through the set of local principles (16). It has been applied in the field of infectious disease control (17–19). Classical epidemic models based on differential equations may be unsuitable for simulating small-scale outbreaks of COVID-19, given the lack of flexibility when

simulating local characteristics of infectious diseases. Cellular automata may have better performance in the simulation of small-scale virus outbreaks.

In previous studies (13, 20), most of the methods have been directed at the large-scale outbreak of SARS or COVID-19. Most predictive studies based on cellular automata focused on H1N1 and the small spread of chickenpox (19, 21). At present, the epidemic situation in China is generally stable, with rebounds in some provinces. In this contribution, we used classic compartmental models (SIR and $SEI_UI_DR_UR_D$) and CA to simulate the small-scale outbreaks of COVID-19 in different provinces of China, a time matrix is set to optimize the cellular automata.

Alongside CA, many researchers have implemented an agent-based model (ABM) in simulating the pandemic (22, 23). ABM to some extent evolved from CA, they are a class of agents, and each of them contains variable information. Each agent can interact with their neighbors and transform their state. The major difference between ABM and CA is that each agent within ABM can move their position as well as change state and interact with neighbors. However, the cell in the class of CA will not be able to transform their physic position. ABM is more intuitive than mathematical or statistical models because it represents objects as individual things in the world. In previous studies, ABM models have been used in searching for cost-effective proactive testing strategies and simulating the effects of health policy (24, 25).

3. METHODS

3.1. SIR Model

In the SIR model, individuals are assigned to three compartments or categories: susceptible(S), infectious(I), and recovered(R). S compartment represents the susceptible individuals that are not immune to the virus and might get infected when exposed to it. I compartment stands for those individuals who are carrying the virus and can spread it. R compartment indicates those infected with the virus and have successfully recovered after treatment or died. Suppose that the recovered individuals will not be re-infected or spread the virus.

As a result of the China's public health emergency system and strict traffic controls, the number of deaths is close to zero and population migration with neighboring provinces is negligible. It is reasonable to suppose that the population remains constant during the outbreak, and the birth, death, and migration rates are zero. SIR model can be described by the following set of differential equations.

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{S}{N} I \\ \frac{dI}{dt} &= \beta \frac{S}{N} I - \gamma I \\ \frac{dR}{dt} &= \gamma I \\ N &= S(t) + I(t) + R(t) \end{aligned} \tag{1}$$

where N is the total population of an area and it remains a constant during the spread of the virus and $S(t), I(t), R(t)$ represent the number of individuals in a different compartment at the time t. β is the infection rate, which means the transition probability from S to I. Similarly, γ is the removal rate, which represents the transition probability from I to R. They are often regarded as constants for simplicity of calculation. $\frac{dS}{dt}$ is the changing rate of susceptible individuals. The number of susceptible individuals decreases with the increment of infected individuals. $\frac{dI}{dt}$ is the changing rate of infected individuals. $\frac{dR}{dt}$ is the changing rate of recovered individuals. The framework of SIR is shown in **Figure 2**.

FIGURE 2 | The framework of SIR model.

3.2. $SEI_UI_DR_UR_D$ Model

Based on the SIR model, we further analyzed China's epidemics prevention and set up the SEIR model. Due to the characteristic of COVID-19, there will be a latency when individuals are exposed to the virus (9). During the latency, exposed individuals are incapable of transmitting the virus and the illness did not deteriorate to the infected stage. Susceptible individuals may become exposed. Exposed individuals (E) will eventually evolve into the infected. Therefore, after being exposed to the virus, patients usually turn into I after latency. However, in SIR the exposed individual is not modeled. In addition, there are two types of infectious diseases: symptomatic infectious (I_D) and asymptomatic infectious (I_U) (6). Due to the strict prevention and control measures, a newly detected I_D will get a strict quarantine, and the transmission of the virus will be cut off. Consequently, the original SEIR model is extended to the $SEI_UI_DR_UR_D$ model. The $SEI_UI_DR_UR_D$ model can be described as follow.

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{S\left(I_U + \sigma I_D\right)}{N} \\ \frac{dE}{dt} &= \beta \frac{S\left(I_U + \sigma I_D\right)}{N} - \eta E \\ \frac{dI_U}{dt} &= \phi \eta E - \gamma I_U \\ \frac{dI_D}{dt} &= (1-\phi)\,\eta E - \gamma I_D \\ \frac{dR_U}{dt} &= \gamma I_U \\ \frac{dR_D}{dt} &= \gamma \mathrm{I_D} \end{aligned} \tag{2}$$

Compartments definition (**Figure 3**):

- Susceptible (S) is the part of the population that could be potentially subjected to the infection.
- Exposed (E) is the fraction of the population that has been infected but does not show symptoms yet: it can be

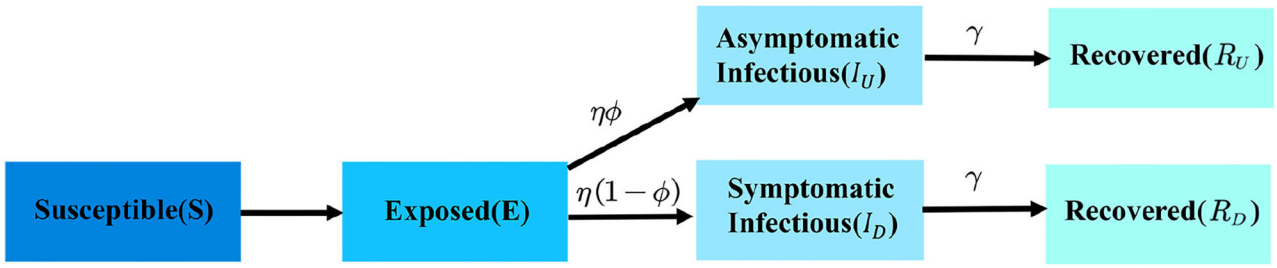


FIGURE 3 | The framework of $SEI_UI_DR_UR_D$ model.

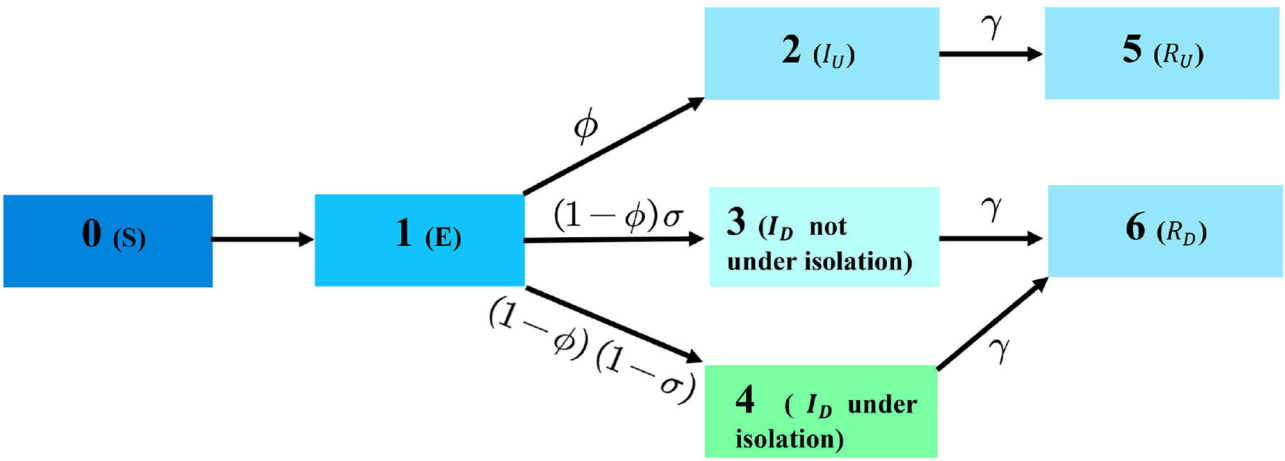


FIGURE 4 | The framework of CA.

called a latent phase. and at this stage, we define it to be not infectious.

- Asymptomatic Infectious (I_U) is people infected with a novel coronavirus that does not exhibit symptoms at any time during the course of infection, and are capable of spreading the virus. They are a potential source of substantial spread within the community (6). Due to the undetectable character of I_U, we assume all the infected cases collected by the health department are symptomatic infectious (I_D).
- Symptomatic Infectious (I_D) stands for population infected with the virus and exhibit a verity of symptoms: fever or chills, cough, shortness of breath, or difficulty breathing. Isolation is needed to cut off the spread of the virus according to the local health policies. In the later experiment, we assume that confirmed cases collected by the public health department represent I_D only.
- Undetected Recovered (R_U) are the people healed from I_U, they have become immune to the virus and will no be reintroduced into the susceptible category.
- Detected Recovered (R_D) are the people healed from I_D, similar to R_U they are immune to the virus, but stand for recovered cases that are in the health department record.

The framework of the $SEI_UI_DR_UR_D$ model is shown in **Figure 3**. Parameters value definition:

(a) β infection rate. It is the number of people that a patient can infect each day, which transports people from the S category to the E category. We define it as a constant and the estimation is in the setting of a later parameter.

(b) η transform rate from E to I_D or I_U, which represents the incubation and is defined as a constant.

(c) ϕ percentage of infections that are asymptomatic stands for the proportion of asymptomatic infectious individual (I_U) in all infections. For example, a parameter value of 0.5 represents that half of the exposed population will transform into I_U.

(d) σ the non-isolation rate of the symptomatic infectious individual (I_D). In some countries, the σ is set to 0, which

TABLE 1 | Daily confirmed and recovered cases of Heilongjiang during the outbreak in January 2021.

Time (day)	Daily confirmed cases	Recovered cases	Time (day)	Daily confirmed cases	Recovered cases
7-Jan	1	0	21-Jan	47	6
8-Jan	0	0	22-Jan	56	8
9-Jan	0	0	23-Jan	29	8
10-Jan	0	0	24-Jan	35	8
11-Jan	1	0	25-Jan	53	9
12-Jan	16	1	26-Jan	29	16
13-Jan	43	1	27-Jan	28	34
14-Jan	43	2	28-Jan	21	56
15-Jan	23	3	29-Jan	27	79
16-Jan	12	3	30-Jan	9	87
17-Jan	7	4	31-Jan	22	99
18-Jan	27	4	1-Feb	8	100
19-Jan	16	5	2-Feb	6	127
20-Jan	68	6	3-Feb	4	170

TABLE 2 | Daily confirmed and recovered cases of Hebei during the outbreak in January 2021.

Time (day)	Daily confirmed cases	Recovered cases	Time (day)	Daily confirmed cases	Recovered cases
2-Jan	1	0	16-Jan	72	13
3-Jan	4	0	17-Jan	54	13
4-Jan	14	0	18-Jan	35	17
5-Jan	20	0	19-Jan	19	18
6-Jan	53	0	20-Jan	20	26
7-Jan	33	0	21-Jan	18	39
8-Jan	14	0	22-Jan	15	56
9-Jan	46	0	23-Jan	19	73
10-Jan	82	0	24-Jan	11	115
11-Jan	40	0	25-Jan	5	148
12-Jan	90	0	26-Jan	7	218
13-Jan	81	13	27-Jan	3	275
14-Jan	90	13	28-Jan	1	310
15-Jan	90	13	29-Jan	1	404

means that all symptomatic infectious are isolated and incapable of transmitting the virus.

(e) γ stands for recovery rate. It gives information about how fast people may recover from the disease or pass away during the treatment ($1/\gamma$ is the average recovery time).

3.3. Cellular Automata

During the transmission, the relationships between data of infectious diseases are extremely complex. However, cellular automata can predict the epidemic through multi-step iteration and parallel evolution only by determining relatively simple evolution rules (18). In the case of traditional dynamics models that are unsuitable for modeling the spread of COVID-19 in China nowadays, we try to use cellular automata to carry out the experiment. Cellular automata are a dynamic system discrete in time, space, and state, different kinds of cells represent different groups of people: S-cell represents susceptible, E-cell represents exposed, I_U-cell is asymptomatic infectious, I_D-cell is symptomatic infectious, R_U-cell is Undetected recovered, R_D-cell the detected recovered. The transmission rules of COVID-19 in cellular automata are the same as $SEI_UI_DR_UR_D$.

When cellular automata are used to simulate the transformation from the Infected (I_U *and* I_D) to Recovered (R_U *and* R_D), previous studies often generate a random number and make a comparison with the removal rate, γ. If the random number is less than γ, the infected cell will turn into recovered (1). However, it is not satisfactory for the real scene. Therefore, we introduce a time matrix to record the time of virus infection of each cell as defined in Equation (3).

TABLE 3 | The daily successful infection rate of Hebei province during the outbreak in January 2021.

Time (day)	$F(t)$	Time (day)	$F(t)$
5-Jan	0.60938	13-Jan	0.20057
6-Jan	0.58974	14-Jan	0.19598
7-Jan	0.44484	15-Jan	0.17769
8-Jan	0.32936	16-Jan	0.14816
9-Jan	0.30050	17-Jan	0.13243
10-Jan	0.29280	18-Jan	0.10498
11-Jan	0.24036	19-Jan	0.08197
12-Jan	0.21479	20-Jan	0.06205

$$T_{n\times n} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{bmatrix} \quad (3)$$

At first, all elements in the time matrix, $T_{n\times n}$, are set to 0. The state of the cell is 0 and there is a 0.5% probability of a cell turning into 1, which represents exposure at the beginning of the outbreak. The structure of CA will be defined as follows:

- The two dimensional lattice of square cells in an orthogonal grid. The size of the orthogonal grid is $n \times n$, a vector (i, j) represents the position of the cell in the grid.
- The size of the grid is theoretical infinity, but in the experiment, we set it as $n^2 = 300^2$. Each cell's neighborhood is composed of all its eight neighboring cells (the Moore neighborhood).
- Each cell has six states, we can picture 0 as the state of being susceptible to a given cell, 1 as the state of being exposed,

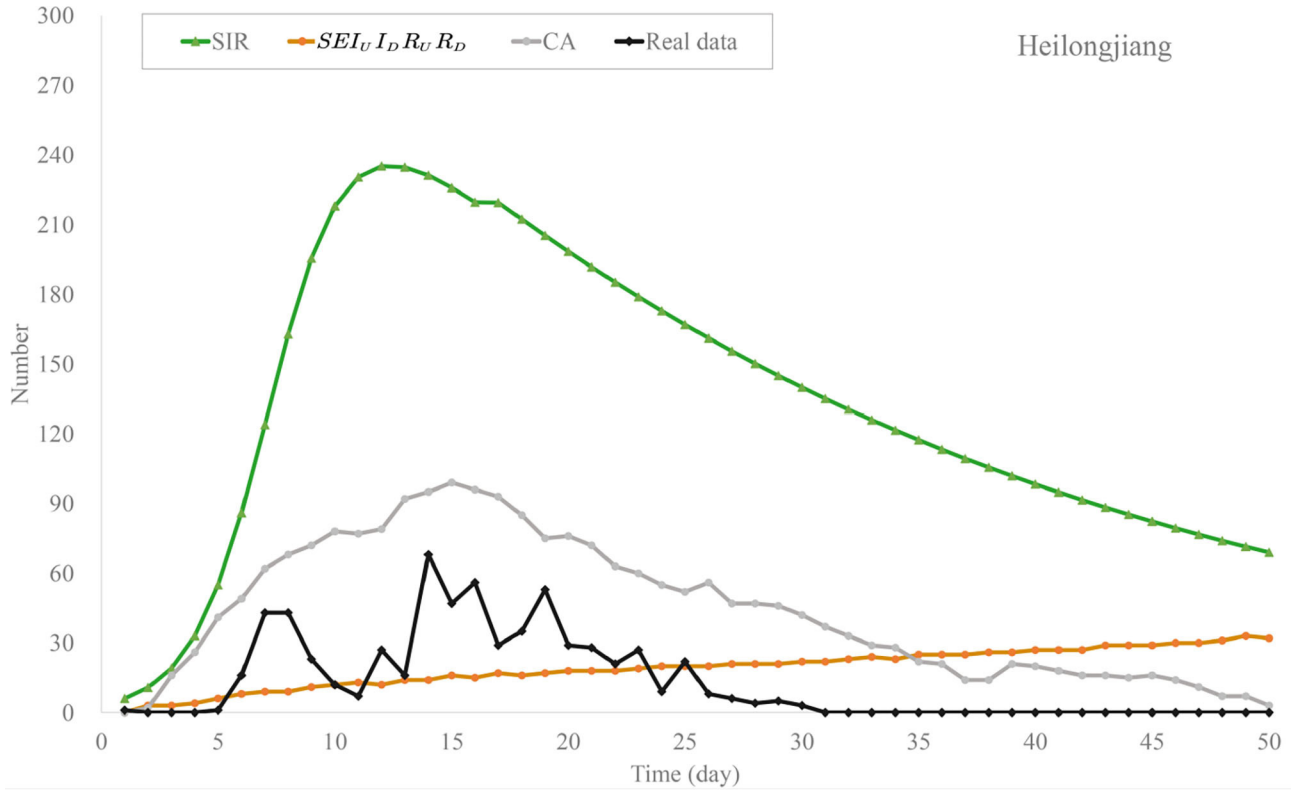


FIGURE 5 | Trend chart of daily confirmed cases of the models in Heilongjiang province.

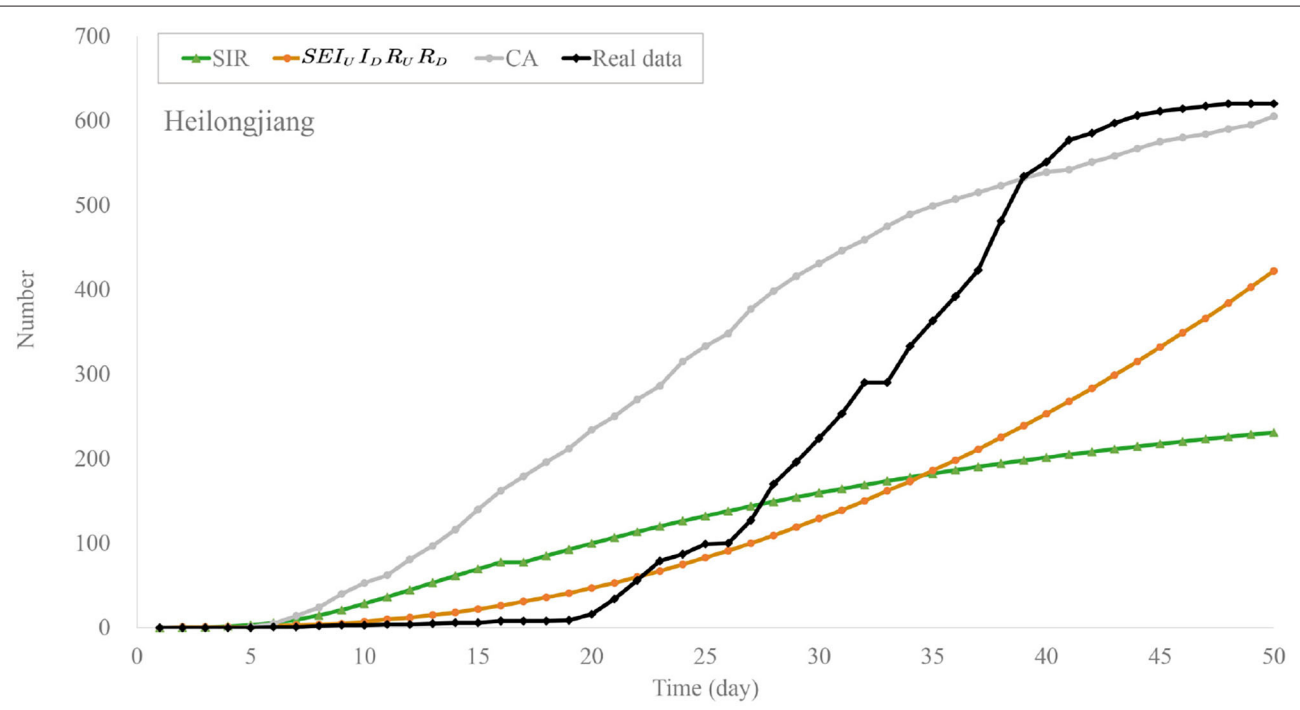


FIGURE 6 | Trend chart of recovered cases of the models in Heilongjiang province.

2 as being asymptomatic infectious (I_U), 3 as symptomatic infectious individuals (I_D) that are **not** being isolated, 4 as symptomatic infectious individuals (I_D) that are being isolated, last 5 and 6 as R_U and R_D, respectively.

- The time matrix T as defined in the previous section will record the time when cells turn into the Exposed. COVID-19's transition rule goes as follows. At each time step t exactly one of five things can happen to a cell.

The structure of CA is shown in **Figure 4**.

(a) Expose: If the cell state at $t-1$ was 0 (susceptible), the cell state has a possibility to become 1 (exposed) if any neighbors were 2 or 3 at $t-1$;

(b) Infect into I_U: If the cell state at $t-1$ was 1 (exposed), the cell has a chance of ϕ to become 2 if the corresponding number in time matrix, $T_{i,j}$ is greater than the average confirmed time;

(c) Infect into I_D: If the cell state at $t-1$ was 1 (exposed), the cell has a chance of $(1-\phi)\,\sigma$ to become 3 (I_D not under isolation) meantime a probability of $\boldsymbol{(1-\phi)\,(1-\sigma)}$ turn into 4 ($\boldsymbol{I_D}$ under isolation), if the corresponding number in time matrix, $\boldsymbol{T_{i,j}}$ is greater than average confirmed time;

(d) Recover: If the cell state at $\boldsymbol{t-1}$ was 2, 3, or 4, the cell state becomes 5 (recovered) if the corresponding element in the time matrix, $\boldsymbol{T_{i,j}}$ is greater than the duration of treatment.

(e) Stay: If the cell state and its corresponding number in $\boldsymbol{T_{i,j}}$ can not meet any of the transmission rules that were previously defined, the cell state remains the same, and the number in the time matrix, $\boldsymbol{T_{i,j}}$ will plus a random number from a normal distribution with mean 1 and variance 1.

TABLE 4 | MAE of the models on the outbreak data of Heilongjiang.

Model	MAE
SIR	122.00
SEIR	95.38
$SEI_UI_DR_UR_D$	37.05
CA without $T_{i,j}$	99.72
CA	30.37

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Data and Parameters Setting

Similar to previous studies (1, 26, 27), we obtained the COVID-19 epidemic data from the COVID-19 Data Repository managed by the local public health agency. The number of confirmed and recovered cases is updated once a day and includes all provinces of China. In this article, we use the data on COVID-19 in the Heilongjiang and Hebei provinces of China in January 2021 to conduct experiments. These data are shown in **Tables 1**, **2**.

According to the previous study (28), the value of the infection rate β in Equations (1) and (2) can be computed as follows:

$$\beta = k \times F \tag{4}$$

where F represents the number of people that a patient has close contact with. According to the data from the National Health Commission, during the first 15 days of the outbreak, the average number of people a patient has close contact with is 12 per day. Then with the implementation of restrictive measures, k drops to 5. The parameter F is the median of the time-dependent successful infection rate, $F(t)$. It can be described as follows:

$$F(t) = \frac{M_n(t)}{M_s(t)} \tag{5}$$

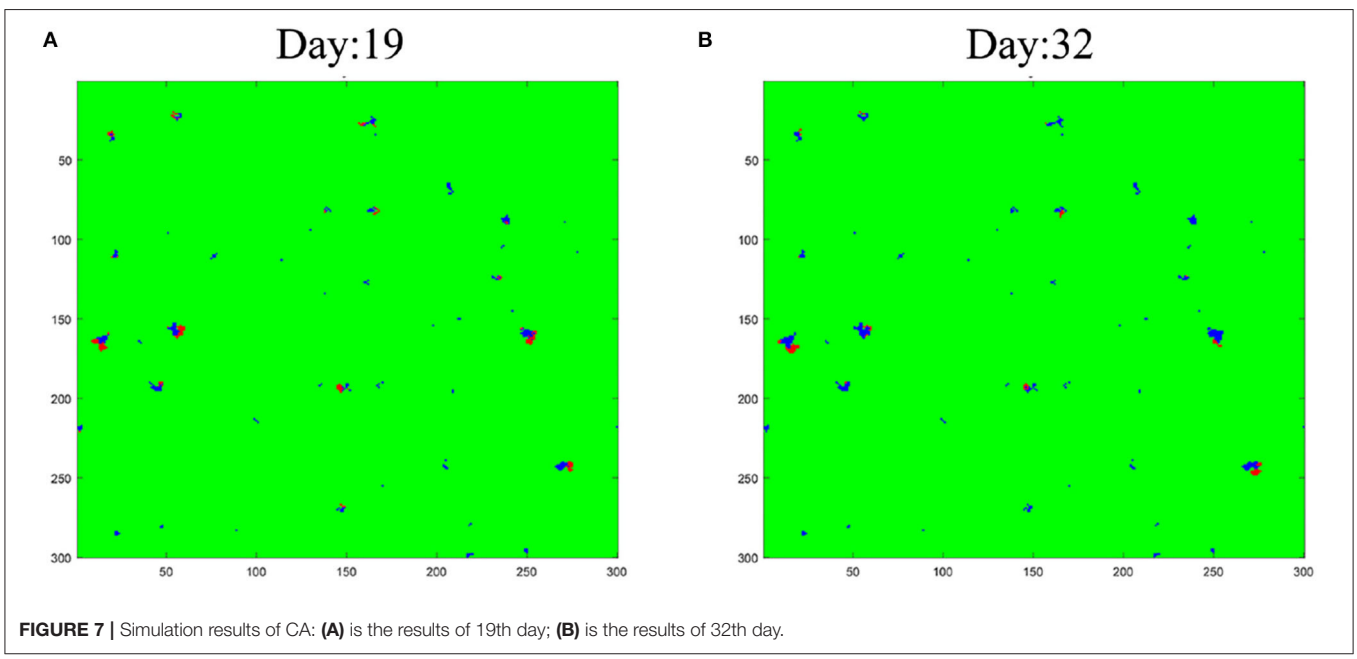


FIGURE 7 | Simulation results of CA: **(A)** is the results of 19th day; **(B)** is the results of 32th day.

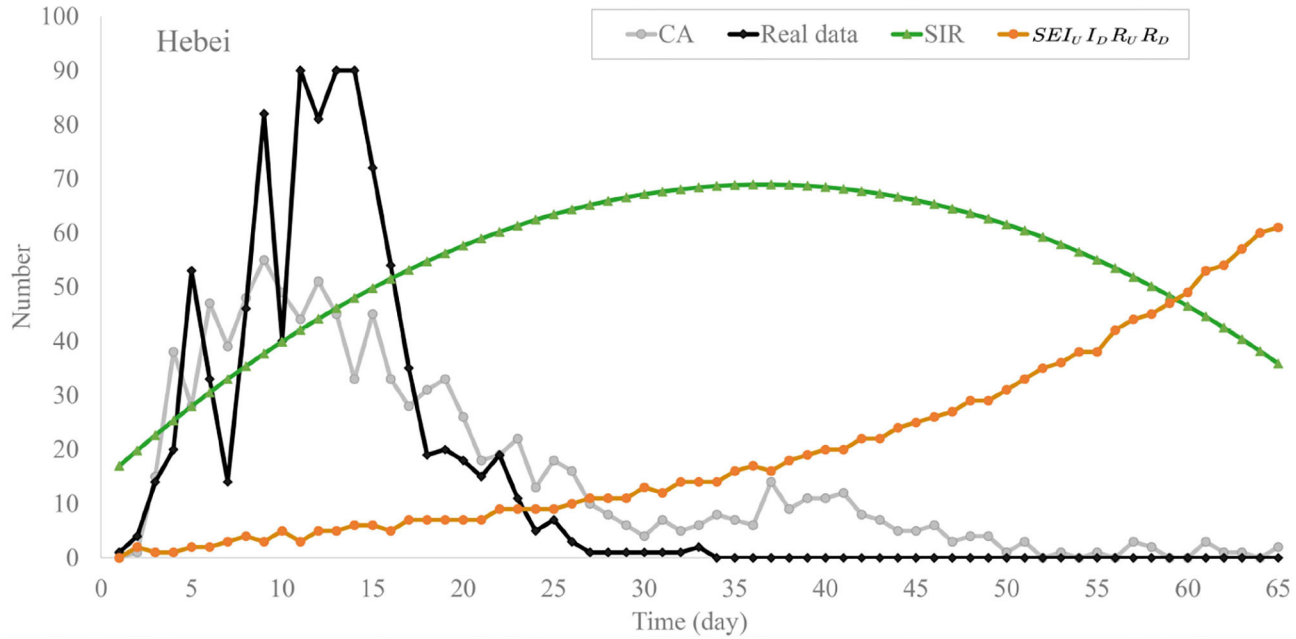


FIGURE 8 | Trend chart of daily confirmed cases of the models in Hebei province.

There is an incubation period between getting infected with the virus and being confirmed as infected. Some previous studies have performed a simulation of the incubation period of COVID-19 (28), and the result shows that the median incubation period of COVID-19 is 6 days. So, in this study, the parameters $M_n(t)$ and $M_s(t)$ represent the sum of daily confirmed cases and the sum of confirmed cases in 6 days preceding time t respectively. We calculate the $F(t)$ of each day during the outbreak using Equation (5) The parameter k is the median of $F(t)$.

According to the data collected by the National Health Commission of China, the daily successful infection rate, $F(t)$ is calculated by the data of the COVID-19 outbreak in Hebei Province in January 2021. The results are shown in **Table 3**.

The median of $F(t)$ is 0.054, so as the value of F. Finally, we can calculate the infection rate β:

$$\beta = k \times F = \begin{cases} 0.648 & (t \leq 15) \\ 0.270 & (t > 15) \end{cases} \tag{6}$$

According to the law of the PRC on the Prevention and Treatment of Infectious Diseases, the isolation will be immediately implemented once the individual is showing the symptoms of COVID-19, thus the non-isolation rate (σ) is set to 0 in outbreaks that take place in China. The percentage of the asymptomatic individual in all infections (ϕ) is a strongly debated aspect, the value of this parameter shows a great difference in outbreaks of COVID-19 that take place in different areas. This phenomenon may due to the definition of asymptomatic infectious has not reached an international agreement and obstacles to fully understanding the virus (29, 30). In this experiment, we choose 0.59 as the value of ϕ according to the prediction of a China medical team. Based on the recovery data of 364 patients in Mobile cabin hospital (13), the average treatment time G is 28.1 days. Thus, we take removal rate γ as a constant during the spread of disease and can be defined as $\gamma = \frac{1}{G} = \frac{1}{28.1}$. Now the value of parameters in SIR has all been set. As mentioned above, the virus has an incubation period of 6 days. In the SEIR model, the transmission rate from exposed to infected η is regarded as a constant and can be defined as $\eta = \frac{1}{6}$.

According to the previous study carried out by (31), the infection rate of close contacts is 0.04. During the outbreak, nucleic acid tests will be carried out every 4 days, therefore, the confirmed time of COVID-19 is set to 4 (9, 16).

In this experiment, we use Matlab to develop these models. Additionally, mean absolute error (MAE) is used to evaluate the performance of these models.

4.2. Results and Analysis

4.2.1. Small-Scale Outbreaks in China

We predict the outbreaks of COVID-19 in Heilongjiang and Hebei provinces in January 2021. First, **Figures 5**, **6** show the predicted numbers of the confirmed and recovered cases for Heilongjiang province. These results indicate that CA has a better performance in simulating the outbreak of COVID-19 in China nowadays. Classical compartmental models may no longer be suitable for modeling small-scale outbreaks in China. The MAE values of these methods are shown in **Table 4**. From these compared results, it can be observed that the proposed CA has the smallest errors. In addition, compared with the traditional CA, the MAE value of the CA model with time matrix is 30.37, which is reduced by 66.35.

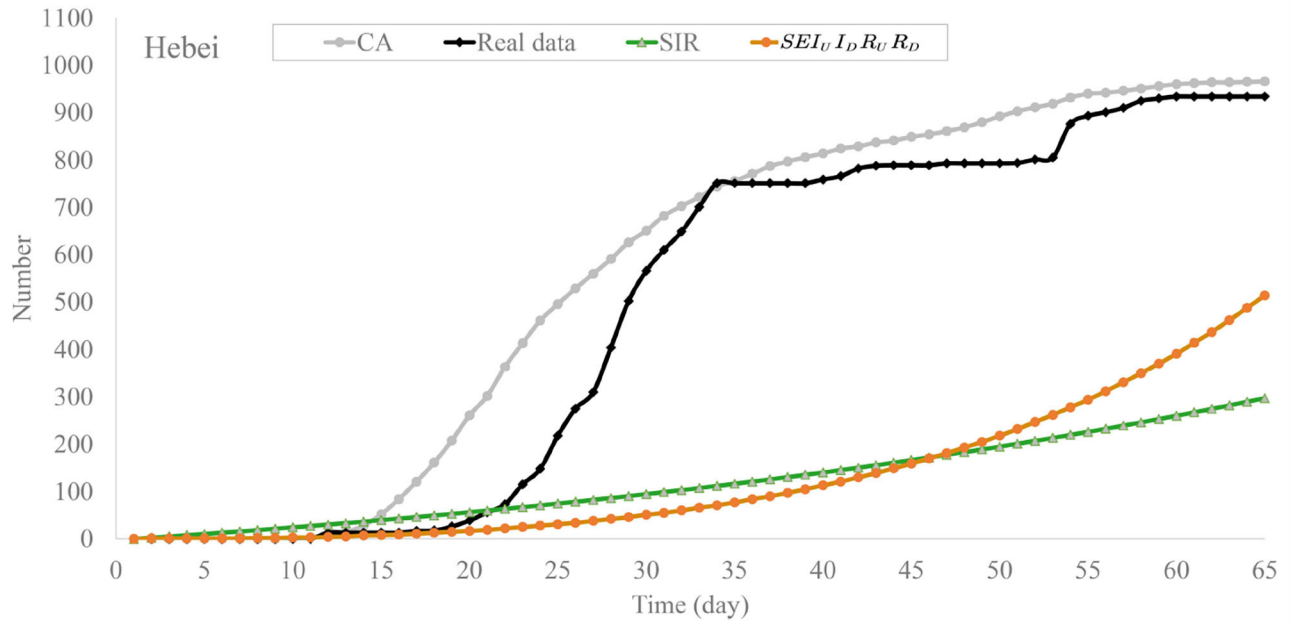


FIGURE 9 | Trend chart of recovered cases of the models in Hebei province.

We further visualize the simulation results of CA on the 19th day and the 32th day of Heilongjiang province. The red spots represent I_D-cell, and the blue is R_D-cell. In **Figure 7A**, there is only a marginal amount of R_D-cell and some red spots. In **Figure 7B**, one can also see that most red spots have turned blue which means that the outbreak is coming to an end. This is consistent with the actual situation of the outbreak in China. The transmission of COVID-19 can get under control within a month. More results about CA can be found in the **Supplementary Materials**.

Once again, we analyzed the data of outbreaks in Hebei province in January 2021 and evacuated these models. The experimental results are shown in **Figures 8**, **9**. In **Figure 8**, the daily confirmed cases of the SIR model grew rapidly and reached 350 on the 100th day, which makes it deviate from official data. Compared with SIR, the SEIR model has a significant improvement in the fitting, the daily confirmed cases of $SEI_UI_DR_UR_D$ slowly rose to 10 person at 35th day and climbed steadily to 60 at the end of outbreak. Compared with the first two models, the cellular automata can fit in the data with high accuracy. The difference between the maximum time of cellular automata and real data is about 5 days, and the trends of the two curves are roughly the same.

In **Figure 9**, the recovered cases of SIR keep increasing and reach about 270 on the 65th day. The recovered cases of $SEI_UI_DR_UR_D$ remained at a low value during the former part of outbreak, and grow from around 70 person at 35th day to 500 person at 65th day. We can see that recovered cases of CA are close to 0 from 0 to 18 days then grow dramatically to 800 in the 45th day and slowly climb to around 1,000 in the remaining time. The predicted results of CA are close to the official data.

TABLE 5 | MAE of the models on the outbreak data of Hebei.

Model	MAE
SIR	47.98
SEIR	44.34
$SEI_UI_DR_UR_D$	30.35
CA without $T_{i,j}$	40.93
CA	9.43

The MAE values of these methods of the outbreak in Hebei province are shown in **Table 5**. The MAE of SIR, SEIR, $SEI_UI_DR_UR_D$, CA without $T_{i,j}$ and CA are 47.98, 44.34, 30.35, 40.93, and 9.43, respectively. It is clear that the MAE of SIR and $SEI_UI_DR_UR_D$ is more than 4 times CA, they performed poorly in both long-term and short-term fitting. These differential models based on compartments may not be suitable for fitting the small-scale outbreaks.

4.2.2. The Outbreak in Potter County Texas US

Furthermore, we performed our CA in the simulation of the small-scale outbreak in Potter County, to test its reliability in a different country. Unfortunately, the databases of Potter County health department just maintained confirmed cases and death cases, and the collection of recovered cases has been stopped since April 2021, which means that the, M_s (t), sum of confirmed cases in 6 days preceding time t are unknown. Because daily recovered cases are not in the record, which makes the existing confirmed cases of each day stay unclear. As a consequence, Equation (4), which defined to calculate the infection rate β, is

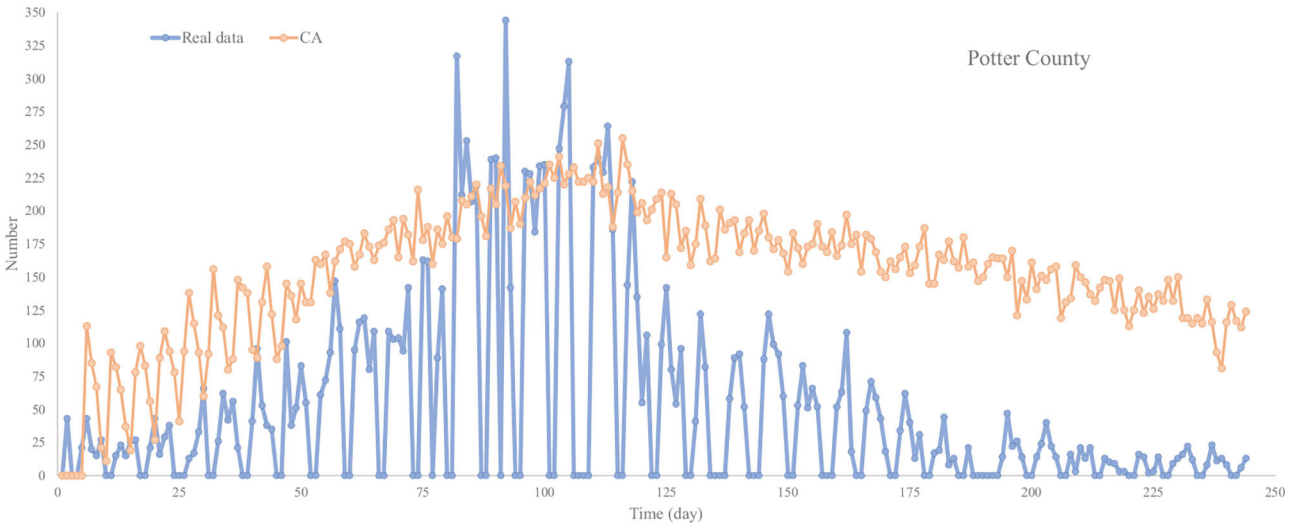


FIGURE 10 | Trend chart of daily confirmed cases of CA in Potter County.

TABLE 6 | MAE for models in Potter County.

Model	MAE
SIR	201.36
SEIR	92.23
$SEI_UI_DR_UR_D$	96.87
CA	79.25

out of work. In that case, we refer to the previous study and determine the value of infection rate, β, is 0.4428. According to the Centers for Disease Control and Prevention (CDC), the current best estimated average time from exposure to symptom onset is 6 days. Therefore, parameter $\eta = 1/6$. The percentage of asymptomatic infections in the US is 30% ($\phi = 0.3$). The non-isolation rate of the symptomatic infectious individual (σ) is 0.5. The estimation of the average treatment time is 24.7 (32) $\left(\gamma = \frac{1}{G} = \frac{1}{24.7}\right)$.

The result of daily new cases can be seen in **Figure 10**. **Table 6** shows the MAE for each model in Potter County. The **Figure 10** gives a breakdown of the trend of CA and real data. The real data of daily cases rise dramatically during the former part of the outbreak and reach the beak in around 92 days with 350 new cases. CA performs well in the former part; it also reaches a peak of 250 new cases at around the same time as real data. However, the daily new cases decrease severely to around 100 shortly after the peak, and slowly down to 0 during the later part of the outbreak. The result of CA in the later part of the outbreak is relatively unsatisfactory. It falls to simulate the sharp decrease and there are still around 100 new cases at the end of the outbreak. But on the bright side of our model, it simulates the former part of the outbreak in Potter County with relatively high accuracy, which means that our CA can roughly simulate the trend in small-scale outbreaks outside China.

5. DISCUSSION

With appropriate parameters and rules, compared with SIR, SEIR, and $SEI_UI_DR_UR_D$, our CA can simulate the small-scale outbreaks of COVID-19 in nowadays China more effectively. The MAE of CA in the outbreak that took place in Hebei reached a value of 9.43, it provides valuable information about the decision on medical policy. Classic compartmental models have been widely used in modeling the transmission dynamics with numerous infected cases, and have gained great success (7, 33). One major drawback of those compartmental models is the hiking of the number and complexity of parameters (6). The parameters of these models had to be more precise and complex to achieve better performance (26). Although many researchers hold the belief that non-linearities in CA alongside ABM destroyed any attempt to use the predicatively, they are oversimplified from realistic words (34, 35). However, this study has proved that small-scale outbreaks can be modeled through a relatively simple abstract model.

5.1. Strengths and Weaknesses

In this contribution, we proposed an improved CA to carry out experiments that introduced a time matrix to have a precise simulation of the outbreak. As the results shown, CA simulated the outbreak accurately which suggests that researchers can consider using it to study the current epidemics in China. However, our rules and parameters of CA are far from perfect. There have been numerous methods to estimate the value of parameters (9, 11, 14). It is true that all parameters are set according to the best simulation rest from the local health department or CDC in the real world situation. But they may not be the perfect values we need according to the structure of our model. During the experiments, we use a time matrix to record the time of virus infection of each cell. Only if the $t_{(i,j)}$ is greater

than the average treatment time G shall the cell state turn into 3, which means that a patient can only recover from COVID-19 after 28 days of treatment. However, this assumption is an oversimplification, as young patients may get recovered before 28 days of treatment, while the aged typically need more time during treatment (9).

5.2. Further Study

In further study, the recovered rate γ will no longer be regarded as a constant in CA. At different times of treatment, the recovery rate will be different. Combined with the time matrix, the transition rules of COVID-19 in CA will be updated. In addition, the structure of CA is a 300 square static orthogonal matrix. Each cell is adjacent to 8 others. In further study, the best adjacent number is needed to be determined, since each cell may have interactions with 4, 6, or more neighbors. In the real world, the number of people an individual have contact with is different from each day, as a result, the cell in the CA may intricate with a different number of neighbor each day at further study. Another major improvement in the future is that we will change our CA into an ABM model. Because ABM is developed from CA, they share many similar characters (36, 37). The existing platforms for ABM are fundamentally helpful when setting up transform rules (38–40).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

Z-YW: methodology and writing—original draft preparation. H-BZ and H-FZ: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the National Key Research and Development Program of China (No. 2019YFC1604700) and the Promotion Program for Young and Middle-Aged Teachers in Science and Technology Research of Huaqiao University (ZQN-YX601).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.907814/full#supplementary-material

REFERENCES

1. Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv:2004.00958*. (2020). doi: 10.2196/preprints.19406
2. Bjornstad ON, Shea K, Krzywinski M, Altman N. The SEIRS model for infectious disease dynamics. *Nat Methods*. (2020) 17:557–8. doi: 10.1038/s41592-020-0856-2
3. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals*. (2020) 139:110057. doi: 10.1016/j.chaos.2020.110057
4. Din RU, Algehyne EA. Mathematical analysis of COVID-19 by using SIR model with convex incidence rate. *Results Phys*. (2021) 23:103970. doi: 10.1016/j.rinp.2021.103970
5. Yang CY, Wang J. A mathematical model for the novel coronavirus epidemic in Wuhan, China. *Math Biosci Eng*. (2020) 17:2708–24. doi: 10.3934/mbe.2020148
6. Kronbichler A, Kresse D, Yoon S, Lee KH, Effenberger M, Shin JI. Asymptomatic patients as a source of COVID-19 infections: a systematic review and meta-analysis. *Int J Infect Dis*. (2020) 98:180–6. doi: 10.1016/j.ijid.2020.06.052
7. Hethcote HW, Lewis MA, Driessche P. An epidemiological model with a delay and a nonlinear incidence rate. *J Math Biol*. (1989) 27:49–64. doi: 10.1007/BF00276080
8. Gao BJ, Zhang T, Xuan HY, Yang J. A heterogeneous cellular automata model for SARS transmission. *J Syst Manage*. (2006) 15:205–9.
9. Yuan X, Chen S, Yuwen L, An S, Mei S, Chen T. An improved SEIR model for reconstructing the dynamic transmission of COVID-19. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Seoul, Korea (2020). p. 2320–7. doi: 10.1109/BIBM49941.2020.9312981
10. Tan XH, Liu Q, Jian-Feng HE. A study on the transmission dynamic model of SARS in Guangdong province. *Chin J Dis Control Prevent*. (2006) 10:560–3. doi: 10.3969/j.issn.1674-3679.2006.06.004
11. Wu Z. Asymptomatic and pre-symptomatic cases of COVID-19 contribution to spreading the epidemic and need for targeted control strategies. *Zhonghua liuxingbingxue zazhi*. (2020) 41:E036. doi: 10.1186/s40249-020-00679-2
12. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, et al. Risk for transportation of coronavirus disease from Wuhan to other cities in China. *Emerg Infect Dis*. (2020) 26:1049–52. doi: 10.3201/eid2605.200146
13. Zhixin W, Zhi L, Zhaojun L. COVID-19 analysis and forecast based on machine learning. *J Biomed Eng Res*. (2020) 39:1–5. doi: 10.19529/j.cnki.1672-6278.2020.01.01
14. Yu G, Yu W, Jianyong W. Epidemics trend prediction model of COVID-19. *CAAI Trans Intell Syst*. (2021) 2021:528–36. doi: 10.11992/tis.202008037
15. Stefano BD. Objectoriented implementation of CA/LGCA modelling applied to the spread of epidemics. In: *Proceedings of Canadian Conference on Electrical and Computer Engineering*. Halifax, NS, Canada (2000). p. 2–6.
16. Wolfram S. Computation theory of cellular automata. *Commun Math Phys*. (1984) 96:15–57. doi: 10.1007/BF01217347
17. Guang-Liang LI, Wen LH, Yan JX. Research on the spread of infectious diseases based on constraints. *J Yangtze Univers*. (2013) 10:85–7+6. doi: 10.16772/j.cnki.1673-1409.2013.25.037
18. Xue BI, Hong-Bo SU, Guo JH, Huang DS, Guan P, University CM. Cellular automata and its application in the field of infectious disease prevention and control. *Pract Prev Med*. (2015). 22:766–9. doi: 10.3969/j.issn.1006-3110.2015.06.044
19. Guan P, Xue BI, Fei LR, Huang DS, Liu L. Simulation of the epidemic of influenza A(H1N1) in a University using cellular automata model. *Chin J Infect Control*. (2016). 15:79–82. doi: 10.3969/j.issn.1671-9638.2016.02.002
20. Yu J, Song Y, Fei F, Weiqiang S, Song Z. Analysis of COVID-19 in Hubei province based on machine learning and dynamics model. *J Biomed Eng Res*. (2020) 39:294–9,305. doi: 10.19529/j.cnki.1672-6278.2020.03.13
21. You AL, Yan P. Swine influenza model based on cellular automata. *J Xinjiang Univ*. (2010). 27:56–9. doi: 10.3969/j.issn.1000-2839.2010.01.010
22. Du Z, Wang L, Pandey A, Lim WW, Chinazzi M, Piontti APy, et al. Modeling comparative cost-effectiveness of SARS-CoV-2 vaccine dose fractionation in India. *Nat Med*. (2022) 28:934–8. doi: 10.1038/s41591-022-01736-z
23. Du Z, Wang L, Bai Y, Wang X, Pandey A, Fitzpatrick MC, et al. Cost-effective proactive testing strategies during COVID-19 mass

vaccination: a modelling study. *Lancet Region Health Am*. (2022) 8:100182. doi: 10.1016/j.lana.2021.100182
24. Silva PCL, Batista PVC, Lima HS, Alves MA, Guimaraes FG, Silva RCP. COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos Solitons Fractals*. (2020) 139:110088. doi: 10.1016/j.chaos.2020.110088
25. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol*. (2021). 17:e10 09149. doi: 10.1371/journal.pcbi.1009149
26. He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn*. (2020) 101:1667–80. doi: 10.1007/s11071-020-05743-y
27. Annas S, Pratama MI, Rifandi M, Sanusi W, Side S. Stability Analysis and Numerical Simulation of SEIR Model for pandemic COVID-19 spread in Indonesia. *Chaos Solitons Fractals*. (2020) 139:110072. doi: 10.1016/j.chaos.2020.110072
28. Huang DS, Guan P. Research on fitting of SIR model on prevalence of SARS in Beijing city. *Chin J Dis Control Prev*. (2004). 5:398–401. doi: 10.3969/j.issn.1674-3679.2004.05.005
29. Qiu J. Covert coronavirus infections could be seeding new outbreaks. *Nature*. (2020). doi: 10.1038/d41586-020-00822-x
30. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. *J Med Virol*. (2021). 93:820–30. doi: 10.1002/jmv.26326
31. Chen Y, Wang A, Yi B, Ding K, Wang H, Wang J, et al. The epidemiological characteristics of infection in close contacts of COVID-19 in Ningbo city. *Chin J Epidemiol*. (2020) 41:E026. doi: 10.3760/cma.j.cn112338-20200304-0 0251
32. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. (2020) 20:669–77. doi: 10.1016/S1473-3099(20)3 0243-7
33. Mwalili S, Kimathi M, Ojiambo V, Gathungu D, Mbogo R. SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Res Notes*. (2020) 13:352. doi: 10.1186/s13104-020-0 5192-1
34. Ferrin D, Miller M, McBroom D. Maximizing hospital financials impact and emergency department throughput with simulation. In: *Proceedings of the Winter Simulation Conference, WSC 2007*. Washington, DC (2007). p. 1566–73. doi: 10.1109/WSC.2007.4419774
35. Squazzoni F, Polhill JG, Edmonds B, Ahrweiler P, Antosz P, Scholz G, et al. Computational models that matter during a global pandemic outbreak: a call to action. *J Artif Soc Soc Simul*. (2020) 23:10. doi: 10.18564/jasss.4298
36. Parry HR, Evans AJ. A comparative analysis of parallel processing and super-individual methods for improving the computational performance of a large individual-based model. *Ecol Modell*. (2008) 214:141–52. doi: 10.1016/j.ecolmodel.2008.02.002
37. Parry HR, Evans AJ, Morgan D. Aphid population response to agricultural landscape change: a spatially explicit, individual-based model. *Ecol Modell*. (2006) 199:451–63. doi: 10.1016/j.ecolmodel.2006.01.006
38. Monserud RA, Leemans R. Comparing global vegetation maps with the Kappa statistic. *Ecol Modell*. (1992) 62:275–93. doi: 10.1016/0304-3800(92)90003-W
39. Hagen-Zanker A. An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. *Int J Geograph Inform Sci*. (2009) 23:61–73. doi: 10.1080/13658810802570317
40. Heppenstall AJ, See LM, Crooks AT. *Agent-Based Models of Geographical Systems*. Dordrecht, Heidelberg, London, New York: Springer (2012). doi: 10.1007/978-90-481-8927-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 01 August 2022
doi: 10.3389/fmed.2022.962101

Influence of Co-morbidities During SARS-CoV-2 Infection in an Indian Population

Adrian Matysek[1,2], Aneta Studnicka[3], Wade Menpes Smith[1,2], Michał Hutny[4], Paweł Gajewski[5], Krzysztof J. Filipiak[6], Jorming Goh[7,8,9] and Guang Yang[10,11]*

[1] Immunidex Ltd., London, United Kingdom, [2] Cognescence Ltd., London, United Kingdom, [3] Clinical Analysis Laboratory, Silesian Centre for Heart Diseases, Zabrze, Poland, [4] Faculty of Medical Sciences in Katowice, Students' Scientific Society, Medical University of Silesia, Katowice, Poland, [5] AGH University of Science and Technology, Krakow, Poland, [6] Maria Skłodowska-Curie Medical Academy, Warsaw, Poland, [7] Healthy Longevity Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [8] Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [9] National University Health System (NUHS), Centre for Healthy Longevity, Singapore, Singapore, [10] Cardiovascular Research Centre, Royal Brompton Hospital, London, United Kingdom, [11] National Heart and Lung Institute, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:
Khalid Hattaf, Centre Régional des Métiers de l'Education et de la Formation, Morocco

Reviewed by:
Ahmed Mohsen, University of Baghdad, Iraq
Ahmed Elaiw, King Abdulaziz University, Saudi Arabia
Jun Liu, Carnegie Mellon University, United States

***Correspondence:**
Adrian Matysek
adrian@cognescence.com

Specialty section:
This article was submitted to Infectious Diseases – Surveillance, Prevention and Treatment, a section of the journal Frontiers in Medicine

Received: *05 June 2022*
Accepted: *23 June 2022*
Published: *01 August 2022*

Citation:
Matysek A, Studnicka A, Smith WM, Hutny M, Gajewski P, Filipiak KJ, Goh J and Yang G (2022) Influence of Co-morbidities During SARS-CoV-2 Infection in an Indian Population. Front. Med. 9:962101. doi: 10.3389/fmed.2022.962101

Background: Since the outbreak of COVID-19 pandemic the interindividual variability in the course of the disease has been reported, indicating a wide range of factors influencing it. Factors which were the most often associated with increased COVID-19 severity include higher age, obesity and diabetes. The influence of cytokine storm is complex, reflecting the complexity of the immunological processes triggered by SARS-CoV-2 infection. A modern challenge such as a worldwide pandemic requires modern solutions, which in this case is harnessing the machine learning for the purpose of analysing the differences in the clinical properties of the populations affected by the disease, followed by grading its significance, consequently leading to creation of tool applicable for assessing the individual risk of SARS-CoV-2 infection.

Methods: Biochemical and morphological parameters values of 5,000 patients (Curisin Healthcare (India) were gathered and used for calculation of eGFR, SII index and N/L ratio. Spearman's rank correlation coefficient formula was used for assessment of correlations between each of the features in the population and the presence of the SARS-CoV-2 infection. Feature importance was evaluated by fitting a Random Forest machine learning model to the data and examining their predictive value. Its accuracy was measured as the F1 Score.

Results: The parameters which showed the highest correlation coefficient were age, random serum glucose, serum urea, gender and serum cholesterol, whereas the highest inverse correlation coefficient was assessed for alanine transaminase, red blood cells count and serum creatinine. The accuracy of created model for differentiating positive from negative SARS-CoV-2 cases was 97%. Features of highest importance were age, alanine transaminase, random serum glucose and red blood cells count.

Conclusion: The current analysis indicates a number of parameters available for a routine screening in clinical setting. It also presents a tool created on the basis of these parameters, useful for assessing the individual risk of developing COVID-19 in patients. The limitation of the study is the demographic specificity of the studied population, which might restrict its general applicability.

Keywords: SARS-CoV-2, blood biomarkers, COVID-19, machine learning, vitamin D

INTRODUCTION

The severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) that causes the coronavirus disease 2019 (COVID-19) is a serious threat to human health and life. Due to the ease of spread and mutation of the virus, the World Health Organization (WHO) has declared COVID-19 a pandemic. Viral infection may be asymptomatic or symptomatic with varying degrees of severity. In some cases, SARS-CoV-2 infection leads to death of the patient (1).

The course of the SARS-CoV-2 infection may be influenced by several factors, including the presence of comorbidities in patients. Comorbidities, such as type 2 diabetes, affect the host immune response, which may be associated with a severe course of SARS-CoV-2 infection. In the context of type 2 diabetes, there is an increased release of pro-inflammatory cytokines, which can lead to cytokine storms in SARS-CoV-2 (+) patients. The occurrence of a cytokine storm correlates with a worse course of infection (1–4). Moreover, some changes in blood parameters have been observed in patients with SARS-CoV-2 (+), which may additionally affect the severity of the infection. Changes for example in the values of liver and kidney parameters, morphology and inflammatory markers in SARS-CoV-2 (+) patients may indicate a complex mechanism of infection and its long-term consequences (5).

Considering the wide disparity in the course of SARS-CoV-2 infection in patients and the need to define an effective therapy, the primary aim of our study was to define the physiological characteristics in patients infected with the SARS-CoV-2 virus, while presenting with concomitant diseases. The secondary aim of our study was to elucidate the underlying mechanisms that influence multiorgan dysfunction in COVD-19 infection. The contributing physiological parameters were ranked in terms of their relevance to the power of predictive model using machine learning.

MATERIALS AND METHODS

Dataset

The analysed dataset consisted of blood test results from 5,000 patients from a digital healthcare system Curisin Healthcare (India). Two thousand five hundred of them were from patients (female, n = 1667; 67% and male, n = 833; 33%, age 25 – 78 years, mean = 51.6 years, median = 51.0 years) hospitalised due to infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); the other two thousand and five hundred blood test results belonged to the control group (female, n = 755; 30%, male, n = 1,745; 70%; age 20 – 55, mean = 30.5 years, median = 30.0 years) and were collected before the coronavirus pandemic outbreak (2016-2018).

In blood biomarkers and parameters were excluded ones that are known to correlate with inflammatory responses such as C-reactive protein (CRP). Biomarkers that were not present in both datasets were also ruled out. The resulting set of data features enrolled in this study consisted of anthropometric parameters: age, gender; biochemical serum parameters: random glucose (RG), urea, alanine transaminase (ALT), cholesterol, creatinine, vitamin D; morphological parameters: red blood cell count (RBC Count); and functional parameters: estimated glomerular filtration rate (eGFR).

Glomerular filtration rate was not directly measured in the blood samples, instead the eGFR score was estimated using patient's age, gender and serum creatinine level according to the simplified modification of diet in renal disease (MDRD) formula, described by the National Kidney Foundation (6), listed in Equation 1. below.

Equation 1. GFR definition.

$$\mathrm{GFR} = \begin{cases} \mathrm{C}^{-1.154}\ x\ 186.3\ x\ \mathrm{A}^{-0.203}\ x\ 0.742, & G = Female \\ \mathrm{C}^{-1.154}\ x\ 186.3\ x\ \mathrm{A}^{-0.203}, & G = Male \end{cases}$$

C, creatinine; A, age; G, gender.

The reference ranges for the analysed parameters are listed below in **Table 1**.

Correlations

The correlations between each of the features in the population and the presence of the SARS-CoV-2 infection were calculated using Spearman's method (see **Equation 3**).

TABLE 1 | Reference ranges of the analysed parameters.

Property	Reference Range
RG	74 – 100 mg/dl
Serum Urea	13 – 40 mg/dl
RBC Count	3.8 – 4.8 million/cumm
ALT	$\leq$ 34 U/L
Cholesterol	$\leq$ 200 mg/dl
Creatinine	0.6 – 1.1 mg/dl
Vitamin D	30 – 100 ng/dl

ALT, alanine transaminase; RBC Count, red blood cells count; RG, random serum glucose.

Equation 2. Pearson's correlation.

$$\text{for } x \in X, \quad y \in Y$$

$$\rho(x, y) = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sigma x * \sigma y}$$

$\bar{x}$, mean of X; $\bar{y}$, mean of Y; σx, standard deviation of X; σy, standard deviation of Y.

Equation 3. Spearman's correlation.

$$\text{for } x \in X, \quad y \in Y$$

$$r_s(x, y) = \rho_{R(X),R(Y)} = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}}$$

ρ, Pearson's correlation (**Equation 2**) applied to rank variables; cov (R(X), R(Y)), covariance of rank variables; $\sigma_{R(X)}$, standard deviation of rank variable of X; $\sigma_{(R(Y)}$, standard deviation of rank variable of Y.

Feature Importance

To discover how each feature affects the possibility of severe SARS-CoV-2 infection, the importance of each feature was evaluated by fitting a Random Forest (7) machine learning model to the data and examining which features had the highest predictive value. The accuracy measure of the predictive model was the F1 Score as defined by **Equation 4**. Results of this analysis were presented graphically in section "Results" in **Figure 2**.

Equation 4. F1 Score measuring the accuracy of the predictive model.

$$F_1 = \frac{tp}{tp + \frac{1}{2}(fp + fn)}$$

tp, number of true positives; fp, number of false positives; fn, number of false negatives.

Vitamin D in Relation to NL Ratio and SII

Additional analysis of the possible correlation between the concentration of vitamin D in SARS-CoV-2 (+) patients and the values of parameters such as Neutrophil-Lymphocyte Ratio (NL Ratio) and Systemic immune-inflammation index (SII) was conducted in our study. For the purpose of the analysis the patients were divided into tertiles in terms of Vitamin D level: I tertial (I T): Vit D (1.11 - 9.40], II tertial (II T): Vit D (9.40 - 17.69], III tertial (III T): Vit D (17.69 - 25.97]; followed by assessment of the NL Ratio values and of the systemic immune inflammation index, which was calculated as defined in **Equation 5**.

Equation 5. Calculation of systemic immune inflammation index.

$$SII = \frac{P \text{ x } N}{L}$$

P, N, and L represent platelet, neutrophil and lymphocyte counts in cells/L.

RESULTS

Results of biochemical examinations performed on the material collected from both SARS-CoV-2 (+) and SARS-CoV-2 (−) groups are presented below in **Table 2**. as the number of patients with results falling into the respective groups created in relation to the reference range (below, above or in reference range).

The percentage of patients in each reference-range-related group in terms of serum vitamin D levels is presented in form of a circle chart for both infection-related groups of patients in **Figure 1**.

To allow a more profound insight into the distribution of parameters values, the number of patients in relation to the exact values of parameters are presented as the histograms for each respective parameter in the **Supplementary Materials** in **Supplementary Figure 1**.

The correlations between the studied parameters and SARS-CoV-2 infection in patients were calculated as described in Section "Materials and Methods." It must be noted that due to simultaneous influences of each parameter, none of them may offer sufficient reliability in terms of prediction of SARS-CoV-2 infection. Results of correlations analysis are listed below in **Table 3**.

To further examine the influence of respective features on infection severity in relation to the complexity of simultaneous influences of studied features, the predictive model was created as described in Section "Feature Importance" and presented below in **Figure 2**. The model itself was a successful one, with an

TABLE 2 | Patients' outcomes in SARS-CoV-2 (–) (n = 2,500) and SARS-CoV-2 (+) group (*n* = 2,500).

Property	Below reference range N;%		Above reference range N;%		Within reference range N;%	
	(–)	(+)	(–)	(+)	(–)	(+)
Random Glucose	286; 11	5; ∼0	780; 31	2162; 86	1434; 57	333; 13
Blood Urea	995; 40	0; 0	0; 0	88; 4	1505; 60	2412; 96
RBC Count	34; 1	651; 26	1707; 68	560; 22	759; 30	1289; 52
ALT	–	–	1384; 55	0; 0	1116; 45	2500; 100
Cholesterol	–	–	799; 32	1414; 57	1701; 68	1086; 43
Creatinine	123; 5	0; 0	541; 21	0; 0	1836; 74	2500; 100
Vitamin D	2324; 93	2500; 100	3; ∼0	0; 0	173; 7	0; 0

ALT, alanine transaminase; RBC Count, red blood cells count.

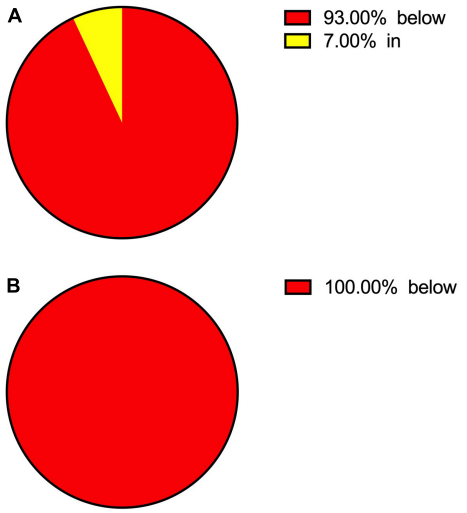


FIGURE 1 | Patients' vitamin D status in: **(A)** SARS-CoV-2 (–) group; **(B)** SARS-CoV-2 (+) group.

TABLE 3 | Correlations between features and SARS-CoV-2 features.

Property	Correlation with SARS-CoV-2
Age	0.671854
Gender	0.364978
RG	0.652085
Serum Urea	0.551667
RBC Count	−0.591819
ALT	−0.662445
Serum Cholesterol	0.255992
Serum Creatinine	−0.368583
Vitamin D	−0.148394
eGFR	−0.067112
Neutrophils	0.229344
Lymphocytes	−0.087234
NL Ratio	0.138331
Platelet Count	0.020377
SII	0.103531

ALT, alanine transaminase; eGFR, estimated glomerular filtration rate; NL Ratio, Neutrophil-lymphocyte ratio; RBC Count, red blood cells count; RG, random serum glucose; SII, Systemic immune-inflammation index.

accuracy of 97% in differentiating positive from negative cases of SARS-CoV-2 infection.

The results of analysis of NL Ratio values structure in relation to the vitamin D levels are presented, respectively, below in **Table 4** and **Figure 3**, whereas relation between SII and vitamin D levels in patients is summarised in **Table 5**.

DISCUSSION

The collected data indicate that the course of SARS-CoV-2 infection may depend on numerous factors. In the present study, the influence of gender, age and changes in blood parameters values of Indian patients were addressed and examined.

Age

A strong correlation between age and severe course of SARS-CoV-2 (0.671007) was observed, which suggests that the risk of developing severe SARS-CoV-2 increases with age. This observation was already reported in the early meta-analyses concerning risk factors for mortality in the course of COVID-19 infection (8). A similar association was also found in study (9) conducted on a group of patients from the New York City metropolitan area and in another study involving data from two Wuhan hospitals – results of these studies showed that age correlates with the need for hospitalization and severity of SARS-CoV-2 infection (10). This dependence is determined by several factors that characterize the ageing process - the percentage of diagnosed comorbidities in those patients, i.e., diabetes and cardiovascular diseases, increases with age. Ageing is also naturally associated with the decreasing functioning of the immune system, which in turn leads to changes in adaptive and innate immunity (11, 12).

Gender

In the presented analysis of the influence of the patient's gender during SARS-CoV-2, a significant correlation between adverse prognosis and female gender was observed, contrary to the results of a previous report (13) that reported higher mortality among men (8). The differences in the course of SARS-CoV-2 infection among women and men may result from the influence of social and biological factors in the latter, e.g., higher expression of angiotensin-converting enzyme 2 (ACE2) found in male patients, which may facilitate entry of the virus into the host cell (14, 15). Due to the global spread of the COVID-19, the differences in societies-specific characteristics in terms of gender may additionally hinder the analysis of this factor. As case fatality ratio for males is higher in COVID-19 than for females in global data analysis, it is worthy further research (16).

The distribution of NL Ratio values in patients grouped according to their SII index (high: SII $<$ 410/low: SII $\geq$ 410) is presented below in **Figure 4**.

Biochemical Parameters

Random Blood Glucose

The results of RG examination in recruited patients indicate that hyperglycaemia correlates with a severe course of SARS-CoV-2. This finding was already reported in an Italian study (17). The random blood sugar level was elevated in over 30% of patients in the control group and over 86% of patients in the SARS-CoV-2 (+) group. Another study also reported that 51.5 and 57.4% of severely and critically ill patients were diagnosed with diabetes (18). High glucose concentration directly impacts the course of the infection, and also through the development of further complications, such as diabetic ketoacidosis and vascular comorbidities (atherosclerosis, peripheral arteriosclerosis) it influences this process indirectly (19).

Vitamin D

The association of vitamin D concentrations and the course of SARS-CoV-2 was assessed due to the high importance of Vitamin D in the activation of immune defence. Vitamin D status was

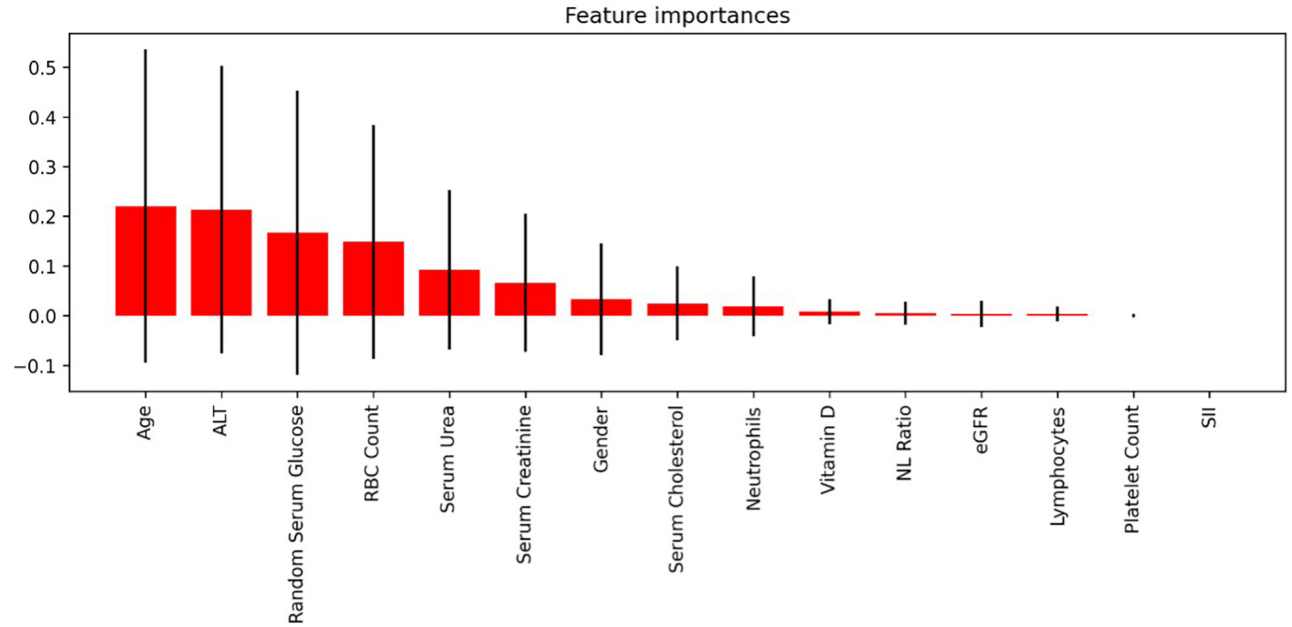


FIGURE 2 | Feature importance of each parameter assessed for the predictive model.

below the reference range threshold in almost all patients in the control group and all SARS-CoV-2 (+) group patients (see **Figure 1**). This could be a possible explanation for a relatively low correlation parameter between SARS-CoV-2 infection and Vitamin D in the presented results. Other studies have shown that groups of patients with severe course of SARS-CoV-2, requiring hospitalization in the ICU and showing higher mortality, have decreased serum concentrations of Vitamin D (20, 21).

There is a significant influence of changes in both blood sugar and Vitamin D status in the immune system (22, 23). Insulin resistance and type 2 diabetes mellitus are two conditions associated with a pro-inflammatory status. Consequently, elevated systemic concentrations of inflammatory markers, including cytokines, are observed in the blood, which in turn correlates with a higher probability of a cytokine storm (2, 4). Also, normal concentrations of vitamin D reduce the risk of a cytokine storm by reducing the systemic concentration of pro-inflammatory cytokines such as IL-6 and TNF-α. The significantly higher concentration of pro-inflammatory cytokines in circulation correlates with the severe course of SARS-CoV-2 (24, 25). Nonetheless, results similar to the presented data in terms of differences in vitamin D status between COVID-19 patients and population-based controls were reported, suggesting that despite the proven potential causality in interactions of vitamin D levels and COVID-19 pathogenesis, the alterations of vitamin D status do not influence the severity of disease course (26). On the other hand, there are some meta-analyses showing that vitamin D supplementation in SARS-CoV-2 positive patients has the potential to positive impact their outcomes (27).

TABLE 4 | Quantities of patients divided according to the Vitamin D levels in terms of NL Ratio.

NL Ratio	Patients Count		
	I T (1.11 – 9.40]	II T (9.40 – 17.69]	III T (17.69 – 25.97]
0.0 - 0.5	0	0	0
0.5 - 1.0	0	0	0
1.0 - 1.5	387	389	381
1.5 - 2.0	206	206	233
2.0 - 2.5	105	131	123
2.5 - 3.0	78	79	93
3.0 - 3.5	20	20	13
3.5 - 4.0	0	0	2

NL Ratio; neutrophil-lymphocyte ratio.

NL Ratio is regarded as a prognostic marker of systemic inflammation. It is an important element in the diagnosis of SARS-CoV-2(+) patients. Some studies indicate that its value correlates with the severe course of the disease and mortality (28). Apart from the assessment of inflammation, the NL Ratio is also used to analyse the course of other diseases, including cardiovascular diseases and cancer. Due to its numerous benefits, this parameter is assessed more frequently during SARS-CoV-2, compared to other inflammatory parameters. Its value can be assessed on the basis of one of the basic tests, i.e., morphology, taking into account the values of relative neutrophils and lymphocytes count. NL Ratio is characterized by a higher specificity and sensitivity compared to blood cell count. Moreover, its value increases faster, which is useful in the early assessment of the patient's condition. Other studies shows that a higher value of NL Ratio correlates with a more severe course of SARS-CoV-2. It has been shown that higher NL Ratio values

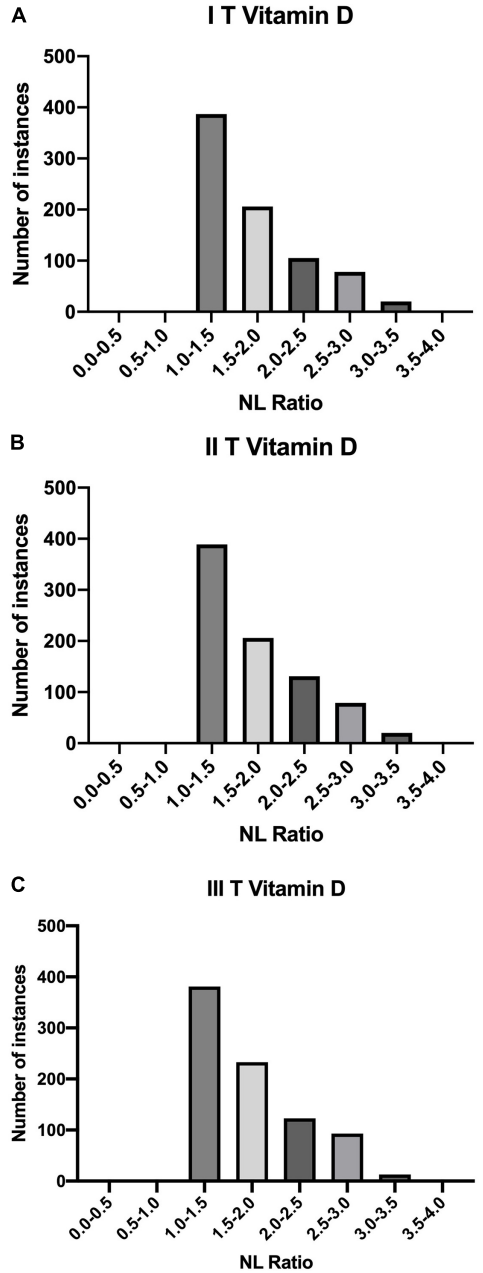


FIGURE 3 | Quantities of patients in buckets of NL Ratio values, grouped according to their Vitamin D status: **(A)** I T (1.11 – 9.40]; **(B)** II T (9.40 – 17.69]; **(C)** III T (17.69 – 25.97].

TABLE 5 | The number of patients divided according to the Vitamin D levels in terms of SII.

SII	Patients Count		
	I T (1.11 – 9.40]	II T (9.40 – 17.69]	III T (17.69 – 25.97]
Low	315	315	321
High	493	519	535

SII, systemic immune inflammation index.

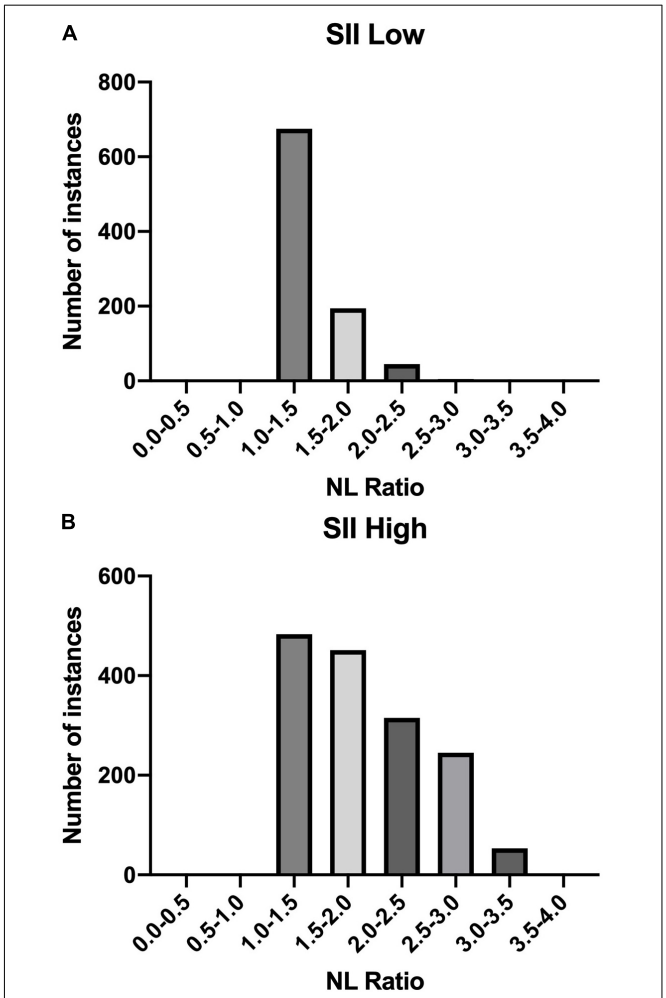


FIGURE 4 | NL Ratio values structure in the studied population. The patients were grouped according to the SII index: **(A)** Low SII; **(B)** High SII.

occur in people with a significant amount of comorbidities (29–31). In this study it was shown that among the tertiles of Vitamin D values, the biggest number of patients with SARS-CoV-2 (+) had an NL Ratio value of 1.0-1.5.

SII is another important predictive marker in SARS-CoV-2 (+) patients. It is used to assess the prognosis in people diagnosed with pulmonary embolism, obesity, sepsis, cancer and diseases accompanied by inflammation (32–36)., SII has been used as an indicator of severe course of SARS-CoV-2 (37, 38). In our study, each of the tertiles of vitamin D was correlated with SII high value (SII High: $\geq$ 410). In each of the tertiles of vitamin D, we included the division into low and high SII values. Our results showed that patients with high SII values predominated in each tertiles (SII High: $\geq$ 410).

Another analysis concerned the correlation between the SII and NL Ratio values. The highest percentage of patients (n = 675) in the SII-Low value group had an NL Ratio of 1.0 - 1.5. However, in the SII High group, the distribution of patients according to the

NL Ratio value was as follows: the largest number of patients are in the NL Ratio ranges: 1.0 - 1.5 (n = 483) and 1.5 - 2.0 (n = 451). In SII-High values, the distribution of patients on NL Ratio values was more diverse.

Our results indicate that SII can be a reliable biomarker that is easily tested than those based on cytokines or coagulation system markers (39, 40).

Cholesterol

Our data confirmed the association between elevated cholesterol levels and the severity of course of SARS-CoV-2 infection. A study of a cohort of 3,988 critically ill Italian patients demonstrated that hypercholesterolaemia may be classified as a risk factor for increased mortality in patients SARS-CoV-2(+) admitted to the ICU (41). This observation may be associated with an increased risk of developing cardiovascular diseases in patients with hypercholesterolaemia, as well as with its influence on one of the components of lung surfactant, which is an important protection in the respiratory system (42–45).

Morphologic Parameters

In an examination of blood morphology, a focus was put on the basic parameters, such as the number of RBC count, due to their accessibility in common clinical practice. The present results show that a low RBC count correlates in studied groups with a higher percentage of hospitalized patients and with a more severe course of infection. This observation may be related to the numerous important functions performed by RBC in the human body. Similar results were presented in a study performed in hospitals in Dhaka, Bangladesh, which indicated the likely impact of a reduction in red blood cell counts on the severity of SARS-CoV-2. The patients in the quoted study, who had a worse prognosis, were found to be present with decreased RBC count. This report also suggests a relationship between the number of RBCs and the concentration of haemoglobin (46). Nonetheless, the impact of changes in RBC Count over the course of SARS-CoV-2 infection requires more research.

Renal and Hepatic Parameters

In the population of the present study, the patients hospitalized for SARS-CoV-2 were presented with changes in renal parameters such as serum urea, creatinine and GFR. The analysis discovered an increase in the level of serum urea and a decrease in the concentration of creatinine, which consequently influenced the GFR values. The strongest correlation among these parameters was observed for serum urea (0.555459). A study by Hachim et al. (47) showed that elevated urea levels correlated with a higher probability of patient admission to the ICU and a more severe course of the infection. Increased creatinine levels were also observed in this group of patients. The results presented by Chen et al. (13) agree with the assumption that patients developing kidney disorders during SARS-CoV-2 (+) are more likely to develop more severe complications in the course of the disease. Higher levels of both blood urea nitrogen and creatinine were described in patients who died of SARS-CoV-2 than in the recovered ones.

The impact of liver disease on the course of SARS-CoV-2 infection was also a scope of the presented study. The liver is an important organ for maintaining the homeostasis of the organism, therefore the disturbance of its function correlates with the occurrence of complications (48, 49). For this purpose, the changes in the level of standard liver marker ALT were examined among the studied groups. The analysis of the collected data shows that a decreased level of ALT is observed among the SARS-CoV-2 (+), contrary to results of a study performed by Chen et al. (13), which showed an increased concentration of this liver parameter in the deceased. The median (IQR) value in this study was for dead patients: $\leq$ 41 U/L: 23.0 (15.0-38.0); and for the recovered: $>$ 41 U/L: 60.

Two crucial results differentiating the present study from others relate to our findings on the concentrations of ALT and creatinine. To reconcile those contradictory findings, the percentage of patients in study groups, both present and previous, with co-existing liver disease before developing SARS-CoV-2 needs to be determined. Another possible reason for such results could be the characteristic of the groups in terms of SARS-CoV-2 severity at the moment of taking blood tests. Patients enrolled in the Chen et al. (13) study were described as moderate to critically ill. 41.2% of the laboratory results presented in their study came from patients who eventually died, which could strongly influence the results. The present study recruited hospitalized patients, though the more detailed information on their clinical condition, exact severity of the disease and medical history concerning chronic diseases are lacking. Further research is required to assess the influence of liver and kidneys' function on the severity of the COVID19. Therefore, whether acute liver disease and acute kidney injury during the course of SARS-CoV-2 could be developed at the more advanced and severe stadium as well as whether the potential distinction in ALT and creatinine outcomes in both studies could be influenced by the presence of chronic liver diseases before the studies remain unknown.

Predictive Model

Mathematical models are valuable tools for the prevention and control of infectious diseases, including SARS-CoV-2. Thanks to them, it is possible to determine the relationship between various processes and to assess the dynamics of the disease spread. It also helps to establish a vaccination strategy. For this reason, during the SARS-CoV-2 pandemic, many researchers have proposed numerous mathematical models based on various correlations. The mathematical models took into account, *inter alia*, the method of transmission, the impact of quarantine and isolation, pharmaceutical and non-pharmaceutical interventions (50, 51). Most of the models developed during the SARS-CoV-2 pandemic were based on the Susceptible-infected-recovered (SIR) model and the Susceptible-Exposed-Infected-Removed (SEIR) model. These models are used to assess the spread of the virus (52). Researchers introduced numerous modifications to the classic SIR model due to the dynamics of SARS-CoV-2 spreading. Cooper et al. (53) adapted the SIR model to the increasing over time number of susceptible individuals. In the Susceptible-Exposed-Infected-Removed model, each person can be assigned to the following category susceptible (S),

exposed (E), infected (I) and recovered/removed (R) at any time during the epidemic. The researchers also made modifications to the SEIR model. One of them concerns the impact of human migration on the spread of the virus (54). Other modifications include assessing the impact of using face masks (55).

The biochemical parameters were studied in terms of disease prediction and triage already in previous coronavirus outbreak – Middle East respiratory syndrome coronavirus (MERS-CoV). The results showed that the routine parameters may turn out to be useful in clinical decision making (56). Similarly, the basic blood biomarkers were a focus of research in early stages of COVID-19 pandemic, though the reports indicated a rather moderate efficiency of these parameters in differentiating infected patients from healthy ones (57, 58). More optimistic results concerned the estimation of mortality risk (59).

The recent advances in bioinformatics enable creation of complex models, which based on the patient's data assess the risk of infection or severity of potentially acquired disease. By harnessing the power of machine learning and artificial intelligence, it is possible to offer an alternative to standard biochemical diagnostic tools, such as reverse transcription-polymerase chain reaction (RT-PCR) assays, which in some parts of the world remain expensive and difficult to obtain (60). It may also be useful in analysing a dynamic of current clinical state of patients, and on this basis nowcasting the further disease development (61).

Analysis of feature importance in the current study returned a valuable insight into the potential of studied parameters in terms of assessing the patient's risk of SARS-CoV-2 infection severity, providing a satisfying prediction accuracy of 97%. The main features of the predictive model created in this study were basic parameters obtained during the regular blood test – RBC count, serum glucose, serum urea, as well as the anthropometric measure – age. These results expand the spectrum of features useful in the COVID-19 prediction, as the previous study harnessed the methods of deep learning and the Random Forest machine learning model reported the efficacy of other parameters: lactate, the absolute level of immature granulocytes, respiratory rate, haemoglobin, procalcitonin, hematocrite (62). The only matching parameter between this and previous studies was serum urea, which in both studies was recognized as an important factor. A model presented in mentioned above study showed lower precision (current study: 0.98 vs. Aktar et al.: 0.90) and recall (0.97 vs. 0.90), F1 score (0.98 vs. 0.90) accuracy (0.97 vs. 0.89) and AUC (0.99 vs. 0.89). Another earlier report also differed in terms of included biomarkers, as despite the similarities concerning included blood count parameters, the biochemical parameters set was different for previous and current study. Interestingly, in the cited study the blood count parameters such as neutrophil, basophil and eosinophil count had similar feature importance as age, which in the current study showed the highest importance. The most important features were lactate dehydrogenase and C-reactive protein, which were not included in the current study (63). It must be noted, however, that this study was conducted on smaller population (2,500 patients in current study vs. 1,455 cases in Goodman-Meza et al.) and led to creation of model characterised by lower AUC value (0.99 vs. 0.91) and area under precision recall curve (1.00 vs. 0.76). The above studies were all based on the biochemical parameters, nonetheless the created models could be enhanced by including an analysis X-ray or computer tomography images, as suggested by the previous reports (64, 65). Further research in this matter is necessary.

As shown with above results, the machine learning may be a key for assessing the SARS-CoV-2 infection risk without the necessity of applying the standard diagnostic measures. Other studies explored also the possibilities of estimating a mortality risk (66–68), as well as the severity of the COVID-19 course (69) based on the simple biochemical parameters. The cardiovascular component of COVID-19 may also be evaluated using machine learning models (70).

LIMITATIONS

The difference in our analysis compared to other studies may be associated with a disproportionate number of women compared with men with SARS-CoV-2 who were enrolled in this study. The higher number of women may influence the observation that the female gender correlates with a more severe course of infection. The demographic variables of the studied population may also be a factor causing differences in the results regarding some of the parameters. Moreover, the analysed data is limited to the residents of India, which may also affect the obtained results. The influence of additional factors on the obtained results should also be taken into account.

The differences in presented ALT and creatinine levels compared to other studies may stem from the analysis of the different demographic groups included in the study. The influence of inter-individual variability, the treatment regimens used and the stage of the disease at which patients were admitted should also be taken into consideration.

Further research is necessary to adequately respond to the influence of limitations on the results and to evaluate the observations of the present study.

CONCLUSION

Assessment of the impact of pre-existing comorbidities and changes in the biochemical and morphological parameters observed in SARS-CoV-2 patients in the course of the disease may contribute to a better understanding of the influence of each of these individual factors on the pathology. Therefore, it could consequently affect the selection of appropriate therapy and the reduction of possible complications. Presented results indicate the importance of adequate vitamin D supplementation and maintaining the physiological functions of the liver and kidneys for reducing the risk of severe COVID-19 course. Standard serum parameters, such as red blood cell count, serum glucose, urea, ALT, cholesterol and creatinine, are efficient in predicting the patient's condition in terms of SARS-CoV-2 infection.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AM designed and coordinated the project. PG did the AI experiments. AM, AS, MH, and WS wrote the manuscript. KF, JG, and GY revised it critically for important intellectual content. All authors approved the version to be published.

FUNDING

GY was supported by the ERC IMI (101005122), the H2020 (952172), the MRC (MC/PC/21013), the Royal Society (IEC\NSFC\211235), the SABER project supported by Boehringer Ingelheim Ltd, NIHR Imperial Biomedical Research Centre (RDA01), and the UKRI Future Leaders Fellowship (MR/V023799/1). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.962101/full#supplementary-material

REFERENCES

1. Suleyman G, Fadel RA, Malette KM, Hammond C, Abdulla H, Entz A, et al. Clinical characteristics and morbidity associated with coronavirus disease 2019 in a series of patients in metropolitan detroit. *JAMA Netw Open.* (2020) 3:e2012270. doi: 10.1001/jamanetworkopen.2020.12270
2. Al-Kuraishy HM, Al-Gareeb AI, Alblihed M, Guerreiro SG, Cruz-Martins N, Batiha GES. COVID-19 in relation to hyperglycemia and diabetes mellitus. *Front Cardiovasc Med.* (2021) 8:644095. doi: 10.3389/fcvm.2021.644095
3. Dana T, Graves RAK. Diabetic complications and dysregulated innate immunity. *FBL.* (2008) 13:1227–39. doi: 10.2741/2757
4. Landstra CP, de Koning EJP. COVID-19 and Diabetes: understanding the Interrelationship and Risks for a Severe Course. *Front Endocrinol.* (2021) 12:649525. doi: 10.3389/fendo.2021.649525
5. Lotfi M, Hamblin MR, Rezaei N. COVID-19: transmission, prevention, and potential therapeutic opportunities. *Clin Chim Acta.* (2020) 508:254–66. doi: 10.1016/j.cca.2020.05.044
6. Kopple JD. National kidney foundation K/DOQI clinical practice guidelines for nutrition in chronic renal failure. *Am J Kidney Dis.* (2001) 37:S66–70.
7. Ho TK. Random decision forests. *Proc 3rd Int Confer Document Anal Recognit.* (1995) 1:278–82.
8. Parohan M, Yaghoubi S, Seraji A, Javanbakht MH, Sarraf P, Djalali M. Risk factors for mortality in patients with Coronavirus disease 2019 (COVID-19) infection: a systematic review and meta-analysis of observational studies. *Aging Male.* (2021) 23:1416–24. doi: 10.1080/13685538.2020.1774748
9. Mikami T, Miyashita H, Yamada T, Harrington M, Steinberg D, Dunn A, et al. Risk factors for mortality in patients with COVID-19 in New York City. *J General Internal Med.* (2021) 36:17–26. doi: 10.1007/s11606-020-05983-z
10. Du Y, Tu L, Zhu P, Mu M, Wang R, Yang P, et al. Clinical features of 85 fatal cases of COVID-19 from Wuhan a retrospective observational study. *Am J Respirat Crit Care Med.* (2020) 201:1372–9. doi: 10.1164/rccm.202003-0543OC
11. Cunha LL, Perazzio SF, Azzi J, Cravedi P, Riella VL. Remodeling of the immune response with aging: immunosenescence and its potential impact on COVID-19 immune response. *Front Immunol.* (2020) 11:1748. doi: 10.3389/fimmu.2020.01748
12. Meftahi GH, Jangravi Z, Sahraei H. Zahra Bahari. The possible pathophysiology mechanism of cytokine storm in elderly adults with COVID-19 infection: the contribution of "inflame-aging.". *Inflamm Res.* (2020) 69:825–39. doi: 10.1007/s00011-020-01372-8
13. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ.* (2020) 368:m1091. doi: 10.1136/bmj.m1091
14. Alwani M, Yassin A, Al-Zoubi RM, Aboumarzouk OM, Nettleship J, Kelly D, et al. Sex-based differences in severity and mortality in COVID-19. *Rev Med Virol.* (2021) 31:e2223. doi: 10.1002/rmv.2223
15. Biswas M, Rahaman S, Biswas TK, Haque Z, Ibrahim B. Association of sex, age, and comorbidities with mortality in COVID-19 patients: a systematic review and meta-analysis. *Intervirology.* (2021) 64:36–47. doi: 10.1159/000512592
16. Tazerji SS, Shahabinejad F, Tokasi M, Rad MA, Khan MS, Safdar M, et al. Global data analysis and risk factors associated with morbidity and mortality of COVID-19. *Gene Rep.* (2022) 26:101505. doi: 10.1016/j.genrep.2022.101505
17. Fadini GP, Morieri ML, Boscari F, Fioretto P, Maran A, Busetto L, et al. Newly-diagnosed diabetes and admission hyperglycemia predict COVID-19 severity by aggravating respiratory deterioration. *Diabetes Res Clin Pract.* (2020) 168:108374. doi: 10.1016/j.diabres.2020.108374
18. Al Argan R, Alkhafaji D, Al Elq A, Albaker W, Alqatari S, Alzaki A, et al. The impact of diabetes mellitus and hyperglycemia on the severity and outcome of patients with COVID-19 disease: a single-center experience. *Int J Gen Med.* (2021) 14:9445–57. doi: 10.2147/IJGM.S338800
19. Wang W, Shen M, Tao Y, Fairley CK, Zhong Q, Li Z, et al. Elevated glucose level leads to rapid COVID-19 progression and high fatality. *BMC Pulm Med.* (2021) 21:64. doi: 10.1186/s12890-021-01413-w
20. Merzon E, Tworowski D, Gorohovski A, Vinker S, Golan Cohen A, Green I, et al. Low plasma 25(OH) vitamin D level is associated with increased risk of COVID-19 infection: an Israeli population-based study. *FEBS J.* (2020) 287:3693–702. doi: 10.1111/febs.15495
21. Campi I, Gennari L, Merlotti D, Mingiano C, Frosali A, Giovanelli L, et al. Vitamin D and COVID-19 severity and related mortality: a prospective study in Italy. *BMC Infect Dis.* (2021) 21:566. doi: 10.1186/s12879-021-06281-7
22. Peng MY, Liu WC, Zheng JQ, Lu CL, Hou YC, Zheng CM, et al. Immunological aspects of sars-cov-2 infection and the putative beneficial role of vitamin-d. *Int J Mol Sci.* (2021) 22:5251. doi: 10.3390/ijms22105251
23. Mercola J, Grant WB, Wagner CL. Evidence regarding vitamin d and risk of covid-19 and its severity. *Nutrients.* (2020) 12:3361. doi: 10.3390/nu12113361

24. Kenneth Weir E, Thenappan T, Bhargava M, Chen Y. Does vitamin D deficiency increase the severity of COVID-19? *Clin Med.* (2020) 20:E107–8. doi: 10.7861/clinmed.2020-0301
25. Charoenngam N, Holick MF. Immunologic effects of vitamin d on human health and disease. *Nutrients.* (2020) 12:2097. doi: 10.3390/nu12072097
26. Hernández JL, Nan D, Fernandez-Ayala M, García-Unzueta M, Hernández-Hernández MA, López-Hoyos M, et al. Vitamin D status in hospitalized patients with SARS-CoV-2 infection. *J Clin Endocrinol Metab.* (2021) 106:e1343–53. doi: 10.1210/clinem/dgaa733
27. Szarpak L, Filipiak KJ, Gasecka A, Gawel W, Koziel D, Jaguszewski MJ, et al. Vitamin D supplementation to treat SARS-CoV-2 positive patients. Evidence from meta-analysis. *Cardiol J.* (2022) 29:188–96. doi: 10.5603/CJ.a2021.0122
28. Yılmaz E, Ak R, Doğanay F. Usefulness of the neutrophil-to-lymphocyte ratio in predicting the severity of COVID-19 patients: a retrospective cohort study. *Sao Paulo Med J.* (2022) 140:81–6. doi: 10.1590/1516-3180.2021.0298.R1.27052021
29. Wang Y, Zhao J, Yang L, Hu J, Yao Y. Value of the neutrophil-lymphocyte ratio in predicting COVID-19 severity: a meta-analysis. *Dis Markers.* (2021) 2021:2571912. doi: 10.1155/2021/2571912
30. Prozan L, Shusterman E, Ablin J, Mitelpunkt A, Weiss-Meilik A, Adler A, et al. Prognostic value of neutrophil-to-lymphocyte ratio in COVID-19 compared with Influenza and respiratory syncytial virus infection. *Sci Rep.* (2021) 11:21519. doi: 10.1038/s41598-021-00927-x
31. Karimi A, Shobeiri P, Kulasinghe A, Rezaei N. Novel systemic inflammation markers to predict COVID-19 prognosis. *Front Immunol.* (2021) 12:741061. doi: 10.3389/fimmu.2021.741061
32. Xue J, Ma D, Jiang J, Liu Y. Diagnostic and prognostic value of immune/inflammation biomarkers for venous thromboembolism: is it reliable for clinical practice? *J Inflamm Res.* (2021) 14:5059–77. doi: 10.2147/JIR.S327014
33. Ji Y, Wang H. Prognostic prediction of systemic immune-inflammation index for patients with gynecological and breast cancers: a meta-analysis. *World J Surg Oncol.* (2020) 18:197. doi: 10.1186/s12957-020-01974-w
34. Huang Y, Gao Y, Wu Y, Lin H. Prognostic value of systemic immune-inflammation index in patients with urologic cancers: a meta-analysis. *Cancer Cell Int.* (2020) 20:499. doi: 10.1186/s12935-020-01590-4
35. Hu B, Yang XR, Xu Y, Sun YF, Sun C, Guo W, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. *Clin Cancer Res.* (2014) 20:6212–22. doi: 10.1158/1078-0432.CCR-14-0442
36. Pace NP, Vassallo J. Association between neutrophil-lymphocyte ratio and gestational diabetes - a systematic review and meta-analysis. *J Endocr Soc.* (2021) 5:bvab051. doi: 10.1210/jendso/bvab051
37. Nalbant A, Demirci T, Kaya T, Aydın A, Altındiş M, Güçlü E. Can prognostic nutritional index and systemic immune-inflammatory index predict disease severity in COVID-19? *Int J Clin Pract.* (2021) 75:e14544. doi: 10.1111/ijcp.14544
38. Ozdemir A, Kocak SY, Karabela SN, Yılmaz M. Can systemic immune inflammation index at admission predict in-hospital mortality in chronic kidney disease patients with SARS-CoV-2 infection? *Nefrología.* (2021) [Online ahead of print]. doi: 10.1016/j.nefro.2021.09.001
39. Szarpak Ł, Nowak B, Kosior D, Zaczynski A, Filipiak KJ, Jaguszewski MJ. Cytokines as predictors of COVID-19 severity: evidence from a meta-analysis. *Pol Arch Intern Med.* (2021) 131:98–9. doi: 10.20452/pamw.15685
40. Ruetzler K, Szarpak L, Ladny JR, Gasecka A, Malinowska NG, Pruc M, et al. D-dimer levels predict COVID-19 severity and mortality. *Kardiol Pol.* (2021) 79:217–9. doi: 10.33963/KP.15830
41. Grasselli G, Greco M, Zanella A, Albano G, Antonelli M, Bellani G, et al. Risk factors associated with mortality among patients With COVID-19 in intensive care units in Lombardy, Italy. *JAMA Intern Med.* (2020) 180:1345–55. doi: 10.1001/jamainternmed.2020.3539
42. Tang Y, Hu L, Liu Y, Zhou B, Qin X, Ye J, et al. Possible mechanisms of cholesterol elevation aggravating COVID-19. *Int J Med Sci.* (2021) 18:3533–43. doi: 10.7150/ijms.62021
43. Frostegård J. Immunity, atherosclerosis and cardiovascular disease. *BMC Med.* (2013) 11:117. doi: 10.1186/1741-7015-11-117
44. Schmidt NM, Wing PAC, McKeating JA, Maini MK. Cholesterol-modifying drugs in COVID-19. *Oxford Open Immunol.* (2020) 1:iqaa001. doi: 10.1093/oxfimm/iqaa001
45. Gowdy KM, Fessler MB. Emerging roles for cholesterol and lipoproteins in lung disease. *Pulm Pharmacol Ther.* (2013) 26:430–7. doi: 10.1016/j.pupt.2012.06.002
46. Rahman MA, Shanjana Y, Tushar MI, Mahmud T, Rahman GMS, Milan ZH, et al. Hematological abnormalities and comorbidities are associated with COVID-19 severity among hospitalized patients: experience from Bangladesh. *PLoS One.* (2021) 16:e0255379. doi: 10.1371/journal.pone.0255379
47. Dai RP, Tonetti T, Donati A, Hachim MY, Hachim IY, bin Naeem K, et al. D-dimer, Troponin, and Urea Level at Presentation With COVID-19 can Predict ICU Admission: a Single Centered Study. *Front Med.* (2020) 7:585003. doi: 10.3389/fmed.2020.585003
48. Roedl K, Jarczak D, Drolz A, Wichmann D, Boenisch O, de Heer G, et al. Severe liver dysfunction complicating course of COVID-19 in the critically ill: multifactorial cause or direct viral effect? *Ann Intens Care.* (2021) 11:44. doi: 10.1186/s13613-021-00835-3
49. Singh S, Khan A. Clinical characteristics and outcomes of coronavirus disease 2019 among patients with preexisting liver disease in the United States: a multicenter research network study. *Gastroenterology.* (2020) 159:768.e–71.e. doi: 10.1053/j.gastro.2020.04.064
50. Ali Z, Rabiei F, Rashidi MM, Khodadadi T. A fractional-order mathematical model for COVID-19 outbreak with the effect of symptomatic and asymptomatic transmissions. *Eur Phys J Plus.* (2022) 137:395. doi: 10.1140/epjp/s13360-022-02603-z
51. Kifle ZS, Obsu LL. Mathematical modeling for COVID-19 transmission dynamics: a case study in Ethiopia. *Results Phys.* (2022) 34:105191. doi: 10.1016/j.rinp.2022.105191
52. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Appl Intell.* (2020) 50:3913–25. doi: 10.1007/s10489-020-01770-9
53. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals.* (2020) 139:110057. doi: 10.1016/j.chaos.2020.110057
54. Zhan C, Tse CK, Fu Y, Lai Z, Zhang H. Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data. *PLoS One.* (2020) 15:e0241171. doi: 10.1371/journal.pone.0241171
55. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect Dis Model.* (2020) 5:293–308. doi: 10.1016/j.idm.2020.04.001
56. Park GE, Kang CI, Ko JH, Cho SY, Ha YE, Kim YJ, et al. Differential cell count and CRP level in blood as predictors for middle East respiratory syndrome coronavirus infection in acute febrile patients during nosocomial outbreak. *J Korean Med Sci.* (2017) 32:151–4. doi: 10.3346/jkms.2017.32.1.151
57. Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A. The role of biomarkers in diagnosis of COVID-19 – A systematic review. *Life Sci.* (2020) 254:117788. doi: 10.1016/j.lfs.2020.117788
58. Arnold DT, Attwood M, Barratt S, Morley A, Elvers KT, McKernon J, et al. Predicting outcomes of COVID-19 from admission biomarkers: a prospective UK cohort study. *Emerg Med J.* (2021) 38:543–8. doi: 10.1136/emermed-2020-210380
59. Loomba RS, Villarreal EG, Farias JS, Aggarwal G, Aggarwal S, Flores S. Serum biomarkers for prediction of mortality in patients with COVID-19. *Ann Clin Biochem.* (2022) 59:15–22. doi: 10.1177/00045632211014244
60. Gangloff C, Rafi S, Bouzillé G, Soulat L, Cuggia M. Machine learning is the key to diagnose COVID-19: a proof-of-concept study. *Sci Rep.* (2021) 11:7166. doi: 10.1038/s41598-021-86735-9
61. Garcia LP, Gonçalves AV, Andrade MP, Pedebôs LA, Vidor AC, Zaina R, et al. Estimating underdiagnosis of COVID-19 with nowcasting and machine learning. *Rev Bras Epidemiol.* (2021) 24:e210047. doi: 10.1590/1980-549720210047
62. Aktar S, Ahamad MM, Rashed-Al-Mahfuz M, Azad A, Uddin S, Kamal A, et al. Machine learning approach to predicting COVID-19 disease severity based on clinical blood test data: statistical analysis and model development. *JMIR Med Inform.* (2021) 9:e25884. doi: 10.2196/25884
63. Goodman-Meza D, Rudas A, Chiang JN, Adamson PC, Ebinger J, Sun N, et al. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity. *PLoS One.* (2020) 15:e0239474. doi: 10.1371/journal.pone.0239474

64. Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F. A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdiscip Sci Comput Life Sci.* (2021) 13:103–17. doi: 10.1007/s12539-020-00403-6
65. Zhang F. Application of machine learning in CT images and X-rays of COVID-19 pneumonia. *Medicine.* (2021) 100:e26855. doi: 10.1097/MD.0000000000026855
66. Sánchez-Montañés M, Rodríguez-Belenguer P, Serrano-López AJ, Soria-Olivas E, Alakhdar-Mohmara Y. Machine learning for mortality analysis in patients with COVID-19. *Int J Environ Res Public Health.* (2020) 17:8386. doi: 10.3390/ijerph17228386
67. Altini N, Brunetti A, Mazzoleni S, Moncelli F, Zagaria I, Prencipe B, et al. Predictive machine learning models and survival analysis for covid-19 prognosis based on hematochemical parameters. *Sensors.* (2021) 21:8503. doi: 10.3390/s21248503
68. Ikemura K, Bellin E, Yagi Y, Billett H, Saada M, Simone K, et al. Using automated machine learning to predict the mortality of patients with COVID-19: prediction model development study. *J Med Internet Res.* (2021) 23:e23458. doi: 10.2196/23458
69. Chen Y, Ouyang L, Bao FS, Li Q, Han L, Zhang H, et al. A multimodality machine learning approach to differentiate severe and nonsevere COVID-19: model development and validation. *J Med Internet Res.* (2021) 23:e23948. doi: 10.2196/23948
70. Zimmerman A, Kalra D. Usefulness of machine learning in COVID-19 for the detection and prognosis of cardiovascular complications. *Rev Cardiovasc Med.* (2020) 21:345–52. doi: 10.31083/j.rcm.2020.03.120

Conflict of Interest: AM and WS were employed by Cognescence Ltd., and Immunidex Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 09 August 2022
DOI 10.3389/fpubh.2022.927387

OPEN ACCESS

EDITED BY
Samer A. Kharroubi,
American University of Beirut, Lebanon

REVIEWED BY
Lin Wang,
University of Cambridge,
United Kingdom
Jun Tanimoto,
Kyushu University, Japan

*CORRESPONDENCE
Jiaoling Huang
jiaoling_huang@sina.com

SPECIALTY SECTION
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

RECEIVED 24 April 2022
ACCEPTED 18 July 2022
PUBLISHED 09 August 2022

CITATION
Qian Y, Cao S, Zhao L, Yan Y and Huang J (2022) Policy choices for Shanghai responding to challenges of Omicron.
Front. Public Health 10:927387.
doi: 10.3389/fpubh.2022.927387

Policy choices for Shanghai responding to challenges of Omicron

Ying Qian[1], Siqi Cao[1], Laijun Zhao[1], Yuge Yan[2] and Jiaoling Huang[2]*

[1]Business School, University of Shanghai for Science and Technology, Shanghai, China, [2]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: A new wave of Coronavirus disease 2019 (COVID-19) infection driven by Omicron BA.2 subvariant hit Shanghai end of February 2020. With higher transmissibility and milder symptoms, the daily new confirmed cases have soared to more than 20 K within one and a half months. The greatest challenge of Omicron spreading is that the rapidly surging number of infected populations overwhelming the healthcare system. What policy is effective for huge cities to fight against fast-spreading COVID-19 new variant remains a question.

Methods: A system dynamics model of the Shanghai Omicron epidemic was developed as an extension of the traditional susceptible-exposed-infected-susceptible recovered (SEIR) model to incorporate the policies, such as contact tracing and quarantine, COVID-19 testing, isolation of areas concerned, and vaccination. Epidemic data from Shanghai Municipal Health Commission were collected for model validation.

Results: Three policies were tested with the model: COVID-19 testing, isolation of areas concerned, and vaccination. Maintaining a high level of COVID-19 testing and transfer rate of the infected population can prevent the number of daily new confirmed cases from recurring growth. In the scenario that 50% of the infected population could be transferred for quarantine on daily bases, the daily confirmed asymptomatic cases and symptomatic cases remained at a low level under 100. For isolation of areas concerned, in the scenario with most isolation scope, the peak of daily confirmed asymptomatic and symptomatic cases dropped 18 and 16%, respectively, compared with that in the scenario with least isolation. Regarding vaccination, increasing the vaccination rate from 75 to 95% only slightly reduced the peak of the confirmed cases, but it can reduce the severe cases and death by 170%.

Conclusions: The effective policies for Omicron include high level of testing capacity with a combination of RAT and PCR testing to identify and quarantine the infected cases, especially the asymptomatic cases. Immediate home-isolation and fast transfer to centralized quarantine location could help control the spread of the virus.

Moreover, to promote the vaccination in vulnerable population could significantly reduce the severe cases and death. These policies could be applicable to all metropolises with huge population facing high transmissible low severity epidemic.

KEYWORDS

COVID-19, Omicron, system dynamics modeling, intervention policies, simulation

Introduction

A new wave of Coronavirus disease 2019 (COVID-19) infection driven by the new variant, Omicron, started in January 2020. Global countries are now facing new challenges brought by Omicron. Current studies consistently showed that this new variant is substantially more transmissible than the Delta variant, capable of significant immune evasion, and with milder or even without symptoms (1–3). The effectiveness of vaccination against infection dropped dramatically suggested by current studies (3, 4), even to 33% reported by one study (5). WHO reported this February that 130 million cases and 500,000 deaths had been recorded globally since Omicron was declared a variant of concern in late November, calling the count "beyond tragic." Latest data showed that the WHO's Europe region accounted for 58% of new confirmed cases, and 35% of new deaths, and the Americas made up 23% of new cases and 44% of new deaths (6). The huge challenge with the steep rise of infections of Omicron is the overwhelmed and understaffed healthcare system. Specialists warned that a considerable overload of the hospitals is to be expected (7). Based on a mathematical model of SARS-CoV-2 transmission tailored to the unique immunization and epidemiological situation of China, Yu concluded that the level of immunity induced by the current vaccination campaign would be insufficient to prevent overwhelming the healthcare system and major losses of human lives (8). Take the UK as an example, in England, more than 650,000 people have probably been infected twice; most of them were re-infected in the past 2 months, according to data collected by the UK Health Security Agency (9, 10). The British Medical Association reported that the facts, figures, and the living reality for thousands of patients and the National Health Service (NHS) staff daily demonstrate undoubtedly that the NHS is currently already overwhelmed (11). Alderwick pointed out that millions of people are already feeling the unbearable strain in the UK health system (12).

Before Omicron, strict intervention protection controls (IPCs) were adopted by China after the outbreak of COVID-19, such as active case surveillance at fever clinics, immediate isolation of cases, quarantine of close contacts and high-risk groups, polymerase chain reaction (PCR) testing, and compulsory use of masks in the general population (13). China's strict policy preference, called zero-COVID strategy, was corroborated by WHO of the outbreak dynamic and case count reported by the Chinese government (14), demonstrating that a strict and rapid response to an emerging epidemic can halt the spread of a new virus (15). The previous study reported that the zero-COVID strategy was estimated to have saved one million lives, compared with the global average mortality of COVID-19 (as of 16 February 2022) (16). This zero-COVID policy is undoubtedly successful containing the pandemic prior to the current outbreak caused by Omicron. However, this policy is being challenged as a new wave of Omicron hits. Recently, China's National Health Commission changed its rules so that mild cases could be isolated in centralized locations, rather than treated in hospitals, and the criteria for a patient to be discharged from isolation have also been lowered, which has aroused extensive attention and discussion. There have been arguments that it is time to prepare for relaxing the policy. Chen and Chen believed that this change will happen sooner or later, as SARS-CoV-2 will probably become a seasonal infection in 2022 and circulate in humans indefinitely (17).

In Shanghai, the timely and precise strategies taken to prevent the spread of COVID-19 had been successful in the past 2 years (18). Active monitoring, precise and fast contact tracing, and timely PCR testing for related population helped interrupt the transmission of virus with minimal social economic impact. However, in this wave of Omicron BA.2 subvariant infection, which started from February 24th 2022 with one asymptomatic case, the number of daily new confirmed cases soared to more than 20 K within one and a half months. From March 12th, all primary and secondary schools stopped offline classes and switched to online learning. On March 27th, extending the IPC policy of isolating high-risk areas, the Shanghai government announced a two-stage home isolation of the whole city, while PRC testing was carried out city-wide. About 9 million residents of Pudong, the eastern half of Shanghai, have been home isolated since March 28th. On the other side of the bund, roughly 15 million people in the west of the city, initiated a home isolation policy since April 1st 2021. After almost 2-week home isolation, on April 9th, the authority announced that the city's residential compounds. Villages and business locations will be classified into three types of zones: isolation, control, and precaution. Shanghai, as a metropolis with huge population, faced extreme

challenges in this wave of Omicron BA.2 subvariant. A series of questions remained unanswered, for example, under current regulation, how will the epidemic develop? More importantly, what is the effective policy for huge cities to fight against fast-spreading COVID-19 new variant? Responding to these questions, we constructed a system dynamic model to predict the COVID-19 trend in Shanghai and tested different policy tools to figure out effective measures for future policy decision-making.

Materials and methods

Data collection

Two kinds of data were used in this research. First, we tracked the policies implemented, such as contact tracing and quarantine, isolation of areas concerned, and PCR testing, of which the implementation time schedule and the scope were collected in order to set model parameters accordingly. Second, the epidemic data of COVID-19 for this wave (Mar 24th to Apr 10th) were collected from the website of the Shanghai Municipal Health Commission, which was publicly published (19). The data included the following items: (1) daily counts of new confirmed cases, such as asymptomatic cases and symptomatic cases; (2) cumulative counts of confirmed cases; and (3) the number of people in hospital for symptomatic cases and the number of people under observation for asymmetrical cases. These data were required for model validation.

Model structure

Susceptible-infected-recovered (SIR), and the related susceptible-infected-susceptible (SIS), susceptible-exposed-infected-susceptible-recovered (SEIR), models are typical mathematical models investigating the evolution of a disease over time. These models are based on a set of ordinary differential equations (ODEs), which presume a well-mixed population (20). The features of the epidemic determine which model to use. If the recovered populations do not become immune to the disease, the SIS model should be considered. If the epidemic has a significant incubation period, the exposed population should be added. Extensions of these classic models have been developed to incorporate human behavior, government policies, or other specific conditions (21). For example, the SVMBIR model includes an individual's decision on vaccination (V), intermediate-defense-measure (M), and both of them (B) (22). The SVEIR model has been used to study quarantine or isolation (23). The SVnIR2n model is used to study the impact of waning effect of vaccination on the spread of the disease (24).

To incorporate Shanghai's policy of central quarantine of close contacts and isolation of areas concerned such as residence building or working place, we extended the SEIR model to disaggregate the total population into three layers, people who are in normal condition, people who are isolated with certain areas and people who are in central quarantine. Moreover, considering the enormous asymmetric cases caused by Omicron, in this model, we also differentiated two types of infected population A (asymptomatic cases) and I (cases with COVID-19 symptoms). Therefore, in each layer of population, there were S (the susceptible), E (the exposed), A (the asymptomatic cases), and I (the infected), as illustrated in Figure 1.

Model equations

The Omicron spread process was represented by V_E and V_{Ei} where the S or Si contact with E and A or Ei and Ai. We noticed that people under isolation still contacted other persons within the isolation area, such as getting deliveries, taking part in PCR testing, and other necessary living or working activities. This made it possible for the virus to spread within the isolation area even though the contact rate was lower than that in outside the isolation areas. People under quarantine did not spread the virus anymore.

$$V_E = c * \beta * S * (A + I)/N * \theta * (1 - \eta_i) + c * \beta * S * (A + I)/N * (1 - \theta) \quad (1)$$

$$V_{Ei} = ci * \beta * Si * Ai/N * \theta * (1 - \eta_i) + ci * \beta * Si * Ai/N * (1 - \theta) \quad (2)$$

where c and ci represent the contact rate of the people outside isolation and under isolation, respectively; β represents the probability of transmission; N presents the total population; θ represents the vaccination rate; and η_i represents the effectiveness of vaccination against infection.

The E, will developed into asymptomatic cases, A, and symptomatic cases, I, and similarly for Ei and Eq expect that Ei, as already in isolation, when develop into symptomatic cases, will be immediately quarantined, which transferred into Iq.

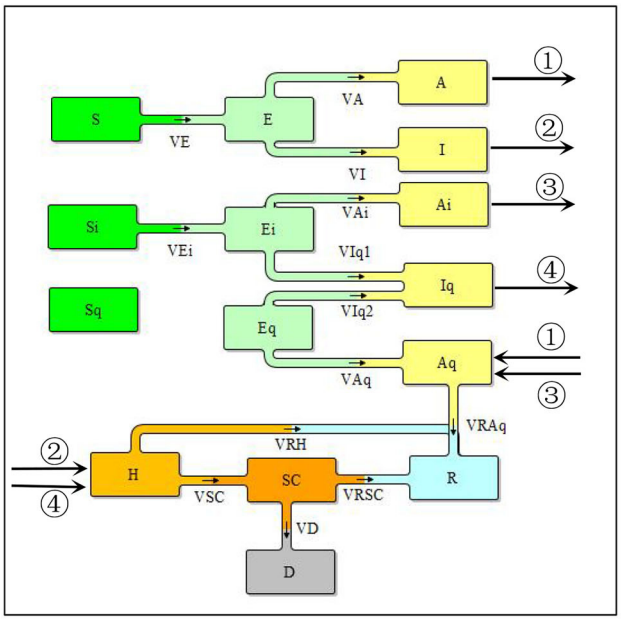
$$V_A = \mathrm{E}^*\alpha/\mathrm{T_{in}} \qquad V_I = \mathrm{E}^*(1\text{-}\alpha)/\mathrm{T_{in}} \quad (3)$$

$$V_{Ai} = \mathrm{Ei}^*\alpha/\mathrm{T_{in}} \qquad V_{Iq1} = \mathrm{Ei}^*(1\text{-}\alpha)/\mathrm{T_{in}} \quad (4)$$

$$V_{Aq} = \mathrm{Eq}^*\alpha/\mathrm{T_{in}} \qquad V_{Iq2} = \mathrm{Eq}^*(1\text{-}\alpha)/\mathrm{T_{in}} \quad (5)$$

where α represents the percentage of asymptomatic cases; therefore (1-α), was the percentage of cases with symptom, and $\mathrm{T_{in}}$ represents the incubation period.

The infected people without symptoms, A and Ai, could only be identified with COVID-19 testing. When confirmed to be infected, they would be transferred to cabin hospitals for quarantine and when they recover, they will be released from the cabin hospitals. $\mathrm{V_{Aq1}}$ and $\mathrm{V_{Aq2}}$ represent the flow of ① and ③,



The first layer S-E-A-I represents people neither in isolation nor in quarantine.

The second layer Si-Ei-Ai represents people in isolation. With isolation policy, people in areas concerned will be isolated. Several PCR testings will be performed.

The third layer Sq-Eq-Aq-Iq represents people in quarantine. The quarantine policy puts close contacts into quarantine for 14 days.

①,③ represents the process that PCR testings are performed to identify asymptomatic cases in A and Ai and then transfer them to cabin hospital for quarantine, in Aq.

②,④ represents the process that symptomatic cases in I and Iq are accepted to hospital, becoming H, some of whom recover later and some develop into SC (the server cases) and then recover or die.

FIGURE 1
The structure of Shanghai Omicron epidemic model.

respectively, in Figure 1.

$$V_{Aq1} = A * \tau_1 \quad V_{Aq2} = Ai * \tau_2 \tag{6}$$

$$V_{RAq} = Aq1 * \gamma_{Aq} \tag{7}$$

where τ_1 *and* τ_2 represents the transfer rate of A and Ai, respectively, which are related to the testing capacity and time for other related transfer process; and γ_{Aq} represents recover fraction of asymptomatic cases.

The people with symptom would be received in hospital. Overtime, some would recover and some would deliver into severe cases, who would get recover or die. V_{IH} and V_{IqH} represent the flow of ② and ④, respectively, in Figure 1.

$$V_{IH} = \mathrm{I}^{*}\kappa \quad V_{IqH} = \mathrm{Iq}^{*}\kappa \tag{8}$$

$$V_{SC} = H * \upsilon * \theta * (1 - \eta_s) + H * \upsilon * (1 - \theta) \tag{9}$$

$$V_{RH} = H1 * \gamma_H \quad V_{RSC} = SC * \gamma_{SC} \tag{10}$$

$$V_D = SC * \phi \tag{11}$$

where κ represents the hospital acceptance rate, which is related to the hospital capacity; ν represents severe case fraction and η_s represents the effectiveness of vaccination against severe case; γ_H represents recover fraction of hospital patients; γ_{SC} represents recover fraction of severe cases; ϕ represents death fraction. Details of the model equations are listed in Appendix section model structure and parameter settings are listed in Appendix section parameter setting, Tables 2, 3 in Appendix.

Model validation

The SEIR model and its extension have been widely used for the studying the spread of epidemic (25), especially for COVID-19 (26–29). This extension of the model and the parameter setting are based on the policies implemented in Shanghai, with details in Appendix. Finally, the model simulation results are compared with the epidemic data, as shown in Figure 2. Fitting the real data adds confidence to model validity.

Results

Based on the Shanghai Omicron epidemic model, we performed simulation to investigate the effect of three policies, COVID-19 testing, isolation, and vaccination.

Policy 1 COVID-19 testing

Four scenarios were tested with identification rate set at 10, 20, 30, and 50% after May 22nd. Results showed that if the identification rate dropped to 10%, within a month, the new infected asymptomatic cases would surge again and approach 8,000, as shown in Figure 3. With higher identification rate, the new infected asymptomatic cases would not reach that high. For the identification rate at 20 or 30%, the asymptomatic cases would increase to around

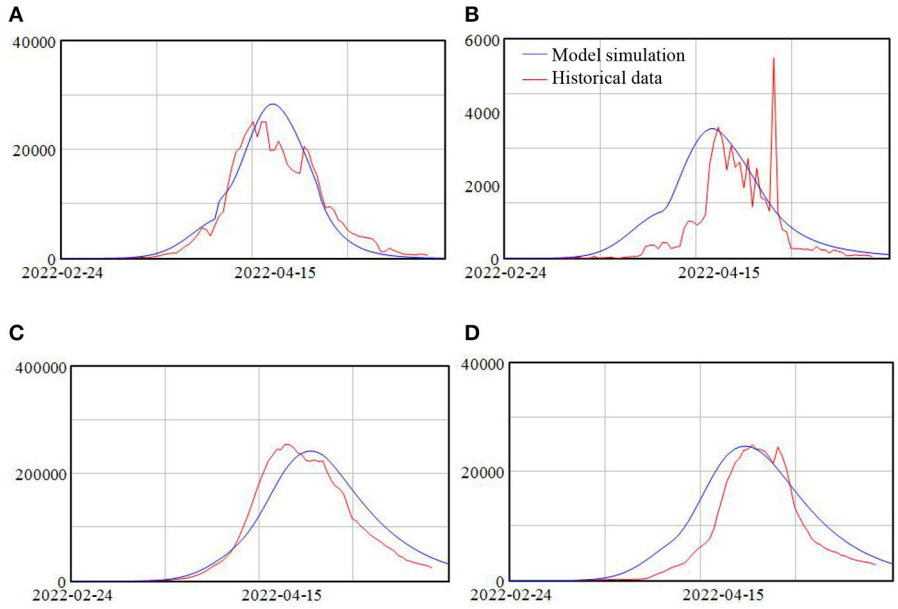


FIGURE 2
(A–D) Model simulation result and historical data.

2,600 and 770, respectively. When the identification rate remained at 50%, the asymptomatic cases would be <100. Similar simulation results were identified for symptomatic cases. This simulation result illustrated that the possible way of sustainable control of the COVID-19 was high identification rate.

Policy 2 isolation of areas concerned

In this model, we used the average number of people isolated for each confirmed case or close contact as an index of the scope of the isolation policy. Three scenarios were tested, with the average number of people isolated at 250, 200, and 150, respectively. Simulation results showed that the patterns of the spread of the Omicron remained the same for three scenarios, with a slow increase at the beginning, following a fast increase, and then a decrease, as shown in Figure 4. Yet the higher the average number of people isolated, the lower the peak of the daily confirmed cases. In the scenario with the least isolation scope that is on average only 150 people isolated per each risky source, the numbers of daily confirmed asymptomatic and symptomatic cases reached more than 30 and 3.7 K, respectively. While in the scenario with the most isolation scope, the number of daily confirmed asymptomatic and symptomatic cases reduced to 25.5 and 3.2 K, representing a drop of 18 and 16%, respectively, compared with the scenario with least isolation.

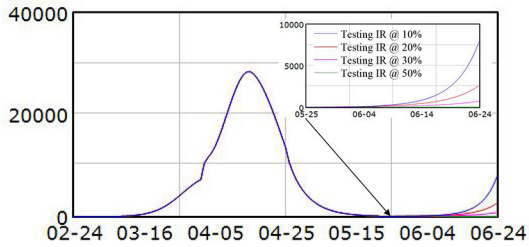


FIGURE 3
Simulation results for asymptomatic cases under different testing policies.

Policy 3 vaccination rate

Vaccination has been widely implemented in Shanghai and we investigate the impact of various vaccination rates on this wave of Omicron variant. Three scenarios were simulated with vaccination rate at 75, 85, and 95%. The simulation results are shown in Figure 5.

Due to the fact that the effectiveness of vaccination against infection for this new variant Omicron is not high, increasing the vaccination rate from 75 to 95% only slightly reduced the peak of the confirmed cases, with around 19% drop, as shown in Figures 5A,B. However, because of the high effectiveness of vaccination against severe cases, the peak of severe cases reduced significantly, from around 700 cases when the vaccination rate was 75% to around 260 cases when the

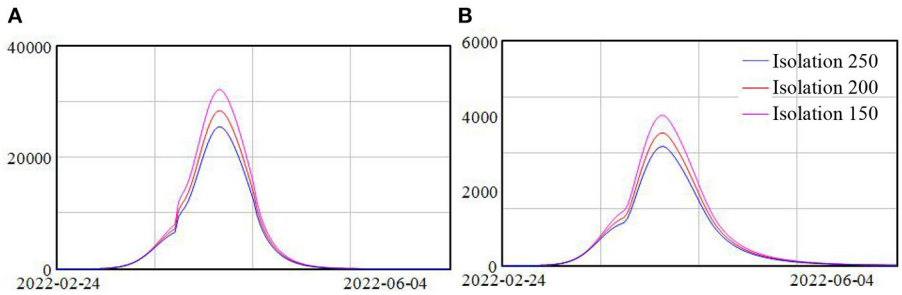


FIGURE 4
(A,B) Model scenarios with different isolation policy.

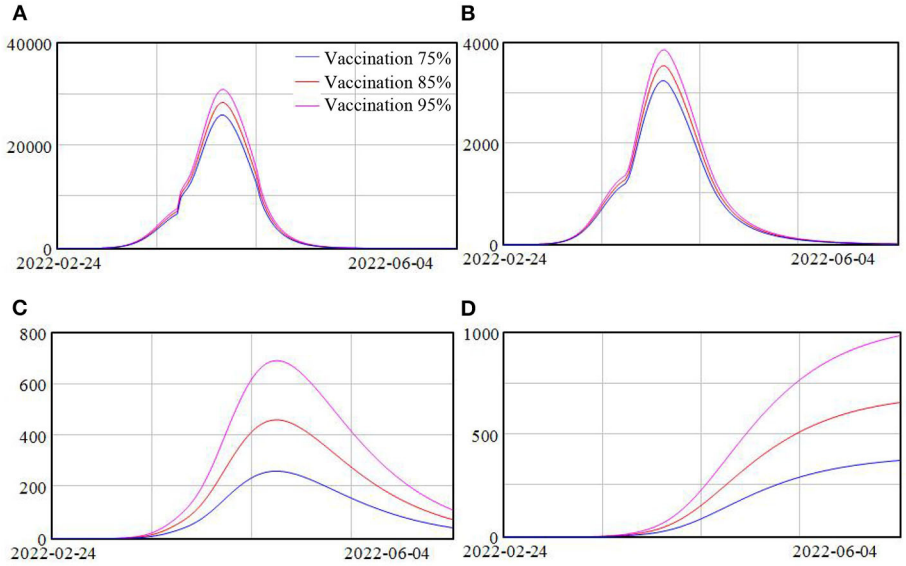


FIGURE 5
(A–D) Model scenarios with different vaccination rate.

vaccination rate was 95%, representing almost 170% drop. And the death population reduced from around 1,000 to 370 when the vaccination rate increased from 75 to 95%, as shown in Figure 5C.

Discussion

Faced with the Omicron new variant, especially with BA. 2 subvariant, the most effective policy is testing. As the Omicron BA.2 is more transmissible and milder in symptoms, the former COVID-19 control essentials "early detection, early reporting, early isolation, early treatment" become harder to achieve. Fever clinics become less effective in identifying infected populations as increasing number of people do not have fever or only have a little bit fever without the need to visit clinics. The main method for identifying the infected population is now COVID-19 testing (30). In Shanghai, PCR used to be the only method for COVID-19 testing. PCR testing provides relatively accurate results but is time- and people-consuming. For a metropolis like Shanghai with 24 million residents, even with the support of medical staffs from other provinces, it took 2 days to conduct COVID-19 testing for the entire population and at least another day to get all the results. As the serial interval of Omicron might be <3 days (31), during the time period of PCR, the infected population would have already been doubled, with new exposed population remained in the population to further develop into infected population and spread the virus around. RAT has recently been applied for identifying infected people as it is fast, inexpensive, and laboratory-independent (30). However, the sensitivity of RAT to detect the COVID-19 variants has been questionable. Independent laboratory

evaluations showed variable results (32–34). Caution should be taken when using RAT as a virus detection strategy. A combination of RAT with PCR test should be considered. On one hand, those who are tested positive with RAT should be re-examined using PCR. On the other hand, populations with high contagious risks should use PCR tests. In this way, early detection of the people infected by Omicron could be better achieved.

Isolation of the areas concerned has helped to reduce the peak of the daily confirmed cases, but the effectiveness of this policy is limited. Several reasons exist: (1) the isolation of the involved areas cannot fully cut the contact with the outside, for example, delivery of living necessities might bring the virus to the receiver; (2) internal movement still occurs, for instance, community workers or volunteers might get into contact with community residents when providing living support, such as maintenance or delivering services; (3) It takes one-or-two-day's time to re-confirm the COVID-19 testing result if one's test turns out to be positive, during which in the infection people might spread the virus to their family members or nearby neighbors. Besides, in the early stage, the healthcare system could not provide for enough hospital beds for the daily new confirmed cases, which might delay the transfer process as well. Our model implicated that immediate isolation of the infected people, either by transferring to a centralized quarantine location or by home isolation, can cut the infection chain more effectively.

Simulation results illustrated that increasing vaccination will not affect the new confirmed cases much, but has a significant impact on reducing the severe cases and death. The antigenic changes of Omicron reduced the sensitivity of this virus to antibody neutralization, which is responsible for the breakthrough cases (30). In this case, even though Shanghai has a high vaccination rate, the spread of the virus was still fast. In Shanghai, currently, more than 95% of the residents have completed 2 doses of vaccination and more than 10 million people have taken the booster shot (25). It is the similar case in Hong Kong, where more than 80% of the population has completed the second dose and more than one-third had taken the booster shot. The fifth wave of Omicron hit Hong Kong badly. Moreover, the death rate in Hong Kong was extraordinary high due to the low vaccination rate of the elderly population. Approximately, 95.8% of deaths were from ages above 60, a large portion of whom has not completed the second dose (35). Vaccination has proved to be effective against severe cases. It is important to improve the vaccination rate, especially for the vulnerable population group, such as the elderly.

Conclusions

Currently, many countries have gradually ceased their restrictive policies, considering the enormous impact on the national economy and daily life. Facing the future, China also needs to look for policies that better balance COVID-19 and other economic and social issues, especially when SARS-CoV-2 new variants become less severe. The major challenge is the rapidly surging infected patients overwhelming the healthcare system. This study showed that high level of testing capacity with a combination of RAT and PCR testing to identify the asymptomatic cases. A combination of immediate home-isolation and fast transfer of confirmed cases is the key to cut the infection chain and reduced the daily infection. Moreover, to promote the vaccination in vulnerable populations could significantly reduce the severe case. These policies could be applicable to all metropolises with huge populations facing high transmissible low severity epidemic.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

YQ and JH: conceptualization and writing—original draft. SC and YY: data curation. YQ, LZ, and JH: formal analysis. JH: funding acquisition and project administration. YQ and LZ: investigation and methodology. YQ and SC: software, supervision, validation, and visualization. LZ, SC, and YY: writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Science and Technology Committee of Shanghai Municipality (grants 22692192300 and 21692190200 to JH), the Science and Technology Innovation Project of Shanghai Jiao Tong University School of Medicine-Humanities and Social Sciences (grants WK2102 to JH), and Shanghai Jiao Tong University China Hospital Development Institute 2022 Hospital Management Project (grants CHDI-2022-B-31 to JH). The funder had no role in the design, data collection, analysis, interpretation, or writing of the report.

Acknowledgments

We would like to express our gratitude to a big hand from our partner, Deyong Shi, on the improvement of the article's expression.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.927387/full#supplementary-material

References

1. Del Rio C, Omer SB, Malani PN. Winter of Omicron-The evolving COVID-19 pandemic. *JAMA*. (2022) 327:319–20. doi: 10.1001/jama.2021.24315

2. Burki TK. Omicron variant and booster COVID-19 vaccines. *Lancet Respir Med*. (2022) 10:e17. doi: 10.1016/S2213-2600(21)00559-2

3. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N Engl J Med*. (2022) 386:1532–46. doi: 10.1056/NEJMoa2119451

4. Mahase E. Covid-19: Omicron and the need for boosters. *BMJ*. (2021) 375:n3079. doi: 10.1136/bmj.n3079

5. Tian D, Sun Y, Xu H, Ye Q. The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant. *J Med Virol*. (2022) 94:2376–83. doi: 10.1002/jmv.27643

6. NDTV. Half A Million People Killed Since Omicron, It's "Really Something": WHO (2022). Available online at: https://www.ndtv.com/world-news/coronavirus-who-says-tragic-500-000-deaths-reported-since-covid-19-omicron-2757388 (accessed April 16, 2022).

7. Vogel G, Kupferschmidt K. Early lab studies shed light on Omicron's behavior. *Science*. (2021) 374:1543–4. doi: 10.1126/science.acz9878

8. Yu HJ, Cai J, Deng XW, Yang J, Sun KY, Liu HC, et al. Projecting the impact of the introduction of SARS-CoV-2 Omicron variant in China in the context of waning immunity after vaccination. *Res Square*. (2022). doi: 10.21203/rs.3.rs-1478539/v1

9. GOV.UK. Coronavirus (COVID-19) in the UK (2022). Available online at: https://coronavirus.data.gov.uk (accessed April 16, 2022).

10. Nature. *COVID Reinfections Surge During Omicron Onslaught*. (2022). Available online at: https://www.nature.com/articles/d41586-022-00438-3 (accessed April 16, 2022).

11. The Guardian. *Parts of NHS May Be Overwhelmed by COVID Wave, Admits Boris Johnson*. (2022). Available online at: https://www.theguardian.com/world/2022/jan/04/parts-of-nhs-may-be-overwhelmed-by-covid-wave-admits-boris-johnson (accessed April 16, 2022).

12. Alderwick H. Is the NHS overwhelmed? *BMJ*. (2022) 376:o51. doi: 10.1136/bmj.o51

13. Azman AS, Luquero FJ. From China: hope and lessons for COVID-19 control. *Lancet Infect Dis*. (2020) 20:756–7. doi: 10.1016/S1473-3099(20)30264-4

14. WHO. *Infection Prevention and Control During Health Care When Novel Coronavirus (nCoV) Infection Is Suspected*. (2020). Available online at: https://www.who.int/publications/i/item/10665-331495 (accessed April 16, 2022).

15. Salzberger B, Glück T, Ehrenstein B. Successful containment of COVID-19: the WHO-Report on the COVID-19 outbreak in China. *Infection*. (2020) 48:151–3. doi: 10.1007/s15010-020-01409-4

16. Worldometer. *COVID-19 Coronavirus Pandemic: Reported Cases and Deaths by Country or Territory*. (2022). Available online at: https://www.worldometers.info/coronavirus/#countries (accessed April 16, 2022).

17. Chen JM, Chen YQ. China can prepare to end its zero-COVID policy. *Nat Med*. (2022) 28:1104–5. doi: 10.1038/s41591-022-01794-3

18. Shi Y, Jiang HL, Yang MX, Dong LJ, Chen Y, Zhou YB, et al. The precision of epidemiological investigation of COVID-19 transmission in Shanghai, China. *Infect Dis Poverty*. (2021) 10:58. doi: 10.1186/s40249-021-00849-w

19. Shanghai Municipal Health Commission. *COVID-19 Bulletin (in Chinese)*. (2020). Available online at: https://wsjkw.sh.gov.cn/yqtb/index.html (accessed April 16, 2022).

20. Jun Tanimoto. *Sociophysics Approach to Epidemics*. Singapore (2021).

21. Tanimoto, J. *Evolutionary Games with Sociophysics*. Singapore (2018). doi: 10.1007/978-981-13-2769-8

22. Alam M, Kuga K, Tanimoto J. Three-strategy and four-strategy model of vaccination game introducing an intermediate protecting measure. *Appl Math Comput*. (2019) 346:408–22. doi: 10.1016/j.amc.2018.10.015

23. Alam M, Kabir KMA, Tanimoto J. Based on mathematical epidemiology and evolutionary game theory, which is more effective: quarantine or isolation policy? *J Stat Mech Theory Exp*. (2020) 033502. doi: 10.1088/1742-5468/ab75ea

24. Kabir KMA, Tanimoto J. Analysis of individual strategies for artificial and natural immunity with imperfectness and durability of protection. *J Theoret Biol*. (2021) 509:110531. doi: 10.1016/j.jtbi.2020.110531

25. Halfmann PJ, Kuroda M, Maemura T, Chiba S, Armbrust T, Wright R, et al. Efficacy of vaccination and previous infection against the Omicron BA.1 variant in Syrian hamsters. *Cell Rep*. (2022) 28:110688. doi: 10.1016/j.celrep.2022.110688

26. Darabi N, Hosseinichimeh N. System dynamics modeling in health and medicine: a systematic literature review. *Sys Dyn Rev*. (2020) 36:29–73. doi: 10.1002/sdr.1646

27. Qian Y, Xie W, Zhao J, Xue M, Liu S, Wang L, et al. Investigating the effectiveness of re-opening policies before vaccination during a pandemic: SD modelling research based on COVID-19 in Wuhan. *BMC Public Health*. (2021) 21:1638. doi: 10.1186/s12889-021-11631-w

28. Zhao J, Jia J, Qian Y, Zhong L, Wang J, Cai Y. COVID-19 in Shanghai: IPC policy exploration in support of work resumption through system dynamics modeling. *Risk Manag Healthc Policy*. (2020) 13:1951–63. doi: 10.2147/RMHP.S265992

29. Hou C, Chen J, Zhou Y, Hua L, Jia E. The effectiveness of quarantine of Wuhan city against the Corona virus disease 2019 (COVID-19): a well-mixed SEIR model analysis. *J Med Virol*. (2020) 92:841–8. doi: 10.1002/jmv.25827

30. Osterman A, Badell I, Basara E, et al. Impaired detection of omicron by SARS-CoV-2 rapid antigen tests. *Med Microbiol Immunol*. (2022) 211:105–17. doi: 10.1007/s00430-022-00730-z

31. Song JS, Lee J, Kim M, Jeong H, Kim MS, Kim SG, et al. Serial intervals and household transmission of SARS-CoV-2 Omicron variant, South Korea, 2021. *Emerg Infect Dis*. (2022) 28:756–9. doi: 10.3201/eid2803.212607

32. Bekliz M, Adea K, Essaidi-Laziosi M, Sacks JA, Escadafal C, Kaiser L, et al. SARS-CoV-2 rapid diagnostic tests for emerging variants. *Lancet Microbe.* (2021) 2:e351. doi: 10.1016/S2666-5247(21)00147-6

33. Schildgen V, Demuth S, Lüsebrink J, Schildgen O. Limits and opportunities of SARS-CoV-2 antigen rapid tests: an experienced-based perspective. *Pathogens.* (2021) 10:38. doi: 10.3390/pathogens10010038

34. Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev.* (2021) 3:CD013705. doi: 10.1002/14651858.CD013705

35. Cheung PH, Chan CP, Jin DY. Lessons learned from the fifth wave of COVID-19 in Hong Kong in early 2022. *Emerg Microbes Infect.* (2022) 11:1072–8. doi: 10.1080/22221751.2022.2060137

frontiers | Frontiers in Public Health

TYPE Systematic Review
PUBLISHED 10 August 2022
DOI 10.3389/fpubh.2022.897784

OPEN ACCESS

EDITED BY
Omar El Deeb,
Lebanese American University, Lebanon

REVIEWED BY
Keun Hwa Lee,
Hanyang University, South Korea
Silvia – Giono Cerezo,
Instituto Politécnico Nacional (IPN), Mexico

*CORRESPONDENCE
Qiwei Liu
qiweiliu@hdu.edu.cn

SPECIALTY SECTION
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

RECEIVED 16 March 2022
ACCEPTED 18 July 2022
PUBLISHED 10 August 2022

CITATION
Zhao Y and Liu Q (2022) Analysis of distribution characteristics of COVID-19 in America based on space-time scan statistic. *Front. Public Health* 10:897784. doi: 10.3389/fpubh.2022.897784

Analysis of distribution characteristics of COVID-19 in America based on space-time scan statistic

Yuexu Zhao and Qiwei Liu*

College of Economics, Hangzhou Dianzi University, Hangzhou, China

Based on the epidemic data of COVID-19 in 50 states of the United States (the US) from December 2021 to January 2022, the spatial and temporal clustering characteristics of COVID-19 in the US are explored and analyzed. First, the retrospective spatiotemporal analysis is performed by using SaTScan 9.5, and 17 incidence areas are obtained. Second, the reliability of the results is tested by the circular distribution method in the time latitude and the clustering method in the spatial latitude, and it is confirmed that the retrospective spatiotemporal analysis accurately measures in time and reasonably divides regions according to the characteristics in space. Empirical results show that the first-level clustering area of the epidemic has six states with an average relative risk of 1.28 and the second-level clustering area includes 18 states with an average relative risk of 0.86. At present, the epidemic situation in the US continues to expand. It is necessary to do constructive work in epidemic prevention, reduce the impact of epidemic, and effectively control the spread of the epidemic.

KEYWORDS

the spatiotemporal analysis, COVID-19, scan statistic, spatial aggregation, Omicron

Introduction

Coronavirus disease 2019 (COVID-19) refers to the new coronavirus infection in 2019 caused by acute respiratory infectious diseases in the majority of patients; some of them will develop as severe cases and even results in death. Since the large-scale outbreak of new corona pneumonia in 21 January 2020, the economy in the US has been gradually affected. Besides, the epidemic is a public emergency that all countries in the world have to face. As one of the most serious epidemic countries, the US has accumulated 74,741,586 new coronavirus cases as on 31 January 2022, and there is a certain aggregation tendency in time and space. Scan statistic is a method to test whether there is an aggregation of diseases, and detect whether the abnormal increase of diseases in time and space is caused by random variation. It has been widely used in infectious diseases, cardiovascular diseases, and other fields as a spatial statistical method in epidemic statistics.

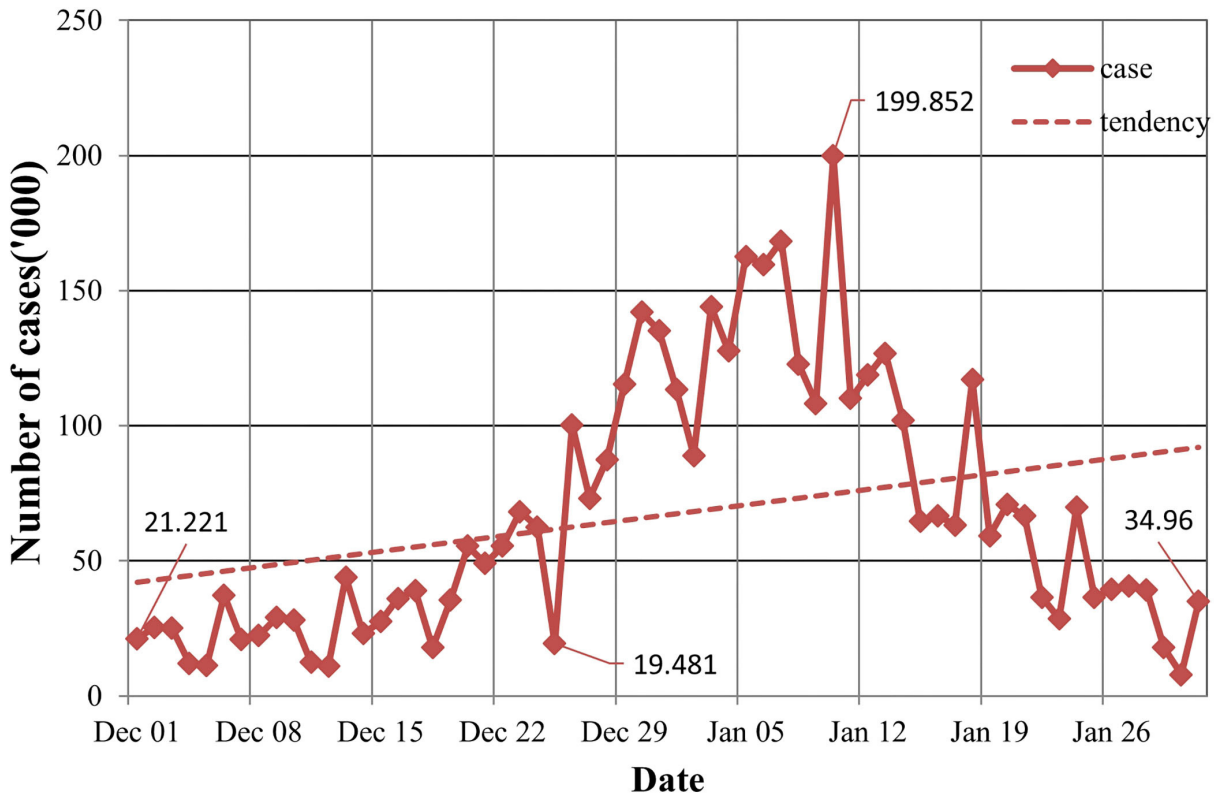


FIGURE 1
The number of COVID-19 cases.

In 1965, Joseph (1) proposed the concept of scan statistic. Kulldorff et al. (2–5) proposed the spatial scan statistic, and applied scan statistic to analyze the breast cancer mortality in the US. For example, they utilized the dynamic variable scanning window to detect the leukemia data in northern New York, and used the log-likelihood ratio to determine the cluster with the highest degree of aggregation. They also proposed many statistical models of spatiotemporal scanning, such as retrospective space-time scan statistic in Bernoulli model or Poisson model, prospective space-time scan statistic, space-time rearrangement scan statistic, and elliptical spatial scan statistic in periodic geographic disease monitoring. As the research goes further, Jung et al. (6, 7) proposed ordinal model scan statistic in 2007, which had excellent performance compared with Bernoulli scan statistic for binary classification of prostate cancer data. Huang et al. (8, 9) proposed the spatial scan statistic based on the exponential model for the male survival data with prostate cancer in the US in 2007; this method could be applied to the survival data and pure spatial data. In order to study the spatial heterogeneity continuously measured in the population data, the weighted normal spatial scan statistic was proposed and applied to the two-stage lung cancer survival research in 2009. Barbara (10) found that Cutl's method was more effective than Kulldorff's scan statistic for irregular shape spatiotemporal clusters, and for cylindrical spatiotemporal clusters; these two methods had similar results. Li et al. (11) analyzed the fund sustainability. Yin (12) carried out the research on application in early warning of infectious diseases, and graded the data of provinces and cities. Ma et al. (13) selected the optimal spatial scale through the number of signals in the monitoring of infectious diseases. So far, scan statistic has been widely used in disease prevention, including tuberculosis, schistosomiasis, and hand, foot, and mouth disease.

The majority of the abovementioned studies explore the spatial aggregation of various infectious diseases. COVID-19 is a highly contagious disease, which has seriously affected people's lives since its outbreak, and has a great threat to people's health. Hohl et al. (14) used the daily new coronavirus case data provided by the John Hopkins University at the county level, and applied SaTScan to conduct a prospective space-time analysis, and detected the active clusters in various provinces and cities in the US. To avoid using prospective space-time scan statistic to identify emergence of COVID-19 disease

TABLE 1 Retrospective spatiotemporal analysis.

Cluster	Start date	The number of states	The actual value	Value of expectation	$\tilde{R}$	*LLR*	*p*
1	Dec 1	6	4,157,904	3,363,765.04	1.28	101,152.5	0.0000
2	Dec 1	18	5,817,663	6,526,216.68	0.86	52,598.68	0.0000
3	Dec 1	1	1,891,720	2,301,321.62	0.81	42,332.33	0.0000
4	Dec 1	5	1,114,238	1,418,944.71	0.78	37,219.81	0.0000
5	Dec 1	4	1,965,177	2,274,140.59	0.85	24,001.90	0.0000
6	Dec 1	2	945,695	1,169,284.65	0.80	23,886.30	0.0000
7	Dec 1	1	672,175	845,809.54	0.79	19,780.36	0.0000
8	Dec 1	6	882,793	1,034,449.98	0.85	12,161.49	0.0000
9	Dec 1	2	3,455,831	3,236,835.98	1.08	8,298.06	0.0000
10	Dec 1	1	1,844,383	1,700,649.87	1.09	6,333.58	0.0000
11	Dec 1	2	1,351,620	1,228,440.97	1.11	6,284.42	0.0000
12	Dec 1	1	535,773	465,366.48	1.15	5,172.76	0.0000
13	Dec 1	1	1,110,569	1,011,672.39	1.10	4,878.70	0.0000
14	Dec 1	1	423,426	485,990.37	0.87	4,288.31	0.0000
15	Dec 1	3	1,056,440	990,205.76	1.07	2,254.76	0.0000
16	Dec 1	3	2,193,108	2,263,168.50	0.97	1,198.58	0.0000
17	Dec 1	1	381,903	355,779.67	1.07	949.66	0.0000

groups, Beard et al. (15) proposed the COVID-19 monitoring method, which was based on spatiotemporal event sequence similarity. Hohl et al. (16) used prospective Poisson space-time scan statistic to detect daily clusters of COVID-19 at successive county levels in 48 states and Washington DC, which was helpful to facilitate decision-making and public health resource allocation. Pei et al. (17) found that the epidemic distribution had obvious space-time heterogeneity, and the spatial-temporal transmission had typical network characteristics.

In this paper, we will study the spatial aggregation of COVID-19 in the US from the following aspects. First, we construct a dynamic scanning window, calculate the relative risk to measure the intensity of aggregation, and utilize the scan statistical analysis through SaTScan9.5 based on the retrospective spatiotemporal analysis method. Second, we analyze the rational treatment of SaTScan9.5, and innovatively use circular distribution method (time latitude) and cluster analysis method (spatial latitude) to test the reliability of spatiotemporal scanning results. Through horizontal comparison, it is found that spatiotemporal scan analysis not only accurately measures in time but also reasonably divides regions according to characteristics in space. Finally, we take into account the data and how the COVID-19 pandemic changes on the ground, locating the gathering area and span period on time. At the same time, according to the scanning results, it not only provides an important theoretical basis for the relevant epidemic prevention work, but also has crucial importance for the establishment of an early warning system for the corresponding disease, ultimately playing a positive role in strengthening prevention and resolving the risk of major diseases in the world.

Methodology

Retrospective spatiotemporal analysis needs to build a scanning window to judge the number of diseases inside and outside the window. Since the scanning statistics involve time and space, the scanning window is in the form of a cylinder; the height of the cylinder represents the time, and the bottom area of the cylinder represents the area. The location and size of the scan window are dynamic, as it is unknown when and where the COVID-19 outbreak will occur.

In the analysis process, a position is randomly selected as the scanning center, and then, the cylindrical scanning window changes continuously. The cluster of geographic size of the scanning window ranges between zero and a predefined upper limit. There are several ways to determine the value of upper bound, for example, one can take the percentage of number of people at risk of disease or radius value of circle as the upper bound. In this article, we use the former method. The time length of the scan window specifies the maximum time frame according to the percentage of the entire study cycle or the specific number of days.

To determine the possibility of aggregation, the actual number of patients and the number of regional populations are calculated to obtain the theoretical number of patients, and the log-likelihood ratio (*LLR*) is constructed by using the actual and theoretical number of patients inside and outside the window; the relative risk ($\tilde{R}$) is calculated to evaluate the strength of aggregation. Since the scanning window undergoes a dynamic change, numerous scanning windows will be generated during the scanning process. For controlling the false-positive rate at a certain level, the window with the largest *LLR* is

TABLE 2 Numbers of cases and morbidities in the top 20 states.

State	Population	Case	Morbidity	Rank
Rhode Island	1,097,379	152,016	0.138526434	1
New York	20,201,249	2,060,920	0.102019435	2
Massachusetts	7,029,917	692,497	0.098507137	3
Delaware	989,948	94,921	0.095884834	4
New Jersey	9,288,994	859,151	0.092491286	5
South Carolina	5,118,425	466,698	0.091180002	6
Wisconsin	5,893,718	535,773	0.090905775	7
Kansas	2,937,880	264,696	0.090097621	8
Alaska	733,391	65,639	0.089500689	9
Hawaii	1,455,271	129,489	0.088979304	10
Utah	3,271,616	289,753	0.088565712	11
Illinois	12,812,508	1,110,569	0.086678502	12
Louisiana	4,657,757	399,318	0.085731823	13
Florida	21,538,187	1,844,383	0.085633159	14
North Carolina	10,439,388	884,922	0.084767613	15
Kentucky	4,505,836	381,903	0.084757412	16
West Virginia	1,793,716	151,977	0.08472746	17
Vermont	643,077	54,458	0.084683483	18
California	39,538,223	3,326,342	0.08412978	19
Arizona	7,151,502	600,864	0.084019273	20

TABLE 3 Results of circular distribution analysis.

Cluster	r	r_0	p
1	0.4392	0.0013	0.001
2	0.4769	0.0011	0.000
3	0.5398	0.0019	0.001

selected as the clustering area among all scanning windows. The statistical significance of *LLR* is tested by Monte Carlo stochastic simulation method.

We then give hypothesis test as follows:

Null Hypothesis (H_0): *The spatial and temporal distribution of newly confirmed cases of COVID-19 in the US is completely random;*

Alternative Hypothesis (H_1): *The spatial and temporal distribution of newly confirmed cases of COVID-19 in the US is not completely random.*

Assuming that the number of cases in window A is n_A, the population is m_A, $E(A)$ is the expected number of cases in the scanning window based on the original assumption and adjusted by covariates, the total number of cases in the total region is n_T, the total population is m_T, and the expected number of cases is $E(T)$, then

$$E(A) = \frac{n_I}{m_I} \times m_A \tag{1}$$

$$E(T) = \sum E(A) \tag{2}$$

The probability density function of specific points observed at region x is as follows:

$$f(x) = \begin{cases} \frac{pE(x)}{pE(A)+q[E(T)-E(A)]}, x \in A \\ \frac{qE(x)}{pE(A)+q[E(T)-E(A)]}, x \notin A \end{cases} \tag{3}$$

where p is the ratio of actual incidence to expected incidence in window A, q is the ratio of actual incidence to expected incidence outside window A, and the probability of any specific point is independent of all other points, one can also refer to Tang et al. (18) and Yang (19).

If $p > q$, the likelihood function $LR(A, p, q)$ is denoted by:

$$LR(A,p,q) = \frac{e^{-n_T}}{n_T!}\left(\frac{n_A}{E(A)}\right)^{n_A}\left(\frac{n_T - n_A}{E(T) - E(A)}\right)^{n_T - n_A}\prod_{x_i \in A} E(x_i) \tag{4}$$

Otherwise, the likelihood function LR_0(based on invalid hypothesis) is

$$LR_0 = \frac{e^{-n_T}}{n_T!}\left(\frac{n_T}{E(T)}\right)^{n_A}\prod_{x_i\in A}E(x_i) \quad (5)$$

Test statistic for spatiotemporal scan λ is defined as follows:

$$\lambda : = \frac{\underset{A,p>q}{Sup}\ LR(A,p,q)}{\underset{A,p=q}{Sup}\ LR(A,p,q)} \quad (6)$$

According to Equations (4) and (5), we have

$$\lambda = \underset{A}{Sup}\left(\frac{n_A}{E(A)}\right)^{n_A}\left(\frac{n_T-n_A}{E(T)-E(A)}\right)^{n_T-n_A}\left(\frac{n_T}{E(T)}\right)^{-n_T}I(B) \quad (7)$$

$$B : = \left\{\frac{n_A}{E(A)} > \frac{n_T-n_A}{E(T)-E(A)}\right\} \quad (8)$$

In formula (7), $I(\cdot)$ is a characteristic function. The ratio of the actual incidence to the expected incidence in window A is greater than the ratio of the actual incidence to the expected incidence outside window A. The is a measure of how risk within a cylinder differs from risk outside.

Next, we use Monte Carlo random method to simulate the p value of LLR to determine whether the aggregation is statistically significant. First, we simulate c random datasets, calculate the maximum LLR for each dataset, and rank it with the real LLR from big to small. If the real value rank is R, then we have

$$p = R(c+1)^{-1} \quad (9)$$

If $p < 0.05$, we reject the original assumption. The relative risk of each aggregation is as follows:

$$\tilde{R} = \left(\frac{n_A}{E(A)}\right)\left(\frac{n_T-n_A}{E(T)-E(A)}\right)^{-1} \quad (10)$$

Empirical analysis

Source of the data

This paper selects 50 states from the US for research. The data of the COVID-19 mainly came from the data of the New York Times, including the date of diagnosis, the current area, and the source of infection of patients with COVID-19. Demographic data mainly came from the US 2020 census data, the basic geographic information data of each state were derived from Google satellite map data, and the latitude and longitude coordinates mainly chose the state capital as the center position.

The variant data are derived from the Centers for Disease Control and Prevention (CDC) study that tracks the proportion of variants estimated from weekly random sampling in the Department of Health and Human Services region followed by gene sequencing tests across the region, which we use to estimate the number of variant infections in the state over a week with new cases per day.

Parameter setting

We use the retrospective spatiotemporal analysis method, and choose the discrete Poisson model. The scanning time is set to be from 1 December 2021 to 31 January 2022, and the time interval is 1 day. As COVID-19 is highly contagious, the population with a ceiling of 50% in the space window is at risk, and the maximum circle size file is set at 30% of the population, rather than 30% of the regular population, and the regional overlap is set at zero. Referring to a large number of relevant literatures, combined with the actual situation, it is known that the outbreak of COVID-19 is fast, and the incubation period is short. Besides, the inaction of the US government to manage the outbreak makes the cycle longer. The daily pattern of COVID-19 changes rapidly, so the minimum time cluster is set to 1 day. In the test window, the number of Monte Carlo random simulation is set to 999.

Description analysis

In the population distribution, the US COVID-19 has nothing to do with gender, and included patients mainly in the age group of 44 to 59 years. In terms of time distribution, the US had the largest number of new cases on 10th January, with 1,420,374 cases. On 3rd, 18th, and 24th January, more than 1 million new cases were added daily with 1,003,751, 1,173,885, and 1,025,999 cases, respectively. In terms of the overall trend, the outbreak in the early stages of each state is relatively serious, and the number of confirmed cases has experienced a short lag and rapid growth. From Figure 1, we can see that the overall epidemic situation has not been effectively controlled, so the number of confirmed cases has increased cumulatively, having a certain increasing tendency. In the regional distribution, cases were mainly concentrated in the east and west of the US, and California has the largest number of confirmed cases, followed by New York.

Space-time analysis

A retrospective spatiotemporal analysis is carried out in the US. After SaTScan 9.5 is run, 17 clustering areas are obtained and arranged from large to small according to the log-likelihood

ratio, and we obtain $p < 0.01$. The clustering areas are tested by the aboriginality test. The specific data are summarized in Table 1.

The cluster areas are mainly concentrated in Connecticut, Rhode Island, New York, Massachusetts, New Hampshire, and New Jersey from 1 December 2021 to 31 January 2022. The log-likelihood ratio is 101,152.56, and the aggregation is the highest, with a relative risk of 1.28. It also shows that the aggregation of COVID-19 in the six places during this period is strong. From 1 December 2021 to 31 January 2022, 18 states, such as Colorado, become the second agglomeration, with a log-likelihood ratio of 52,598.68 and a relative risk of 0.86. Texas from 1 December 2021 to 31 January 2022 is one of the three types of gathering areas, with a log-likelihood ratio of 42,332.33 and a relative risk of 0.81.

Combined with the daily incidence of each state, it can be observed that the starting time of the gathering area is just the time for the sudden increase of the confirmed cases of COVID-19 in the region, and the end time is the time for the growth rate of the confirmed cases to begin to decline. Combined with Table 2, the incidence of the four states involved, Rhode Island, New York, Massachusetts, and New Jersey, accounts for the top five regions of the incidence of COVID-19 in the US, and New York is the city with the second largest number of confirmed cases. Although Massachusetts and New Hampshire have fewer confirmed cases than New York, they are geographically close to New York, where the epidemic is relatively serious.

Circular distribution analysis

Since the research time is 62 days, we divide 360° evenly over each day, then 1 day is equivalent to 5.81°, and 1 h is equivalent to 0.21°. To avoid the infinite calculation, the calculation time of

TABLE 4 Peak day and peak period of incidence of COVID-19 in each cluster area.

Cluster	$\overline{\alpha}$	s	Peak incidence	Epidemic peak period	Peak period span (Day)
1	32.7887	73.5005	Dec 6	Dec 6–Dec19	14
2	87.1093	69.7240	Dec 16	Dec 4–Dec28	25
3	77.1640	63.6210	Dec 14	Dec 3–Dec24	12

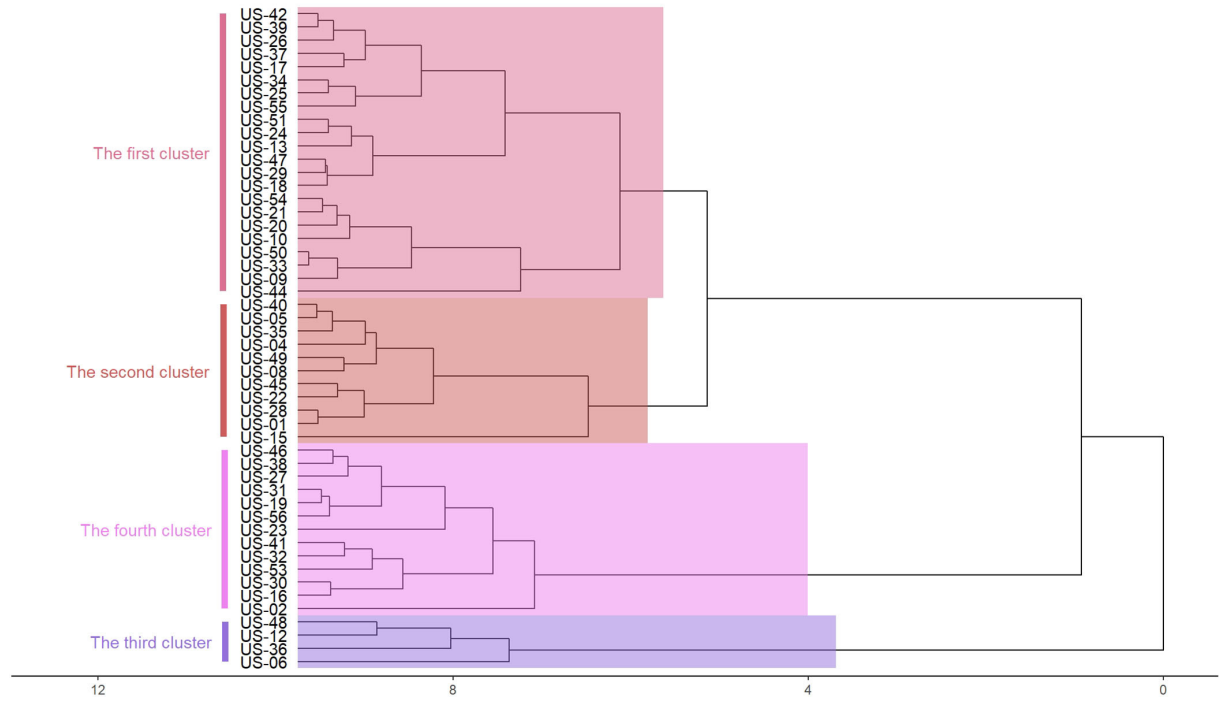


FIGURE 2
Coordinated hierarchical cluster analysis.

TABLE 5 Correspondence tables of states.

Code	State	Code	State
US-01	Alabama	US-30	Montana
US-02	Alaska	US-31	Nebraska
US-04	Arizona	US-32	Nevada
US-05	Arkansas	US-33	New Hampshire
US-06	California	US-34	New Jersey
US-08	Colorado	US-35	New Mexico
US-09	Connecticut	US-36	New York
US-10	Delaware	US-37	North Carolina
US-12	Florida	US-38	North Dakota
US-13	Georgia	US-39	Ohio
US-15	Hawaii	US-40	Oklahoma
US-16	Idaho	US-41	Oregon
US-17	Illinois	US-42	Pennsylvania
US-18	Indiana	US-44	Rhode Island
US-19	Iowa	US-45	South Carolina
US-20	Kansas	US-46	South Dakota
US-21	Kentucky	US-47	Tennessee
US-22	Louisiana	US-48	Texas
US-23	Maine	US-49	Utah
US-24	Maryland	US-50	Vermont
US-25	Massachusetts	US-51	Virginia
US-26	Michigan	US-53	Washington
US-27	Minnesota	US-54	West Virginia
US-28	Mississippi	US-55	Wisconsin
US-29	Missouri	US-56	Wyoming

each day is 8:00 a.m., that is, the one-third corresponding degree of 1 day is taken as the degree of the day. By the spatiotemporal scanning analysis, we obtain 17 clustering areas, and take the first three clustering areas as example. In order to compare the following analysis results to previous ones, we combine the daily newly confirmed cases according to the clusters. The r, r_0, and p-values of Rayleigh test are obtained through calculation, as summarized in Table 3.

The peak day and peaks of each cluster are summarized in Table 4.

Clustering analysis

Hierarchical cluster analysis method is commonly used in classification research. This method can overcome the shortcomings of qualitative classification. According to the index characteristics of the classification object, the total feature similarity is divided into a class. In this case, the cumulative confirmed cases, the regional population, and the incidence rate are used as the indicators of each region, and imported into R software for standardization. The deviation square and clustering analysis are used to divide them into four categories. Since the latitude and longitude coordinates are involved in the spatiotemporal scanning analysis, the central coordinates of the capital are added to the index, as shown in Figure 2. Due to the mess up of text and pictures as displayed in the diagram, it should be replaced with a geographical code (US-01), as shown in Table 5.

From Figure 2, we can see that there are highly correlated with geographical location. The same category of states are adjacent states, and the case information is not reflected. Therefore, the clustering method cannot well-balance the relationship between the number of cases and their geographical locations.

Results comparison

The peak periods calculated by spatial-temporal scanning analysis are compared with those calculated by circular distribution method, as shown in Table 6. Combined with the actual situation, the peak period of the disease obtained by the circular distribution method is similar to the epidemic situation of COVID-19 in the region. The peak period of the disease obtained by the spatial-temporal scanning method is longer than that of the circular distribution method, and the time is generally advanced. Spatiotemporal scanning analysis can send early warning signals for COVID-19, which is a kind of fulminant and fast infectious disease, and has a higher practical value for disease prevention and control.

The clustering areas obtained by spatiotemporal scanning analysis are compared with the classification results obtained by the system clustering method, as shown in Table 7. In the Spatiotemporal Scanning Analysis (STSA), Hierarchical Cluster Analysis (HCA), and Coordinated Hierarchical Cluster Analysis (CHCA), US-44 (Rhode Island), US-25 (Massachusetts), US-09 (Connecticut), US-34 (New Jersey), and US-33 (New Hampshire) are classified into the first category, while US-36 (New York) is classified into the first category by spatiotemporal scanning. Combined with the actual situation, it can be seen that the results have a great relationship with the cumulative confirmed cases. After adding the coordinate index, the results are highly correlated with the geographical location, and the case information is weakened. The spatiotemporal scanning method makes good use of the information of regional population, case information, geographical location, and other information to give a reasonable clustering area. In terms of disease prevention and control, spatiotemporal scanning method can better provide theoretical basis for its adaptation to local conditions.

Through the abovementioned comparative analysis, it can be seen that the circular distribution method and the space-time scan method have a certain overlap interval in the peak

TABLE 6 Comparison of STSA and CDA.

Cluster	Fastigium (STSA)	Span (Day)	Fastigium (CDA)	Span (Day)
1	Dec 1–Jan 31	62	Dec 6–Dec 19	14
2	Dec 1–Jan 31	62	Dec 4–Dec 28	25
3	Dec 1–Jan 31	62	Dec 3–Dec 24	12

TABLE 7 Comparison of STSA and (C)HCA.

Cluster	State (STSA)	State (HCA)	State (CHCA)
1	US-09 US-44 US-36 US-25 US-33 US-34	US-15 US-02 US-10 US-54 US-33 US-50 US-38 US-35 US-28 US-40 US-09 US-05 US-21 US-22 US-55 US-20 US-45 US-49 US-04 US-34 US-25 US-44	US-09 US-10 US-13 US-17 US-18 US-20 US-21 US-24 US-25 US-26 US-29 US-33 US-34 US-37 US-39 US-42 US-44 US-47 US-50 US-51 US-54 US-55
2	US-08 US-56 US-35 US-49 US-46 US-31 US-20 US-40 US-38 US-04 US-30 US-19 US-16 US-29 US-27 US-48 US-05 US-32	US-08 US-01 US-27 US-29 US-47 US-18 US-53 US-51 US-13 US-37 US-26 US-17 US-42 US-39	US-01 US-04 US-05 US-08 US-15 US-22 US-28 US-35 US-40 US-45 US-49
3	US-48	US-36 US-12 US-48 US-06	US-12 US-36 US-06 US-48
4	US-53 US-41 US-16 US-30 US-32	US-32 US-19 US-41 US-24 US-46 US-31 US-56 US-23 US-16 US-30	US-02 US-16 US-19 US-23 US-27 US-30 US-31 US-32 US-38 US-41 US-46 US-53 US-56

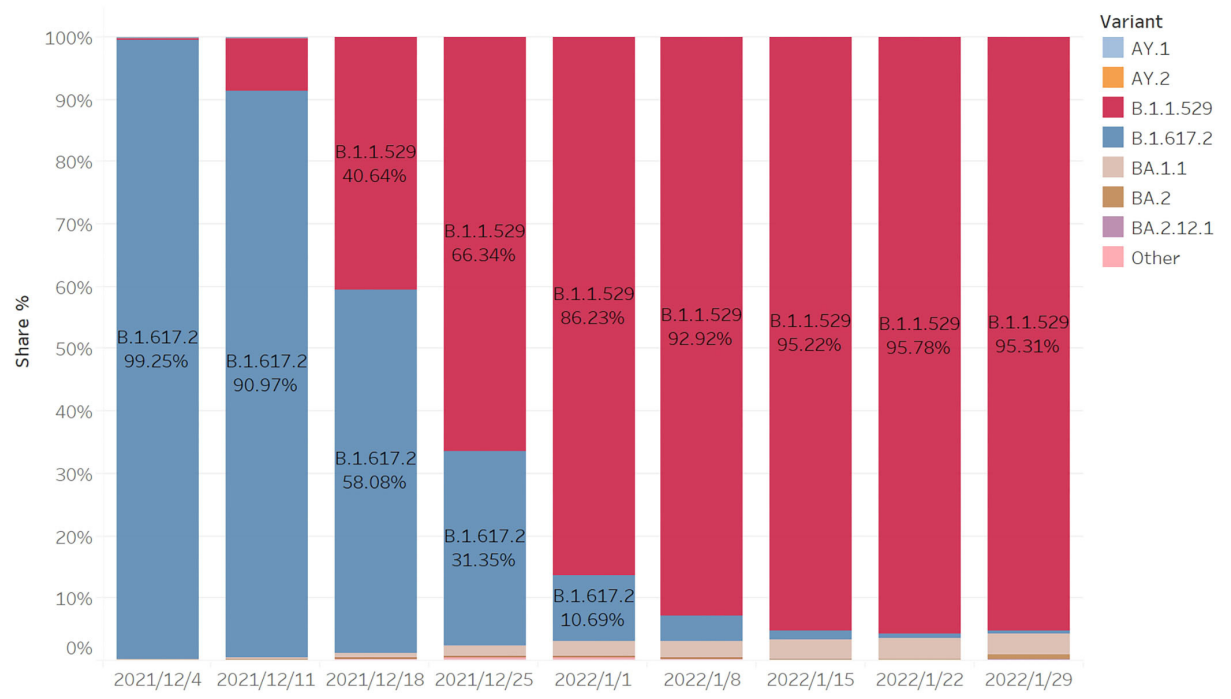


FIGURE 3
Proportion of COVID-19 variants.

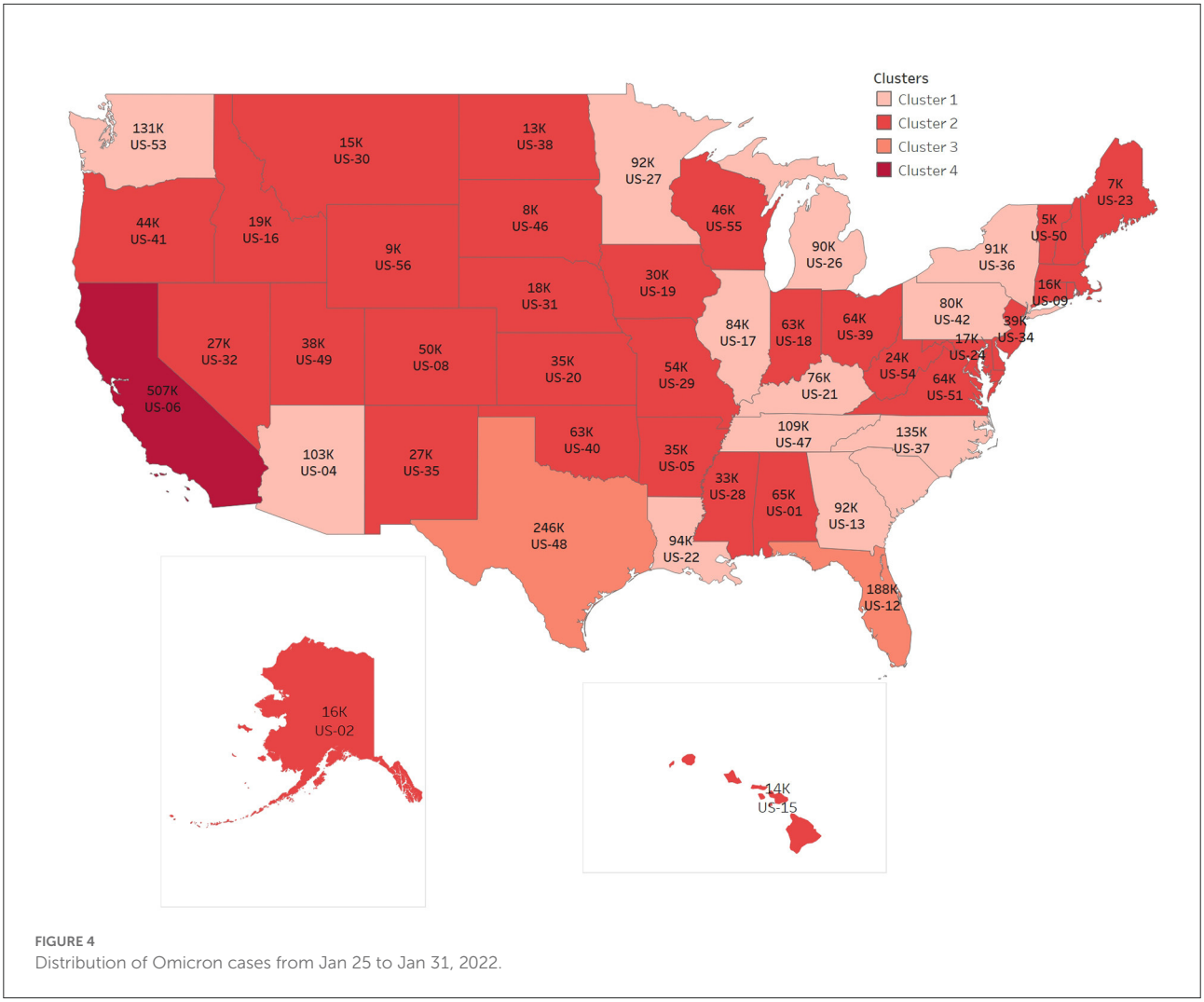


FIGURE 4
Distribution of Omicron cases from Jan 25 to Jan 31, 2022.

period of disease onset, and the clustering analysis method is certainly similar with its regional aggregation. However, the spatiotemporal scanning method can provide early warning and make better use of geographical factors to determine disease outbreak areas in detail, which is more instructive for the early warning and prevention and control of COVID-19.

The spatiotemporal scanning method can provide more objective grouping basis for the further model establishment of related research. According to the epidemic law of different regions, different groups can be included in different covariate modeling. The qualitative and quantitative research on the influencing factors of COVID-19 will provide an important basis for the development of effective epidemic prevention measures by health institutions such as disease control centers in the region by analyzing the incidence characteristics of patients with COVID-19 in different regions and at different times, and combining the economic level, population flow, medical conditions, and other factors in the region.

Omicron variation

In the study of infectious diseases, we cannot ignore the situation of some variants. Based on the time node selected in this paper, the first Omicron case was reported in the US on 1 December, so we are paying attention to Omicron at this stage. Next, we need to know more about Omicron. In fact, Omicron has a significant growth advantage over Delta, leading to rapid spread in the community with higher levels of incidence than previously seen in this pandemic. With the sharp increase of cases and the scarcity of medical resources, we should also give importance to its dissemination.

From Figure 3, we can see that the B.167.2 (Delta) accounted for 99.25% on 4 December, while B.1.1.529 (Omicron) accounted for a low proportion. After 2 weeks, the proportion of Omicron increased rapidly, reaching 40.64%, while the corresponding Delta decreased to 58.08%. After another week, the proportion of Omicron exceeded Delta, becoming the largest

variant of infection. After the following 5 weeks, the proportion reached 95.31%. Within 2 months, Omicron became the mutant with the largest proportion of infection, and its propagation speed was very fast.

Based on the CDC's tracking data of variants and the prediction of the proportion of variants, we calculate the number of variants per week in different regions according to the new cases per day and the proportion of variants per week in the corresponding region. The following figure clearly shows the cumulative number of Omicron cases in the last week. In order to show the map integrally, US-02 and US-15 have changed the actual location in the map. According to the number clustering, the map is divided into four categories. We can find the features in Figure 4.

The numbers of items in clusters 1–4 are 18, 29, 2, 1, respectively, and the cluster centers are 89,703, 24,703, 217,010, 507,330 respectively, in Figure 4. The number of Omicron cases increases rapidly in 2 months, with the largest number of cases in one state, US-06, accumulating to 507,330 per week, and also with the largest number in neighboring states.

To summarize, it can be clearly seen that the rate of infection of Omicron increased rapidly, and there is a trend of diffusion from the middle to the surrounding. Population flow is one of the reasons for the rapid spread of virus. The reality of the spread from densely populated cities to other cities can also be observed from the distribution map of Omicron cases.

Conclusion

In this paper, a retrospective spatiotemporal analysis of confirmed cases of COVID-19 in 50 states of the US is carried out. The first cluster is Connecticut, Rhode Island, New York, Massachusetts, New Hampshire, and New Jersey. The second cluster comprises 18 states, and the three types of gathering area is Texas. Through observation, it can be seen that the geographical location of the capital belonging to the same type of gathering area is relatively close. There is minimal difference between the gathering time and the peak time of newly confirmed cases daily, and the incidence of prominent gathering areas is higher. The reliability test of space-time scan results show that space-time scan has the advantages of accurate measurement in time and reasonable division of regions according to characteristics in space. On the basis of making full use of the existing time and spatial information, a spatiotemporal scanning analysis accurately locates the clustering area, timing and quantifying the corresponding clustering area, and evaluating the risk degree of the region, as we know that a high level of economic development and perfect medical conditions have played a positive role in the recovery of patients. From the analysis of this paper, spatiotemporal scanning analysis has greatly improved the timeliness and effectiveness of early warning of diseases, and can provide scientific basis for early prevention and control of diseases.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

YZ conceived and designed the study. QL analyzed the data. YZ and QL contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Naus J. The distribution of the size of the maximum cluster of points on a line. *J Am Stat Assoc.* (1965) 60:532–8. doi: 10.2307/2282688

2. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *J Royal Statist Soc.* (2001) 164:61–72. doi: 10.2307/2680534

3. Kulldorff M, Heffernan R, Hartman J, et al. A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Med.* (2005) 2:e59. doi: 10.1371/journal.pmed.0020059

4. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med.* (2006) 25:3929–43. doi: 10.1002/sim.2490

5. Kulldorff M, Mostashari F, Duczmal L, Katherine Yih W, Kleinman K, Platt R. Multivariate scan statistic for disease surveillance. *Stat Med.* (2007) 26:1824–33. doi: 10.1002/sim.2818

6. Jung I, Kulldorff M, Klassen A, A. spatial scan statistic for ordinal data. *Stat Med.* (2007) 26:1594–607. doi: 10.1002/sim.2607

7. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Stat Med.* (2010) 29:1910–8. doi: 10.1002/sim.3951

8. Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. *Biometrics.* (2007) 63:109–18. doi: 10.1111/j.1541-0420.2006.00661.x

9. Huang L, Tiwari RC, Zou Z, Kulldorff M, Feuer EJ. Weighted normal spatial scan statistic for heterogeneous population data. *J Am Stat Assoc.* (2009) 104:886–98. doi: 10.1198/jasa.2009.ap07613

10. Wieckowska B, Górna I, Trojanowski M, Pruciak A, Stawińska-Witoszyńska B. Searching for space-time clusters: the CutL method compared to Kulldorff's scan statistic. *Geospatial Health.* (2019) 14:314–20. doi: 10.4081/gh.2019.791

11. Li D, Fang Z, Yu Y. Scan statistic: a new method to detect the persistence of fund performance. *Operat Res Manag.* (2006) 15:82–7. doi: 10.3969/j.issn.1007-3221.2006.01.018

12. Yin F. *Application of Spatio-Temporal Scan Statistic in Early Warning of Infectious Diseases*. Chengdu: Sichuan University (2007).

13. Ma Y, Li X, Zhang Y. Spatial scale selection of scan statistic in infectious disease surveillance. *Modern Prevent Med.* (2011) 38:1601–4.

14. Hohl A, Delmelle EM, Desjardins MR, Lan Y. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spat Spatiotemporal Epidemiol.* (2020) 34:100354. doi: 10.1016/j.sste.2020.100354

15. Xu FY, Beard K. A comparison of prospective space-time scan statistic and spatiotemporal event sequence based clustering for COVID-19 surveillance. *PLoS ONE.* (2021) 16:e0252990. doi: 10.1371/journal.pone.0252990

16. Hohl A, Delmelle E, Desjardins M. Rapid detection of COVID-19 clusters in the United States using a prospective space-time scan statistic: an update. *SIGSPATIAL Special.* (2020) 12:27–33. doi: 10.1145/3404111.3404116

17. Pei T, Wang X, Song C, Liu Y, Huang Q, Shu H, et al. Research progress on spatiotemporal analysis and modeling of COVID-19 epidemic. *J Geo-Inform Sci.* (2021) 23:188–210. doi: 10.12082/dqxxkx.2021.200434

18. Tang X, Zhou H. Scanning statistics and its application in epidemiology. *China Health Stat.* (2011) 28:332–7. doi: 10.3969/j.issn.1002-3674.2011.03.042

19. Yang L. *The Distribution Characteristics of COVID-19 in Wenzhou City Based on Scan Statistic*. Hangzhou: Hangzhou Dianzi University (2020).

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 12 September 2022
DOI 10.3389/fpubh.2022.993745

OPEN ACCESS

EDITED BY
Khalid Hattaf,
Centre Régional des Métiers de l'Education et de la Formation (CRMEF), Morocco

REVIEWED BY
Ahmed Elaiw,
King Abdulaziz University, Saudi Arabia
Ahmed Mohsen,
University of Baghdad, Iraq

*CORRESPONDENCE
Chang Hyeong Lee
chlee@unist.ac.kr

SPECIALTY SECTION
This article was submitted to Infectious Diseases–Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

RECEIVED 14 July 2022
ACCEPTED 23 August 2022
PUBLISHED 12 September 2022

CITATION
Kim JE, Choi H, Choi Y and Lee CH (2022) The economic impact of COVID-19 interventions: A mathematical modeling approach. *Front. Public Health* 10:993745. doi: 10.3389/fpubh.2022.993745

The economic impact of COVID-19 interventions: A mathematical modeling approach

Jung Eun Kim[1], Heejin Choi[1], Yongin Choi[2] and Chang Hyeong Lee[1]*

[1]Department of Mathematical Sciences, Ulsan National Institute of Science and Technology, Ulsan, South Korea, [2]Busan Center for Medical Mathematics, National Institute of Mathematical Sciences, Daejeon, South Korea

Prior to vaccination or drug treatment, non-pharmaceutical interventions were almost the only way to control the coronavirus disease 2019 (COVID-19) epidemic. After vaccines were developed, effective vaccination strategies became important. The prolonged COVID-19 pandemic has caused enormous economic losses worldwide. As such, it is necessary to estimate the economic effects of control policies, including non-pharmaceutical interventions and vaccination strategies. We estimated the costs associated with COVID-19 according to different vaccination rollout speeds and social distancing levels and investigated effective control strategies for cost minimization. Age-structured mathematical models were developed and used to study disease transmission epidemiology. Using these models, we estimated the actual costs due to COVID-19, considering costs associated with medical care, lost wages, death, vaccination, and gross domestic product (GDP) losses due to social distancing. The lower the social distancing (SD) level, the more important the vaccination rollout speed. SD level 1 was cost-effective under fast rollout speeds, but SD level 2 was more effective for slow rollout speeds. If the vaccine rollout rate is fast enough, even implementing SD level 1 will be cost effective and can control the number of critically ill patients and deaths. If social distancing is maintained at level 2 at the beginning and then relaxed when sufficient vaccinations have been administered, economic costs can be reduced while maintaining the number of patients with severe symptoms below the intensive care unit (ICU) capacity. Korea has wellequipped medical facilities and infrastructure for rapid vaccination, and the public's desire for vaccination is high. In this case, the speed of vaccine supply is an important factor in controlling the COVID-19 epidemic. If the speed of vaccination is fast, it is possible to maintain a low level of social distancing without a significant increase in the number of deaths and hospitalized patients with severe symptoms, and the corresponding costs can be reduced.

KEYWORDS

COVID-19, mathematical model, cost estimation, vaccination, social distancing

Introduction

The outbreak of coronavirus disease 2019 (COVID-19) occurred in 2019, and it is still affecting the world in 2022. When there was no effective vaccine or treatment in the early stage of COVID-19, the only control strategy against the disease involved non-pharmaceutical interventions, such as social distancing and lockdowns.

The method of implementing these non-pharmaceutical interventions varied from country to country, but most such strategies have had a significant effect in reducing the number of confirmed cases, severe cases, and deaths. However, as the spread of COVID-19 became prolonged, the economic damage caused by the non-pharmaceutical interventions increased significantly. Thus, when implementing an effective control policy, it is very important to estimate not only the effect on the reduction of confirmed cases, severe cases, and deaths, but also the economic damage.

In Korea, the vaccination rate has steadily increased since vaccination began on 26 February 2021 and, as of June 2022, 86% of the total population had completed their second vaccination, while the number with a third vaccination has reached 65% of the total population (1). However, when the vaccines were being introduced, vaccination was not implemented quickly due to supply restrictions, and the number of vaccinations only increased rapidly after June 2021. Koreans have a favorable attitude toward vaccination and possess the infrastructure for large-scale vaccination. This study investigates the effect of the vaccination rollout speed at the beginning of vaccination, based on various scenarios.

Estimation of the time-dependent dynamics of the number of confirmed cases, severe cases, and deaths due to COVID-19, as well as economic damage, can be conducted using mathematical modeling, which has been previously used to describe the transmission of COVID-19. Based on an age-structured compartmental model, the effects of non-pharmaceutical interventions (NPIs) and vaccination policies have been studied (2–4). Estimation of the number of confirmed cases has been investigated through simulation of various scenarios based on the vaccination plan and social distancing (5–7). In some studies (8, 9), the cost-effectiveness of control strategies against COVID-19 has been studied through the use of mathematical models. As the vaccines against COVID-19 were developed at the end of 2020, cost-effective control including vaccination has become important and cost-effective control strategies based on different vaccine allocation policies have been studied (10, 11). There have been studies (12–14) using mathematical models that considered mild, hospitalized, and critical symptom cases without an age structure. However, since the important parameters of COVID-19 such as transmission rate, severity and death rate are closely related to age (1, 7, 15), it is necessary to apply an age-structured model for analyzing the effect of the transmission and control of the disease. In order to study the economic effect of COVID-19, we constructed a mathematical model considering both the age-specific structure and the subdivided severity.

In this paper, we use age-specific mathematical models to estimate the cost of COVID-19 in terms of medical expenses, wage loss, cost due to deaths, vaccination cost, and gross domestic product (GDP) loss. Efficient policies, according to vaccination rollout speed and social distancing levels, are investigated by considering both cost reduction and the control of the number of patients with severe symptoms. We also investigate the mitigation effect of social distancing policies on the total cost due to COVID-19 and the number of critically ill patients.

Methods

Epidemiological data

Since the first confirmed case was reported on 19 January 2020, Korea has steadily reported cases of COVID-19. Figure 1A shows the daily number of confirmed cases, by age group, from 1 February 2020 to 31 December 2021. In November 2021, the vaccination coverage rate approached 80% of the total population, and the social distancing phase-easing policy was implemented on 1 November 2021. However, this resulted in a significant increase in the daily number of confirmed cases. In particular, the number of confirmed cases in patients under the age of 18 who were not vaccinated increased significantly. In addition, the number of confirmed cases has increased rapidly due to the prevalence of the Omicron variant and waning vaccine-based immunity. As of 17 March 2022, the number of daily confirmed cases exceeded 600,000 (1). Figure 1B shows the first, second, and third vaccination doses per day from 26 February 2021, when the vaccination program was started in Korea. It shows that, until May 2021, the daily dose was relatively small, due to the limited vaccine supply. A significant increase in the number of vaccinations was observed after July 2021, which can be seen as indicating the favorable reception of the Korean public for vaccination and the sufficient vaccination infrastructure. The third vaccination was also actively promoted and, as of May 2022, more than 65% of the population received booster shots (1).

In this work, we investigated the importance of the vaccination rollout speed for disease outbreak control. The horizontal dashed lines in the figure indicate the daily dose ($\nu = 100000, 200000, \ldots, 500000$) used for model simulation. The maximum daily dose ($\nu = 500{,}000$) was set based on the average weekly second vaccination doses in Korea. We also considered the social distancing levels implemented in Korea in order to determine the effect of non-pharmaceutical intervention policies. Details of the social distancing polices used in Korea are given in Supplementary material section 1.

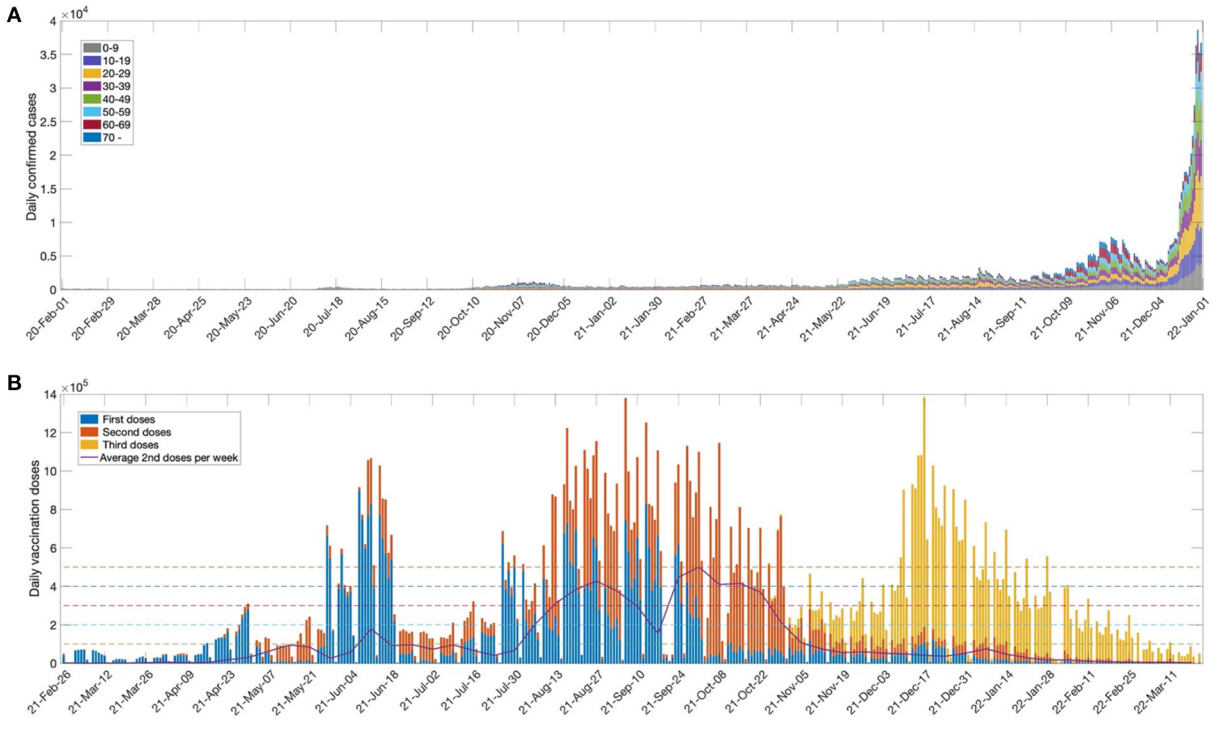


FIGURE 1
(A) Number of weekly confirmed cases for age groups (0–9,10–19,20–29,30–39,40–49,50–59,60–69, 70+) and **(B)** Daily first, second and third vaccination doses. The solid curve indicates average second doses per week. The horizontal dashed lines represent daily vaccination doses (v=100000, 200000, . . ., 500000) used in the model simulation.

Mathematical model

We developed an age-structured mathematical model to describe the transmission dynamics of COVID-19 with vaccination. In the model, the population was separated into compartments, based on their characteristics, for each age group i: $S_i(t)$ denotes susceptible, $V_i(t)$ denotes completely vaccinated, $E_i(t)$ denotes exposed, $A_i(t)$ denotes asymptomatic, $I_i(t)$ denotes infectious, $H_i^M(t)$ denotes having mild symptoms, $H_i^H(t)$ denotes hospitalized without intensive care, $H_i^I(t)$ denotes hospitalized with intensive care, $R_i(t)$ denotes recovered, and $D_i(t)$ denotes deceased. The H^M, H^H, and H^I groups correspond to the highest severity level of an individual during the quarantine period. The age classes $i = 1, 2, \cdots, 8$ represent individuals aged $0-9$, $10-19$, $20-29$, $30-39$, $40-49$, $50-59$, $60-69$, and older than 70 years, respectively. A schematic diagram of the model is shown in Figure 2. The system of differential equations that describes the model dynamics is as follows:

$$
\begin{aligned}
\dot{S}_i &= -\Lambda_i S_i - \phi_i v \\
\dot{V}_i &= \phi_i v - (1-\tau)\Lambda_i V_i \\
\dot{E}_i &= \Lambda_i S_i + (1-\tau)\Lambda_i V_i - \alpha E_i \\
\dot{A}_i &= \rho\alpha E_i - \gamma^A A_i \\
\dot{I}_i &= (1-\rho)\alpha E_i - qI_i \\
\dot{H_i^M} &= q\delta_i^M I_i - \gamma_i^M H_i^M \\
\dot{H_i^H} &= q\delta_i^H I_i - \gamma_i^H H_i^H \\
\dot{H_i^I} &= q\delta_i^I I_i - \left(1-\kappa_i^I\right)\gamma_i^I H_i^I - \kappa_i^I \gamma_i^I H_i^I \\
\dot{R}_i &= \gamma^A A_i + \gamma_i^M H_i^M + \gamma_i^H H_i^H + \left(1-\kappa_i^I\right)\gamma_i^I H_i^I \\
\dot{D}_i &= \kappa_i^I \gamma_i^I H_i^I
\end{aligned}
$$

Where the force of infection for age group i is obtained by $\Lambda_i = \sum_j \left[\beta_{ij}\left(I_j + \theta_A A_j\right)/N_j\right]$ with the disease transmission rates β_{ij}, for $N_j = S_j + V_j + E_j + A_j + I_j + R_j$.

The disease transmission rates (β_{ij}) of a person in age group i per contact for each age group j are estimated as 8×8 matrices, determined by minimizing the difference between confirmed case data and the simulated results from the model (in the least-squares sense) for social distancing levels (LV) 0, 1, 2, and 3 in Korea. The transmission rate matrices and the data-fitting results for all ages are provided in the Supplementary material section 2. Other parameters used in the model and the baseline values are given in Table 1.

The effective reproduction number (R_t) measures the average number of secondary cases per infectious

10.3389/fpubh.2022.993745

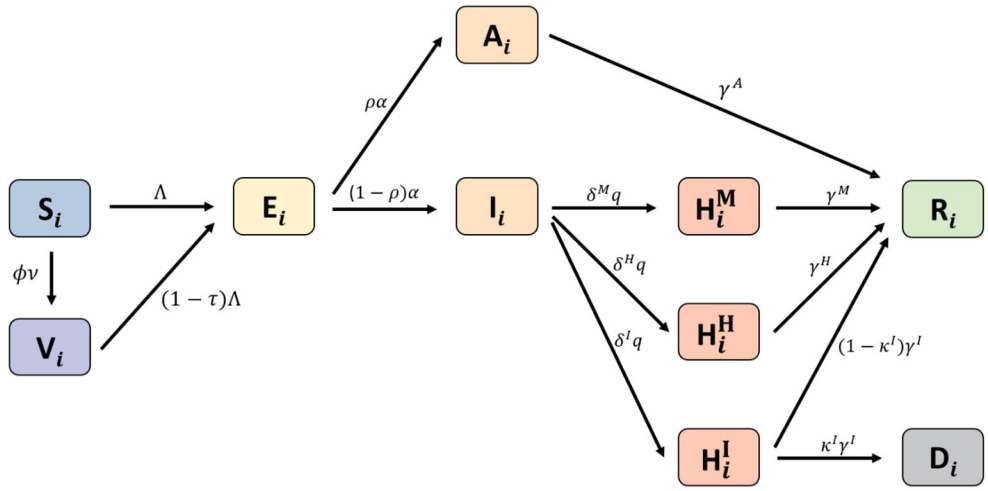


FIGURE 2
Schematic diagram for the proposed model.

TABLE 1 Parameter definitions and baseline values used in numerical simulations.

Parameter	Description	Value	Reference
Λ_i	The force of infection for age group i		Estimated
β_{ik}	Transmission rate from age group k to i	Given in Supplementary material section 2	Estimated
ϕ_i	Vaccination allocation vector	Vary	
ν	Daily vaccination doses	Vary	
τ	Vaccine efficacy	0.79, vary	(7)
ρ	Probability of unconfirmed asymptomatic cases	0.16	(7)
$1/\alpha$	Latent period	5.2	(16)
δ_i^M	Probability of cases having mild symptoms	1, 0.996, 0.991, 0.984, 0.971, 0.924, 0.854, 0.694	(15)
δ_i^H	Probability of hospitalization without intensive care	0, 0.002, 0.007, 0.012, 0.027, 0.068, 0.123, 0.275	(15)
δ_i^I	Probability of hospitalization with intensive care	$1-\delta_i^M-\delta_i^H$	
$1/q$	Mean duration of the case confirmation	3	(6)
$1/\gamma^A$	Recovery period of asymptomatic cases	3.5	(7)
$1/\gamma^M$	Recovery (or quarantine) period of mild symptom cases	14 (treatment center) 7 (home treatment)	(1)
$1/\gamma_i^H$	Recovery period of hospitalized without intensive care cases	15.32, 15.99, 18.66, 17.70, 17.84, 18.44, 19.77, 23.79	(1)
$1/\gamma_i^I$	Recovery period of hospitalized with intensive care cases	15.32, 15.99, 18.66, 17.70, 17.84, 18.44, 19.77, 23.79	(1)
μ_i	Death rate of groups in confirmed cases	0, 0, 0.0001, 0.0004,0.0007, 0.0027, 0.0108, 0.0513	(17)
r^H	Probability of death from hospitalization without intensive care	0.4216	Estimated
r^I	Probability of death from hospitalization with intensive care	0.5786	Estimated
κ^I	Mortality rate of hospitalization with intensive care cases	μ/δ^I	
θ_A	Relative infectiousness of asymptomatic infections	0.51	(18)

individual at time t, which is obtained by calculating the spectral radius of the next-generation matrix. The details of the derivation of R_t of the model are given in Supplementary material section 3.

Cost estimation

The total cost due to COVID-19 was estimated in terms of medical expenses, wage loss, cost due to deaths, vaccination cost,

TABLE 2 Formulae for the cost estimation.

Cost factor	Formula
Medical expenses	
Mild symptom case (H^M)	$\sum_{i=1}^{8}$[Average daily cost of treatment for mild patients (C_M) × Recovery period of mild symptom cases $\left(1/\gamma_i^M\right)$ × Number of mild patients $\left(H_i^M\right)$]
Hospitalized case without intensive care (H^H)	$\sum_{i=1}^{8}$[Average daily cost of treatment for hospitalized patients without intensive care (C_H) × Recovery period of hospitalized patients without intensive care $\left(1/\gamma_i^H\right)$ × Number of hospitalized patients without intensive care $\left(H_i^H\right)$]
Hospitalized case with intensive care (H^I)	$\sum_{i=1}^{8}$[Average daily cost of treatment for hospitalized patients with intensive care (C_I) × Recovery period of hospitalized patients with intensive care $\left(1/\gamma_i^I\right)$ × Number of hospitalized patients with intensive care $\left(H_i^I\right)$]
Wage loss	
Older than 20 years	$\sum_{x=M,H,I}$[$\sum_{i=1}^{8}$[Average daily income in age group i (W_i) × Employment rate in age group i (E_i) × Recovery period of cases $\left(1/\gamma_i^x\right)$ × Number of patients $\left(H_i^x\right)$]]
Younger than 19 years	Average daily income of women in their 30s and 40s $\left(W_f\right)$ × Female employment rate of children younger than 19 years $\left(E_f\right)$ × Average recovery period × Number of patients younger than 19 years
Death	Average funeral cost (FC)+Present value of the predicted future income (PV) $\left(PV = \sum_{j=1}^{N_i} \frac{W_i(1+g)^j}{(1+r)^j}\right)$
Vaccination cost	[Vaccination cost per person (VC) + Vaccination procedure cost (PC) + Logistics cost (LC)] × Population × Vaccination rate (VR)
GDP loss	GDP loss rate for social distancing level j (GDP_{loss_j}) × GDP in 2019 × Simulation period (year)

TABLE 3 Descriptions and values of parameters for cost estimation.

	Description	Value	Reference
C_M	Medical cost for mild patients per day	$160.8 (treatment center) $5 (home treatment)	(20, 21)
C_H	Medical cost for number of hospitalized patients without intensive per day	$432.5	(20, 21)
C_I	Medical cost for hospitalized patients with intensive per day	$1,129.5	(20, 21)
E_i	Employment rate in age group i	0, 04, 0.557, 0.753, 0.771, 0.743, 0.566, 0.23	(22)
E_f	Female employment rate with children younger than 19 years	0.555	(22)
W_i	Average daily income in age group i	0, 55.47, 76.63, 109.40, 129.33, 127.14, 82.14, 68.08	(23)
W_f	Average daily income of women in 30s and 40s	99.3500	(23)
g	Average annual salary increase rate	0.02	(24)
r	Social discount rate	0.04	(24)
N_i	Average working period in age group i	50, 50, 45, 35, 25, 15, 5, 0	Assumed
FC	Average funeral cost	$9,405.38	(25)
VC	Vaccination cost per person	$17.89	(26)
PC	Vaccination procedure cost	$16.87	(27)
LC	Logistics cost	$1	Assumed
VR	Vaccination rate	0.8	Assumed
GDP	GDP per capita in 2019	$31,929	(28)
GDP_{loss_j}	GDP loss rate for social distancing level $j = 1, 2, 3$	0.002, 0.018, 0.064	(29)

and the GDP loss due to implementing the social distancing polices. The cost of medical expenses was computed considering the average daily cost of treatment and the recovery period with respect to the status of patients (i.e., mild symptom cases, hospitalized cases without intensive care, or hospitalized cases with intensive care). In Korea, before 10 February 2022, even patients with mild symptoms were admitted to community treatment centers and quarantined; however, after that, the policy was changed to give priority to home treatment. As a result, the quarantine period and cost decreased. The costs according to these two policies were calculated and compared. The cost of wage loss for patients older than 20 years was computed considering the average daily income, the employment rate, and the recovery period for each age group. The cost of wage loss for patients younger than 19 years was computed as the income decrease for females with infected children younger than 19 years. The cost due to death was calculated by summing the average funeral cost per capita and the present value of the predicted future income for the potential economic production loss using the forgone labor output equation (19). The vaccination cost included the average vaccine price, the procedure cost, and the logistics cost. The GDP loss was estimated by multiplying the GDP reduction rate according to the SD level with the 2019 GDP. The formulae for the estimation of each factor of the total cost are given in Table 2, and the parameters used for cost estimation are given in Table 3.

Results

The effect of rollout speed of vaccination on disease transmission

The effects of the daily vaccination doses on the number of confirmed cases, cumulative deaths, and hospitalized population with intensive care unit (ICU) care were investigated through numerical simulation using the mathematical model. As the COVID-19 vaccination program began on 26 February 2021, we set the initial date of the simulation as 1 April 2021, and the duration of the simulation was set as 365 days. The initial conditions were determined with respect to the confirmed case data in Korea at the start date of the simulation (1). At the beginning of vaccination in Korea, vaccination was implemented for those aged 19 years and older; however, in October 2021, the vaccine target was expanded to those aged 12 years and older.

In this work, it was assumed that vaccination was administered to those aged 10 years or older, and vaccination was terminated when 80% of the total population of Korea

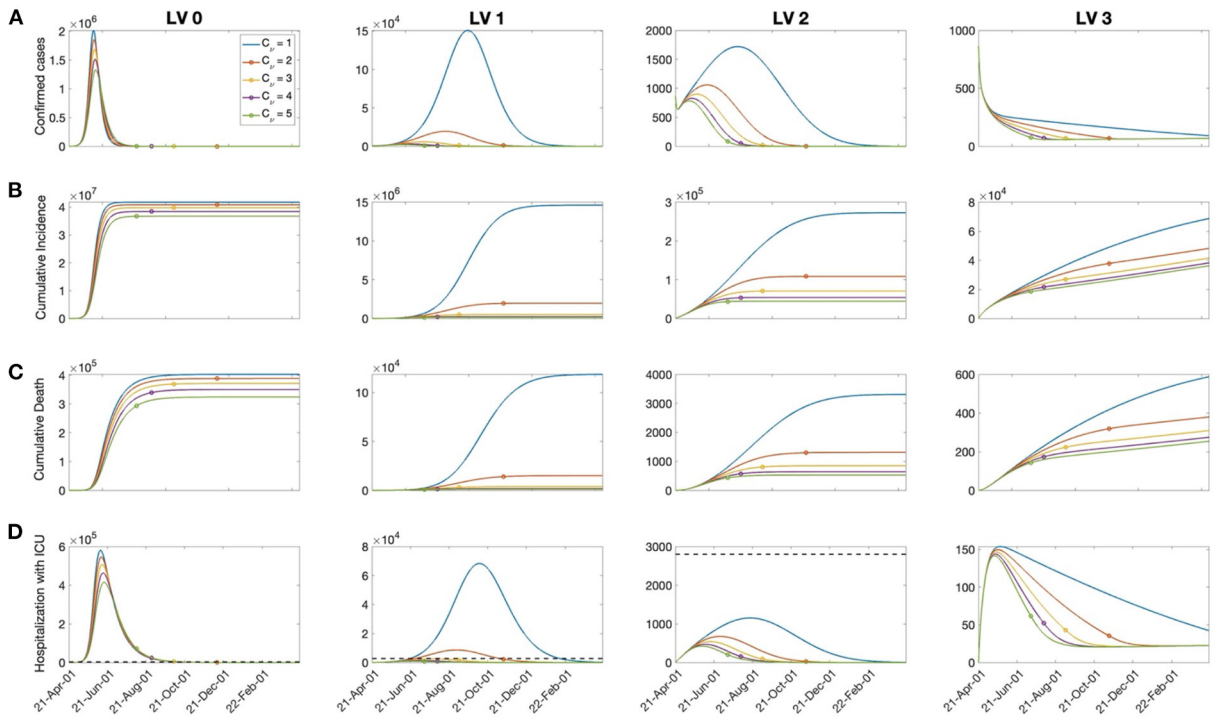


FIGURE 3
The effects of vaccination on: **(A)** Confirmed cases, **(B)** cumulative confirmed cases, **(C)** death, and **(D)** hospitalized population in the ICU for C_v=1,2,$\cdots$, 5 and the SD level 0, 1, 2, 3. The dashed line in the bottom panels represents the capacity of ICU bed for COVID-19 patients in Korea. The circle on each curve represents the time at which the vaccination coverage rate reaches 80%.

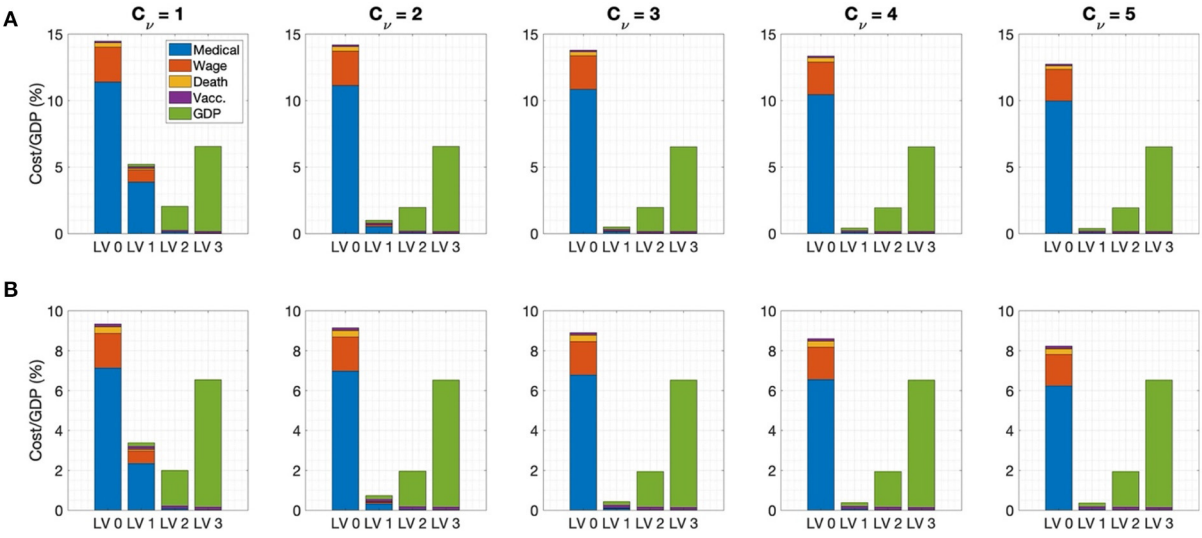


FIGURE 4
Ratio of cost to GDP for C_ν=1,2,$\cdots$, 5 for **(A)** admission to the treatment center for mild patients **(B)** home treatment for mild patients.

was vaccinated. It was assumed that inoculation was evenly distributed according to the number of people in each age group. We used $\nu_0 = 100,000$ as the baseline value for the daily vaccination dose, based on the average daily completed vaccinations in Korea (see Figure 1B). The daily vaccination dose varied as $\nu = C_\nu \nu_o$ for $C_\nu = 1, 2, \cdots, 5$. Figure 3 shows the time-series of daily new confirmed cases, cumulative confirmed cases, cumulative deaths, and the hospitalized population in intensive care, with respect to the rollout speed ($C_\nu = 1, 2, \cdots, 5$) under SD levels 0, 1, 2, and 3. It can be seen that, under SD levels 1 and 2, the rollout speed significantly affected the reduction in the number of confirmed cases. Therefore, when a moderate level of social distancing is implemented, the speed of vaccine release becomes even more important. It has been reported that the number of available ICU beds for COVID-19 patients in Korea is about 2800, as indicated by the dashed lines in Figure 3D (17). Figure 3D shows that, under SD LV0 and LV1 with slow vaccination speed ($C_\nu = 1, 2$), the maximum number of ICU patients exceeds 2800, thus posing a burden on the Korean medical system. The numbers of cumulative confirmed cases, cumulative death, and maximum hospitalized population in ICU for various social distancing level, rollout speed ($C_\nu = 1, 2, \cdots, 5$), and vaccine efficacy ($\tau = 0.79, 0.6$) are given in Supplementary material section 4.

Cost estimation due to COVID-19

The cost due to COVID-19 was estimated based on medical costs, wage losses, deaths costs, immunization costs, and GDP losses under various social distancing policies considering different rollout speeds and SD levels. Figure 4 shows the cost of each factor calculated over a one-year period with rollout speed $C_\nu = 1, 2, \cdots, 5$ and SD level 0, 1, 2, 3 under (A) treatment center admission and (B) home treatment for patients with mild symptoms. When the rollout rate was low ($C_\nu = 1$), implementing SD LV 2 was the least costly; otherwise, SD LV 1 was more cost effective. The medical expenses accounted for the largest proportion of the total cost at SD level 0 or 1, while the GDP loss accounted for the largest proportion at SD level 2 or 3. Home treatment, instead of being admitted to a treatment center, for the patients having mild symptoms reduces medical costs by about 40%. The cost values for each factor are given in the Supplementary material section 5. If the rollout speed is fast, the number of fatalities and hospitalizations with severe symptoms is not significantly different under SD levels 1 and 2. Therefore, increasing the vaccine supply rate can bring economic cost savings while maintaining a low SD level, without a significant increase in the number of fatalities and patients with severe symptoms.

The effects of social distancing level mitigation

As the vaccination rate in Korea increased and the economic and psychological damage to people caused by the long-term implementation of social distancing strengthened, the necessity for mitigating the social distancing level emerged. However, the number of confirmed cases surged as social distancing was eased in accordance with the "Reorganization

for the Step-by-step Daily Recovery" implemented in November 2021 in Korea.

In this section, the economic effects of social distancing mitigation policies are investigated. Social distancing mitigation scenarios were assumed as a relaxation from SD LV 2 to LV 1, SD LV 2 to LV0, or SD LV 1 to LV 0 when the vaccination rate of the total population reached 60, 70, or 80%. Figure 5 shows the cost of each factor for each social distancing mitigation scenario in the case of admission to a treatment center for patients with mild symptoms in the top panel for $C_\nu = 3$, and the total cost for $C_\nu = 1, 2, ..., 5$ *in the bottom panel*. In addition, cases with (A) high vaccine efficiency ($\tau = 0.79$) and (B) low vaccine efficiency ($\tau = 0.6$) were compared.

It can be seen that the most cost-effective scenario was to implement SD mitigation from LV2 to LV1 when the vaccination rate reached 60%. At this time, there was no significant increase in the number of deaths and the number of patients with severe symptoms. However, with SD LV2 to LV0 mitigation in the early stage, the cost increased significantly. In particular, the lower the vaccine efficiency, the greater the cost due to early mitigation. Figure 5 shows that, in the SD LV2 to LV0 and LV2 to LV1 cases, the faster the rollout speed, the higher the total cost under the same vaccine coverage. Therefore, it can be seen that a certain level of social distancing is necessary even if vaccination has sufficiently progressed. A time-series of the number of confirmed cases under each SD level mitigation scenario is provided in the Supplementary material section 6.

The vaccination effects on SARS-CoV-2 variants

It has been shown that vaccine efficacy may decrease due to COVID-19 variants, such as Omicron, and waning vaccine-based immunity (30–32). In addition, the disease trans-mission rate can change depending on the characteristics of the variant and non-pharmaceutical intervention policies. In this section, we investigate the effects on the total cost due to COVID-19 and number of hospitalized patients in the ICU due to changes in vaccine efficacy and the disease transmission rate.

In Figure 6, the total costs are computed for the vaccine efficacy varying from $\tau = 0.3 - 0.8$ and for the R_t corresponding to the transmission rate matrix β, which changes as $\beta \times C_\beta$ for $C_\beta = 0.7 - 1.3$ under SD LV 0, 1, 2, and 3, and the rollout speed $C_\nu = 1, 2, \ldots, 5$. The figure also represents the cases where the maximum number of hospitalized patients with severe symptoms is greater than the capacity of the intensive care unit for COVID-19 patients in Korea, shown as red curves.

It was found that, under SD LV 1 and 2, the total cost changed more sensitively with changes in τ and R_t. In these cases, the total cost and the number of hospitalized patients with severe symptoms were also greatly affected by changes in the rollout speed. The number of hospitalized patients with severe symptoms depended more on the change in R_t than on the change in τ. If the vaccine efficiency is high and the rollout speed is sufficiently fast, the number of hospitalized patients with severe symptoms can be kept below the ICU bed capacity, even under SD LV 1. When the vaccine efficacy τ is low, the effect of a change in β on the total cost is greater. Therefore, the less effective the vaccine, the more effort is needed to reduce the transmission rate of the disease. The cumulative number of confirmed cases and the maximum number of hospitalized patients with severe symptoms in each case are included in the Supplementary material section 7.

Sensitivity analysis

Sensitivity analyses were conducted to determine the relative importance of the parameters related to cost—that is β, τ, ρ, ν, the vaccine cost, and GDP loss rate—with respect to the disease transmission dynamics. We performed further sensitivity analyses on the model parameters described in Table 1. We used the normalized forward sensitivity index of the total cost (TC) on parameter p, defined as $\mathrm{TC}_p = \frac{\partial(TC)}{\partial p} \times \frac{p}{TC}$ (33). The TC for one year was computed by varying one parameter by 5% from the baseline value while the rest of the parameters were fixed at their baseline values. For the transmission rate matrix β, all components were increased simultaneously. Figure 7 shows that increases in β, VC, and GDP_{loss} affected the total cost positively, meaning that when those parameters increased, the total cost also increased; however, an increase in ν_0 and τ negatively affected the total cost for all cases. TC_ρ for the probability of unconfirmed asymptomatic cases ρ was either positive or negative, depending on the SD level and rollout speed C_ν. If the number of confirmed cases was high, such as in the case of SD LV 0, a negative value was displayed. This can be understood as, when the number of un-confirmed cases increases, a relatively decreased effect on confirmed cases is seen. However, when the number of confirmed cases is small, an increase in the number of unconfirmed asymptomatic patients leads to an increase in the infection rate in the susceptible population, such that the sensitivity index becomes positive. The change in total cost, according to the change in the transmission rate and vaccine efficacy, was greatest under SD level 1. When the SD level was higher, the change in total cost according to the GDP loss rate was greater.

Discussion

In this study, we developed an age-structured mathematical model to describe the transmission dynamics of COVID-19 considering vaccination. In the model, we divided the

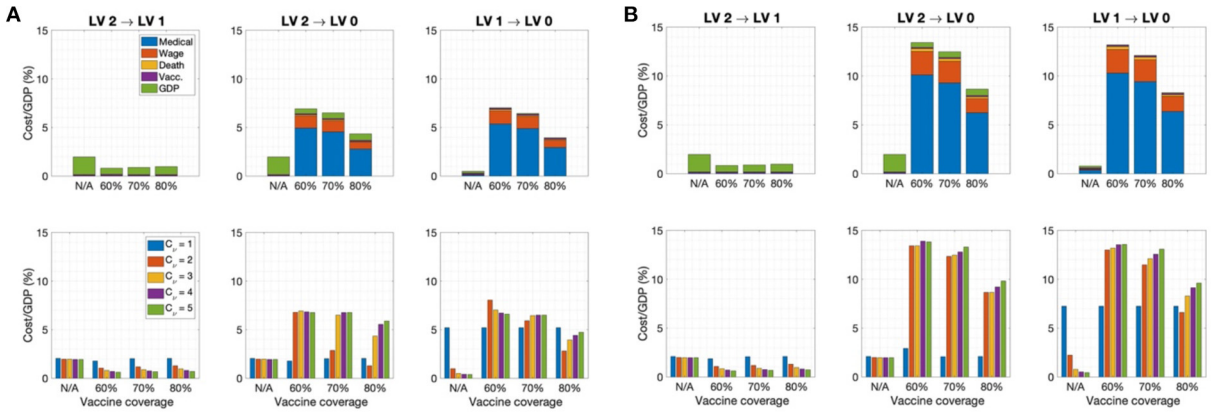


FIGURE 5
Ratio of cost to GDP for each SD level mitigation scenario when the vaccination coverage rate reaches 60, 70, or 80% in case of admission to the treatment center for mild patients for **(A)** $\tau = 0.79$ and **(B)** $\tau = 0.6$ for (Top) $C_\nu = 3$. (Bottom) $C_\nu = 1, 2, ..., 5$. (N/A indicates that social distancing easing was not implemented).

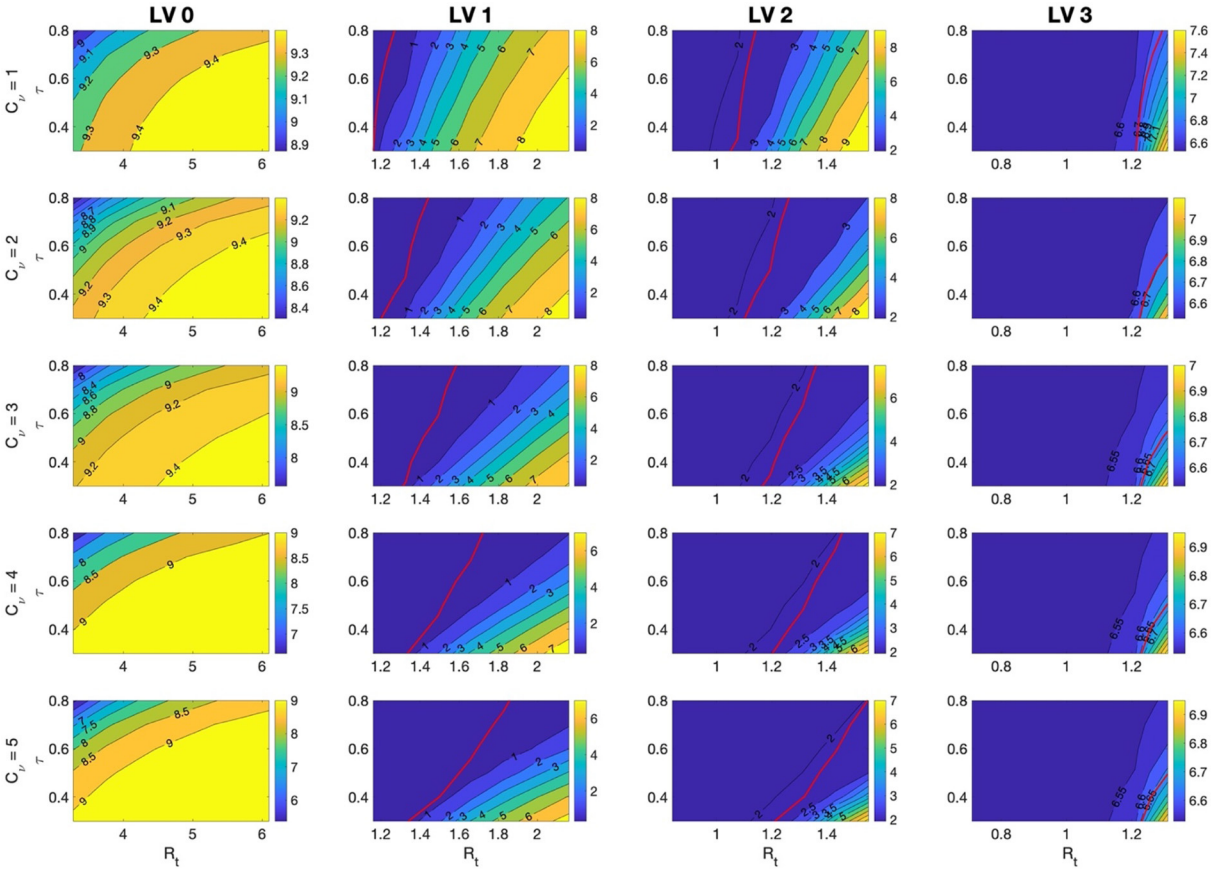


FIGURE 6
The ratio of total costs to GDP for τ=0.3 − 0.8 and for R_t corresponding to β that is varied as $\beta \times C_\beta$ for C_β=0.7 − 1.3 under SD LV 0, 1, 2, and 3, and the rollout speed C_ν=1, 2, 3, 4, 5. Red lines indicate the case when the maximum number of hospitalized patients with severe symptoms is the capacity of the intensive care unit, 2,800, for COVID-19 patients in Korea. It means that the maximum number hospitalized patients with severe symptoms is lower than capacity on the left side of the red curves.

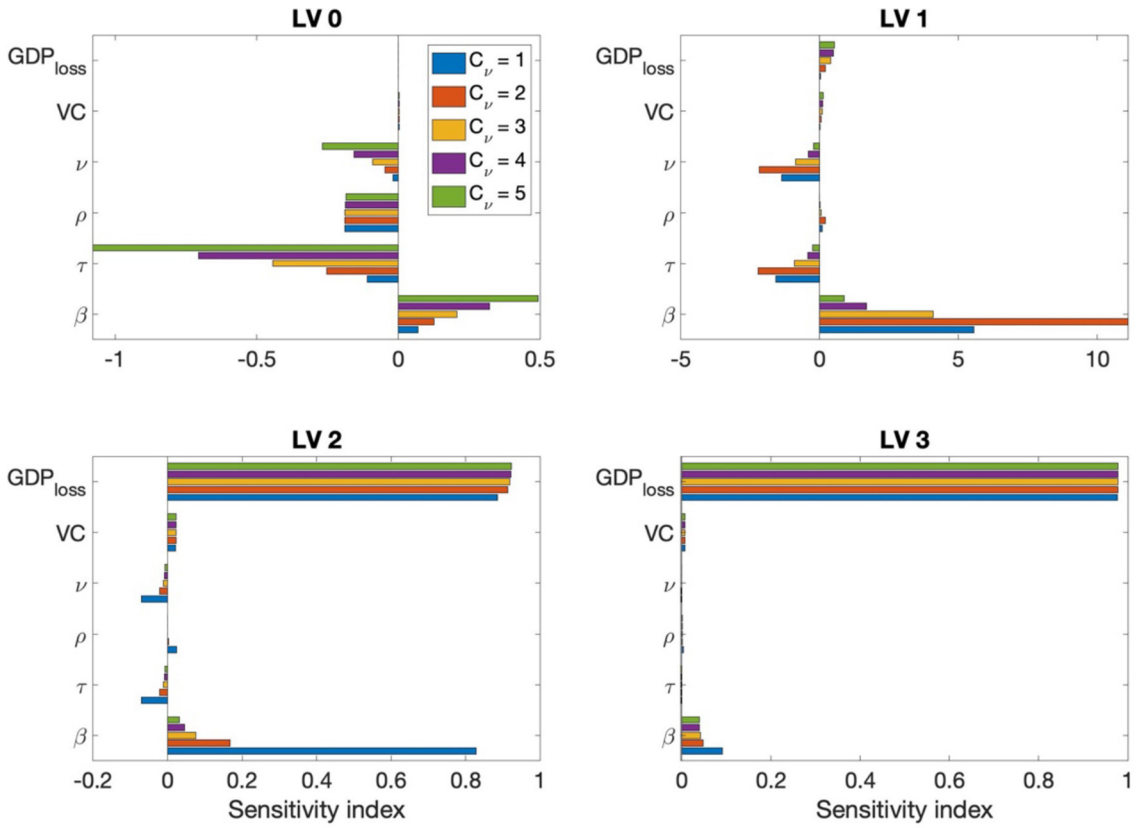


FIGURE 7
Sensitivity index of β, τ, ρ, ν, the vaccine cost, and GDP loss rate for C_ν=1,2,..., 5 and each SD level.

population into eight distinct age groups and estimated the transmission rate for each age group under different levels of social distancing, formulated as an 8×8 matrix. Using the model, we investigated the time-dependent dynamics of the number of infected, severe cases, and deaths under varying social distancing levels, vaccine efficiencies, and vaccination rates.

Furthermore, we estimated the total cost of COVID-19, including medical expenses, wage loss, cost due to death, vaccination cost, GDP loss under various vaccine rollout speeds, and vaccine efficiencies under different levels of SD. We found that a faster vaccination rollout speed and higher efficiency resulted in a lower overall cost.

We also investigated the change in the total cost under variation of the rollout speed and disease transmission rate. Figure 4 shows that, when the rollout speed was slow (C_ν =1), SD LV 2 minimized the total cost; otherwise, the total cost was minimized in the case of SD LV1.

Moreover, if the vaccination rate was fast enough ($C_\nu \geq 3$), when a sufficient number of vaccines was available, maintaining SD LV1 was the best strategy in terms of both cost minimization and proper medical care of seriously ill patients.

Figure 5 demonstrates that, in order to mitigate the social distancing level, the vaccination coverage rate must be as high as 80%, and the total cost will decrease according to the coverage rate. In particular, when the SD level is alleviated from LV 2 to LV 1, the total cost can be sufficiently reduced even when the coverage rate is about 60%; however, it can be seen that some degree of social distancing must be maintained to prevent further outbreaks.

As it is very important to prevent the collapse of the medical system by reducing the number of seriously ill patients and the number of deaths, it is necessary to devise a control strategy that minimizes damage from a comprehensive perspective, rather than simply planning a strategy from a cost-effective perspective. Figure 3 shows that the rollout speed is important in reducing the number of confirmed cases and serious cases. In the case of SD LV 1, the rollout speed must be as high as $C_\nu \geq 3$ in order to sufficiently care for critically ill patients. In case of SD LV 2, even if the rollout speed is $C_\nu = 1$, severe patients can still receive sufficient medical support.

Omicron variations were not taken into account in the model. Concerning the vaccine efficacy, we assumed two values (0.79 and 0.6) based on the actual efficacy for COVID-19 before Omicron emerged. The results presented in this paper may differ for other efficiency values, but the overall conclusion will not be different.

We assumed the vaccine rollout speed values based on the actual data regarding vaccinations conducted in Korea. In other countries, the vaccination rollout speed may vary more than our hypothesized values.

In this paper, we demonstrated that the vaccination rollout speed is important for both controlling the spread of COVID-19 and reducing costs. In the case of Korea, as the medical infrastructure is solid and the voluntary vaccination rate of the people is high, if the rollout speed of a high-efficiency vaccine is fast enough, economic costs can be reduced by lowering the social distancing level.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JEK and CHL: conceptualization, funding acquisition, and writing—review & editing. JEK and YC: formal analysis, visualization, and methodology. JEK, HC, and CHL: validation. JEK and HC: writing—original draft. All authors contributed to the article and approved the submitted version.

Funding

JEK was supported by a National Research Foundation of Korea (NRF) grant and funded by the Korean government (MSIT) (Grant Number NRF-2021R1I1A1A01044426). HC and CHL were supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1F1A1064487) and the BK21 Program (Next Generation Education Program for Mathematical Sciences, 4299990414089) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea (NRF).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.993745/full#supplementary-material

References

1. Korea Disease Control and Prevention Agency. Available online at: https://www.kdca.go.kr/index.es?sid=a2 (accessed June 21, 2022).

2. Moore S, Hill EM, Tildesley MJ, Dyson L, Keeling MJ. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. *Lancet Infect Dis.* (2021) 21:793–802. doi: 10.1016/S1473-3099(21)00143-2

3. Canabarro A, Tenório E, Martins R, Martins L, Brito S, Chaves R. Data-driven study of the COVID-19 pandemic *via* age-structured modelling and prediction of the health system failure in Brazil amid diverse intervention strategies. *Plos one.* (2020) 15:e0236310. doi: 10.1371/journal.pone.0236310

4. Ram V, Schaposnik LP. A modified age-structured SIR model for COVID-19 type viruses. *Scientific Rep.* (2021) 11:1–15. doi: 10.1038/s41598-021-94609-3

5. Min KD, Tak S. Dynamics of the COVID-19 epidemic in the post-vaccination period in South Korea–a rapid assessment. *Epidemiol Health.* (2021) 21:e2021040. doi: 10.4178/epih.e2021040

6. Choi Y, Kim JS, Choi H, Lee H, Lee CH. Assessment of social distancing for controlling COVID-19 in Korea: an age-structured modeling approach. *Int J Environ Res Public Health.* (2020) 17:7474. doi: 10.3390/ijerph17207474

7. Choi Y, Kim JS, Kim JE, Choi H, Lee CH. Vaccination prioritization strategies for COVID-19 in Korea: a mathematical modeling approach. *Int J Environ Res Public Health.* (2021) 18:4240. doi: 10.3390/ijerph18084240

8. Asamoah JKK, Owusu MA, Jin Z, Oduro FT, Abidemi A, Gyasi EO. Global stability and cost-effectiveness analysis of COVID-19 considering the impact of the environment: using data from Ghana. *Chaos Solitons Fractals.* (2020) 140:110103. doi: 10.1016/j.chaos.2020.110103

9. Asamoah JKK, Jin Z, Sun GQ, Seidu B, Yankson E, Abidemi A, et al. Sensitivity assessment and optimal economic evaluation of a new COVID-19 compartmental epidemic model with control interventions. *Chaos Solitons Fractals.* (2021) 146:110885. doi: 10.1016/j.chaos.2021.110885

10. Hagens A, Inkaya AÇ, Yildirak K, Sancar M, van der Schans J, Acar Sancar A, et al. COVID-19 vaccination scenarios: a cost-effectiveness analysis for Turkey. *Vaccines (Basel).* (2021) 9:399. doi: 10.3390/vaccines9040399

11. Debrabant K, Grønbæk L, Kronborg C. The cost-effectiveness of a COVID-19 vaccine in a Danish context. *Discussion Papers on Business and Economics, University of Southern Denmark.* (2021), p. 2.

12. Pinto Neto O, Kennedy DM, Reis JC, Wang Y, Brizzi ACB, Zambrano GJ, et al. Mathematical model of COVID-19 intervention scenarios for São Paulo—Brazil. *Nat Commun.* (2021) 12:1. doi: 10.1038/s41467-020-20687-y

13. Padula W v., Malaviya S, Reid NM, Cohen BG, Chingcuanco F, Ballreich J, et al. Economic value of vaccines to address the COVID-19 pandemic: a U.S. cost-effectiveness and budget impact analysis. *J Med Econ.* (2021) 21:32. doi: 10.1080/13696998.2021.1965732

14. Khajanchi S, Sarkar K, Mondal J, Nisar KS, Abdelwahab SF. Mathematical modeling of the COVID-19 pandemic with intervention strategies. *Results Physic.* (2021) 21:285. doi: 10.1016/j.rinp.2021.104285

15. Korea Disease Control and Prevention Agency. *Public health and weekly report, PHWR.* (2020), p. 85.

16. Abrams S, Wambua J, Santermans E, Willem L, Kuylen E, Coletti P, et al. Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. *Epidemics.* (2021) 35:100449. doi: 10.1016/j.epidem.2021.100449

17. Coronavirus (COVID-19), Republic of Korea (2022). Available online at: http://ncov.mohw.go.kr/ (accessed June 30, 2022).

18. Buckner JH, Chowell G, Springborn MR. Dynamic prioritization of covid-19 vaccines when social distancing is limited for essential workers. *Proc Natl Acad Sci U S A.* (2021) 118:16. doi: 10.1073/pnas.2025786118

19. Ghaffari Darab M, Keshavarz K, Sadeghi E, Shahmohamadi J, Kavosi Z. The economic burden of coronavirus disease 2019 (COVID-19): evidence from Iran. *BMC Health Services Res.* (2021) 21:1–7. doi: 10.1186/s12913-021-06126-8

20. Choi E. *1.13 trillion won in treatment costs for COVID-19. 75.1% of health insurance financial statements. News The Voice Healthcare [Internet].* (2021). Available online at: https://www.newsthevoice.com/news/articleView.html?idxno=21357 (accessed June 30, 2022)

21. Health Insurance Review & Assessment Service. *Guidance on the application standards and claim method for medical care benefits for residents of the COVID-19 Living Treatment Center (additional)* (2020). Available online at: http://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020002000100&brdScnBltNo=4&brdBltNo=8499#none (accessed June 23, 2022).

22. Employed persons by gender/age group [Internet]. *Korean Statistical Information Service*. Available online at: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1DA7024S&vw_cd=MT_ETITLE&list_id=B17&scrId=&language=en&seqNo=&lang_mode=en&obj_var_id=&itm_id=&conn_path=MT_ETITLE&path=%252Feng%252FstatisticsList%252FstatisticsListIndex.do (accessed June 30, 2022).

23. Age, Days, Hours, Payments, Workers by Size, Education, Age [Internet]. *Korean Statistical Information Service*. Available online at: https://kosis.kr/statHtml/statHtml.do?orgId=118&tblId=DT_PAY0004&vw_cd=MT_ETITLE&list_id=D_6&scrId=&language=en&seqNo=&lang_mode=en&obj_var_id=&itm_id=&conn_path=MT_ETITLE&path=%252Feng%252FstatisticsList%252FstatisticsListIndex.do (accessed June 30, 2022).

24. Bank W, for Health Metrics I, Evaluation. The cost of air pollution: strengthening the economic case for action. *World Bank*. (2016) 16:13. doi: 10.1596/25013

25. Lee PD, Kim MH. A study of standard funeral service charge. *Health Soc Welfare Rev.* (2009) 29:98–125. doi: 10.15709/hswr.2009.29.2.98

26. COVID-19 Vaccination (2022). Available online at: https://ncv.kdca.go.kr/ (accessed June 23, 2022).

27. Korea Disease Control and Prevention Agency. *Announcement of vaccination cost according to Regulation on Consignment of Vaccination Service*. (2021). Available online at: https://www.kdca.go.kr/board/board.es?mid=a20505000000&bid=0017 (accessed June 30, 2022).

28. GDP per capita (current year price) (OECD) [Internet]. *Korean Statistical Information Service*. Available online at: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2KAA904_OECD (accessed June 30, 2022)

29. Bank of Korea. *Monetary Policy Report*. (2020), p. 107.

30. Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *N Engl J Med.* (2021) 21:585–94. doi: 10.1056/NEJMc2113090

31. Gardner BJ, Kilpatrick AM. 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 [Preprint] (2021).

32. Hansen CH, Schelde AB, Moustsen-Helms IR, Emborg HD, Krause TG, Mølbak K, et al. Vaccine effectiveness against SARS-CoV-2 infection with the Omicron or Delta variants following a two-dose or booster BNT162b2 or mRNA-1273 vaccination series: A Danish cohort study. *MedRxiv [Preprint].* (2021). Available at: https://www.medrxiv.org/content/10.1101/2021.12.20.21267966v3 (accessed July 13, 2022).

33. Rodrigues HS, Monteiro MTT, Torres DFM. Sensitivity analysis in a dengue epidemiological model. *Conference Papers Math.* (2013) 2013:1–7. doi: 10.1155/2013/721406

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 09 November 2022
DOI 10.3389/fpubh.2022.1009152

OPEN ACCESS

EDITED BY
Khalid Hattaf,
Centre Régional des Métiers de l'Education et de la Formation (CRMEF), Morocco

REVIEWED BY
Muhammad Ozair,
COMSATS University Islamabad, Attock, Pakistan
Rashid Jan,
University of Swabi, Pakistan
Fahad Al Basir,
Kazi Nazrul University, India

*CORRESPONDENCE
Huixuan Zhou
chouhuixuan@live.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Public Health Policy, a section of the journal Frontiers in Public Health

RECEIVED 01 August 2022
ACCEPTED 24 October 2022
PUBLISHED 09 November 2022

CITATION
Liu Z, Zhou H, Ding N, Jia J, Su X, Ren H, Hou X, Zhang W and Liu C (2022) Modeling the effects of vaccination, nucleic acid testing, and face mask wearing interventions against COVID-19 in large sports events.
Front. Public Health 10:1009152.
doi: 10.3389/fpubh.2022.1009152

Modeling the effects of vaccination, nucleic acid testing, and face mask wearing interventions against COVID-19 in large sports events

Zeting Liu[1†], Huixuan Zhou[2,3*†], Ningxin Ding[4], Jihua Jia[2], Xinhua Su[1], Hong Ren[2], Xiao Hou[2,3], Wei Zhang[5] and Chenzhe Liu[2]

[1]Department of Mathematic Science, School of Sport Engineering, Beijing Sport University, Beijing, China, [2]Department of Physical Fitness and Health, School of Sport Science, Beijing Sport University, Beijing, China, [3]Key Laboratory of Sports and Physical Health, Ministry of Education, Beijing Sport University, Beijing, China, [4]School of Government, Wellington School of Business and Government, Victoria University of Wellington, Wellington, New Zealand, [5]Department of Chemical Drug Control, China National Institute for Food and Drug Control, Beijing, China

The transmission of SARS-CoV-2 leads to devastating COVID-19 infections around the world, which has affected both human health and the development of industries dependent on social gatherings. Sports events are one of the subgroups facing great challenges. The uncertainty of COVID-19 transmission in large-scale sports events is a great barrier to decision-making with regard to reopening auditoriums. Policymakers and health experts are trying to figure out better policies to balance audience experiences and COVID-19 infection control. In this study, we employed the generalized SEIR model in conjunction with the Wells–Riley model to estimate the effects of vaccination, nucleic acid testing, and face mask wearing on audience infection control during the 2021 Chinese Football Association Super League from 20 April to 5 August. The generalized SEIR modeling showed that if the general population were vaccinated by inactive vaccines at an efficiency of 0.78, the total number of infectious people during this time period would decrease from 43,455 to 6,417. We assumed that the general population had the same odds ratio of entering the sports stadiums and becoming the audience. Their infection probabilities in the stadium were further estimated by the Wells–Riley model. The results showed that if all of the 30,000 seats in the stadium were filled by the audience, 371 audience members would have become infected during the 116 football games in the 2021 season. The independent use of vaccination and nucleic acid testing would have decreased this number to 79 and 118, respectively. The combined use of nucleic acid testing and vaccination or face mask wearing would have decreased this number to 14 and 34, respectively. The combined use of all three strategies could have further decreased this number to 0.

According to the modeling results, policymakers can consider the combined use of vaccination, nucleic acid testing, and face mask wearing to protect audiences from infection when holding sports events, which could create a balance between audience experiences and COVID-19 infection control.

KEYWORDS

COVID-19 vaccine, COVID-19 nucleic acid testing, mathematical model, SEIR model, Wells–Riley model

Introduction

The outbreak of COVID-19 has affected 222 countries and regions worldwide and has led to 102 million cases and 2.2 million deaths (1). Its pathogen, SARS-CoV-2, has a higher transmissibility than H1N1 and SARS-CoV (2). The pathogen mainly transmits *via* respiratory droplets and contact. Social gathering and migration are the primary reasons for the spread of SARS-CoV-2. To prevent the transmission of COVID-19, the governments of some countries have implemented a series of emergency measures, such as travel restrictions and the cessation of social gathering events, which successfully contain the transmission of COVID-19 while inevitably limiting the development of some of the industries relying on social gathering and casting a pall over participants (3).

Sports events are one of the subgroups facing great challenges in the post-pandemic era. Some games were played in an empty stadium, while other stadiums opened a small part of their auditorium to maintain safe social distancing. Home-and-away games were prohibited. These measures have had some negative effects on the development of sports economics and have impaired audience experiences. For example, the Chinese Football Association Super League (hereinafter referred to as CSL), one of the most prosperous professional sports events in China, attracted a total of 5.6 million audience members (25,000–30,000 audience members for each game) in the 2019 game season (4). However, the CSL has been played in an empty stadium since the outbreak of COVID-19. Without match day revenue, 16 professional football teams announced their intention to disband, and Jiangsu Suning F.C., the champion team of 2020 CSL, announced its closing down due to financial crisis in 2021. The experience of millions of fans has been impaired at the same time. Therefore, it is an urgent task for policymakers to establish some policies other than social distancing to control COVID-19 infection and reopen sports events to the audience.

Vaccination against COVID-19 and nucleic acid testing provide the potential to resume attending sports events. The Chinese residents aged 3 years or more have been encouraged to receive a two-dose inoculation by inactivated vaccines since September 2020, based on a high vaccine supply and the safety reported by a series of clinical trials (5–8). The CSL tried to reopen a couple of games to the audience in 2021. The allowed admission was decreased from 30,000 in the pre-pandemic era to 2,000. The audiences were required to show a negative record of nucleic acid testing within 7 days prior to the games (7) and to wear face masks when watching them. Meanwhile, to avoid the transmission of COVID-19 in the transportation of audiences and football teams, games were held in two cities instead of being home-and-away games. As a result, the income associated with the CSL and the audience experiences were not substantially enhanced toward the pre-pandemic levels. The uncertainty of COVID-19 transmission among audiences is a great barrier for policymakers to make a decision with regard to resuming sports events with large crowds.

Dynamic models of infection can be utilized to predict the effects of preventive strategies against transmission (9–12). Compared to the Susceptible, Exposed, Infectious (SIR) model and the classic Susceptible, Exposed, Infectious, Recovered (SEIR) model, the generalized SEIR model (13) is more appropriate to reflect the transmission dynamics of COVID-19 (14), which accounts for some important characteristics of the disease, such as the latency period, death, quarantine for confirmed cases, and the protection rate of susceptible populations. Some efforts have been devoted to predict the effectiveness of various preventive strategies against COVID-19 transmission in the general population by the generalized SEIR model (15–20). By adjusting the protection rate of the SEIR model, the impact of vaccination (15–20) and non-pharmaceutical interventions (NPIs) (15, 16, 18, 20) have been simulated. The results of some studies suggest that the effectiveness of vaccination depends on the efficacy of the vaccines used and the coverage of vaccination (15, 17–19); if the population cannot be fully covered by effective vaccination, combining vaccination and NPIs is an ideal measure to reduce the number of confirmed cases and deaths of COVID-19 (15, 16, 20). These studies have indicated that the generalized SEIR model is useful for simulating COVID-19 dynamics in the general population during a long period of time, and the herd immunity effect of vaccination can be predicted. However, to predict the transmission probability in audiences during a sports event, we need to further simulate a scenario in which susceptible people could be infected through air droplets in a confined space.

The Wells–Riley model is based on the concept of "quantum of infection," which is defined as the dose of pathogens required to infect a susceptible person when he or she inhales in a ventilated room during a time of exposure (21). The Wells–Riley model and its deviations have been extensively employed in studies on the transmission of airborne infectious diseases, including COVID-19 (22). For instance, Wang et al. estimated the infection probability of COVID-19 for a 2-h fight (23) and for Chinese long-distance trains (24) under scenarios at various face mask efficiency levels. Che et al. (25) and Sha et al. (26) predicted the impact of ventilation systems on COVID-19 control and prevention in high-rise buildings.

Hence, in this study, we used the generalized SEIR model in conjunction with the Wells–Riley model to simulate the transmission of COVID-19 in sports stadiums under the scenarios of various preventive strategies, including vaccination, face mask wearing, and nucleic acid testing. We first used the generalized SEIR model developed by Cheynet (13) to simulate the impact of full vaccination delivery in the general population in China. We assumed that a part of the general population would become audience members of the CSL and that vaccination would decrease the number of infectious audience members entering the sports stadiums. Next, we used the Wells–Riley model, involving face mask efficiency (22) and nucleic acid testing, to simulate the transmission of COVID-19 among the audience members under scenarios including (1) without any preventive strategy, (2) with vaccination against COVID-19, (3) with a negative nucleic acid testing result within 7 days before admission, (4) with combined vaccination and nucleic acid testing, (5) with combined nucleic acid testing and face mask wearing, and (6) with combined vaccination, nucleic acid testing, and face mask wearing. By combining those two mathematic models, this study involved the effect of vaccination policy on the general population in a sports stadium and visualized the effects of face mask wearing and nucleic acid testing on COVID-19 transmission in CLS audiences. The study findings indicate the potential infection risk in sports stadium, suggest some effective measures for infection control in the post-pandemic era, and provide some evidence for policymakers to reopen large sports events.

Methods

Generalized SEIR model and fitting

To characterize the epidemic of COVID-19 during the 2021 CSL game season, we used the generalized SEIR model with seven different states, namely, susceptible (S), insusceptible (P), exposed (E, in a latent period, infected but not showing infectiousness), infectious (I, infectious and not yet quarantined), quarantined (Q, confirmed), recovered (R), and death (D). $S(t)$, $P(t)$, $E(t)$, $I(t)$, $Q(t)$, $R(t)$, and $D(t)$ denote, at time t, the number of susceptible, insusceptible, exposed, infectious, quarantined, recovered, and death cases, respectively. Their relations are governed by a serial of equations, which can be formulated through ordinary differential equations (ODEs) as follows (13):

$$\begin{aligned} dS(t)/dt &= -\beta(t)I(t)S(t)/N - \alpha\, S(t), \\ dP(t)/dt &= \alpha S(t), \\ dE(t)/dt &= \beta(t)I(t)S(t)/N - \gamma(t)E(t), \\ dI(t)/dt &= \gamma(t)E(t) - \delta(t)I(t), \\ dQ(t)/dt &= \delta(t)\,I(t) - \lambda(t)\,Q(t) - \kappa(t)Q(t), \\ dR(t)/dt &= \lambda(t)\,Q(t), \\ dD(t)/dt &= \kappa(t)Q(t), \end{aligned} \tag{1}$$

where N is the total population and the coefficients α, β, γ^{-1}, δ^{-1}, $\lambda(t)$, and $\kappa(t)$ are the protection rate, infection rate, average latent time, average quarantine time, cure rate, and mortality rate, respectively. We assumed that $N = S+P+E+I+Q+R+D$ is constant and the model ignores immigration, emigration, birth, and death unrelated to COVID-19; that the infectious population is evenly mixed with the susceptible population; that the people in the model have the same odds of making a decision about watching a game in a stadium.

The seven states $\{S(t), P(t), E(t), I(t), R(t), D(t)\}2$ with fitted parameters $\{\alpha, \beta, \gamma^{-1}, \delta^{-1}, \lambda(t), \kappa(t)\}$ were calculated by a non-linear least-square function based on Chinese data from 20 April to 10 September. The ODEs were written in a matrix form and solved using the classic fourth-order Runge–Kutta method to find the time evolution of the seven states. The matrix form is shown as follows:

$$dY/dt = G \times Y + F \tag{2}$$

where

$$Y = [S, E, I, Q, R, D, P]^T \tag{3}$$

$$G = \begin{pmatrix} \alpha & 0 & 0 & 0 & 0\ 0\ 0 \\ 0 & -\alpha & 0 & 0 & 0\ 0\ 0 \\ 0 & \alpha & -\delta & 0 & 0\ 0\ 0 \\ 0 & 0 & \delta & -\kappa(t) - \lambda(t) & 0\ 0\ 0 \\ 0 & 0 & 0 & \lambda(t) & 0\ 0\ 0 \\ 0 & 0 & 0 & \kappa(t) & 0\ 0\ 0 \\ 0 & 0 & 0 & 0 & 0\ 0\ 0 \end{pmatrix} \tag{4}$$

$$F = S(t) \times I(t) \times \begin{pmatrix} -\beta/N \\ \beta/N \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{5}$$

The simulation time period was set from 20 April (the beginning of the 2021 game season, also the stationary phase of the COVID-19 pandemic in China) to 10 September (the endpoint of an outbreak in this time period). The 2021 game season finished on 5 August, which was included in this simulation time period. The initial numbers of the coefficients N, R_0, D_0, and Q_0 were determined according to the daily briefing of the National Health Commission (NHC) of China (27), and E_0 and I_0 were inferred according to prior studies based on Chinese data (13). The values and sources of the coefficients can be seen in Table 1. For the optimization parameter, the "fit_SEIQRDP" function was used, which was available in the generalized SEIR model package developed by Cheynet in MATLAB (13).

The fitting is shown using the daily number of current quarantined *cases* $Q(t)$ and cumulative recovered cases $R(t)$, which can be observed in the real world and was reported in the daily briefing from the NHC. Susceptible $S(t)$, insusceptible $P(t)$, exposed $E(t)$, and infectious $I(t)$ cases and the parameters α, β, γ^{-1}, δ^{-1}, $\lambda(t)$, and $\kappa(t)$ could also be computed by this package, while $I(t)$ and α were used as targets.

The boundedness and non-negativity of the model variables can be seen in Supplementary materials.

Integrating vaccination into the generalized SEIR model

We simulated the effectiveness of vaccination on I_v (t) by adding the efficacy of inactivated vaccines to α. In the vaccination scenario, people were assumed to have received two doses of inoculation and thus have completed the immunogenic process when entering the model, and the effect of vaccination was to increase the protection rate α and thereby decrease the susceptible population; although the vaccine may offer less protection due to the emergency of new variants and the immunity might wane over long timescales, we assumed that the efficacy of vaccination did not change in the simulation of the 4-month game season. The WHO report on the inactive Sinopharm/BBIBP COVID-19 vaccine showed that the overall efficacy of two doses of the vaccine is 0.78 (28). By definition, vaccine efficacy is the proportional reduction in infection rates. Therefore, we assumed in this study that vaccination would increase α by 0.78. The equation can be written as:

$$\alpha_v = \alpha + 0.78 \tag{6}$$

$I(t)$ and I_v (t) are the numbers of infectious people without and with vaccination in the general population, respectively, which were used as parameters in the next step of the infection simulation in the stadium.

Wells–Riley model and infection probability in the stadium

According to the Wells–Riley equation, the risk of infection by pathogens of a susceptible person through the air in a ventilated room is related to the quanta she or he inhales, the amount of pulmonary ventilation, the concentration of pathogen-bearing particles in the inhaled air, and the time of exposure. Based on these assumptions, the equation can be written as (21, 29):

$$p = 1 - exp(-\frac{Aqvt}{Q})\ (7) \tag{7}$$

where p is the infection probability, A is the number of infectious people entering the room, q is the quanta produced by an infected individual (quanta/h), v is the pulmonary ventilation rate of a person (m^3/h), t is the exposure time (h), and Q is the outdoor air supply rate (m^3/h). This equation is based on the assumptions that infectious particles are well-mixed in airspace, exposure to one quantum of infection provides an average probability of 63.2% based on a Poisson distribution, and the outdoor air supply rate Q remains constant.

Rudnick and Milton proposed a modified Wells–Riley model using the exhaled air volume fraction, which does not require the second assumption (30). Q could be a function of the exhaled air volume fraction, the number of people, and the pulmonary ventilation rate v; the equation can be written as:

$$Q = nv/\bar{f} \tag{8}$$

where f is the exhaled air volume fraction and n is the number of people in the ventilated room (30). Sheng (31) considered that the volume of a sports stadium would affect the transmission of pathogens, and the audience could use personal respiratory protection to decrease the probability of infection. Therefore, Sheng combined the modified Wells–Riley model by Fennelly and Nardell (32), which included personal respiratory protection in the equation, and wrote the equation as (31):

$$B = 1 - exp\{(-\frac{\bar{f}Aqt\theta}{n})[1 - \exp\left(-\frac{nvt}{V\bar{f}}\right)]\} \tag{9}$$

where θ is the penetration ratio of the protective respirator, which reflects the amount of leakage through and around the respirator (32). For instance, θ is 0.25 for an ordinary surgical mask (26), and θ is 1 without any respirator (32); V is the volume of the stadium.

The value of quantum varies significantly in different studies, as it depends on the types of pathogens and the estimation methods used (26). Sheng asserted that the quantum in a sports stadium is influenced by A, the number of infectious people entering the stadium, and assumed that the quantum has a

TABLE 1 Parameters in the SEIR model.

Parameter	Value	Description and source
N	26,418,000	Total population of cities where pandemic outbroke during the 2021 game season (41).
R_0	85,612	Number of cumulative recovered cases on 20 April, 2021 reported by NHC (27).
D_0	4,636	Number of cumulative dead cases on 20 April, 2021 reported by NHC (27).
Q_0	616	Number of quarantined cases on 20 April, 2021 reported by NHC, including confirmed patients and asymptomatic infectors under medical observation (27).
E_0	616	Equal to Q_0 (13).
I_0	123	1/5 of Q_0 (13).

linear relationship with A (31). According to Sheng's method, the equation can be written as:

$$q = 0.16A + 36 \tag{10}$$

where A is estimated based on $I(t)$ on each match day according to the results of the SEIR model and the false negative rate of nucleic acid testing. For instance, A is equal to $0.3nI(t)/N$ when the audience are required to show a negative nucleic acid testing result with a 30% false negative rate.

The parameters of the six scenarios used in the Wells–Riley model are shown in Table 2.

Results

Simulation of infection in the general population with and without vaccination

We applied the above-described SEIR model to fit the public data of the daily current quarantined and cumulative recovered cases from 20 April to 10 September (Figure 1). The epidemic curve during this period was divided into two phases. In the first phase from 20 April to 24 July, the number of quarantined cases was initially low and slowly increased. In the second phase from 25 July to 10 September, there was an outbreak of the pandemic, reaching a peak of over 2,000 quarantined cases in early August.

Through extensive simulations, the values for the unknown model parameters α, β, γ^{-1}, δ^{-1}, $\lambda(t)$, and $\kappa(t)$ were calculated and can be seen in Table 3. The protection rate α in the first phase was 0.04, lower than that in the second phase (0.37), which indicates that before the outbreak, people might have paid less attention to self-protection, and most of the general population were susceptible to SARS-CoV-2. The higher infection rate β and longer latent time γ^{-1} in the first phase showed that people were more likely to be infected by the virus and that it was more difficult to detect in the first phase, which afterward led to the surge of confirmed cases in the second phase. The accessibility of medical treatment for COVID-19 cases in China has maintained a high level since the initial outbreak of the pandemic in early 2020. As a result, the quarantine time δ^{-1} was maintained at 9 days in both phases. In addition to the constant parameters α, β, γ^{-1}, and δ^{-1}, this study showed the function of the cure rate $\lambda(t)$ and the mortality rate $\kappa(t)$, which are the choices of best approximation determined by the generalized SEIR model. The values of these parameters best interpreted the data by showing a perfect fit of the curves for quarantined and recovered cases, which can be observed in the real world.

To understand the impact of vaccination, $I(t)$, the number of infectious people without vaccination was calculated. After adding the vaccine efficacy (0.78) to the initial protection rate, $I_V(t)$, the number of infectious people under vaccination was also calculated. $I(t)$ and $I_V(t)$ are shown in Figure 2, and the daily data from 20 April to 10 September can be seen in Supplementary Table S1. Without vaccination, the daily number of infectious people would have increased from 123 on 20 April to a peak of 1,018 on 7 August, and then decreased to 96 on 10 September. The total number of infectious cases would have been 43,455 during this time period. Under the assumption that the population were vaccinated, the number of infectious people would have constantly decreased from 123 on 20 April to 18 on 10 September. The outbreak of the pandemic in August, which has led to thousands of people becoming infected, could have been constantly controlled at a lower level. The total number of infectious cases during this time period would have been 6,417, which is 85.2% $<$ the number without vaccination.

Simulation of infection in the stadium audiences during the CSL game season

Based on the number of infectious people calculated by the generalized SEIR model, parameter A in the Wells–Riley model with and without vaccination could be identified, and the daily infection probability of the stadium audiences during the 2021 CSL game season under six scenarios was calculated (Figure 3). The infection probabilities of Scenario 1 (p1), p3, and p5 reflect those scenarios without vaccination, and their curves were similar to that of $I(t)$. Meanwhile, the infection probabilities of Scenario 2 (p2), p4, and p6 reflect those scenarios

TABLE 2 Parameters in the Wells–Riley model.

Parameter	Description	S1	S2	S3	S4	S5	S6
A	Number of infectious people entering the stadium	nI(t)/N	$nI_v(t)/N$	0.3nI(t)/N	$0.3nI_v$ (t)/N	0.3nI (t)/N	$0.3nI_v$ (t)/N
θ	Penetration ratio of face mask	1 (32)	1 (32)	1 (32)	1 (32)	0.25 (26)	0.25 (26)
v	Pulmonary ventilation rate of a person	$3m^3$ h (31)					
n	Number of people in the stadium	30000 (4)					
V	Volume of the stadium	30000 m^3 (31)					
t	Exposure time	2 h					
$\bar{f}$	Exhaled air volume fraction	1.1 m^3 h (42)					
q	Quanta produced by infectious people	0.16A+36 (31)					

S1, Scenario 1, without any preventive strategy; S2, Scenario 2, with vaccination against COVID-19; S3, Scenario 3, with a negative nucleic acid testing result within 7 days before admission; S4, Scenario 4, with combined vaccination and nucleic acid testing; S5, Scenario 5, with combined nucleic acid testing and face mask wearing; S6, Scenario 6, with combined vaccination, nucleic acid testing, and face mask wearing.

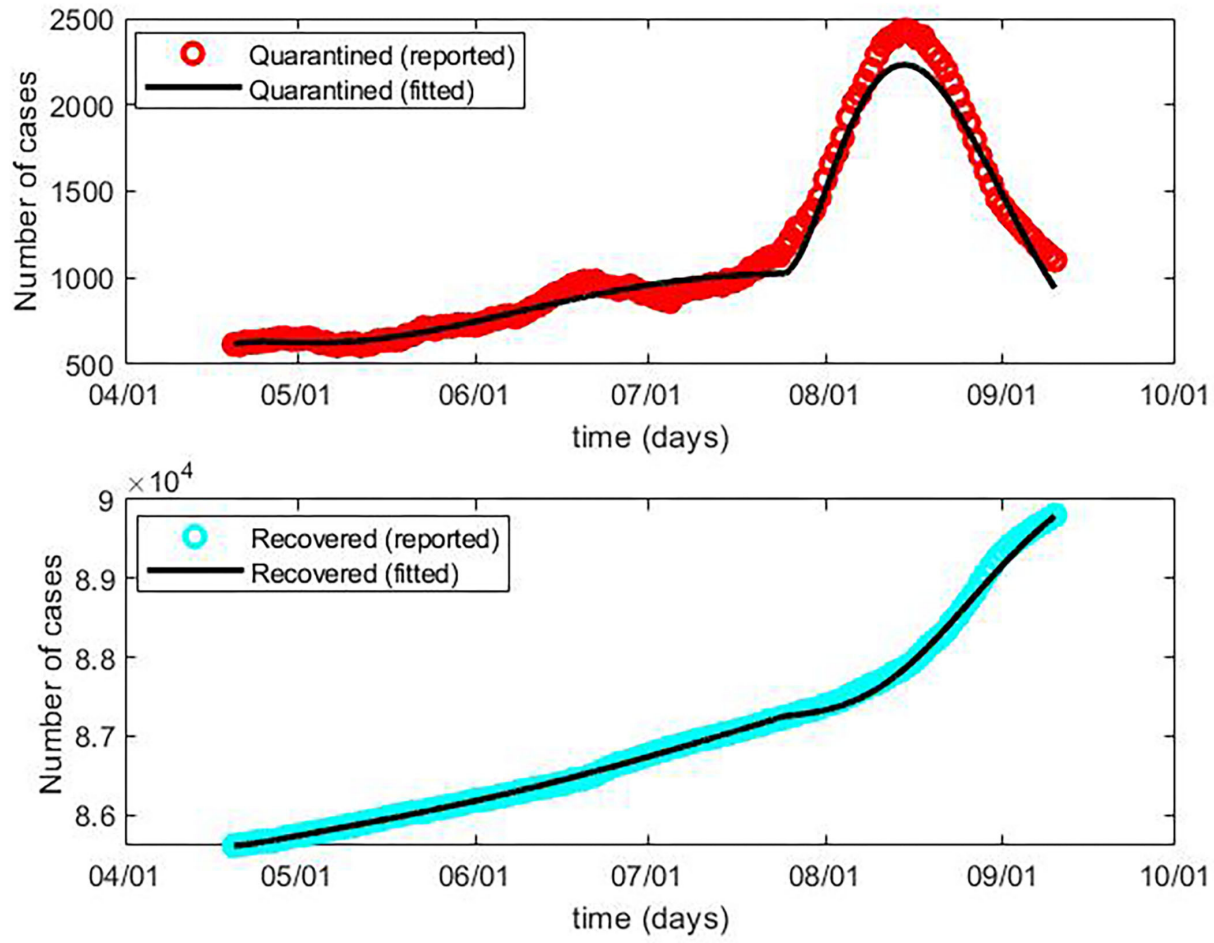


FIGURE 1
The reported and fitted current quarantined cases and cumulative recovered cases of COVID-19, China, 20 April to 10 September.

with vaccination, which decreased progressively from 20 April to 10 September in accordance with the curve of I_v *(t)*.

According to the game schedule, there were 116 games held on 58 game days in the 2021 season. The daily and the total number of infected people for the game season are shown in Table 4. If all of the 30,000 seats in the stadium were taken by audience members, just as it was before the COVID-19 pandemic, 3–25 audience members would have

10.3389/fpubh.2022.1009152

TABLE 3 Values for unknown parameters in the fitted SEIR model.

Parameter	Description	20 April to 24 July	25 July to 10 September
α	Protection rate	0.04	0.37
β	Infection rate	0.81	0.29
γ^{-1}	Average latent time (day)	87.0	6.9
δ^{-1}	Average quarantine time (day)	9.3	9.3
$\lambda(t)$	Cure rate	0.02/{1+exp[−0.46*(t−0.78)]}	0.07/{1+exp[−0.09*(t−26.11)]}
$\kappa(t)$	Mortality rate	$[10^{(}-4)]$*{exp[−0.06*(t-69350)]^2}	$[10^{(}-3)]$/{exp[0.69*(t−4.49)]+exp[−0.69*(t−4.49)]}
α_v	Protection rate with vaccination	0.82 (28)[a]	

[a] αv= α+0.78, 0.78 is the efficacy of Sinopharm/BBIBP COVID−19 vaccine reported by WHO.

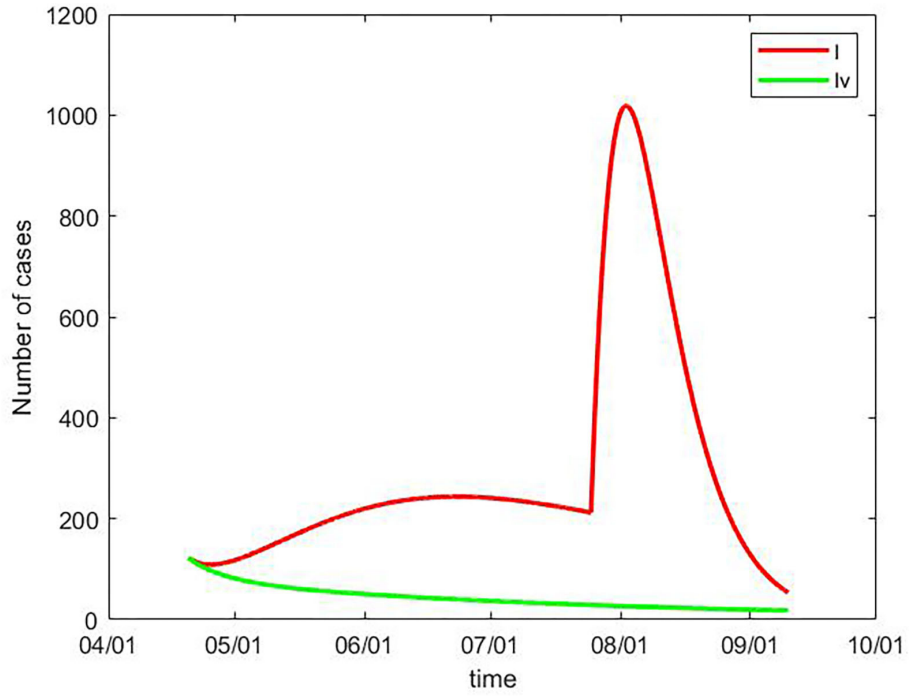


FIGURE 2
Number of infectious people with vaccination I_v *(t)* and without vaccination *I(t)*, China, 20 April to 10 September.

been infected with COVID-19 for each game day without any preventive strategies, and a total of 371 audiences would have been infected during this game season (Scenario 1). With the requirement of a negative nucleic acid test of SARS-CoV-2 within 7 days before admission, the daily number of infected audience members would have ranged from one to seven, and the total number of infected people during this game season would have been 118 (Scenario 3). The combination of nucleic acid testing and face mask wearing would have further decreased this number to 34, meaning 27 game days would be "completely safe" for the audiences and one or two audience members would have been infected for each of the other 31 game days (Scenario 5). If the general population were vaccinated before the game season began, the number of infected audiences would have been 79 during this season, with one to three audience members would have been infected on each game day (Scenario 2). The combination of vaccination and nucleic acid testing would have decreased the number of infected audience members to 14, with 42 safe game days (Scenario 4). The number of infected people would have been zero after rounding off when combining all three preventive strategies (Scenario 6).

The daily infection probability and the number of infected audiences can be seen in Supplementary Table S1.

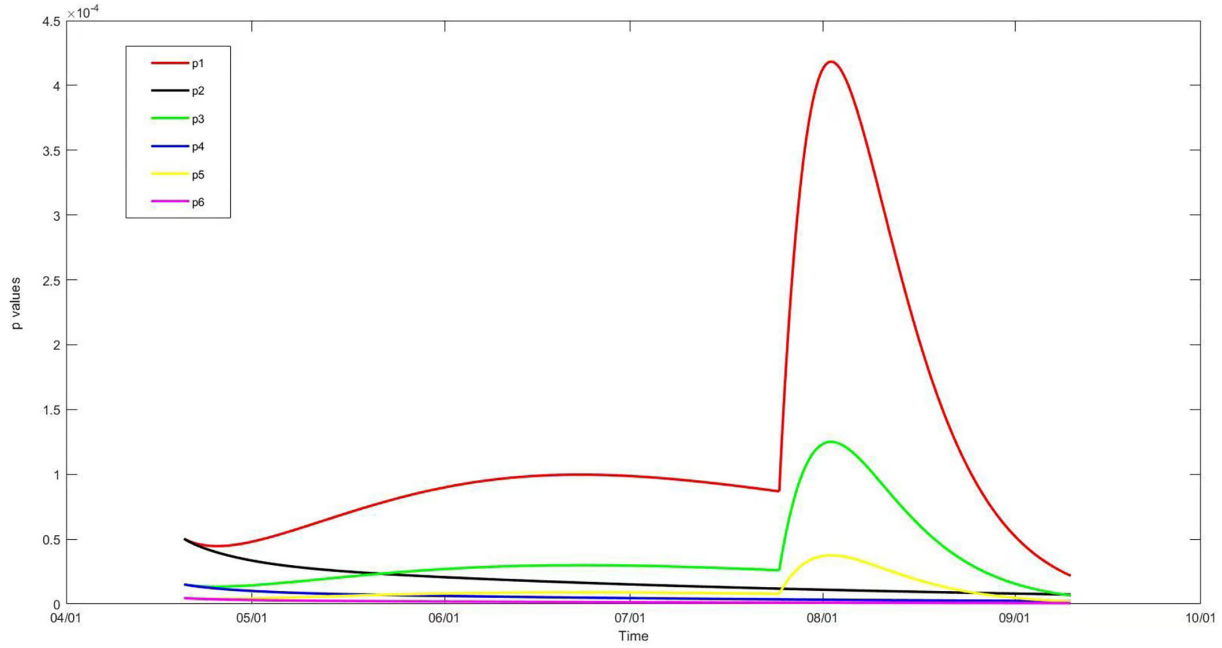


FIGURE 3
Infection probability (p) in stadium, 20 April to 10 September. p1, without any preventive strategy; p2, with vaccination against COVID-19; p3, with a negative nucleic acid testing result within 7 days before admission; p4, with combined vaccination and nucleic acid testing; p5, with combined nucleic acid testing and face mask wearing; p6, with combined vaccination, nucleic acid testing, and face mask wearing.

Discussion

We conducted mathematic modeling analyses to assess the effect of vaccination, nucleic acid testing, and face mask wearing on the transmission of COVID-19 in stadium audiences during the 2021 CSL game season. By using these modeling approaches, this study incorporated the natural history of COVID-19 infection and predicted the exact number of infected audience members upon using various preventive strategies, which may guide the resumption of large-scale sports events with audiences.

The generalized SEIR model interpreted the dynamics of COVID-19 transmission over 5 months, from a relatively stationary phase to an outbreak of cases. The values of the parameters in this study share some similarities with those in previous studies. The protection rate was 0.04 in the first phase, which is close to Cheynet's estimate of 0.06 (13), and a little bit less than the estimate of 0.17 by Peng et al. (33). These two studies used the generalized SEIR model to simulate Chinese data from January 2020 onward. The scatter outbreak in the first phase may have alerted people and the government to conduct some preventive interventions (33), and led to an increase in the protection rate to 0.37 in the later outbreak from 25 July to 10 September. The infection rate was 0.81 in the first phase, which is equal to the estimate in Cheynet's study (13) and close to the estimate of 1.0 in Peng et al.'s study (33), which indicates that most people were susceptible to the virus at the outbreak of COVID-19. As the protection rate increased, the infection rate decreased to 0.29 in the second phase. According to the clinical study of Jiang et al. on the characteristics of COVID-19 cases, the average latent time is 5.2 days, and 95% of cases develop clinical symptoms in 12.5 days (2), which is similar to the latent time of 6.9 days in the later phase in our study. The latent time of 6.9 days in the second phase and the quarantine time of 9.3 days estimated in this study are also close to the estimates of 5 and 10 days in Cheynet's study, and of 2 and 6.6 days in Peng et al.'s study, respectively. Simulating the COVID-19 pandemic in different time periods may lead to minor variants in parameter estimation. However, the latent time of 87 days for the first phase was much longer than the latent time shown in previous studies (2, 13, 33). This extensive simulation tried to fit the real-world data in a stationary stage by extending the latent time and, therefore, showed a flat curve of quarantined cases. This indicates that less outbreaks scattered across the country may be difficult to predict using the SEIR model. This mathematical model is more useful for interpreting a typical pandemic with a high peak value.

As the SARS-CoV-2 variants become more transmissible, the introduction of SARS-CoV-2 in audiences in sports stadiums is almost inevitable (34). The resumption of sports events with audiences, which is undoubtedly important for the audience experience and the development of the sports industry, can greatly increase the risk of infection due to mass gathering and transportation (35). In this study, we showed that without any preventive strategies, over 300 audience members would have

10.3389/fpubh.2022.1009152

TABLE 4 Number of infected audiences during 2021 CLS game season.

Date	Number of infected spectators					
	S1	S2	S3	S4	S5	S6
20 April	3	3	1	1	0	0
21 April	3	3	1	1	0	0
22 April	3	3	1	1	0	0
23 April	3	3	1	1	0	0
26 April	3	2	1	1	0	0
27 April	3	2	1	1	0	0
28 April	3	2	1	1	0	0
29 April	3	2	1	1	0	0
2 May	3	2	1	1	0	0
3 May	3	2	1	1	0	0
4 May	3	2	1	1	0	0
5 May	3	2	1	1	0	0
8 May	3	2	1	1	0	0
9 May	4	2	1	1	0	0
10 May	4	2	1	0	0	0
11 May	4	2	1	0	0	0
14 May	4	2	1	0	0	0
15 May	4	2	1	0	0	0
16 May	4	2	1	0	0	0
17 May	4	1	1	0	0	0
21 June	6	1	2	0	1	0
22 June	6	1	2	0	1	0
23 June	6	1	2	0	1	0
24 June	6	1	2	0	1	0
26 June	6	1	2	0	1	0
27 June	6	1	2	0	1	0
28 June	6	1	2	0	1	0
29 June	6	1	2	0	1	0
2 July	6	1	2	0	1	0
3 July	6	1	2	0	1	0
4 July	6	1	2	0	1	0
5 July	6	1	2	0	1	0
7 July	6	1	2	0	1	0
8 July	6	1	2	0	1	0
9 July	6	1	2	0	1	0
10 July	6	1	2	0	1	0
12 July	6	1	2	0	1	0
13 July	6	1	2	0	1	0
14 July	6	1	2	0	1	0
15 July	6	1	2	0	1	0
17 July	6	1	2	0	1	0
18 July	6	1	2	0	1	0
19 July	6	1	2	0	0	0
20 July	5	1	2	0	0	0
23 July	5	1	2	0	0	0
24 July	5	1	2	0	0	0
25 July	5	1	2	0	0	0
26 July	5	1	2	0	0	0
28 July	5	1	2	0	0	0
29 July	10	1	3	0	1	0
30 July	14	1	4	0	1	0
31 July	17	1	5	0	2	0
2 August	22	1	7	0	2	0
3 August	24	1	7	0	2	0
4 August	24	1	7	0	2	0
5 August	25	1	7	0	2	0
Total	371	79	118	14	34	0

S1, Scenario 1, without any preventive strategy; S2, Scenario 2, with vaccination against COVID−19; S3, Scenario 3, with a negative nucleic acid testing result within 7 days before admission; S4, Scenario 4, with combined vaccination and nucleic acid testing; S5, Scenario 5, with combined nucleic acid testing and face mask wearing; S6, Scenario 6, with combined vaccination, nucleic acid testing, and face mask wearing.

been infected in the 2021 game season. For the sake of the COVID-19 pandemic, the 2021 game schedule was simplified by cutting 50% of games, and this season ended with a premature closing before the major outbreak in August. It can be inferred that more audience members would have been infected if home-and-away games were played all over the country.

The combined use of all three strategies, namely, vaccination, nucleic acid testing, and face mask wearing, was the most effective measure in this study, which would have completely protected audiences from infection. Vaccination can effectively control the transmission of SARS-CoV-2 in the general population and audiences. By simulating a scenario with vaccination, this study showed that high levels of population immunity would be generated if all members of the general population received vaccination with 78% efficiency. As a result, the number of infectious people entering stadiums could be decreased, further leading to a lower infection probability of audience members. Moreover, vaccination was shown to be more effective than nucleic acid testing when independently using either strategies. Given that watching a game while wearing a face mask impairs the audience experience, vaccination and nucleic acid testing may work as a compromise between safety and the audience experience when designing preventive strategies for a sports event.

However, the effectiveness of vaccination against COVID-19 is challenged by some uncertainties. First, the vaccine efficacy has been shown to range from 22% (36) to 100% (37) in clinical trials of different types of vaccines. Second, the Omicron variant significantly increased the transmissibility and immune evasion of the virus. Its influence on vaccine efficacy still holds a great

degree of ambiguity (38), and booster doses may be needed consistently to resist the waning of vaccine efficacy over time (39). Moreover, the vaccine mistrust and anti-vaccine movement in some places are a threat to the coverage of vaccination (40). Given the uncertainties around vaccination strategy, this study also simulated some scenarios in which there was no vaccination. The results showed that the independent use of nucleic acid testing would have decreased the number of infected audience members by two-third, and the combined use of nucleic acid testing and face mask wearing would have further decreased the number of infected audience members. Although nucleic acid testing and face mask wearing are less effective than strategies combined with vaccination, policymakers can consider these strategies to control the infection probability to some degree.

There are several limitations related to modeling assumptions, data limitations, and uncertainty. First, modeling is a process to simplify a real situation. During the 2021 CSL game season, the extensive COVID-19 pandemic that broke out in late 2019 had been controlled and limited to scattered outbreaks in a couple of cities. Home-and-away games were also limited in two cities. Therefore, some assumptions of the SEIR model, including an even mix of the infectious and susceptible populations and people having the same odds ratio of entering the stadium, may not be well fitted to a real-world situation. However, most of the parameter estimates in our study were reasonable and in accordance with those of previous studies based on Chinese data (2, 13, 33). It was indicated that the generalized SEIR model and the initial conditions used in this study are efficient for the prediction of COVID-19 transmission in China. Second, the simulation results based on the Wells–Riley model cannot be validated by real-world data, because most of the 2021 CLS games were played in an empty stadium. Third, there is uncertainty around some of the model parameters, such as the vaccine efficacy, penetration ratio of face mask, and false negative ratio of nucleic acid testing, which were extracted according to previous studies. Future modeling studies could build on this effort when updated data become available for model parameter estimation and calibration targets. The results of the Wells–Riley model could be tested when the auditorium reopening policy was implemented for large sports events in China.

In conclusion, the combined use of the SEIR and Wells–Riley models offers a tool to benchmark the effects of preventive strategies against COVID-19 in activities involving mass gatherings such as large sports events, which include transportation and confined-space gathering. Our findings showed that the reopening of the auditorium would have a potential infection risk without any preventive strategies. The combined use of vaccination, nucleic acid testing, and face mask wearing could effectively protect audience members against infection. The use of any two of these strategies could significantly lower the infection rates. Accordingly, the public can understand the risk of game watching, and policymakers can consider the combined use of preventive strategies when holding sports events, which could create a balance between audience experiences and COVID-19 infection control.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

HZ, ZL, and ND conceptualized the study. XS, HR, XH, WZ, and CL collated the data. ZL and XS analyzed the data. HZ drafted the initial manuscript. ZL and ND reviewed and revised the manuscript. All authors contributed to the interpretation of the results and approved the submitted version.

Funding

This study was supported by the Fundamental Research Funds for the Central Universities, Beijing Sport University (Grant Nos. 2022YB007, 2021TD002, and 3001049).

Acknowledgments

We would like to thank the suggestions for study design from Professor Shuxiang Zhao and Professor Ran Li.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.1009152/full#supplementary-material

References

1. WHO. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update 2021. Available online at: https://www.who.int/publications/m/item/weekly-epidemiological-update-$-$2-february-2021 (accessed January 31, 2021).

2. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z., et al. Review of the clinical characteristics of Coronavirus Disease 2019 (COVID-19). *J Gen Intern Med*. (2020) 35:1545–9. doi: 10.1007/s11606-020-05762-w

3. Jin H, Wang H, Li X, Zheng W, Ye S, Zhang S, et al. Economic burden of COVID-19, China, January-March, 2020: a cost-of-illness study. *Bull World Health Organ*. (2021) 99:112–24. doi: 10.2471/BLT.20.267112

4. Xia J, Han J. Research on the impact of Chinese super league watching experience on the spectators'consumption intention. *J Jilin Sport Univer*. (2020) 36:43–52.

5. WHO. *Evidence Assessment: Sinopharm/BBIBP COVID-19 Vaccine*. SAGE Working Group on COVID-19 vaccines. (2021). Available online at: https://cdn.who.int/media/docs/default-source/immunization/sage/2021/april/2_sage29apr2021_criticalevidence_sinopharm.pdf?sfvrsn=3dfe32c1_5 (accessed May 21, 2021).

6. Jara A, Undurraga EA, González C, Paredes F, Fontecilla T, Jara G, et al. Effectiveness of an inactivated SARS-CoV-2 Vaccine in Chile. *N Engl J Med*. (2021) 385:875–84. doi: 10.1056/NEJMoa2107715

7. Al Kaabi N, Zhang Y, Xia S, Yang Y, Al Qahtani MM, Abdulrazzaq N, et al. Effect of 2 inactivated SARS-CoV-2 vaccines on symptomatic COVID-19 infection in adults: a randomized clinical trial. *JAMA*. (2021) 326:35–45. doi: 10.1001/jama.2021.8565

8. Wu Z, Hu Y, Xu M, Chen Z, Yang W, Jiang Z, et al. Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine (CoronaVac) in healthy children and adolescents: a double-blind, randomised, controlled, phase 1/2 clinical trial. *Lancet Infect Dis*. (2021). doi: 10.2139/ssrn.3820545

9. Jan R, Boulaaras S. Analysis of fractional-order dynamics of dengue infection with non-linear incidence functions. *Trans Institute Meas Control*. (2022) 44:2630–41. doi: 10.1177/01423312221085049

10. Jan R, Boulaaras S, Ali Shah SA. Fractional-calculus analysis of human immunodeficiency virus and CD4+ T-cells with control interventions (accepted). *Commun Theoretical Phy*. (2022) 74:105001. doi: 10.1088/1572-9494/ac7e2b

11. Tang T-Q, Shah Z, Bonyah E, Jan R, shutaywi M, Alreshidiet N et al. Modeling and analysis of breast cancer with adverse reactions of chemotherapy treatment through fractional derivative. *Comput Math Methods Med*. (2022) 2022:5636844. doi: 10.1155/2022/5636844

12. Boulaaras S, Jan R, Khan A, Ahsan, M. Dynamical analysis of the transmission of dengue fever via Caputo-Fabrizio fractional derivative. *Chaos, Solitons & Fractals: X*. (2022) 8:100072. doi: 10.1016/j.csfx.2022.100072

13. Cheynet E. Generalized SEIR Epidemic Model (fitting and computation). *Zenodo*. (2020). doi: 10.5281/ZENODO.3911854

14. Kong L, Duan M, Shi J, Hong J, Chang Z, Zhang Z, et al. Compartmental structures used in modeling COVID-19: a scoping review. *Infect Dis Poverty*. (2022) 11:72. doi: 10.1186/s40249-022-01001-y

15. Wang X, Wu H, Tang S. Assessing age-specific vaccination strategies and post-vaccination reopening policies for COVID-19 control using SEIR modeling approach. *Bull Math Biol*. (2022) 84:108. doi: 10.1007/s11538-022-01064-w

16. Canga A, Bidegain G. Modelling the effect of the interaction between vaccination and nonpharmaceutical measures on COVID-19 incidence. *Glob Health Epidemiol Genom*. (2022) 2022:9244953. doi: 10.1155/2022/9244953

17. Lin L, Chen B, Zhao Y, Wang W, He D, et al. Two waves of COIVD-19 in Brazilian cities and vaccination impact. *Math Biosci Eng*. (2022) 19:4657–71. doi: 10.3934/mbe.2022216

18. Junge M, Li S, Samaranayake S, Zalesak M. Safe reopening of university campuses is possible with COVID-19 vaccination. *PLoS ONE*. (2022) 17:e0270106. doi: 10.1371/journal.pone.0270106

19. Lin L, Zhao Y, Chen B, He D. Multiple COVID-19 Waves and Vaccination Effectiveness in the United States. *Int J Environ Res Public Health*. (2022) 19:2282. doi: 10.3390/ijerph19042282

20. Liu W, Guo Z, Abudunaibi B, Ouyang X, Wang D, Yang T, et al. Model-based evaluation of transmissibility and intervention measures for a COVID-19 outbreak in Xiamen City, China. *Front Public Health*. (2022) 10:887146. doi: 10.3389/fpubh.2022.887146

21. Riley EC, Murphy GR, Riley RL. Airborne spread of measles in a suburban elementary school. *AmJ Epidemiol*. (1978) 107:421–32. doi: 10.1093/oxfordjournals.aje.a112560

22. Guo Y, Qian H, Sun Z, Cao J, Liu F, Luo X, et al. Assessing and controlling infection risk with Wells-Riley model and spatial flow impact factor (SFIF). *Sustain Cities Soc*. (2021) 67:102719. doi: 10.1016/j.scs.2021.102719

23. Wang Z, Galea ER, Grandison A, Ewer J, Jia F. Inflight transmission of COVID-19 based on experimental aerosol dispersion data. *J Travel Med*. (2021) 28:23. doi: 10.1093/jtm/taab023

24. Wang Z, Galea ER, Grandison A, Ewer J, Jia F,et al. A coupled Computational Fluid Dynamics and Wells-Riley model to predict COVID-19 infection probability for passengers on long-distance trains. *Saf Sci*. (2022) 147:105572. doi: 10.1016/j.ssci.2021.105572

25. Che W, Ding J, Li L. Airflow deflectors applied to external windows of a classroom ventilated with induced wind: towards COVID-19 prevention and control. *Sustain Cities Soc*. (2021) 77:103548. doi: 10.1016/j.scs.2021.103548

26. Sha H, Zhang X, Qi D. Optimal control of high-rise building mechanical ventilation system for achieving low risk of COVID-19 transmission and ventilative cooling. *Sustain Cities Soc*. (2021) 74:103256. doi: 10.1016/j.scs.2021.103256

27. NHC-China. *Daily Briefing on Novel Coronavirus Cases in China 2021*. Available online at: http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=1be44bd0ea884a6c82f1fe990ead9e56 (accessed April 21, 2021).

28. WHO. *Evidence Assessment: Sinopharm/BBIBP COVID-19 vaccine 2021*. Available online at: https://cdn.who.int/media/docs/default-source/immunization/sage/2021/april/2_sage29apr2021_critical-evidence_sinopharm.pdf?sfvrsn=3dfe32c1_5 (accessed May 4, 2021).

29. Wells W. *Airborne Contagion and Air Hygiene: An Ecological Study of Droplet Infection*. Cambridge, London: Harvard University Press. (1955).

30. Rudnick SN, Milton DK. Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. *Indoor Air*. (2003) 13:237-45 doi: 10.1034/j.1600-0668.2003.00189.x

31. Sheng H. Modeling evaluation on restart of large sports events after COVID-19. *Comput Engineer Appl*. (2020) 56:33–40. doi: 10.3778/j.issn.1002-8331.2006-0148

32. Fennelly KP, Nardell EA. The relative efficacy of respirators and room ventilation in preventing occupational tuberculosis. *Infect Control Hosp Epidemiol*. (1998) 19:754–9 doi: 10.2307/30141420

33. Peng L, Yang W, Zhang D, Zhuge C, Hong L, et al. Epidemic analysis of COVID-19 in China by dynamical modeling. *medRxiv*.

34. Ryckman T, Chin ET, Prince L, Leidner D, Long E, Studdert D, et al. Outbreaks of COVID-19 variants in US prisons: a mathematical modelling analysis of vaccination and reopening policies. *Lancet Public Health*. (2021) 6:e760–e70. doi: 10.1016/S2468-2667(21)00162-6

35. Saidan MN, Shbool MA, Arabeyyat OS, Al-Shihabi ST, Abdallat YA, Barghash MA, et al. Estimation of the probable outbreak size of novel coronavirus (COVID-19) in social gathering events and industrial activities. *Int J Infect Dis*. (2020) 98:321–7. doi: 10.1016/j.ijid.2020.06.105

36. Behera P, Singh AK, Subba SH, et al. Effectiveness of COVID-19 vaccine (Covaxin) against breakthrough SARS-CoV-2 infection in India. *Hum Vaccin Immunother*. (2022) 18:2034456. doi: 10.1080/21645515.2022.2034456

37. SinaNews. *China's Protein Subunit Vaccine Showed Efficacy of 81.76% for Prevention COVID-19, and 77.54% for Delta variant 2021*. Available online at: https://news.sina.com.cn/c/2021-08-28/doc-iktzscyx0875449.shtml (accessed August 28, 2021).

38. Madewell ZJ, Dean NE, Berlin JA, Coplan PM, Davis KJ, Struchiner CJ, et al. Challenges of evaluating and modelling vaccination in emerging infectious diseases. *Epidemics*. (2021) 37:100506. doi: 10.1016/j.epidem.2021.100506

39. Joshi G, Poduri R. Omicron, a new SARS-CoV-2 variant: assessing the impact on severity and vaccines efficacy. *Hum Vaccin Immunother*. (2022) 18:2034458. doi: 10.1080/21645515.2022.2034458

40. Khadafi R, Nurmandi A, Qodir Z, Misran, et al. Hashtag as a new weapon to resist the COVID-19 vaccination policy: a qualitative study of the anti-vaccine movement in Brazil, USA, and Indonesia. *Hum Vaccin Immunother*. (2022) 18:2042135. doi: 10.1080/21645515.2022.2042135

41. China NBoSo. *China Statistical Yearbook*. Beijing: China Statistics Press. (2020).

42. Xu C, Gong C, Wang Y, Nielsen P, Liu L. Effect of air stability on the continuous and periodic exhalation flow of a person. *J Hunan Univer (Natural Sci)*. (2014) 41:106–12. doi: 10.3969/j.issn.1674-2974.2014.09.017

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fpubh.2022.1001340

OPEN ACCESS

EDITED BY
Gopi Battineni,
University of Camerino, Italy

REVIEWED BY
Junxiang Chen,
University of Pittsburgh, United States
Muhammad Imran Khan,
The University of Haripur, Pakistan

*CORRESPONDENCE
Shiren Sun
sunshiren@medmail.com.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

RECEIVED 23 July 2022
ACCEPTED 24 October 2022
PUBLISHED 17 November 2022

CITATION
Yu Z, Li X, Zhao J and Sun S (2022) Identification of hospitalized mortality of patients with COVID-19 by machine learning models based on blood inflammatory cytokines.
Front. Public Health 10:1001340.
doi: 10.3389/fpubh.2022.1001340

Identification of hospitalized mortality of patients with COVID-19 by machine learning models based on blood inflammatory cytokines

Zhixiang Yu[1†], Xiayin Li[1†], Jin Zhao[1] and Shiren Sun[1,2*]

[1]Department of Nephrology, Xijing Hospital, Fourth Military Medical University, Xi'an, China, [2]First Unit, Third Branch of Fangcang Shelter Hospital of National Exhibition and Convention Center, Shanghai, China

Coronavirus disease 2019 (COVID-19) spread worldwide and presented a significant threat to people's health. Inappropriate disease assessment and treatment strategies bring a heavy burden on healthcare systems. Our study aimed to construct predictive models to assess patients with COVID-19 who may have poor prognoses early and accurately. This research performed a retrospective analysis on two cohorts of patients with COVID-19. Data from the Barcelona cohort were used as the training set, and data from the Rotterdam cohort were used as the validation set. Cox regression, logistic regression, and different machine learning methods including random forest (RF), support vector machine (SVM), and decision tree (DT) were performed to construct COVID-19 death prognostic models. Based on multiple clinical characteristics and blood inflammatory cytokines during the first day of hospitalization for the 138 patients with COVID-19, we constructed various models to predict the in-hospital mortality of patients with COVID-19. All the models showed outstanding performance in identifying high-risk patients with COVID-19. The accuracy of the logistic regression, RF, and DT models is 86.96, 80.43, and 85.51%, respectively. Advanced age and the abnormal expression of some inflammatory cytokines including IFN-α, IL-8, and IL-6 have been proven to be closely associated with the prognosis of patients with COVID-19. The models we developed can assist doctors in developing appropriate COVID-19 treatment strategies, including allocating limited medical resources more rationally and early intervention in high-risk groups.

KEYWORDS

COVID-19, outcome, inflammatory cytokines, machine learning, prognostic models

Introduction

Syndrome coronavirus-2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19), which infected more than 180 million people. Compared with a similar acute respiratory syndrome caused by the severe acute respiratory syndrome coronavirus, COVID-19 seems milder but more infectious (1). After being infected with COVID-19, patients' characteristics vary. Some patients became clinically asymptomatic or had mild cases of fever, cough, fatigue, and other symptoms. However, some people became patients with severe COVID-19 and even lost their lives (2–4). Due to inappropriate assessments and treatment, strategies were detrimental to the patient's health and promoted SARS-CoV-2 to become a global societal problem, which has caused over 4 million deaths to date (5). Therefore, identifying patients with poor prognoses as early as possible for necessary interventions is an essential direction for the treatment of COVID-19, which will significantly improve the prognosis of patients and release a tremendous burden on the medical care system.

The immune system and inflammatory syndrome have been proven to play a crucial role in COVID-19 infection (6). Inflammatory cytokines are critical mediators that oversee and regulate immune and inflammatory responses *via* complex networks and serve as biomarkers for many diseases (7). According to previous studies (8–10), inflammatory cytokines were closely related to the progression, complications, and mortality of COVID-19. Universally, these studies paid attention to the relationship between cytokines and disease severity. However, few researchers specifically employed cytokines to construct a model for predicting the prognosis of patients with COVID-19.

Machine learning (ML) algorithms have been widely applied in the medical field, including diagnosing and predicting prognosis. ML models are also used in every aspect of the diagnosis and treatment of COVID-19 due to their fantastic data processing capabilities. Previous ML studies have used multiple indicators, including clinical and blood text indicators, to determine the prognosis of patients with COVID-19. Due to cytokine tests' simplicity, high efficiency, and accuracy, they gradually became an alternative plan for the early prediction of COVID-19 prognosis. Abers et al. (11) fit a Cox proportional hazard to screen the mortality-related inflammatory cytokines. Patterson et al. (12) applied ML methods for the early identification of patients with severe COVID-19 based on cytokines. Mueller et al. (13) classified patients with COVID-19 into different subgroups according to inflammatory cytokines and applied the immunotypes to predict long-term post-COVID-19 complications. However, there are no available models for early prediction of the death of patients with COVID-19 based on blood inflammatory cytokines for clinical work. Constructing a predictive model that can be applied in the clinic seemed urgent. Therefore, in this research, we made use of data collected by Mueller et al. (13), containing 138 inpatients with COVID-19, to construct multiple models based on different algorithms, including logistic regression, random forest (RF), and decision tree (DT) to predict patient deaths.

Methods

Data acquisition

In this research, we obtained COVID-19 data from a dataset of Mueller et al. (13) after obtaining author approval. The data were from a finished cohort study, which has been approved by the Ethics Committee for Research with Medicines of Hospital Universitari Vall d'Hebron and Erasmus University Medical Center (13). Therefore, we do not require reapproval from the ethics committee for this study. The dataset contained 138 patients with COVID-19 from two independent cohorts (Rotterdam cohort $n = 50$ and Barcelona cohort $n = 88$). Clinical parameters and laboratory data for each patient during the first 24 h of hospitalization were included in the dataset. Final clinical outcomes were classified into discharge from the hospital and in-hospital death. All inpatients in the study cohort were older than 18 and SARS-CoV-2 positive diagnosed by reverse transcription-polymerase chain reaction (RT-PCR) test and were sampled at hospital entry. Dataset included cytokines measured through the ELLA Simple Plex system and ELISA kits. The COVID-19 antibody concentration in serum was measured through ELISA, and other clinical indicators were obtained by routine tests. Data acquisition is presented in the study by Mueller et al. (13).

Data preprocessing

First, we binarized the outcome variables into Booleans. We used mean value interpolation to substitute missing values, which was widely used and proved effective for missing values in datasets. Thankfully, no missing values in the COVID-19 patients' cytokines values ensured the model's reliability. We used the Barcelona cohort to construct the model and applied the Rotterdam cohort as an independent external validation set.

Patients' inflammatory cytokine values from the training and validation datasets were loaded into R software (version 4.1.0). R packages "vegan" and "stats" were applied to perform the principal component analysis (PCA) and draw the PCA figure. Besides, we used "ggplot2" for the figure drawing.

Our research selected the following five algorithms: Cox regression, logistic regression, RF, support vector machine (SVM), and decision tree (DT). COVID-19 cases located in different areas were conducted to verify the accuracy of each model to ensure the model's reliability. In the logistic regression

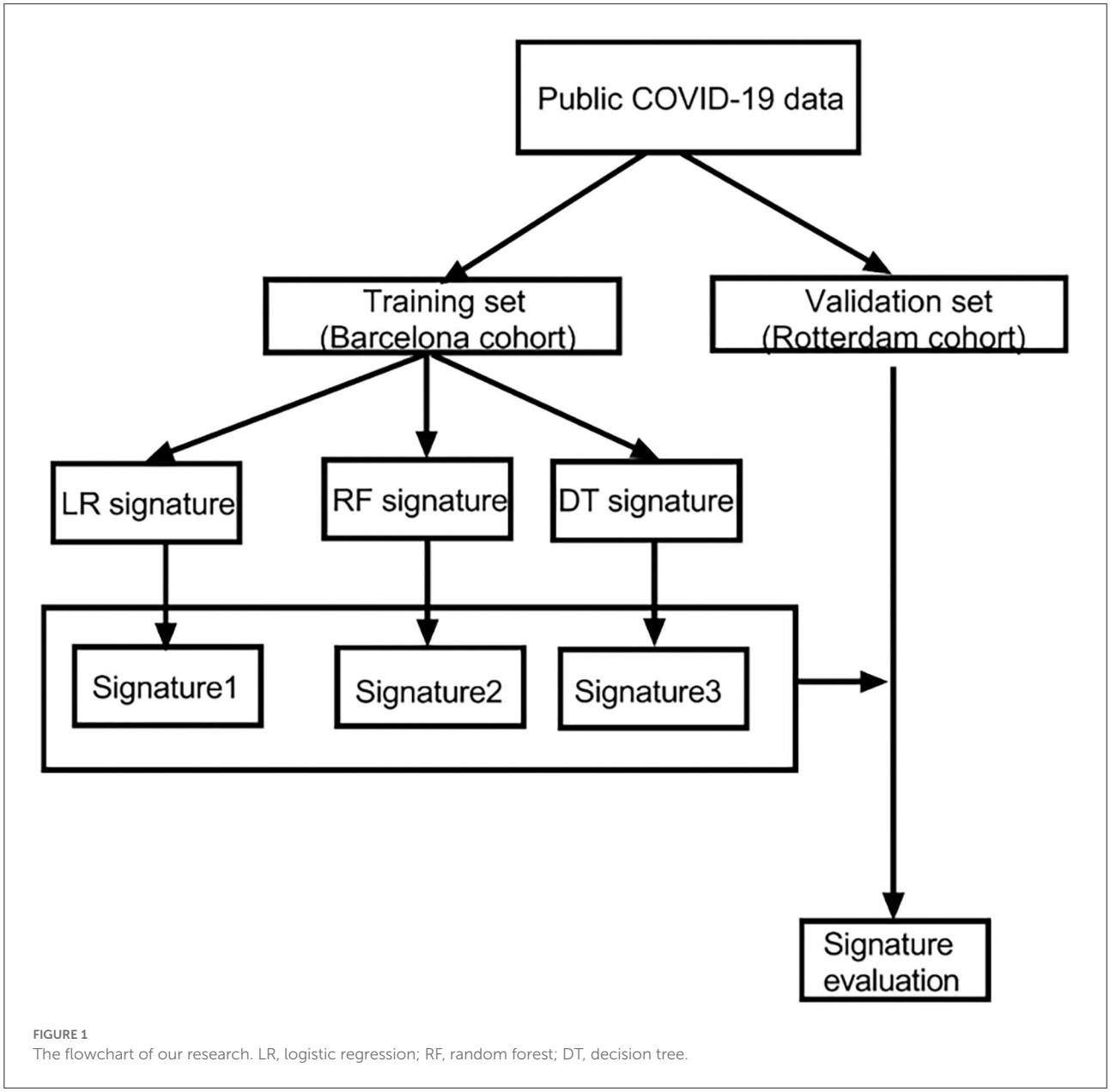


FIGURE 1
The flowchart of our research. LR, logistic regression; RF, random forest; DT, decision tree.

model, we input all the inflammatory cytokine values, age, and sex into the logistic regression in the training set. Then, the variables with $p < 0.1$ in the single logistic regression were filtered for the following research. We applied multiple logistic regression (backward: likelihood ratio method) multivariate analysis for the hub variables, and the coefficients derived were used to generate a prognostic model. A prognostic model was constructed based on the logistic regression coefficients. For verification, we generated ROC and calibration curves to calculate the model results for the training set and validation set. The area under the curve (AUC) and 95% confidence interval (CI) were used to verify the model efficiency. Moreover, to obtain a more comprehensive evaluation of the application value of the signature, decision curve analysis (DCA) was performed, demonstrating the net benefit to the patients with COVID-19 after applying the model for prognosis. These steps above were finished using SPSS (version 26.0) and Stata software (version 16). The Cox regression was performed using STATA (version 16), too. The Cox prognostic model was based on the Cox regression coefficients. The ROC and DCA were both drawn to evaluate the model.

In supervised ML, we evaluated the residuals of both methods for the model's accuracy and compared the residuals in our research. After confirming that the RF model is a better method with fewer residuals, we created an RF model to obtain the variables' significance. Artificial neural network (ANN)

TABLE 1 Characteristics of patients with COVID-19.

Variables		Training cohort and testing cohort		*P* value
	All patients ($n = 138$)	Barcelona cohort ($n = 88$)	Rotterdam cohort ($n = 50$)	
Age	62 (54–70)	61 (50–70)	63 (57.25–69)	0.072
Male (n, %)	91 (65.9%)	58 (65.9%)	33 (66.0%)	0.991
WHO score (Entry)				0.000
3	38 (27.5%)	34 (38.6%)	4 (8.0%)	
4	60 (43.5%)	31 (35.2%)	29 (58.0%)	
5	20 (14.5%)	12 (13.6%)	8 (16.0%)	
6	12 (8.7%)	10 (11.4%)	2 (4.0%)	
7	8 (5.8%)	1 (1.1%)	7 (14.0%)	
Laboratory test				
CRP (mg/L)	113 (70, 117.4)	129.5 (76.7,186.8)	97.5 (57.8, 151.4)	0.126
Ferritin (μg/L)	757 (406.7, 1,031.5)	712 (375.8, 954)	882.5 (494.5, 1,131)	0.122
Leukocytes ($x10^9$/L)	7.6 (6.1, 9.9)	7.0 (5.1, 10.0)	7.5 (5.9,9,8)	0.338
Neutrophils ($x10^9$/L)	6.1 (4.6, 8.0)	5.5 (3.7, 7.7)	5.9 (4.2, 8.0)	0.223
Lymphocytes ($x10^9$/L)	1.1 (0.8, 1.4)	1.1 (0.8, 1.5)	1.1 (0.8, 1.4)	0.506
Monocytes ($x10^9$/L)	0.5 (0.3, 0.7)	0.4 (0.3, 0.6)	0.5 (0.3, 0.6)	0.164
Thrombocytes ($x10^9$/L)	211 (165.5, 288.8)	234 (150.8, 299)	223 (163.5, 292)	0.978
Outcomes				
LOS	13.5 (4, 30.3)	7 (2, 33.25)	14.5 (7.75, 24.75)	0.108
ICU LOS	0 (0, 16.3)	0 (0, 17)	0 (0, 15)	0.759
Mortality	27	18	9	0.825

BRI, balanced response immunotype; EXI, excessive inflammation immunotype; LAI, low antibody immunotype; CRP, C-reactive protein; LDH, lactate dehydrogenase.

model training was performed, and the ROCs were used to verify the model efficiency. What is more, we constructed a DT based on all inflammatory cytokines, sex, and age through the R package "rpart" and calculated the accuracy of the DT in the training set and validation set. The RF model was constructed using R packages "randomForest" and "neuralnet." DT was finished through R package "rpart." Additionally, the R package "pROC" was applied to perform ROC.

A flowchart is shown in the Figure 1 to help readers better understand the complete analysis steps.

Results

Characteristics of the patients

There are 138 patients with COVID-19 participating in our study cohort. There were 27 (19.6%) deceased cases, containing 18 (13.0%) dead cases in the Barcelona cohort and 9 (7.0%) dead cases in the Rotterdam cohort. In Table 1, we listed the characteristics of each variable in both cohorts. We found no significant differences between the two cohorts, which ensures the feasibility and rationality of applying both datasets to model and validate.

Building a logistic regression model to classify patients and evaluate the model

We compared the patients with COVID-19 with different outcomes. There existed a significant difference between survival and death cases in age, some antibodies, and cytokines, indicating that these characteristics might play a key role in promoting death (Table 2).

All the inflammatory cytokines, sex, and age of the training set were used as input variables to perform single logistic regression. Five variables, including IL8, IL6, IFN-α, and IL17-α, and age with $p < 0.1$ might play a vital role in the COVID-19 patients' classification (Table 3). Multivariate analysis revealed that there were four variables with $p < 0.05$. The logistic regression model was constructed based on these four hub variables (Table 3).

The AUC of the model in the training set was 0.919 and in the validation set was 0.7236 (Figures 2A,B). Both

TABLE 2 Characteristics of two cohorts.

Variables	Barcelona cohort ($n = 88$)			Rotterdam cohort ($n = 50$)		
	Survival ($n = 70$)	Mortality ($n = 18$)	*P* value	Survival ($n = 41$)	Mortality ($n = 9$)	*P* value
Age	57.16 ± 13.8	70.4 ± 10.5	0.000	63.4 ± 10.1	71.8 ± 15.0	0.047
Male (n, %)	49 (70%)	9 (50%)	0.095	27 (65.9%)	6 (66.7%)	1.000
Laboratory data						
CRP (mg/L)	125.8 (72.7, 183.7)	129.5 (102.0, 249.7)	0.2	101 (69.3, 148)	108 (43.8, 220.8)	0.97
Ferritin (μg/L)	649.5 (324,958.3)	954 (601.8, 954.0)	0.221	882.5 (539.3, 1,108)	856.5 (485.5, 1,177.5)	0.82
Leukocytes (10^9/L)	7.5 (6.2, 9.6)	8.5 (5.9, 12.4)	0.277	7.2 (5.2, 10.3)	6.4 (4.1, 10.4)	0.419
Neutrophils (10^9/L)	5.9 (4.7, 7.8)	6.6 (4.4, 9.8)	0.238	6.1 (3.8, 8.3)	4.7 (3.3, 5.6)	0.289
Lymphocytes (10^9/L)	1.1 (0.9, 1.4)	0.9 (0.8, 1.3)	0.333	1.1 (0.8, 1.5)	0.9 (0.7, 1.3)	0.254
Monocytes (10^9/L)	0.5 (0.4, 0.7)	0.4 (0.3, 0.8)	0.63	0.4 (0.3, 0.6)	0.3 (0.2, 1.1)	0.361
Thrombocytes (10^9/L)	211 (166.5, 282)	219 (154.5, 300.5)	0.926	244.8 (178.0, 332.8)	126.5 (85.3, 184.8)	0.005
Cytokines and anti-body						
anti-N IgM	11 (11, 13.9)	11 (11, 15.4)	0.911	11 (11.0, 23.7)	11 (11.0, 11.0)	0.045
anti-N IgG	11.6 (11, 27.5)	11 (11, 39.8)	0.819	18.6 (11.0, 33.1)	11.0 (11.0, 11.9)	0.014
anti-N IgA	20.7 (11, 61.7)	11 (11,64.1)	0.415	21.0 (11.0, 49.1)	11.0 (11.0, 12.8)	0.011
TGFb1	32,599.5 (23,850, 40,977.8)	31,597.5 (20,318, 44,350)	0.664	35,709 (26,997.5, 48,028.5)	18,182.5 (10,165.3, 30,744.0)	0.003
IL5	0.3 (0.1, 0.5)	0.4 (0.1, 1.4)	0.259	0.7 (0.4, 1.7)	0.9 (0.1, 4.0)	0.544
IFNg	7.2 (2.6, 18.6)	2.3 (1.1, 7.6)	0.014	7.4 (2.8, 19.3)	14.8 (3.4, 35.1)	0.456
IFNa	7.8 (1.5, 27.0)	13.6 (1.5, 75.9)	0.47	5.1 (1.5, 15.5)	22.7 (2.0, 78.8)	0.038
CCL2	541.5 (381, 804.3)	766.5 (552, 1,024.3)	0.019	671.0 (520.5, 1,198.3)	1,285.0 (839.3, 1,971.3)	0.086
IL6	44.8 (25.2, 87.4)	106 (47.6, 148.5)	0.007	38.3 (20.8,62.8)	375.3 (51.3, 1,495.3)	0.003
TNFa	18.8 (14.8, 23.6)	24.0 (14.7, 28.3)	0.289	18.6 (14.4, 22.0)	23.6 (17.8, 28.7)	0.029
IL1b	0.6 (0.2, 1.1)	0.5 (0.3, 1.0)	0.692	0.4 (0.2, 0.6)	0.5 (0.4, 1.0)	0.087
IL8	71.1 (40.7, 136.3)	90.1 (62.2, 316.5)	0.053	44.6 (27.2,80.4)	76.9 (32.8, 160.3)	0.093
IL18	385 (314.5, 512)	495.5 (343.3, 851.5)	0.044	493.5 (411.3, 673.0)	590.5 (396.8, 711.0)	0.544
IL10	13.6 (6.7, 22.1)	21 (15.7, 27.9)	0.02	10.4 (7.0, 17.9)	24.1 (16.0, 32.9)	0.003
IL4	0.3 (0.3, 0.3)	0.3 (0.3, 0.3)	0.576	0.3 (0.3, 0.3)	0.3 (0.3, 0.3)	1
IL2	0.5 (0.5, 0.5)	0.5 (0.5, 0.5)	0.785	0.6 (0.6, 0.6)	1.2 (0.6, 1.9)	0.005
IL12p70	0.5 (0.5, 0.5)	0.5 (0.5, 0.5)	0.83	0.5 (0.5, 0.5)	0.5 (0.5, 0.5)	1
IL17A	1.1 (1.1, 1.1)	1.1 (1.1, 1.1)	0.246	1.1 (1.1, 1.1)	4.1 (2.7, 9.5)	0

were >0.7, which means that the model possessed good diagnosability. Besides, we found that the different groups were separated based on the hub variables (Figures 2C,D) in the PCA, indicating that these four variables might represent essential differences between different groups (Figure 2E). We drew a nomogram to apply our model in clinical work better. For evaluation, we performed calibration curve plots that fit well with the diagonal reference line (Figure 2F), indicating our model's great performance. DCA is a powerful method for assessing the degree of patient benefits. This research applied the DCA for the model in training and validation sets (Figures 2G,H). The DCA curves revealed that patients with COVID-19 could obtain net benefits through the logistic regression model. To establish a more comprehensive prognosis evaluation system, we applied Cox regression in Supplementary material 1.

Establishing an ML model to classify patients and evaluate the model

To improve the diagnostic performance of the model, we applied ML algorithms, including RF and SVM, to construct a new model. To reduce the subsequent unnecessary workload, first, we evaluated the residuals of the SVM and RF (Figures 3A,B). The results indicated that RF performed better with fewer residuals. Therefore, we decided to choose RF as the main ML algorithm to construct a model.

In the next step, all the variables (including cytokines, sex, and age) were entered into the RF classifier. We set the optimal parameter mtry to 2 as a default setting. The optimal number of trees in the classifier was set as 500, maintaining a low error for the classifier (Figure 4A). We used the perspective of reducing mean square error to measure the variable importance

TABLE 3 Logistic regression analysis for mortality of patients in the training set.

Characteristics	Univariate logistic regression		Multivariate logistic regression	
	OR (95%CI)	*P* value	OR (95%CI)	*P* value
Sex	0.739 (0.259, 2.108)	0.571	-	-
Age	1.107 (1.046, 1.173)	0.000	1.140 (1.051, 1.236)	**0.002**
TGFb1	1.000(1.000, 1.000)	0.153	-	-
Il5	1.194 (0.908, 1.569)	0.204	-	-
IFNg	0.959 (0.892, 1.032)	0.261	-	-
IFNa	1.016 (1.002, 1.030)	0.029	1.023 (1.033, 1.044)	**0.026**
CCL2	1.000 (1.000, 1.000)	0.854	-	-
Il6	1.001 (1.000, 1.002)	0.057	1.002 (0.999, 1.004)	0.179
TNFa	1.037 (0.985, 1.092)	0.167	-	-
IL1b	1.484 (0.741, 2.971)	0.265	-	-
IL8	1.004 (1.001, 1.007)	0.022	1.004 (1.001, 1.008)	**0.015**
IL18	1.001 (0.999, 1.002)	0.458	-	-
IL10	1.008 (0.992, 1.024)	0.337	-	-
IL4	1.941 (0.694, 5.429)	0.206	-	-
IL2	4.727 (0.587, 38.075)	0.145	-	-
IL12p70	0.954 (0.57, 1.595)	0.857	-	-
IL17A	1.689 (1.072, 2.662)	0.024	1.989 (1.140, 3.467)	**0.015**

The bold values mean that the P-value < 0.05.

of the results (Gini coefficient). To keep our model succinct, we identified hub variables using a cutoff of importance >2 (Figure 4B). After obtaining the hub variables, we constructed an ANN model through the R package "neuralnet". Two parallel training processes were used to construct a scoring model based on the training set. The ANN topology of the training set contained eight input layers, five hidden layers, and two output layers (Figure 4C).

The PCA of hub cytokines from RF and age shows that patients with COVID-19 with different outcomes are separate in both training and validation sets, revealing that the model possesses good discrimination (Figures 4D,E). In the training set, the AUC of the model was 0.99 (Figure 4F). Additionally, in the validation set, the AUC came to 0.783 (Figure 4G). The results indicated that the model we built possessed advantages in some situations.

Constructing a DT to classify patients and evaluate the model

To simplify the model and improve the feasibility of models for clinical application, we performed DT for all the variables previously mentioned. DT showed the interrelationship among the selected variables screened by the DT algorithm. The DT became the most straightforward tree when the complexity parameter (CP) was 0.1944444 (Figure 5). The first filtration age was (≥68 years), and the mortality was 48.27%. The second combination was IFN-α≥47 pg/ml. The mortality for patients with COVID-19 within the double combination was 100%. In the training set and validation set, the accuracy of the DT model was 0.875 and 0.86, respectively. The results revealed that the DT model considered both effectiveness and concise usability.

Models' performance evaluation

We listed the results of the three models' performance in overall patients with COVID-19 in our cohort (the data from the training set and the validation set are merged, 138 patients). According to Figure 6, the models developed by logistic regression happened to be the highest in accuracy with 86.96% when compared with other models developed by the RF model and the DT model have 80.43 and 85.51%. While for sensitivity that shows the mortality rate of patients with COVID-19 correctly by the models, the logistic regression model seems to be the beat one with 96.3%, followed by the RF model with 70.37% and the DT model with 25.93%. Additionally, for specificity showing the survival rate of patients with COVID-19 correctly by the models, the DT model emerged to be the best one with 100%, followed by the RF model with 81.98% and the logistic regression model with 63.06%.

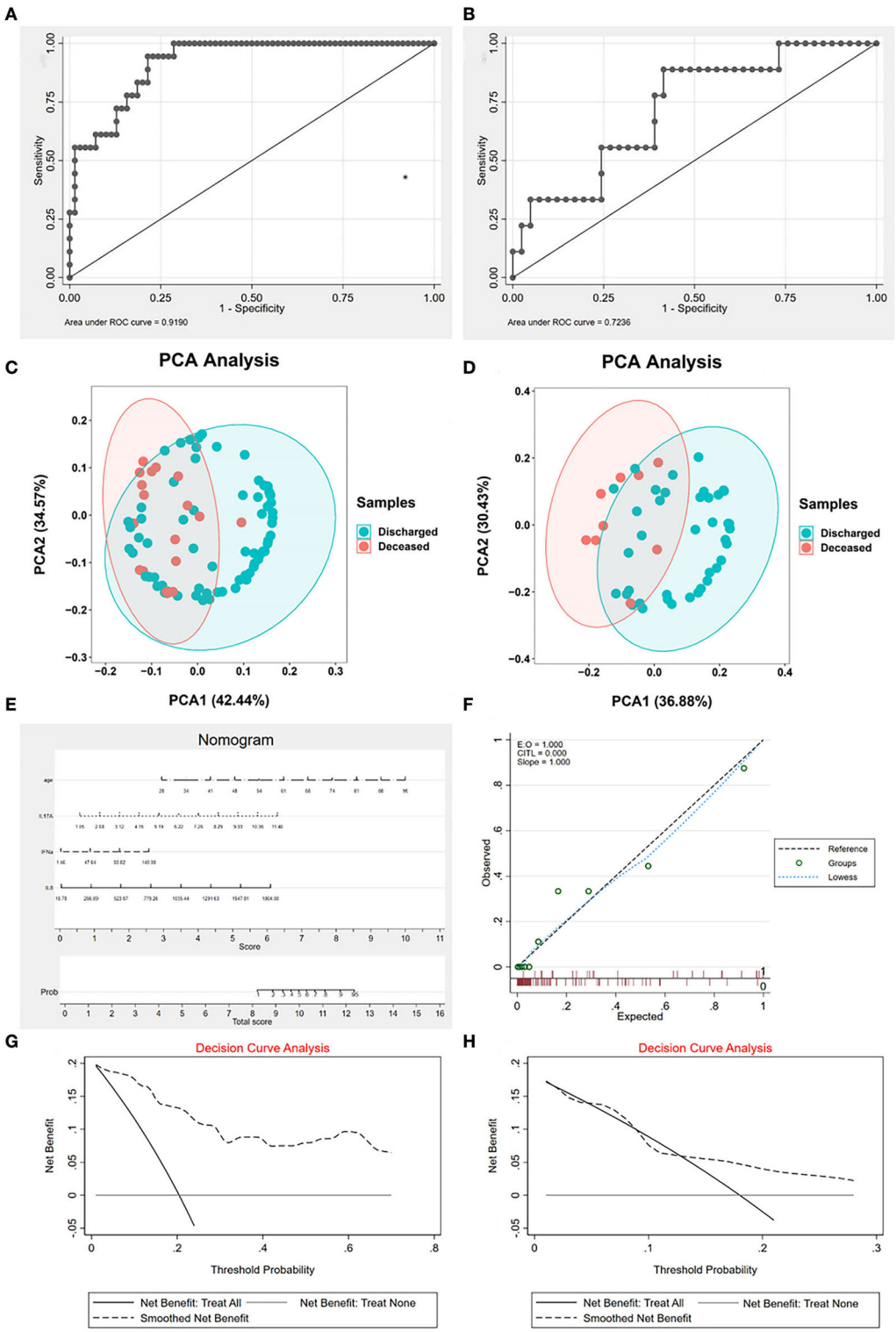


FIGURE 2
The characteristics of logistic regression (LR) signature. **(A)** The receiver operating characteristic curve (ROC) of signature in the training set. The area under the curve (AUC) is 0.919, indicating that the signature works well in the training set. **(B)** The ROC of signature in the validation set. AUC is 0.7236, indicating that the signature is valuable in the validation set. **(C)** The principal component analysis (PCA) of hub cytokines and age shows that patients with COVID-19 with different outcomes are separate in the training set. **(D)** The PCA of hub cytokines and age shows that patients with COVID-19 with different outcomes are separate in the validation set. **(E)** The nomogram of the signature. **(F)** Calibration curve plots of the signature. The lowess fits the reference line well, showing that the signature is effective. **(G)** The decision curve analysis (DCA) of the signature in the training set. **(H)** The DCA of the signature in the validation set. Both decision curve analyses revealed that patients with COVID-19 can benefit from applying the LR signature.

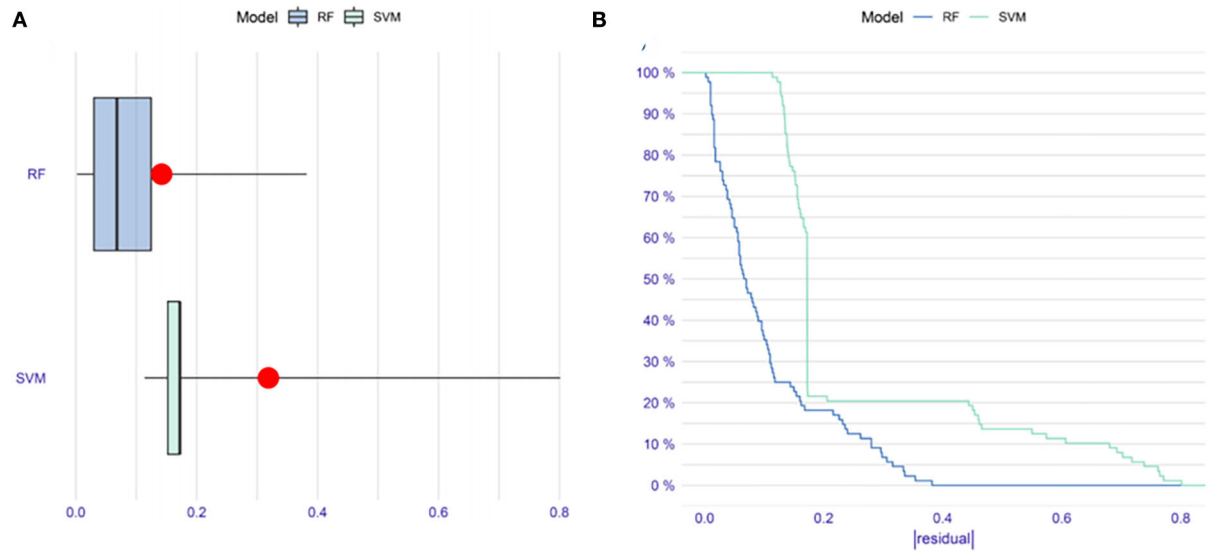


FIGURE 3
The comparison of random forest (RF) and support vector machine (SVM) methods in constructing models. **(A)** Boxplots of |residual|. The red dot stands for the root mean square of residuals. **(B)** Reverse cumulative distribution of |residual|. These results indicate that RF might be more suitable than SVM to construct a model in our research.

Discussion

This research found that old age and several inflammatory cytokines played a crucial role in promoting severe COVID-19. We constructed multiple predictive prognostic models based on these factors to identify patients with COVID-19 at high risk of death at hospital entry. The models were validated using external datasets, and all models' performance is satisfactory. What we did may provide a novel insight into evaluating COVID-19 patients' conditions.

SARS-CoV-2, caused by COVID-19, spread worldwide at an unpredicted speed and brought profound and unfolding impacts on every aspect of human life. Previous studies revealed that patients with COVID-19 expressed huge heterogeneous characteristics, ranging from asymptomatic to losing lives (14, 15). COVID-19 infection affects various systems in the human body, including the immune system, leading to changes in the patient's internal environment and inflammation. Gao et al. (16) reported that pro-inflammatory cytokines were highly associated with severe disease. Previous studies had found that several cytokines were closely related to the development of COVID-19 but simply considered individual cytokines as predictive indicators (17, 18). COVID-19 causes immune dysregulation accompanied by multiple cytokine disturbances, which is hard to evaluate scientifically with one variable. Therefore, an effective inflammatory cytokines signature to comprehensively evaluate the COVID-19 patient's immune status must contain multiple variables. Inflammatory cytokines are texted in serum, which is easy to obtain from inpatients. The 12 inflammatory cytokines, the most essential and common series of cytokines, are sufficient to evaluate the patient's immune status.

In the research, we filtered out some prominent inflammatory cytokines in predicting the prognosis of COVID-19. IFN-α also has immunoregulatory effects, which might activate inflammatory responses and cause uncontrolled pathogenic damage (19). Krämer et al. (20) indicated preferential IFN-α responses in severe COVID-19 and declared that IFN-α was associated with a poorer COVID-19 infection outcome. Our research found that early high IFN-α signatures were hazardous features of poor prognoses for patients with COVID-19. Patients with COVID-19 with old age and elevated IFN-α levels suffered a very extreme risk of death. Systemic and autocrine IL-8 loops were essential neutrophil activation factors for immunopathology, triggering multiple cell dysfunctions (21). A previous study indicated that in patients with severe COVID-19, IL-8 might be a prognostic indicator for in-hospital death and a target for an effective treatment strategy (22). IL-8 was included in both the logistic regression and RF models in our research, which declared a crucial clinical value for this inflammatory cytokine. IL-6 is one of the most prominent inflammatory cytokines. According to Mojtabavi et al. (23), the elevated IL-6 level was an independent risk factor for adverse COVID-19 outcomes. There had already been several treatment strategies based on IL-6. Some results of them were reported as encouraging (24, 25). In our study, IL-6 significantly differed between death and survival cases in training and verification sets. This result

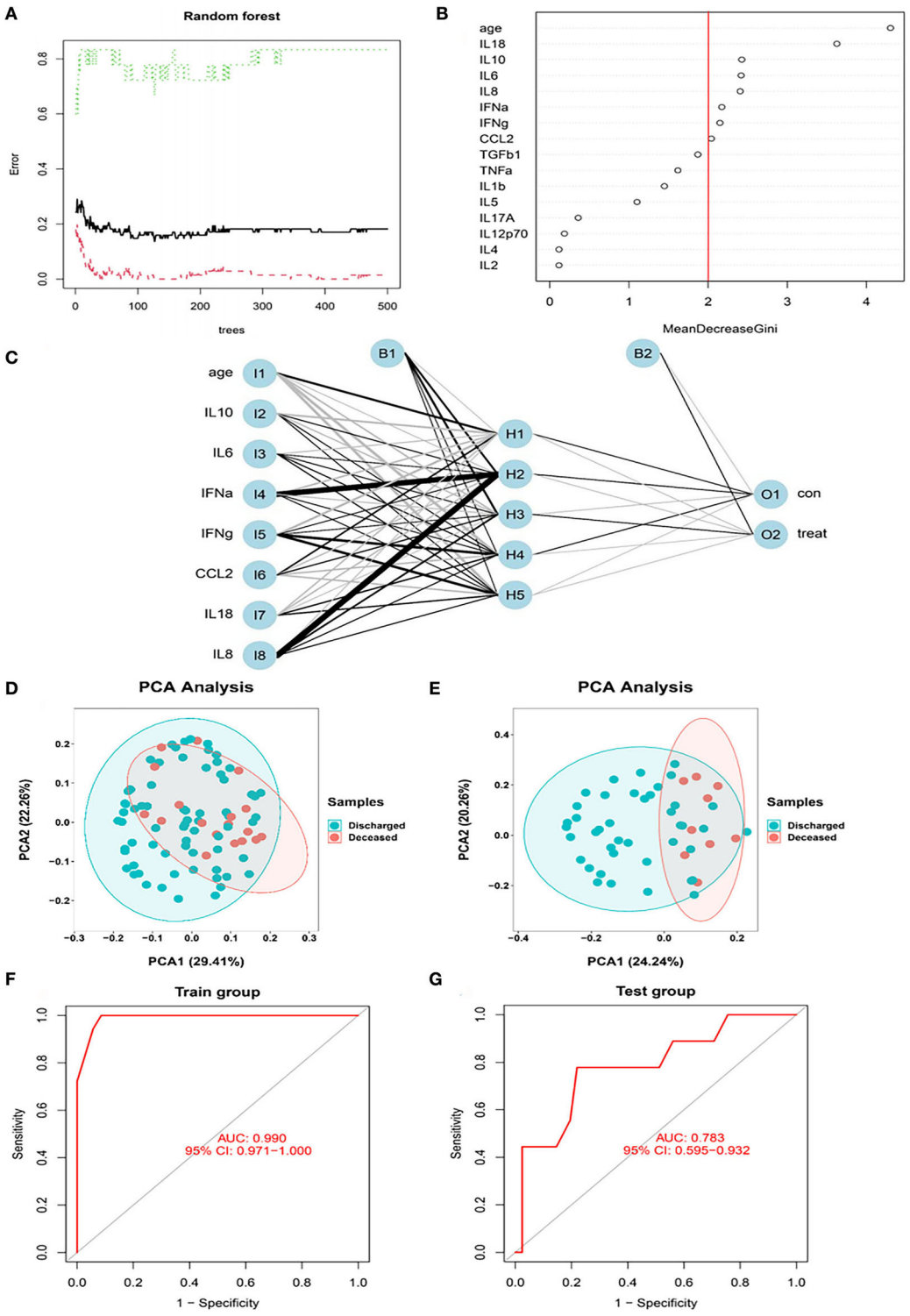


FIGURE 4
The RF modeling processes. **(A)** The influence of the number of decision trees on the error rate. After multiple repetitive operations, the error becomes stable gradually. **(B)** Results of the Gini coefficient method in the random forest classifier. We set importance=2 as a cutoff. **(C)** Neural network topology of the microarray with 8 input layers, 5 hidden layers, and 2 output layers. **(D)** The PCA of hub cytokines and age shows that patients with COVID-19 with different outcomes are separate in the training set. **(E)** The PCA of hub cytokines and age shows that patients with COVID-19 with different outcomes are separate in the validation set. **(F)** The ROC of signature in the training set. AUC is 0.99, indicating that the signature is perfect in the training set. **(G)** The ROC of signature in the validation set. AUC is 0.783, indicating that the signature is good in the validation set.

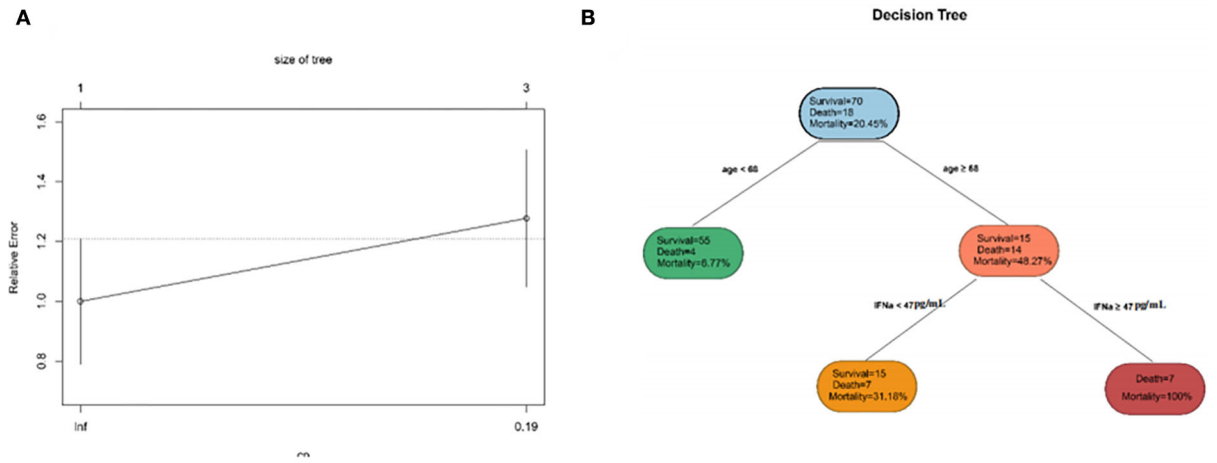


FIGURE 5
The characteristics of decision tree (DT). **(A)** The relative errors of the DT model and tree size when the complexity parameter comes to an ideal value. **(B)** The inter-relationship among selected clinical indicators.

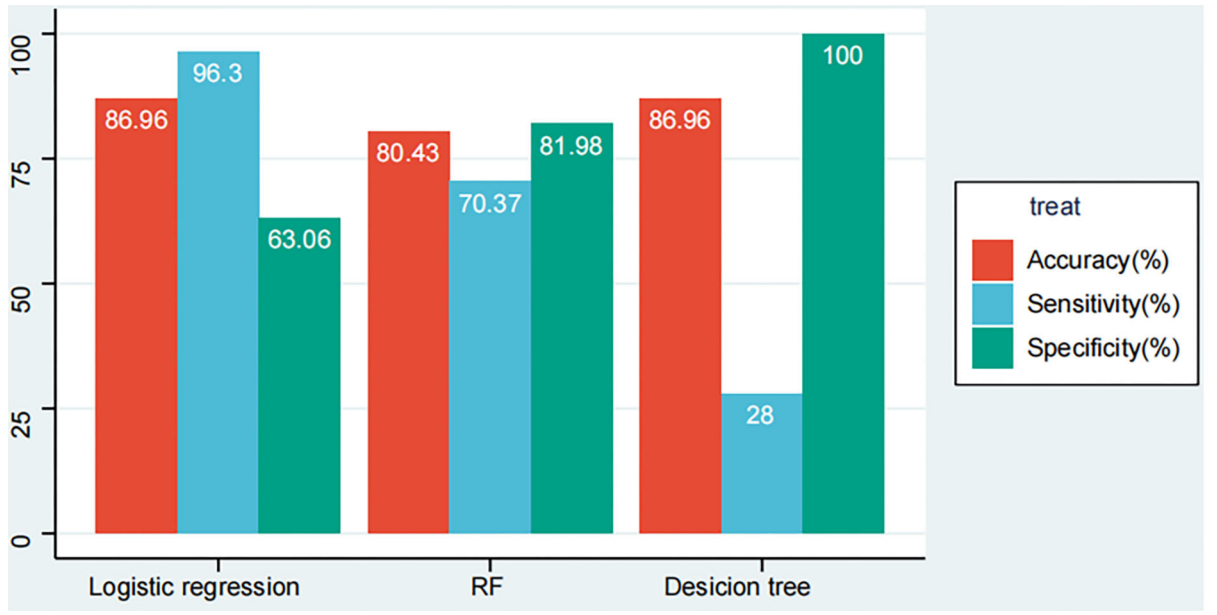


FIGURE 6
LR, RF, and DT models' performance evaluation results in the whole dataset.

revealed that IL-6 was a stable prognostic factor that could be applied on a large scale. Maione et al. (26) indicated IL-17A as a silent amplifier of cytokine storm in patients with COVID-19, activating several inflammatory pathways. In the clinical work, researchers identified IL-17A as a target to develop therapeutic strategies and made some progress (27). In this study, we screened out a possible link between elevated IL-17A levels and COVID-19 mortality.

Due to our work, we found that age was a key factor for all models and scored the highest in the RF model. There exist significant differences in age between survival and death cases. O'Driscoll et al. (28) declared that patients with COVID-19 aged older than 65 years suffered from higher mortality. Chen et al. (29) stated that age was related to declining and dysregulation of immune function, which heightened vulnerability to COVID-19 in elders. Previous researchers set

the age>65 years as a high risk for severe COVID-19 outcomes (28, 30). Our DT sets age>68 years as a cutoff of the first combination, similar to those in previous reports. These results indicated that age played a major role in contributing to COVID-19 mortality.

Considering all models involved in this study comprehensively, the models developed with RF and logistic regression happened to be two well-rounded models, with considerable AUC in the training set and AUC = 0.783 in the validation set and high accuracy. However, the complexity of these models might bring some inconvenience in clinical applications. The model constructed by DT had good accuracy, specificity, and convenience. As for areas with underdeveloped medical conditions, it is an excellent choice to apply the DT model we constructed which was a simple model with good accuracy.

Building COVID-19 prognostic models through blood inflammatory cytokines levels is a novel thought. Thus, our research suggests that more extensive cohort studies should be conducted to reveal the role of inflammatory cytokines in predicting long-term post-COVID-19 complications.

Conclusion

In this research, we identified that advanced age, IFN-α, IL-8, and IL-6 have been identified as potential prognostic predictors of COVID-19 outcomes by multiple models in our research, which indicated that these cytokines might play a vital role in the progression of SARS-CoV-2. Therefore, we advised that accurate and quantitative detection of the inflammatory cytokines could be performed when necessary. The RF, logistic regression, and DT models based on blood inflammatory cytokines performed well in identifying patients with COVID-19 at risk of death. We strongly suggest that the models developed with RF and logistic regression should be applied in the regions with more abundant medical resources, and the model developed by DT could be used in the regions with less abundant medical. The models we developed can assist doctors in applying individual strategies to different risk cohorts to perform early intervention and treatment to benefit patients with COVID-19.

Data availability statement

The data and codes we used in this passage could be obtained from the corresponding author with reasonable requests. Requests to access these datasets should be directed to SS, sunshiren@medmail.com.cn.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

ZY and XL conducted the analysis and prepared an initial draft. JZ collected data. SS offered ideas for the project. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the Special Scientific Research Project of Shanghai Medical Assistance Team from Air Force Military Medical University in 2022, 2022YH-1009.

Acknowledgments

We express our most sincere thanks to Professor Peter D. Katsikis and his team for their generous willingness to share their data. Meanwhile, we show great respect to all who fought against COVID-19.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.1001340/full#supplementary-material

References

1. Alsharif W, Qurashi A. Effectiveness of COVID-19 diagnosis and management tools: a review. *Radiography (London, England: 1995).* (2021) 27:682–7. doi: 10.1016/j.radi.2020.09.010

2. Pascarella G, Strumia A, Piliego C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: a comprehensive review. *J Intern Med.* (2020) 288:192–206. doi: 10.1111/joim.13091

3. Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. *Int J Infect Dis.* (2020) 94:44–8. doi: 10.1016/j.ijid.2020.03.004

4. Guan WJ Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* (2020) 382:1708–20. doi: 10.1056/NEJMoa2002032

5. CDC COVID-19 Response Team, Food and Drug Administration. Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Moderna COVID-19 Vaccine - United States, December 21, 2020-January 10, 2021. *MMWR Morb Mortal Wkly Rep.* (2021) 70:125–9. doi: 10.15585/mmwr.mm7004e1

6. Wack S, Patton T, Ferris LK. COVID-19 vaccine safety and efficacy in patients with immune-mediated inflammatory disease: review of available evidence. *J Am Acad Dermatol.* (2021) 85:1274–84. doi: 10.1016/j.jaad.2021.07.054

7. Liu C, Chu D, Kalantar-Zadeh K, George J, Young HA, Liu G. Cytokines: from clinical significance to quantification. *Adv Sci (Weinheim, Baden-Wurttemberg, Germany).* (2021) 8:e2004433. doi: 10.1002/advs.202004433

8. Kunnumakkara AB, Rana V, Parama D, Banik K, Girisa S, Henamayee S, et al. COVID-19, cytokines, inflammation, and spices: How are they related? *Life Sci.* (2021) 284:119201. doi: 10.1016/j.lfs.2021.119201

9. Kaur S, Bansal R, Kollimuttathuillam S, Gowda AM, Singh B, Mehta D, et al. The looming storm: blood and cytokines in COVID-19. *Blood Rev.* (2021) 46:100743. doi: 10.1016/j.blre.2020.100743

10. Darif D, Hammi I, Kihel A, El Idrissi Saik I, Guessous F, Akarid K. The pro-inflammatory cytokines in COVID-19 pathogenesis: what goes wrong? *Microb Pathog.* (2021) 153:104799. doi: 10.1016/j.micpath.2021.104799

11. Abers MS, Delmonte OM, Ricotta EE, Fintzi J, Fink DL, de Jesus AAA, et al. An immune-based biomarker signature is associated with mortality in COVID-19 patients. *JCI Insight.* (2021) 6. doi: 10.1172/jci.insight.144455

12. Patterson BK, Guevara-Coto J, Yogendra R, Francisco EB, Long E, Pise A, et al. Immune-based prediction of COVID-19 severity and chronicity decoded using machine learning. *Front Immunol.* (2021) 12:700782. doi: 10.3389/fimmu.2021.700782

13. Mueller YM, Schrama TJ, Ruijten R, Schreurs MWJ, Grashof DGB, van de Werken HJG, et al. Stratification of hospitalized COVID-19 patients into clinical severity progression groups by immuno-phenotyping and machine learning. *Nat Commun.* (2022) 13:915. doi: 10.1038/s41467-022-28621-0

14. Badal S, Thapa Bajgain K, Badal S, Thapa R, Bajgain BB, Santana MJ. Prevalence, clinical characteristics, and outcomes of pediatric COVID-19: A systematic review and meta-analysis. *J Clin Virol.* (2021) 135:104715. doi: 10.1016/j.jcv.2020.104715

15. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease (2019). (COVID-19) in China: a systematic review and meta-analysis. *J Infect.* (2020) 80:656–65. doi: 10.1016/j.jinf.2020.03.041

16. Gao YD, Ding M, Dong X, Zhang JJ, Kursat Azkur A, Azkur D, et al. Risk factors for severe and critically ill COVID-19 patients: A review. *Allergy.* (2021) 76:428–55. doi: 10.1111/all.14657

17. Liu F, Li L, Xu M, Wu J, Luo D, Zhu Y, et al. Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. *J Clin Virol.* (2020) 127:104370. doi: 10.1016/j.jcv.2020.104370

18. McElvaney OJ, McEvoy NL, McElvaney OF, Carroll TP, Murphy MP, Dunlea DM, et al. Characterization of the inflammatory response to severe COVID-19 illness. *Am J Respir Crit Care Med.* (2020) 202:812–21. doi: 10.1164/rccm.202005-1583OC

19. Yang L, Wang J, Hui P, Yarovinsky TO, Badeti S, Pham K, et al. Potential role of IFN-α in COVID-19 patients and its underlying treatment options. *Appl Microbiol Biotechnol.* (2021) 105:4005–15. doi: 10.1007/s00253-021-11319-6

20. Krämer B, Knoll R, Bonaguro L, ToVinh M, Raabe J, Astaburuaga-García R, et al. Early IFN-α signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. *Immunity.* (2021) 54:2650–69.e14. doi: 10.1016/j.immuni.2021.09.002

21. Kaiser R, Leunig A, Pekayvaz K, Popp O, Joppich M, Polewka V, et al. Self-sustaining IL-8 loops drive a prothrombotic neutrophil phenotype in severe COVID-19. *JCI Insight.* (2021) 6:e150862. doi: 10.1172/jci.insight.150862

22. Li H, Zhang J, Fang C, Zhao X, Qian B, Sun Y, et al. The prognostic value of IL-8 for the death of severe or critical patients with COVID-19. *Medicine.* (2021) 100:e23656. doi: 10.1097/MD.0000000000023656

23. Mojtabavi H, Saghazadeh A, Rezaei N. Interleukin-6 and severe COVID-19: a systematic review and meta-analysis. *Eur Cytokine Netw.* (2020) 31:44–9. doi: 10.1684/ecn.2020.0448

24. Gordon AC, Mouncey PR, Al-Beidh F, Rowan KM, Nichol AD, Arabi YM, et al. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N Engl J Med.* (2021) 384:1491–502. doi: 10.1056/NEJMoa2100433

25. RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet (London, England).* (2021) 397:1637–45. doi: 10.1016/S0140-6736(21)00676-0

26. Maione F, Casillo GM, Raucci F, Salvatore C, Ambrosini G, Costa L, et al. Interleukin-17A (IL-17A): a silent amplifier of COVID-19. *Biomed Pharmacother.* (2021) 142:111980. doi: 10.1016/j.biopha.2021.111980

27. Ayhan E, Öztürk M, An I, Abdelmaksoud A, Araç E. Potential role of anti-interleukin-17 in COVID-19 treatment. *Dermatol Ther.* (2020) 33:e13715. doi: 10.1111/dth.13715

28. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature.* (2021) 590:140–5. doi: 10.1038/s41586-020-2918-0

29. Chen Y, Klein SL, Garibaldi BT Li H, Wu C, Osevala NM, et al. Aging in COVID-19: Vulnerability, immunity and intervention. *Ageing Res Rev.* (2021) 65:101205. doi: 10.1016/j.arr.2020.101205

30. Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nat Med.* (2021) 27:28–33. doi: 10.1038/s41591-020-01202-8

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 07 December 2022
DOI 10.3389/fpubh.2022.970363

OPEN ACCESS

EDITED BY
Omar El Deeb,
Lebanese American University, Lebanon

REVIEWED BY
Saroj Kumar Chandra,
OP Jindal University Raigarh, India
Ayman Massaoudi,
Carthage University, Tunisia
Monika Yuliarti,
Sebelas Maret University, Indonesia

*CORRESPONDENCE
Jeffrey E. Harris
jeffrey@mit.edu

SPECIALTY SECTION
This article was submitted to Public Health Policy, a section of the journal Frontiers in Public Health

RECEIVED 15 June 2022
ACCEPTED 11 November 2022
PUBLISHED 07 December 2022

CITATION
Harris JE (2022) Concentric regulatory zones failed to halt surging COVID-19: Brooklyn 2020.
Front. Public Health 10:970363.
doi: 10.3389/fpubh.2022.970363

Concentric regulatory zones failed to halt surging COVID-19: Brooklyn 2020

Jeffrey E. Harris[1,2]*

[1]Massachusetts Institute of Technology, Cambridge, MA, United States, [2]Eisner Health, Los Angeles, CA, United States

Methods: We relied on reports of confirmed case incidence and test positivity, along with data on the movements of devices with location-tracking software, to evaluate a novel scheme of three concentric regulatory zones introduced by then New York Governor Cuomo to address an outbreak of COVID-19 in South Brooklyn in the fall of 2020. The regulatory scheme imposed differential controls on access to eating places, schools, houses of worship, large gatherings and other businesses within the three zones, but without restrictions on mobility.

Results: Within the central red zone, COVID-19 incidence temporarily declined from 131.2 per 100,000 population during the week ending October 3 to 62.5 per 100,000 by the week ending October 31, but then rebounded to 153.6 per 100,000 by the week ending November 28. Within the intermediate orange and peripheral yellow zones combined, incidence steadily rose from 28.8 per 100,000 during the week ending October 3 to 109.9 per 100,000 by the week ending November 28. Data on device visits to pairs of eating establishments straddling the red-orange boundary confirmed compliance with access controls. More general analysis of device movements showed stable patterns of movement between and beyond zones unaffected by the Governor's orders. A geospatial regression model of COVID-19 incidence in relation to device movements across zip code tabulation areas identified a cluster of five high-movement ZCTAs with estimated reproduction number 1.91 (95% confidence interval, 1.27–2.55).

Discussion: In the highly populous area of South Brooklyn, controls on access alone, without restrictions on movement, were inadequate to halt an advancing COVID-19 outbreak.

KEYWORDS

SARS-CoV-2, mobility, mobile device tracking, geospatial regression, paired point-of-interest analysis

Introduction

The idea of drawing a series of concentric containment circles around an outbreak is well-established in the control of communicable diseases. The U.S. Department of Agriculture, for example, has adopted the model of three concentric containment zones – the infected zone, the buffer zone, and the surveillance zone – as its standard practice to contain highly contagious animal diseases (1). During the coronavirus disease 2019 (COVID-19) pandemic, the National

Task Force in the Philippines established a four-circle scheme to enforce graded degrees of quarantine: a critical zone subject to complete lockdown, where a cluster of cases had been identified; a surrounding 500-meter-radius containment zone where a modified lockdown prevailed; a surrounding buffer zone subject to community-level quarantine; and surrounding outside area with further relaxation of mobility controls (2, 3).

In their classic form, concentric regulatory zones have served as quarantine boundaries (4). Their use has been especially appealing when attack rates are directly related to the duration of contact and inversely related to the distance from an identifiable source, as they were in the Toronto-area severe acute respiratory syndrome (SARS) outbreak of 2003 (5). It is well-understood, however, that zone boundaries cannot simply be drawn around the areas of highest infection density, but need to take movement patterns into account (6).

On October 6, 2020, then New York Governor Cuomo issued a series of executive orders establishing a novel variation on the classic concentric control scheme (7). Rather than serving as mass quarantine boundaries, the concentric areas would define the extent of access control to restaurants, schools, gyms, houses of worship, and large gatherings generally. While several areas of concern were identified throughout the state of New York, far and away the principal challenge was the surge of new COVID-19 cases in the South Brooklyn area of New York City.

Our task here is to combine data on COVID-19 incidence and testing outcomes with data on the movements of devices equipped with location-tracking software to evaluate what happened over the ensuing months. Relying on geospatial regression analysis and spatial visualization tools (8), we find that the Governor's novel regulatory scheme failed to halt the surge of COVID-19. Our findings appear to go against the well-documented relationship between a reduction in mobility and a subsequent decline in COVID-19 incidence (9–14). To resolve the apparent contradiction, we distinguish between two types of mobility controls: regulations concerning access (15–17) and restrictions on movement (18–20).

Materials and methods

Regulatory background

By mid-September 2020, it had becoming increasingly evident that the recent surge of COVID-19 cases in certain hotspots of New York City was threatening the city's reopening plans. By September 29, then New York City Mayor de Blasio had signaled his intention to close non-essential businesses and all public and private schools in nine zip codes in the boroughs of Queens and Brooklyn for 14–28 days (21). The target zip codes included five in Brooklyn: Borough Park (11219), Gravesend (11223), Midwood (11230), Bensonhurst (11204), Flatlands (11210), and Gerritsen Beach/Homecrest/Sheepshead Bay (11229). Test positivity rates had increased beyond the acceptable threshold of 3% in these areas, exceeding 7% in Gravesend (11223) (22).

On October 6, however, then New York State Governor Cuomo intervened with his own regulatory control strategy, which he termed a "cluster action initiative" (7). Developed in consultation with public health experts, the initiative imposed new local restrictions on activity within "red zones" where clusters of new cases had been identified (23, 24). Recognizing that individuals within these high-risk areas tended to "interface with the surrounding communities," the initiative established two concentric rings – an intermediate orange zone and a peripheral yellow zone – surrounding the high-risk red zone.

Zone boundaries were drawn based on the test positivity rate, that is, confirmed COVID-19 cases as a percentage of all persons presenting for testing. Among highly populated areas, which included the borough of Brooklyn, a red zone was defined as having a sustained test positivity rate above 4 percent, while an orange zone had a rate from 3 to 4 percent, and a yellow zone had a rate from 2.5 to 3 percent (23).

Among the restrictions on activity, a red zone prohibited mass gatherings, allowed only essential businesses to open, closed in-person schooling, and restricted restaurants and other food providers to takeout/delivery only. An orange zone allowed gatherings up to 10 people, closed only high-risk businesses such as gyms and personal care, closed in-person schooling, and allowed outdoor dining with up to 4 persons per table. A yellow allowed gatherings up to 25 people, permitted all businesses to open, permitted indoor as well as outdoor dining up to four persons per table, and opened schools to in-person instruction subject to mandatory testing of students, teachers and staff (23). Restrictions on access to houses of worship were also initially imposed, with limits of 25% capacity in a red zone, 33% capacity in an orange zone, and 50% capacity in a yellow zone, but were subsequently blocked by the United States Supreme Court (25). The sanctions for failure to comply included withholding of funds to localities and schools (26).

Revisions of zone boundaries and changes in zone classification were based principally on the test positivity rate. On October 21, the Governor, citing the early success of the strategy in Brooklyn, reclassified the borough's original orange zone as a yellow zone, while the red zone remained unchanged (24). On November 3, citing further gains, the Governor reduced the size of the red zone by half (27). A few days later, on November 9, the red zone was reclassified as an orange zone (27), and subsequently as a yellow zone on November 18 (28). The zones were eventually dissolved without fanfare in January 2021.

Data sources: Regulatory zone boundaries

We determined regulatory zone boundaries from detailed maps issued by the office of the Governor of New

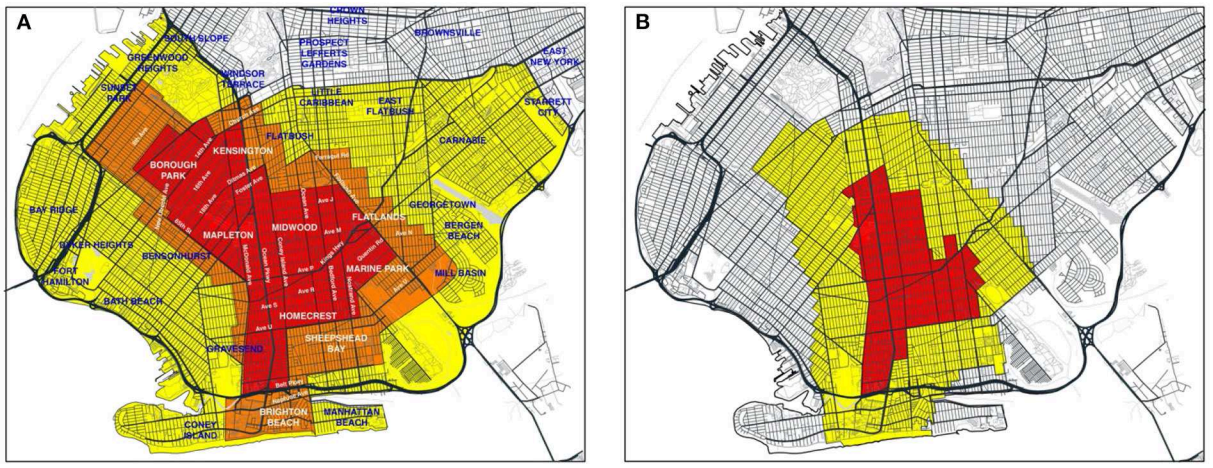


FIGURE 1
(A) Red, orange, and yellow zones effective October 9, 2020 (29), Overlaid on Annotated Map of South Brooklyn. On October 21, the orange zone was re-designated as a yellow zone, while the red zone remained unchanged (30). **(B)** Revised boundaries of red and yellow zones as of November 3 (32). On November 9, the red zone was re-designated as an orange zone (27), and subsequently as a yellow zone on November 18 (28).

York (29–31), along with accompanying announcements of updates (27, 28). Supplementary Figure A1 shows the boundaries of the original concentric red, orange, and yellow zones, effective October 9, 2020, overlaid on a street map of the larger New York City area (29). Figures 1A,B below depict the evolution of the regulatory zone boundaries, overlaid on more detailed street maps of South Brooklyn.

Figure 1A identifies the original red, orange and yellow zones (29). On October 21, the orange zone was incorporated into the existing yellow zone, while the original red zone boundaries remained unchanged (30). Figure 1B shows the contracted red and yellow zones as of November 3 (31). On November 9, the red zone was re-designated as an orange zone (27), and subsequently as a yellow zone on November 18 (28).

Data sources: Zip code tabulation area boundaries

Figures 2A,B superimpose the respective regulatory boundaries of Figures 1A,B on a map of zip code tabulation areas (ZCTAs) in South Brooklyn. The lack of complete congruence between the ZCTA and regulatory boundaries is evident. As discussed below, geographically detailed data on confirmed COVID-19 incidence over time was available only at the ZCTA level. Accordingly, for the purposes of analyzing COVID-19 incidence, we classified any of the nine ZCTAs that even partially overlapped the original red zone as an original red-zone ZCTA. These ZCTAs, indicated in boldface in Figure 2A, included 11204, 11210, 11218, 11219, 11223, 11229, 11230, 11234, and 11235. The remaining 12 ZCTAs, indicated in italics, were classified as original orange-yellow zone ZCTAs. In Figure 2B, the same classification of ZCTAs is shown in the map of the contracted regulatory zones effective November 3.

Data sources: Census block groups

As discussed below, we relied on the Safegraph Social Distancing database (34) to gauge the movements of devices equipped with location-tracking software throughout the greater New York City area. The origin and destination of each device movement in the Social Distancing database are keyed to census block groups (CBGs). Accordingly, we developed a separate correspondence between CBGs and regulatory zones, as shown in Figure 3A. Relying on Quantum Open Source Geographic Information System (QGIS) software (35), we determined the geocoordinates the centroids of all CBGs in South Brooklyn based upon their U.S. Census-defined shape files (36). Using the Stata Statistical Software (Stata) routine *geoinpoly* (37), we then assigned each CBG to the regulatory zone polygon in which its centroid was situated. While not explicitly shown in the figure, we used the same procedure to map CBGs into the ZCTA polygons shown in Figures 2A,B.

Data sources: Points of interest

We relied upon the SafeGraph Patterns database (38) – a source distinct from the Social Distancing database (34) – to analyze visits of devices to points of interest (POIs) within

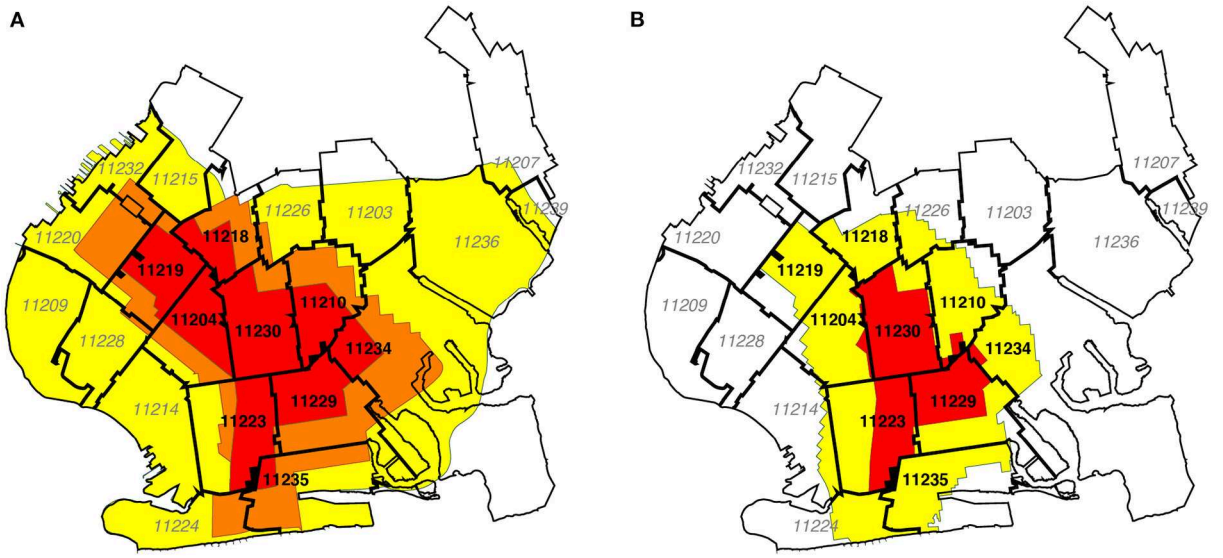


FIGURE 2
(A) Boundaries of the original October 9 zones overlaid on map of zip code tabulation areas (ZCTAs). In our calculations of confirmed COVID-19 incidence rates and test positivity rates, we assigned the nine ZCTAs in black boldface to the original red zone, while the remaining 12 ZCTAs in gray italic were assigned to the combined orange and yellow zones. **(B)** Updated boundaries of the redrawn November 3 zones overlaid on the same ZCTA Map. For base map of New York City ZCTAs, see Ref. (33).

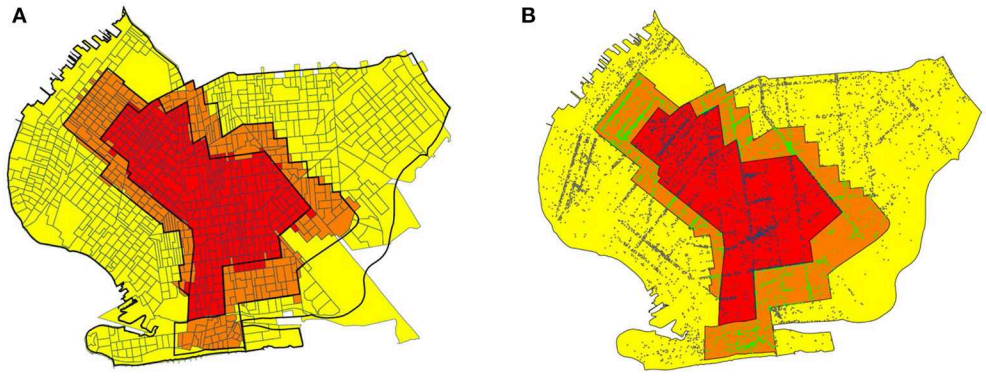


FIGURE 3
(A) Boundaries of the original October 9 zones overlaid on map of census block groups (CBGs). In our calculations of within- and between-zone device movements, we assigned each CBG to the zone containing its geographic centroid. **(B)** Boundaries of the original October 9 zones overlaid on map of Safegraph points of interest (POIs) (38). Each point corresponds to a POI. Points have different colors depending on the regulatory zone in which they were located. For base map of New York City census block groups, see Ref. (39).

the regulated area in South Brooklyn. At the broadest level, Safegraph classifies POIs according to the variable *top_category*, which includes such categories as "Automotive Repair and Maintenance," "Child Day Care Services," "Clothing Stores," "Elementary and Seconday Schools," "Gasoline Stations," and "Health and Personal Care Stores." One of the largest such categories is "Restaurants and Other Eating Places." Taking advantage of the Safegraph-supplied geocoordinates of each POI, and again relying on the Stata *geoinpoly* routine (37), we classified each POI as being located in one of the original three regulatory zones. Based upon this classification, we constructed Figure 3B, which plots the location of every POI as a color-coded point within the original October 9 regulated area.

Data sources: COVID-19 incidence and test positivity

We relied upon data published by the New York City Department of Health on COVID-19 incidence, measured as the number of confirmed cases per 100,000 population, and COVID-19 test positivity, measured as the percentage of positive

tests, broken down by ZCTA and week (40, 41). Both data sources covered the weeks ending August 8 through November 28, 2020.

Statistical methods: Paired point-of-interest analysis

We developed a paired point-of-interest (POI) analysis to test whether the regulations imposed by the Governor were in fact enforced and effective. To that end, we focused on device movements into restaurants and other eating places located in the original red zone, where establishments were restricted to takeout/delivery only, and in the original orange zone, where establishments could also offer outdoor dining with up to four persons per table (23). To avoid potentially biased comparisons between local dining patterns in distinct neighborhoods (such as Borough Park in the original red zone and Brighton Beach in the original orange zone, as shown in Figure 1A), we restricted our comparisons to pairs of nearby eating establishments that straddled the original red-orange border.

We identified all POIs in the SafeGraph Patterns database (38) with a *top_category* designated as "Restaurants and Other Eating Places." Within this restricted dataset, there were 395 POIs in the original red zone and 507 POIs in the original orange zone. (There were also 1,045 POIs in the original yellow zone, but they were not included in our paired POI analysis). Relying on the Stata program *geonear* (42), we isolated 219 pairs of red-zone and orange-zone POIs that were nearest neighbors of each other, where the maximum distance between POIs within each pair was 300 meters. The median distance between paired POIs was 130.9 meters, with 25th and 75th percentiles equal to 70.7 and 218.6 meters, respectively. Because a POI on one side of the red-orange boundary could be the nearest neighbor of multiple POIs on the other side, the resulting dataset contained 145 unique red-zone POIs and 114 unique orange-zone POIs. Supplementary Figure A2 maps two such pairs straddling the red-orange boundary running along Avenue U in Brooklyn.

Let I denote the set of all red-zone POIs, with typical element $i \in I$, and let J denote the set of all orange-zone POIs, with typical element $j \in J$. Then our database consists of a subset of 219 unique pairs (i, j) contained within the larger set $I \times J$. For each POI, we relied on the Safegraph Patterns variable *visits_by_day* to compute the number of visits during each week, starting with the week ending October 1 (designated $t = 0$) and continuing through the week ending December 3 ($t = 9$). We thus had 219 paired observations (y_{Rit}, y_{Ojt}), where y_{Rit} represents the number of visits during week t to red-zone POI i, and where y_{Ojt} represents the number of visits during week t to orange-zone POI j.

Given these data, we ran the following fixed-effects regression model:

$$\log\left(\frac{y_{Rit}}{y_{Ojt}}\right) = \mu + \beta_t + \theta_i + \phi_j + \epsilon_{ijt} \quad (1)$$

In equation (1), the parameter μ is an overall constant term, β_t, θ_i. and ϕ_j are fixed-effect parameters corresponding to each week t, red-zone POI i, and orange-zone POI j, and ϵ_{ijt} are independently distributed spherical error terms. Within this fixed-effects framework, only the contrasts $\beta_t - \beta_0$ $(t = 1,\ldots,9)$ can be identified. If the regulations imposed by the Governor were in fact enforced and effective, then we would expect the estimated parameters β_t to be negative.

Statistical methods: Geospatial analysis

Movements to specific points of interest such as restaurants, auto repair shops and daycare centers are part of a larger set of movements to destinations that include private residences and workplaces. We sought to determine how these more general movement patterns related to the evolution of COVID-19 incidence, particularly during the period from the second half of October to the end of November, when cases of the disease were increasing throughout the regulated area in South Brooklyn. To that end, we developed a geospatial model relating COVID-19 incidence to general device movements. The central feature of this model was that the incidence of the disease in a particular ZCTA during a particular week was related to the incidence in all ZCTAs during the prior week. Moreover, the influence of one ZCTA on another was determined by the volume of device traffic between the two. The details of our model and its implementation are given in Appendix B.

Our model relied on two types of data: COVID-19 incidence and device movements. Because our data on COVID-19 incidence were based upon ZCTAs, we classified device movements between ZCTAs as well. Relying on data published by the New York City Department of Health (40), we constructed a data series $\{y_{kt}\}$ of the incidence of confirmed COVID-19 cases per 100,000 population in ZCTA k during week t, where $k = 1,\ldots,21$ indexes the 21 ZCTAs within the regulated area in Figure 2, and where $t = 1,\ldots,7$ indexes the 7-week period running from the week ending October 17 through the week ending November 28. Relying upon the variables *origin_census_block_group* and *destination_cbgs* in the Safegraph Social Distancing database (34), we constructed a data series $\{n_{k\ell t}\}$ of counts of device movements from ZCTA k into ZCTA ℓ during week t. The counts n_{kkt}, which represented the number of device movements staying within ZCTA k during week t, included those devices homed the ZCTA that made no movements. While we also observed device movements beyond the 21-ZCTA regulated area, as well as movements into the

regulated area from outside, we focused sharply on the regulated area in order to ascertain how the traffic between local ZCTAs influenced the dynamics of COVID-19 transmission.

Based upon our underlying data on device movements, we let V_t be a 21×21 square matrix with typical element $v_{k\ell t}$ measuring the proportion of all devices originating in ZCTA k that moved into ZCTA ℓ during week t. The elements of each row of V_t sum to 1. Likewise referring to week t, we let W_t be a 21×21 square matrix with typical element $w_{k\ell t}$ measuring the fraction of all devices with a destination in ZCTA ℓ that originated in ZCTA k. The elements of each column of W_t sum to 1. Let Y_t denote a 21×1 column vector with elements y_{kt}. We let Y_{t+1} represent the corresponding vector of incidence rates 1 week later, while η_{t+1} is a contemporaneous vector of error terms. Under the strong assumption of *homogeneous mixing*, our geospatial model yields:

$$Y_{t+1} = \alpha\ V_t\ W_t'\ Y_t + \eta_{t+1} \tag{2}$$

The model of equation (2) is an adaptation of the conventional law of mass action implicit in SIR-type compartmental models of the dynamics of contagious disease transmission (43). It allows for a susceptible individual homed in ZCTA k to be infected not only through contact with another resident of the same ZCTA k, but also through contact with a resident of another ZCTA ℓ. The inclusion of both vectors V_t and W_t, moreover, allows for two loci of transmission from an infected individual residing in ZCTA ℓ to a susceptible individual homed in ZCTA k. Either the susceptible individual homed in k had temporarily moved to ℓ, or the infected individual in ℓ had temporarily moved to k. The unknown parameter α in equation (2), to be estimated from the data, represents the uniform reproductive number for COVID-19 transmission throughout the entire regulated area for the 7-week time period under study.

Inhomogeneous mixing

The assumption of homogeneous mixing, with a uniform reproductive number α, is strong. Accordingly, we considered two alternative specifications involving inhomogeneous mixing. First, we assumed instead that movements by individuals who remained within their home ZCTA could have a different influence on COVID incidence. To capture the effect of these within-ZCTA device movements, we defined $D_t = \text{diag}\left(V_t W_t'\right)$ as the $K \times K$ square matrix with the same diagonal elements as $V_t W_t'$ but zero off-diagonal elements, and then introduced the additional regressor $D_t Y_t$ into our model. Defining the $K \times K$ square matrix $X_t = V_t W_t' - D_t$, we have:

$$Y_{t+1} = \alpha_0\ D_t\ Y_t + \alpha_1 X_t Y_L + \eta_{t+1} \tag{3}$$

We refer to this alternative as *inhomogeneous mixing model A*. In equation (3), the parameter α_0 reflects the reproductive number for within-ZCTA movements, while the parameter α_1 reflects the corresponding reproductive number for between-ZCTA movements. We estimated the models of equation (2) with weighted least squares, where the weights were the ZCTA populations derived from the New York City Department of Health data (40).

Second, we relaxed the assumption that all between-ZCTA movements had the same reproductive number α_1. Instead, movements to and from certain high-risk ZCTAs were permitted to exert more influence than movements to and from the remaining lower-risk ZCTAs. To capture such differences in transmission efficiency, we partitioned the set of ZCTAs into two mutually exclusive subsets, L and H, representing the low- and high-transmission ZCTAs, respectively. Conformally partitioning X_t vertically into two matrices, X_{tL} and X_{tH}, and the column vector Y_t horizontally into two vectors: Y_{tL} and Y_{tH}, our equation (3) becomes:

$$Y_{t+1} = \alpha_0\ D_t\ Y_t + \alpha_1 X_{tL} Y_{tL} + \alpha_2 X_{tH} Y_{tH} + \eta_{t+1} \tag{4}$$

The unknown parameters α_0, α_1, and α_2, respectively, represent the reproductive numbers for movements within-ZCTAs, movements to and from low-risk ZCTAs, and movements to and from high-risk ZCTAs. We refer to this alternative as *inhomogeneous mixing model B*.

We performed two tests of inhomogeneous mixing model B, based upon two different partitions of the set of 21 ZCTAs delineated in Figure 2. First, on the basis of our examination of the trends in inter-ZCTA movements, as detailed in the Results section, we identified five high-movement ZCTAs along the southern boundary of the 21-ZCTA area as the most likely elements of the high-risk set H. The estimates based upon this high-low partitioning of ZCTAs were identified as B1. Second, we relied on the original classification of regulatory zones specified in the Governor's order of October 6, with the red-zone ZCTAs specified as high risk (H) and the remaining orange and yellow zones specified as low risk (L). These estimates were identified as B2. We similarly estimated the inhomogeneous geospatial models of Equations (3) and (4) with population-weighted least squares.

The model of equations (1) through (4) does not account for potential confounding factors. In Appendix B, we show how the model can be extended to incorporate such potential confounders, and we test the effects of including three demographic characteristics as covariates: the proportion of persons of Hispanic-Latino origin; the proportion of black non-Hispanics; and the proportion of persons receiving public assistance. While mobile device use has become pervasive in the U.S., the lower rates of smartphone use among the poorest individuals (44), as well as racial and ethnic differences in the patterns of smartphone use (45), could have biased our results.

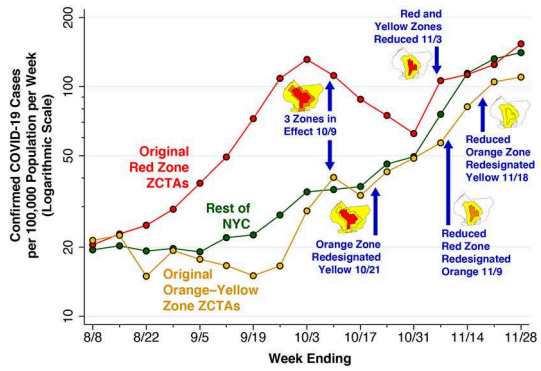


FIGURE 4
Confirmed COVID-19 cases per 100,000 population in the original red zone, original combined orange-yellow zones, and the rest of New York City, weeks ending 8/8 through 11/28/2020. The data points represent population-weighted rates, derived from ZCTA-specific weekly COVID-19 incidence as reported by the NYC Department of Health (40), and aggregated according to the scheme in Figure 2A. Superimposed on the graphic are the dates of five successive regulatory actions: (i) establishment of the original three zones on October 9 (7, 29); (ii) re-designation of the original orange zone as yellow on October 21 (24); (iii) reduction in the size of the red and orange zones on November 3 (27, 32); (iv) re-designation of the reduced red zone as orange on November 9 (27); and (v) re-designation of the reduced orange zone as yellow on November 18 (28).

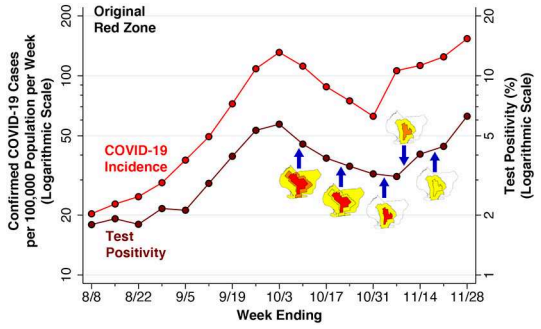


FIGURE 5
Confirmed COVID-19 cases per 100,000 population (Left Axis) and Test Positivity rate (Right Axis) in the Original Red Zone, weeks ending 8/8 through 11/28/2020. As in Figure 4, the data points represent population-weighted rates, derived from ZCTA-specific weekly COVID-19 incidence (40) and test positivity (41) among the ZCTAs identified as covering the red zone in Figure 2A. As in Figure 4, the blue arrows show the dates on which the five successive regulatory actions went into effect, while the accompanying maps show the corresponding zone changes.

Results

COVID-19 incidence and test positivity

Figure 4 below plots the incidence of confirmed cases of COVID-19 per 100,000 population in the original red zone, the original combined orange and yellow zones, and in the rest of New York City during the weeks ending August 8 through November 28. Also noted in the plot are the dates of Governor's five successive regulatory actions, starting with the imposition of the original three concentric zones, effective October 9. While the regulatory zones underwent revisions, the geographic areas used to compute case incidence in Figure 4 remained unchanged.

After rising during August and September, COVID-19 incidence in the ZCTAs comprising the original red zone started to decline during the week ending October 10, a time period that included 5 days before the regulatory scheme took effective. COVID-19 incidence in the red zone continued to decline through the week ending October 31, but then began to rebound. By contrast, COVID-19 incidence in the original orange and yellow zones, as well as the rest of New York City, had been increasing since at least mid-September, and reached approximately the same level as the original red zone by November. By November 21, all three series had exceeded the threshold of 100 cases per 100,000 population per week.

Figure 5 focuses sharply on the original red zone. The incidence of confirmed COVID-19 cases per 100,000 population, measured on the left-hand vertical scale, is reproduced from Figure 4. Superimposed on this time series is the test positivity rate, measured on the right-hand vertical scale. As in Figure 4, the dates when each of the five successive regulatory actions went into effect are noted. Again, while the regulatory zones underwent successive revisions, the geographic area used to compute the positivity rate – namely the original red zone – remained unchanged.

The variable vertical gap between the two time series in Figure 5 corresponds to the changing testing rate for COVID-19. Thus, the testing rate per 1,000 population progressively increased from 11.33 during the week ending August 8 to 17.15 during the week ending September 12, and then further increased to 24.61 by the week ending October 10. By the weeks ending November 7, 14 and 21, respectively, the testing rates had reached 33.99, 27.88, and 28.20 per 1,000 population.

Because the red-zone ZCTAs in Figure 2A only approximate the precise boundaries of the original red zone, the estimated test positivity rates plotted in Figure 5 represent only approximations to the positivity rates that were relied upon by state regulators. Still, during the week ending November 7, there is a striking divergence between the declining test positivity rate and the concurrently rising incidence rate. This finding suggests that regulators, relying on the trend in a test positivity rate that

was biased downward by enhanced testing, relaxed restrictions when in fact the incidence of the disease was rising.

Figure 6 below tracks the detailed evolution of COVID-19 incidence in each of the 21 ZCTAs identified in Figure 2 during the weeks ending September 12 through November 28. The first row, covering the weeks ending September 12 through October 3, shows increasing disease incidence in the central ZCTAs covering what would ultimately be designated as the red zone. In the second row, covering the weeks ending October 10 – 31, the incidence of COVID-19 was declining in the central ZCTAs but increasing in the peripheral ZCTAs, particularly along the southern and western borders of the South Brooklyn region. In the third row, covering the weeks ending November 7 – 28, COVID-19 incidence continued to rise in these peripheral ZCTAs, while resuming its upward trend in the central ZCTAs. By November 28, the original central zone of high-incidence ZCTAs is no longer distinguishable.

Visits to restaurants: Paired POI analysis

Figure 7 below shows the results of our paired POI analysis of visits to restaurants and other eating places. The estimate of $\beta_1 = -0.075$ for the week ending October 8 is negative and significantly different from zero in a two-sided test ($p = 0.0497$). That is, visits to restaurants in the red zone had already declined by 7.5% relative to those in the orange zone during the week before the regulatory scheme went into effect. While the individual estimates of $\beta_2 = -0.059$ and $\beta_3 = -0.067$ are not significantly different from zero, the overall downward trend is evident by the weeks ending October 29 and November 5, where $\beta_4 = -0.114$ ($p = 0.003$) and $\beta_5 = -0.172$ ($p < 0.001$). Thereafter, as the red zone is reduced by half (effective November 3), then reclassified as orange (effective November 9), and then reclassified as yellow (November 18), the estimates of β_t begin to rebound. By the week ending December 3, the

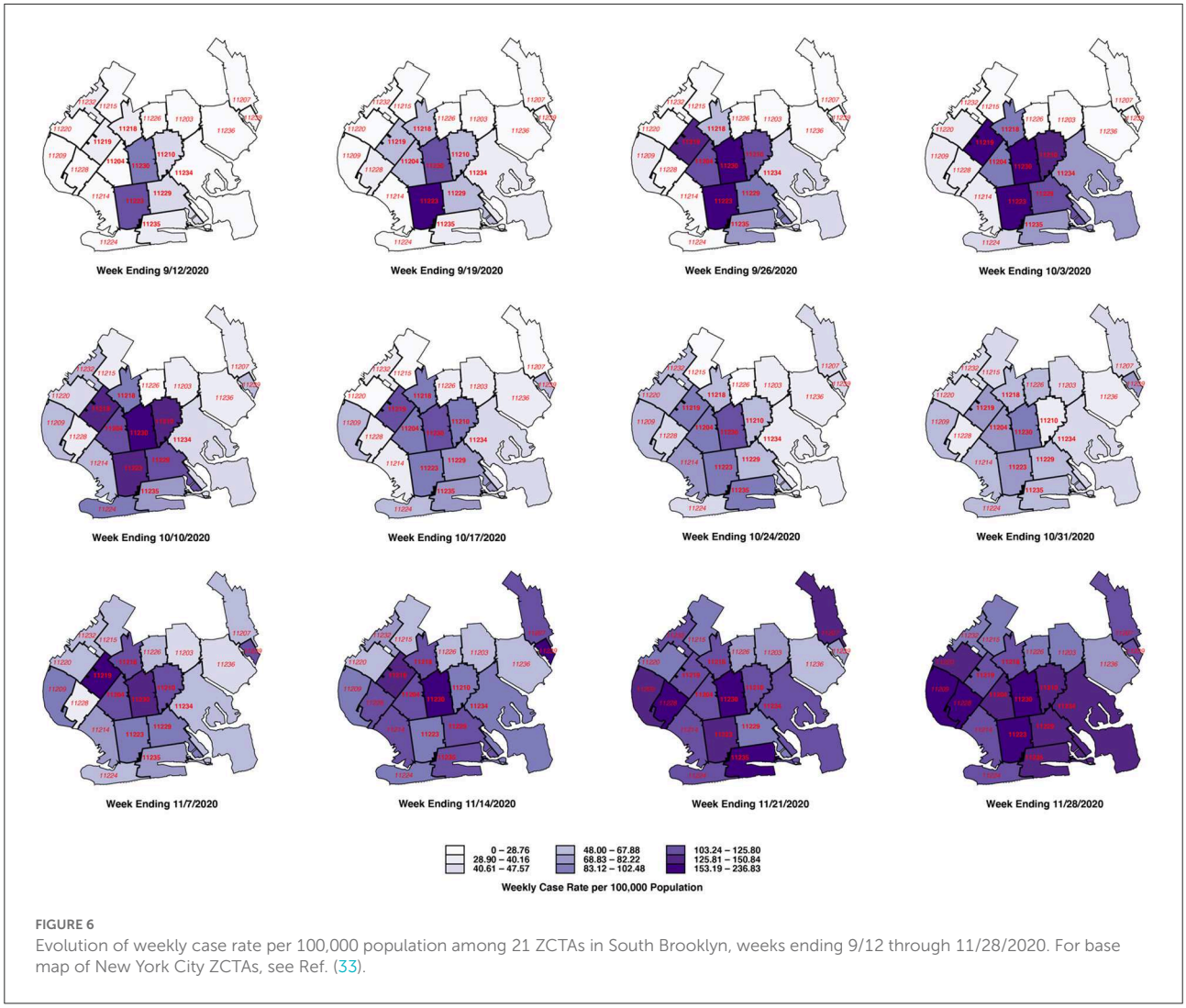


FIGURE 6
Evolution of weekly case rate per 100,000 population among 21 ZCTAs in South Brooklyn, weeks ending 9/12 through 11/28/2020. For base map of New York City ZCTAs, see Ref. (33).

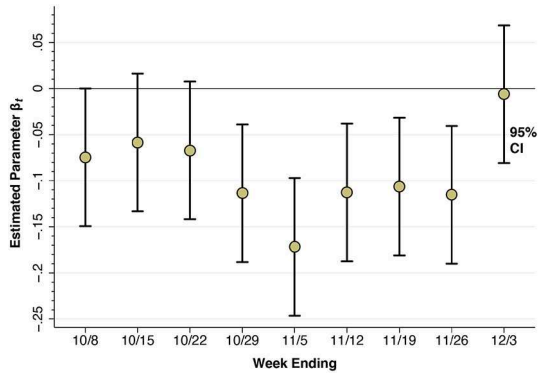


FIGURE 7
Estimated time-specific fixed effects derived from the paired point-of-interest analysis. Each point represents an estimate of the contrast $\beta_t - \beta_0$ for $t = 1,..., 9$, where $t = 0$ represents the week ending 10/1/20 as a reference category. The error bars surrounding each point represent 95% confidence intervals.

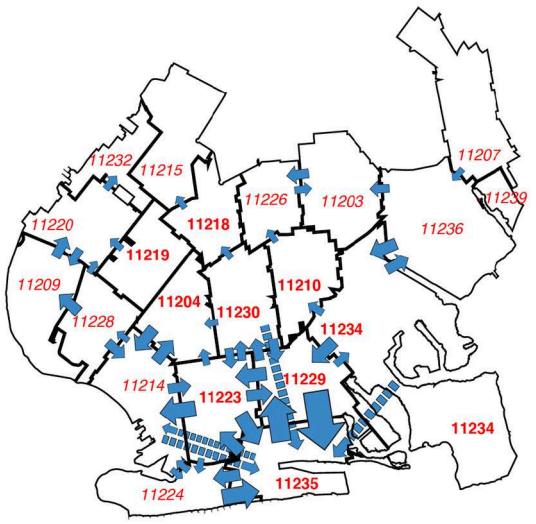


FIGURE 8
Most frequent transits between ZCTAs during 10/25 through 11/28/2020. The width of each blue arrow is proportional to the number of device movements in the Safegraph cohort. The longer arrows with dashed lines represent transits between non-contiguous ZCTAs. The number of device movements ranged from 21139 (corresponding to the transit 11229 → 11235) to 5002 (corresponding to 11214 → 11224). Transits between ZCTAs with fewer than 5,000 movements are not shown. The transits captured by the arrows comprised 44.5% of all device movements between ZCTAs. For base map of New York City ZCTAs, see ref. (33).

estimate $\beta_9 = -0.006$ is no longer significantly different from zero ($p = 0.87$).

Movements within and between regulatory zones

The results of our paired analysis of restaurants and other eating places, as shown in Figure 7, narrowly reflect trips to a specific category of establishments that were subject to specific regulatory controls. They do not necessarily capture broader trends in unregulated movements of individuals within and between zones.

Supplementary Table A1 delineates movements of devices within and between zones, as well as movements outside the regulated area, during the 3 weeks before and the subsequent 3 weeks after the regulations went into effect on October 9. Comparison of the movement matrices during the two time periods (September 18 through October 8, October 9 through October 29) indicates that overall movement patterns remained stable. Among devices homed in the original red zone, only 57–58% of movements were confined to the red zone, while 23–24% of movements were to destinations outside the regulated area entirely. The same pattern is evident in the movements of devices homed in the original orange and yellow zones as well.

The overall stability of within-zone movements seen in Supplementary Table A1 could still obscure significant changes in mobility, particularly in the proportion of devices that made few if any movements. Supplementary Figure A2, however, shows that the percentage of devices that made no movements at all rose only by 1 to 2 percentage points.

Figure 8 tracks more detailed movements between ZCTAs, rather than between regulatory zones, during October 25 through November 28. This time period corresponds to the final 5 weekly maps in Figure 6, when COVID-19 was increasing both in the central ZCTAs and the peripheral ZCTAs along the southern and western edges of the regulated area. The width of each arrow corresponds to the magnitude of the flow between ZCTAs. The figure demonstrates that dominant inter-ZCTA movements were between four red-zone ZCTAs (11229, 11223, 11230, 11204) and three orange- and yellow-zone ZCTAs (11235, 11224, 11214). Moreover, there was significant device traffic between these southern ZCTAs and other peripheral ZCTAs to the west of the regulated area.

Influence of high-risk ZCTAs: Geospatial analysis

Based upon our findings in Figure 8, we identified five high-mobility ZCTAs as potential candidates for high-risk transmission to other surrounding ZCTAs in the area: 11223, 11229, 11235, 11224, and 11214. As seen in Figure 2A, two of the areas, 11224 and 11214, were not in the original red zone,

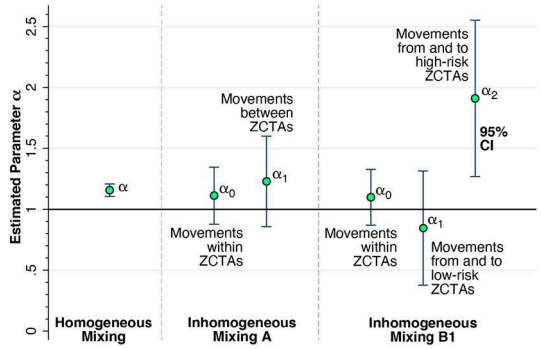


FIGURE 9
Estimates of the reproductive number α from a series of geospatial analyses. **(Left)** Estimate of a uniform parameter $\alpha = 1.157$ with 95% confidence interval [1.106, 1.208] for the entire regulated region, based upon the assumption of homogeneous mixing. The estimate of α is significantly different from 1.0 (2-sided F-test, $p < 0.001$). **(Middle)** Estimates of the parameters α_0 for the effect of movements within ZCTAs and α_1 for between-ZCTA movements, based upon a two-parameter model of inhomogeneous mixing A. **(Right)** Estimates of the parameters α_0 for the effect of movements within ZCTAs, α_1 for the effect of movements from and to low-risk ZCTAs, and α_2 for the effect of movements from and to high-risk ZCTAs 11223, 11229, 11235, 11224, and 11214, based upon a three-parameter model of inhomogeneous mixing B1. The estimated parameter $\alpha_2 = 1.910$ with 95% confidence interval [1.268, 2.552] was significantly different from 1.0 (2-sided F-test, $p = 0.012$). Full details are given in Supplementary Table C1.

while another area, 11235, was excluded from the red zone when the zones were contracted on November 3.

Figure 9 shows the principal results of our geospatial analysis. At the left end of the graphic, in a model of homogeneous mixing with a single, uniform parameter for the entire regulated area (equation 2 in the Methods section), the estimated reproductive number was $\alpha = 1.157$ with 95% confidence interval [1.106, 1.208]. The middle panel shows the results of a two-parameter model of inhomogeneous mixing A, which allowed for movements within ZCTAs to have a different effect on COVID-19 transmission (equation 3). Both parameters α_0 for within-ZCTA movements and α_1 for between-ZCTA movements were in the range of 1.1-—1.2, though they were too imprecise to be distinguishable from each other or from 1.0.

The right-hand panel of Figure 9 shows the estimates of the three-parameter model with inhomogeneous mixing B1 (equation 4), which further distinguished the five high-risk ZCTAs 11223, 11229, 11235, 11224, and 11214 from the remaining sixteen low-risk ZCTAs. The estimate of α_0 was imprecise but consistent with the findings in the two-parameter model A in the middle panel. The estimate of α_1 for low-risk ZCTAs was likewise imprecise, but pointed to a reproductive number <1.0. By contrast, the estimate of α_2 for the effect of device traffic to and from the five high-risk ZCTAs gave a reproductive number of 1.910 with 95% confidence interval [1.268, 2.552]. Moreover, the estimate of α_2 was significantly different from α_0 ($p = 0.045$) and α_1 ($p = 0.012$) based upon 2-sided F-tests.

The alternative test of the three-parameter model with inhomogeneous mixing B2 (equation 4), based upon the classification of all the original red-zone ZCTAs as high risk, did not yield such precise results. The estimate of α_2 for the effect of device traffic to and from these original red-zone ZCTAs gave a reproductive number of 1.485 with 95% confidence interval [0.992, 1.978]. This estimate was not significantly different from α_0 ($p = 0.284$) or from α_1 ($p = 0.124$). The complete results of the geospatial analysis are shown in Supplementary Table C1.

Supplementary Table C2 displays the effects of incorporating three demographic characteristics as covariates in our model: the proportion Hispanic-Latino, the proportion black non-Hispanic, and the proportion receiving public assistance. None of the coefficients of these covariates achieved statistically significance at the 5-percent level. The results were consistent with those reported in Figure 9.

Discussion

Non-identifiability of the effect of the governor's regulatory scheme

Our analysis of paired eating places straddling the red-orange border suggests that the Governor's October 6 regulatory scheme did indeed have some effect. The red-zone rules allowed for only takeout and delivery, while the less stringent orange-zone rules allowed for outdoor dining as well (23). By the week ending November 5, as seen in Figure 7, device visits to establishments on the red side of the red-orange border were down 17.2 percent more than their counterparts on the orange side. Attributing the entire decline to the regulatory scheme is problematic, however, inasmuch as device visits to establishments on the red side of border were already dropping more rapidly during the week ending October 8. Aside from voluntary action on the part of the red zone's residents, the Mayor's threats to impose controls in five key zip codes in South Brooklyn, voiced as early as September 29 (21), may have contributed to a preexisting downward trend.

What's more, a demonstrated narrow effect on a specific endpoint such as restaurant visits does not necessarily imply that the regulatory scheme had an overall deterrent effect on SARS-CoV-2 transmission in South Brooklyn. The COVID-19 incidence data in Figure 4 do show a temporary decline during the weeks ending October 17, 24, and 31, after the regulatory scheme had entered into force. However, as in our interpretation of the restaurant visitation data in Figure 7, COVID-19 incidence was declining during the week before any controls went into effect. And in view of the ∼5-day

incubation period between initial infection and the subsequent development of symptoms warranting testing (46), the incidence of the disease was likely to have been declining even earlier. These considerations reinforce the conclusion that the actual effect of the Governor's regulatory scheme is, strictly speaking, not identifiable from the available data.

Even if the Governor's regulatory scheme was at least partly responsible for retarding the surge of COVID-19 in the central red zone during the last 3 weeks of October, it is evident that the program's success was short-lived. As Figures 4, 6 show, the decline in COVID-19 incidence during October was ultimately reversed by a wave of increasing disease incidence that had similarly overtaken the surrounding orange and yellow zones by the end of November. These observations beg the question: Why did the Governor's scheme ultimately fail?

Overreliance on test positivity

One clue is offered by the striking discordance between the upward surge of disease *incidence* and the continuing downward trend in *test positivity* during the week ending November 9, as shown in Figure 5. The underlying explanation for the discordance of these two trends was the continuing expansion of testing, which diluted the rising incidence with an abundance of negative test results. It was just at this juncture that regulators, overly fixated on the test positivity rate, cut the size of the red zone in half and then converted the red zone to orange. Under this interpretation of the evidence, the Governor's regulatory scheme did have an initial retardant effect on COVID-19 incidence, but subsequent premature withdrawal of regulatory controls neutralized the effect of the initial policy.

Access controls vs. restrictions on population movement

The regulations imposed in the Governor's concentric regulatory zones were fundamentally controls on access – to eating places, to school buildings, to houses of worship, and to large meetings. They differ from the classic remedy of containment, which entails restrictions not just on access, but also on overall population movement (47–49). Although restricting access to some critical locations is indeed likely to reduce disease propagation (15–17, 50, 51), the question here is whether focused controls on access alone were sufficient to alter the underlying population movement patterns that served as the template for a surge in COVID-19 cases. While some have cited increasing indoor activity with the arrival of colder fall weather (52) or a trend toward large family gatherings as the Thanksgiving holiday approached (53), we have in mind more fundamental, well-established contact networks.

The hypothesis that the underlying population movement patterns within the regulated area in South Brooklyn were insufficiently altered by the Governor's regulatory controls on access was supported by the stability of the interzone movement matrices before and after the promulgation of the regulatory scheme (Supplementary Table A1) as well as absence of any significant change in the proportion of devices with no movements (Supplementary Figure A3).

High-movement ZCTAs as drivers of the local COVID-19 surge

One interpretation of the trends in Figure 4 is that the surge in COVID-19 cases that ultimately overran South Brooklyn was a citywide phenomenon, and that the incidence of the disease was simply increasingly uniformly across all ZCTAs. However, the patterns of declining and rising COVID-19 incidence seen in Figure 6 go against this interpretation. Once the incidence reached a low point around the week ending October 17, the subsequent increase was driven by ZCTAs along the southern and western borders of the area. During the resurgence of COVID-19 incidence, movement patterns were hardly uniform, as shown in Figure 8. In fact, the geospatial analysis demonstrates that five high-movement ZCTAs – of which only two were part of the original red zone – were the main drivers of the resurgence in COVID-19, with a reproductive number close to 2, as shown in the panel labeled Inhomogeneous Mixing B1 in Figure 9. That population movement was the critical determinant is further supported by the finding of the inferior performance of Model B2, based upon the Governor's original partitioning of regulatory zones (Supplementary Table C1).

A natural experiment

Our study is fundamentally observational. We did not analyze a macro-experiment in which various communities were randomly assigned to different regulatory controls or no intervention at all (54). Still, our setup has many of the features of a natural experiment. The Governor's announcement of a new regulatory regime on October 6, to become effective by October 9, could reasonably be characterized as an abrupt shock (7). In view of the Mayor's earlier threats to impose controls on certain zip codes in South Brooklyn (21), however, it could hardly be considered an unanticipated shock. The implementation of distinct regulations within each of three concentric regulatory zones provided natural intervention and control groups, and the results of our paired restaurant analysis (Figure 7) suggest that the regulations on access were effective and enforced. While the restaurants stayed put, however, Supplementary Table A1 and Figure 8 show that the experimental participants crossed over from one zone to another. Midway through the intervention, on

November 9, the Governor cut the regulatory zones in half (27), thereby contaminating the original experimental assignments.

Questions of generalizability

Our study has focused on a narrowly defined geographic location during a specific wave of the COVID-19 epidemic. Generalization of the current findings to other settings or other phases of the epidemic needs to proceed with caution. Still, the evidence reported here is broadly consistent with findings reported in Hong Kong (8), in Shenzhen, China (18), in Dublin, Ireland (13), in Cataluña, Spain (16), in various Latin American cities (9), and in other sites in Asia Pacific and Europe (20). The critical role of mobility has also been demonstrated in the most populous counties of the United States during the Omicron variant wave of the COVID-19 epidemic during December 2021 –February 2022 (14).

Conclusions

Combining data on COVID-19 incidence and testing outcomes with data on the movements of devices equipped with location-tracking software, we found that a regulatory scheme of concentric geographic zones imposing graduated restrictions on access did not halt the surge of COVID-19 in South Brooklyn, New York, during October-November 2020.

Beyond this principal conclusion, we can reasonably draw some additional inferences from the accumulated evidence. First, test positivity as a real-time indicator of regulatory effectiveness is fraught with potential biases (55, 56). Here, the Governor and his advisors may have been led astray by a test positivity rate that was kept misleadingly close to three percent by an endogenous increase in testing among COVID-negative persons (Figure 5).

Second, while restrictions on access to eating establishments and other high-risk venues may be narrowly effective (Figure 7), they do not prevent people from moving around (Supplementary Table A1 and Figure 8). In highly populous areas such as South Brooklyn, a halfway strategy of concentric regulatory zones based solely on access restrictions may be no substitute for the classic approach of concentric containment/quarantine areas (6, 49).

Third, overreliance on static rather than dynamic measures of disease burden to draw the boundaries of regulatory zones can prove to be highly misleading. Our geospatial analysis of COVID-19 incidence, entailing a dynamic model of COVID-19 incidence across 21 zip code tabulation areas (Figure 9), identified five high-movement ZCTAs where the reproductive number approached 2. Two of the five were not in the original red zone. Concentric zones may appear to be an effective regulatory approach in principle, but only if the boundaries are drawn correctly.

Fourth, policies restricting movement can take many forms, including controls on transportation networks. There is substantial evidence pointing to the initial widespread dissemination of SARS-CoV-2 *via* New York City's subway-based network during February-March 2020, followed by percolation of new infections within local community hotspots (57). A policy of running express lines with limited density might have been an alternative to the complete shutdown of subway lines adopted in Wuhan (58).

Finally, in extreme cases, it may be necessary to impose stay-at-home restrictions. During an outbreak in September 2020 on the campus of the University of Wisconsin-Madison, the university administration barred students from leaving two highly infected residence halls. By the end of the month, COVID-19 incidence on campus had fallen below that of the surrounding county (51). Even so, such stay-at-home orders may prove incompletely effective when disease propagation is dominated by intrahousehold transmission, as it was during the winter COVID-19 surge in Los Angeles County, a region with a high prevalence of multi-generational households (59). Whether such a stay-at-home order would have been effective or even feasible in the case of South Brooklyn remains an open question.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: Supporting programs and data have been posted at the Open Science Framework https://osf.io/rquyx/.

Author contributions

JH is the sole author of this work. He is responsible for the conceptualization of the study, the data analysis, the drafting of the manuscript, and the creation of the figures.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.970363/full#supplementary-material

References

1. U.S. Department of Agriculture. *Develop a Surveillance Plan: Disease Control Areas / Zones*. (2020). Available online at: https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/ceah-toolbox/section-3.-disease-control-areas-and-zones: Animal and Plant Health Inspection Service, Updated June 2.

2. Department of the Interior and Local Government (DILG), *RIATF-RTF NCR lead zoning containment strategy orientation*. (2020). Available online at: http://ncr.dilg.gov.ph/home/riatf-rtf-ncr-leads-zoning-containment-strategy-orientation/: Republic of the Philippines.

3. Gotinga, J.C. *Zonal lockdowns: Is your house, street, or city a critical zone*? (2020). Available online at: https://www.rappler.com/newsbreak/iq/explainer-zonal-coronavirus-lockdowns

4. Bajardi P, Poletto C, Ramasco JJ, Tizzoni M, Colizza V, Vespignani A. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS ONE*. (2011) 6:e16591. doi: 10.1371/journal.pone.0016591

5. Rea E, Lafleche J, Stalker S, Guarda BK, Shapiro H, Johnson I, et al. Duration and distance of exposure are important predictors of transmission among community contacts of Ontario SARS cases. *Epidemiol Infect*. (2007) 135:914–21. doi: 10.1017/S0950268806007771

6. Kuo FY, Wen TH. Regionalization for infection control: an algorithm for delineating containment zones considering the regularity of human mobility. *Appl Geogr*. (2021) 126:102375. doi: 10.1016/j.apgeog.2020.102375

7. New York Governor, *Governor Cuomo Announces New Cluster Action Initiative*. (2020). Available online at: https://www.governor.ny.gov/news/governor-cuomo-announces-new-cluster-action-initiative

8. Wong PP, Low CT, Lai PC. The impact of geographic mobility on the spread of COVID-19 in Hong Kong. *Geospat Health*. (2022) 17:1022. doi: 10.4081/gh.2022.1022

9. Kephart JL, Delclòs-Alió X, Rodríguez DA, Sarmiento OL, Barrientos-Gutiérrez T, Ramirez-Zea M, et al. The effect of population mobility on COVID-19 incidence in 314 Latin American cities: a longitudinal ecological study with mobile phone location data. *Lancet Digit Health*. (2021). 3:e716–722. doi: 10.1016/S2589-7500(21)00174-6

10. Oh J, Lee HY, Khuong QL, Markuns JF, Bullen C, Barrios OE, et al. Mobility restrictions were associated with reductions in COVID-19 incidence early in the pandemic: evidence from a real-time evaluation in 34 countries. *Sci Rep*. (2021) 11:13717. doi: 10.1038/s41598-021-92766-z

11. Nouvellet P, Bhatia S, Cori A, Ainslie KE, Baguelin M, Bhatt S, et al. Reduction in mobility and COVID-19 transmission. *Nat Commun*. (2021) 12:1090. doi: 10.1038/s41467-021-21358-2

12. He S, Lee J, Langworthy B, Xin J, James P, Yang Y, et al. Delay in the effect of restricting community mobility on the spread of COVID-19 during the first wave in the United States. *Open Forum Infect Dis*. (2022) 9:ofab586. doi: 10.1093/ofid/ofab586

13. Manzira CK, Caulfield B. Assessing the impact of mobility on the incidence of COVID-19 in Dublin City. *Sustain Cities Soc*. (2022) 80:103770. doi: 10.1016/j.scs.2022.103770

14. Harris JE. Mobility was a significant determinant of reported COVID-19 incidence during the Omicron surge in the most populous U.S. counties. *BMC Infect Dis*. (2022) 22:691. doi: 10.1186/s12879-022-07666-y

15. Schnake-Mahl AS, O'Leary G, Mullachery PH, Vaidya V, Connor G, Rollins H, et al. The impact of keeping indoor dining closed on COVID-19 rates among large US cities: a quasi-experimental design. *Epidemiology*. (2022) 33:200–8. doi: 10.1097/EDE.0000000000001444

16. Leon. Evaluating the policy of closing bars and restaurants in Cataluna and its effects on mobility and COVID19 incidence. *Sci Rep*. (2022) 12:9132. doi: 10.1101/2020.04.21.20073049

17. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. *Science*. (2021) 371:9338. doi: 10.1126/science.abd9338

18. Zhou Y, Xu R, Hu D, Yue Y, Li Q, Xia J. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. Lancet Digit Health. (2020). 2(8): p. e417–e424. doi: 10.1016/S2589-7500(20)30165-5

19. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. (2020) 368:395–400. doi: 10.1126/science.aba9757

20. Han E, Tan MM, Turk E, Sridhar D, Leung GM, Shibuya K, et al. Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. *Lancet*. (2020) 396:1525–34. doi: 10.1016/S0140-6736(20)32007-9

21. Lardieri A. *De Blasio Warns of Reimposing Coronavirus Restrictions as New York Positivity Rate Rises*. (2020). Available online at: https://www.usnews.com/news/health-news/articles/2020-09-29/de-blasio-warns-of-reimposing-coronavirus-restrictions-as-new-york-positivity-rate-rises

22. Kim E, Pereira S. *De Blasio Moves To Order Shutdowns In 9 Brooklyn And Queens ZIP Codes With Virus Outbreaks*. (2020). Available online at: https://gothamist.com/news/nyc-moves-order-shutdowns-9-brooklyn-and-queens-zip-codes-virus-outbreaks

23. New York Governor, *New York "Micro-Cluster" Strategy*. (2020). Available online at: https://www.governor.ny.gov/sites/default/files/atoms/files/MicroCluster_Metrics_10.21.20_FINAL.pdf

24. New York Governor. *Governor Cuomo Details COVID-19 Micro-Cluster Metrics*. (2020). Available online at: https://www.governor.ny.gov/news/governor-cuomo-details-covid-19-micro-cluster-metrics

25. U.S. Supreme Court. *Roman Catholic Diocese of Brooklyn, New York v. Andrew M. Cuomo, Governor of New York*. (2020). Available online at: https://www.supremecourt.gov/opinions/20pdf/20a87_4g15.pdf

26. New York Governor. *Governor Cuomo Announces State Will Withhold Funds for Localities and Schools in COVID-19 Cluster Zones That Fail to Enforce Public Health Law*. (2020). Available online at: https://www.governor.ny.gov/news/governor-cuomo-announces-state-will-withhold-funds-localities-and-schools-covid-19-cluster

27. New York Governor, *Governor Cuomo Announces Updated COVID-19 Micro-Cluster Focus Zones*. (2020). Available online at: https://www.governor.ny.gov/news/governor-cuomo-announces-updated-covid-19-micro-cluster-focus-zones

28. New York Governor. *Governor Cuomo Announces Updated COVID-19 Micro-Cluster Focus Zone*. (2020). Available online at: https://www.governor.ny.gov/news/governor-cuomo-announces-updated-covid-19-micro-cluster-focus-zones-2

29. New York Governor. *Brooklyn Large Map*. (2020). Available online at: https://www.governor.ny.gov/sites/default/files/atoms/files/Brooklyn_large_map.pdf

30. New York Governor. *Brooklyn_HiRes2.pdf (Map)*. (2020). Available online at: https://www.governor.ny.gov/sites/default/files/atoms/files/Brooklyn_HiRes2.pdf: October 21.

31. New York Governor. *Brooklyn_New_Zones_11_03_2020 (Map)*. (2020). Available online at: https://www.governor.ny.gov/sites/default/files/atoms/files/Brooklyn_New_Zones_11_03_2020.pdf

32. New York Governor. *Brooklyn New Zones (Map)*. (2020). Available online at: https://www.governor.ny.gov/sites/default/files/atoms/files/Brooklyn_New_Zones_11_03_2020.pdf

33. New York City Department of Health and Mental Hygiene. *Geo Files*. (2022). Available online at: https://github.com/nychealth/coronavirus-data/tree/master/Geography-resources (accessed April 27).

34. SafeGraph Inc. *Social Distancing Metrics*. (2020). Available online at: https://docs.safegraph.com/docs/social-distancing-metrics

35. QGIS Development Team. *QGIS Geographic Information System*. (2021). Available online at: http://qgis.osgeo.org

36. U.S Census Bureau. *TIGER2017 Shape Files for Census Block Groups by State*. (2017). Available online at: https://www2.census.gov/geo/tiger/TIGER2017/BG

37. Picard R. *GEOINPOLY: Stata module to match geographic locations to shapefile polygons*. (2015). Available online at: https://ideas.repec.org/c/boc/bocode/s458016.html

38. SafeGraph Inc. *Places Schema*. (2020). Available online at: https://docs.safegraph.com/docs/places-schema (accessed July 30–31, September 24–26).

39. U.S Census Bureau. TIGER/Line Shapefile. State, New York, Current Block Group State-based (2021). Available online at: https://catalog.data.gov/dataset/tiger-line-shapefile-2017-state-new-york-current-block-group-state-based

40. New York City Department of Health and Mental Hygiene, caserate-by-modzcta.csv (Weeks ending 8/8 through 11/28/2020). (2020). Available online at: https://raw.githubusercontent.com/nychealth/coronavirus-data/84dead220def7b53bcb2d9f9210609dee9446045/trends/caserate-by-modzcta.csv

41. New York City Department of Health and Mental Hygiene, *percentpositive-by-modzcta.csv*. (2021). Available online at: https://github.com/nychealth/coronavirus-data/blob/master/trends/percentpositive-by-modzcta.csv

42. Picard R. GEONEAR: Stata module to find nearest neighbors using geodetic distances. (2019). Available online at: https://ideas.repec.org/c/boc/bocode/s457146.html

43. Kolokolnikov T, Iron D. Law of mass action and saturation in SIR model with application to Coronavirus modelling. *Infect Dis Model.* (2021) 6:91–7. doi: 10.1016/j.idm.2020.11.002

44. Vogels EA. Digital divide persists even as Americans with lower incomes make gains in tech adoption. (2021). Available online at: https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/

45. Anderson M. Racial and ethnic differences in how people use mobile technology. (2015). Available online at: https://www.pewresearch.org/fact-tank/2015/04/30/racial-and-ethnic-differences-in-how-people-use-mobile-technology/

46. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med*. (2020) 172:577–82. doi: 10.7326/M20-0504

47. Chiew. Can we contain the COVID-19 outbreak with the same measures as for SARS? *Lancet Infect Dis*. (2020) 20:e102–7. doi: 10.1016/S1473-3099(20)30129-8

48. Day T, Park A, Madras N, Gumel A, Wu J. When is quarantine a useful control strategy for emerging infectious diseases? *Am J Epidemiol*. (2006) 163:479–85. doi: 10.1093/aje/kwj056

49. Yacoub. The concept of quarantine in history: from plague to SARS. *J Infect*. (2004) 49:257–61. doi: 10.1016/j.jinf.2004.03.002

50. Harris JE. COVID-19, bar crowding, and the Wisconsin supreme court: a non-linear tale of two counties. *Res Int Bus Finance*. (2020) 54:101310. doi: 10.1016/j.ribaf.2020.101310

51. Harris JE. Geospatial analysis of a COVID-19 outbreak at the university of Wisconsin - Madison: potential role of a cluster of local bars. *Epidemiol Infect*. (2022) 3:1–31. doi: 10.1017/S0950268822000498

52. Mallapaty S. Why COVID outbreaks look set to worsen this winter. *Nature*. (2020) 586:653. doi: 10.1038/d41586-020-02972-4

53. Mehta SH, Clipman SJ, Wesolowski A, Solomon SS. Holiday gatherings, mobility and SARS-CoV-2 transmission: results from 10 US states following. *Thanksgiving Sci Rep*. (2021) 11:17328. doi: 10.1038/s41598-021-96779-6

54. Harris JE. Macro-experiments versus micro-experiments for health policy, in *Social Experimentation*, Wise JH editor, Chicago: University of Chicago Press (1985). Available online at: https://www.nber.org/books-and-chapters/social-experimentation/macroexperiments-versus-microexperiments-health-policy. p. 145-186.

55. Harris JE. *Reopening Under COVID-19: What to Watch For*. (2020). Available online at: http://web.mit.edu/jeffrey/harris/HarrisJE_WP3_COVID19_WWF_6-May-2020.pdf. doi: 10.3386/w27166

56. Harris JE. *The Test Positivity Rate is (Nearly) Useless*. (2020). Available online at: https://lifeunderquarantine.blog/2020/12/04/the-test-positivity-rate-is-nearly-useless/

57. Harris JE. Critical role of the subways in the initial spread of SARS-CoV-2 in New York City. *Front Public Health*. (2021) 9:754767. doi: 10.3389/fpubh.2021.754767

58. BC News. *Coronavirus: Wuhan shuts public transport over outbreak*. (2020). Available online at: https://www.bbc.com/news/world-asia-china-51215348

59. Harris JE. Los Angeles county SARS-CoV-2 epidemic: critical role of multi-generational intra-household transmission. *J Bioecon*. (2021) 23:55–83. doi: 10.1007/s10818-021-09310-2

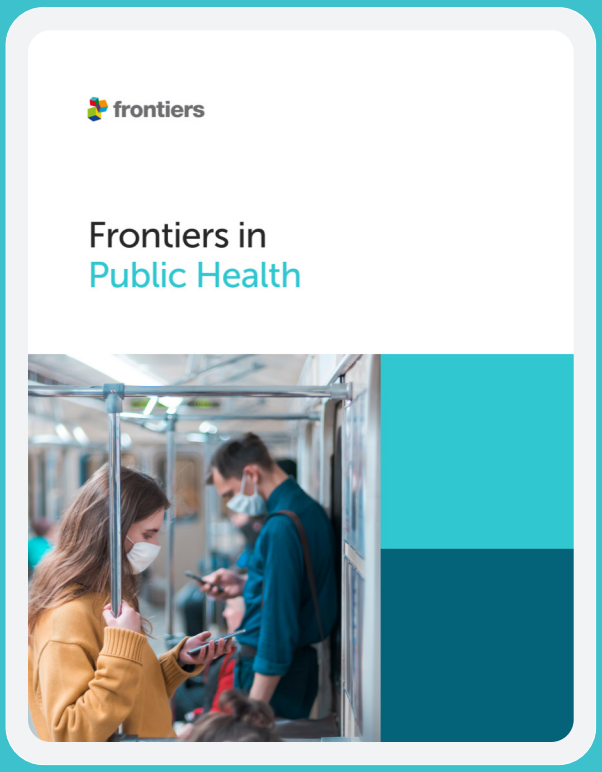
Frontiers in Public Health

Explores and addresses today's fast-moving healthcare challenges

One of the most cited journals in its field, which promotes discussion around inter-sectoral public health challenges spanning health promotion to climate change, transportation, environmental change and even species diversity.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact