Mathematical and statistical modeling of infection and transmission dynamics of viral diseases

Edited by
Pierre Magal, Jacques Demongeot
and Olumide Babatope Longe

Coordinated by
Kayode Oshinubi

Published in
Frontiers in Public Health

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

Mathematical and statistical modeling of infection and transmission dynamics of viral diseases

Topic editors

Pierre Magal — Université de Bordeaux, France
Jacques Demongeot — Université Grenoble Alpes, France
Olumide Babatope Longe — Academic City University College, Ghana

Topic coordinator

Kayode Oshinubi — Northern Arizona University, United States

Table of
contents

frontiers | Frontiers in Public Health

Editorial: Mathematical and statistical modeling of infection and transmission dynamics of viral diseases

Kayode Oshinubi[1], Pierre Magal[2], Olumide Longe[3] and
Jacques Demongeot[4]*

[1]School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ,
United States, [2]Individual Based Modeling, UMR CNRS 5251, University Bordeaux, Talence, France,
[3]Faculty of Computational Sciences and Informatics, Academic City University, Accra, Ghana,
[4]Laboratory AGEIS EA 7407, Team Tools for e-Gnosis Medical, Faculty of Medicine, University Grenoble
Alpes (UGA), La Tronche, France

Editorial on the Research Topic
Mathematical and statistical modeling of infection and transmission
dynamics of viral diseases

Introductory remarks on viral disease dynamics

The articles in this Research Topic show, many advances have been made in public health research, especially in the field of data collection, modeling (deterministic and random), and prediction: (i) public databases on COVID-19 have multiplied [see COVID-19 Open Data Repository (1) for example 301 public sites around the world allowing access to COVID-19 data], (ii) review articles on COVID-19 outbreak including statistics and models are numerous (in Frontiers in Public health alone, the site in WHO COVID-19 Research Database (2) counts 173 articles under heading Systematic review/Meta-Analysis) and (iii) the number of articles using a formal approach predicting COVID-19 waves is constantly growing (Google Scholar lists for example 144,000 articles with the query "modeling and forecasting COVID-19 pandemic").

The article in this Research Topic is useful for the modeling of viral disease dynamics.

The various facets of the recent mathematical and statistical models of viral disease data in this Research Topic

In the Research Topic entitled "*Mathematical and statistical modeling of infection and transmission dynamics of viral diseases*," the twelve papers are representative of different mathematical and statistical approaches to the recent COVID-19 pandemic, Hepatitis A/B virus, and Mpox disease data:

We first present articles that deals with the recent COVID-19 pandemic. In Tormos et al. analyzed the role of in-person school reopening in Spain on the evolution of

COVID-19 infections using an interrupted time-series perspective by exploring the dataset and modeling the dynamics of the disease in different Spanish regions and autonomous communities that reopened schools at varying moments in time during September 2020. Their analysis suggests that school reopening may generate a retro-feedback of the disease spread with parents' return to work and social activity, leading to exponential growth, as observed in Catalonia and other Spanish autonomous communities during September and October of 2020. Naffeti et al. established that the geography of the COVID-19 pandemic in Africa largely overlaps with the geography of the wealth of the 30 countries considered by using spatiotemporal evolution techniques, which take into consideration demographic, economic, and environmental aspects that can better explain the geographical variations of the basic reproduction rate at the early beginning of each wave of the pandemic. BuHamra et al. used self-developed natural language processing to automate the extraction of causes of death and comorbidities from the electronic health records of COVID-19 outbreak from the start of the pandemic until the end of all major epidemic waves. The research findings proposed that to reduce misspellings or incorrect forms, organizing the electronic health record with well-defined sections and giving menu-driven options for reporting causes of death and comorbidities. Majeed et al. aim to examine the impact of SARS-CoV-2 during an influenza season and quantify the effects of the two respiratory infections co-circulating using a mathematical model. The research investigates the best techniques to delay and split the peaks of the influenza outbreak and the COVID-19 wave among numerous scenarios and interventions. This study demonstrates that effectively managing and controlling both influenza and COVID-19 outbreaks during the same season depends on establishing optimal vaccine coverage techniques. KhudaBukhsh et al. investigated COVID-19 epidemic in Marion Correctional Institution in the spring of 2020 using a thorough and statistically sound technique. The analysis is based on a compartmental mathematical transmission model that is fitted to data using dynamical survival analysis, which permits the computation of explicit likelihoods to summarize uncertainty. The research findings underscore the tremendous potential for respiratory infection transmission in prisons, as well as the crucial need for improved infection monitoring and reporting in correctional facilities.

Chu et al. identify a random long-term pattern of biweekly global new COVID-19 cases with a seasonal feature. Most countries have co-integration linkages of newly reported instances of distinct varieties of concerns, regardless of their demographics or responses to the virus. The findings suggested that consistent techniques may be used to limit the spread. Furthermore, drastic eradication attempts may be ineffective, and there is a substantial risk that the COVID-19 pandemic may become an endemic. Janko et al. developed a framework to assist policymakers in developing plausible intervention tactics by dynamically changing non-pharmaceutical interventions using artificial intelligence to forecast the infection trends, aggregated the socioeconomic costs from the literature and expert knowledge, and used a multi-objective optimization algorithm to find and evaluate various non-pharmaceutical

intervention plans. The model generates efficient intervention plans to fight a pandemic and can evaluate their effect and costs.

Furthermore, we present articles in this research theme that deals with Hepatitis A/B virus modeling. In Sun et al. work focuses on a fractional-order differential equation model with time delay and logistic proliferation to better understand the transmission mechanism of the Hepatitis B virus in the human body. Their research findings revealed that immune response time delay and fractional order can substantially impact the dynamic behavior of the Hepatitis B virus infection transmission. As a result, while modeling and investigating Hepatitis B virus infection, temporal delay and fractional order should be considered. Jeong et al. used a flexible spatio-temporal model to analyze the spatio-temporal fluctuations of the hepatitis A virus in Korea and the influence of socioeconomic and weather-related parameters. To evaluate the effects of risk factors, the authors developed a Bayesian spatiotemporal zero-inflated Poisson regression model of weekly hepatitis A virus incidence in Korea. This is the first study to build a spatiotemporal model of hepatitis A virus occurrence in Korea, considering numerous socioeconomic parameters. The proposed model will be beneficial in forecasting, preventing, and regulating the spread of the hepatitis A virus.

Mpox disease modeling is also considered by some research presented by authors in this Research Topic. Ngungu et al. provided a brief overview of the Mpox virus and its transmission dynamics by investigating its spread and the effect of a non-pharmaceutical intervention (quarantine). The work provides insight into the exponential growth rate of the Mpox virus dynamics prediction and how to stop it from spreading and understand the effects of non-pharmaceutical intervention on infected individuals, which will guide how to deploy intervention resources to contain the disease's spread. Yuan et al. developed a Susceptible-Exposed-Infected-Recovered (SEIR) modeling framework to evaluate the impact of vaccination and other disease control strategies. The vaccination of a high-risk group and ring vaccination strategy, as well as testing, isolating patients and contact tracing, are all explored, as are various interventions during gathering occasions. The authors primarily focused on evaluating the effectiveness of public health control methods, such as preventive vaccination or immunization post-exposure, to assist public health decision-making by simulating scenarios of gatherings with varying numbers of attendance and levels of intervention. The research findings indicated that reactive ring vaccination may not be sufficient in and of itself; however, if close contacts of cases can be identified, vaccinated, and isolated, an outbreak following a mass gathering event may be avoided.

Lastly, other infectious disease modeling is considered by N'konzi et al. modified the basic deterministic Susceptible-Exposed-Infected-Recovered (SEIR) model to account for the effect of disease control measures as well as the feedback loop between non-pharmaceutical interventions adherence and disease dynamics. The model is used to study the impact of temporal fluctuations in non-pharmaceutical intervention adherence levels on infectious disease dissemination. To capture the dynamics of the public level of adherence to

non-pharmaceutical interventions, authors leverage on the simulation of disease dynamics and expand on the health belief model. The model implies that treatments aimed at enhancing non-pharmaceutical interventions adherence may be far more valuable than raising overall non-pharmaceutical interventions stringency.

Concluding remarks

All the contributions to the Research Topic on "*Mathematical and statistical modeling of infection and transmission dynamics of viral diseases*" have a focus on developing mathematical and statistical models that are well suited for real data, which are gathered through a process filtered by modeling constraints. The idea is to obtain the optimal model with strong predictive power that will match the collected data and can forecast the future evolution of the observed disease trends while applying the most appropriate and scalable inquiry techniques.

Author contributions

Funding

Acknowledgments

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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. COVID-19 Open Data Repository (2022). Available online at: https://health.google.com/covid-19/open-data/ (accessed August 29, 2022).

2. WHO COVID-19 Research Database (2022). Available online at: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/ (accessed August 29, 2022).

Check for updates

In-person school reopening and the spread of SARS-CoV-2 during the second wave in Spain

Raül Tormos[1,2]*, Pau Fonseca i Casas[3] and
Josep Maria Garcia-Alamino[4]

[1]Centre d'Estudis d'Opinió - Generalitat de Catalunya, Barcelona, Spain, [2]Department of Law and
Political Science, Open University of Catalonia, Barcelona, Spain, [3]Universitat Politecnica de
Catalunya, Barcelona, Spain, [4]Global Health, Gender and Society, Open University of Catalonia,
Barcelona, Spain

We investigate the effects of school reopening on the evolution of COVID-19 infections during the second wave in Spain studying both regional and age-group variation within an interrupted time-series design. Spain's 17 Autonomous Communities reopened schools at different moments in time during September 2020. We find that in-person school reopening correlates with a burst in infections in almost all those regions. Data from Spanish regions gives a further leverage: in some cases, pre-secondary and secondary education started at different dates. The analysis of those cases does not allow to conclude whether reopening one educational stage had an overall stronger impact than the other. To provide a plausible mechanism connecting school reopening with the burst in contagion, we study the Catalan case in more detail, scrutinizing the interrupted time-series patterns of infections among age-groups and the possible connections between them. The stark and sudden increase in contagion among older children (10−19) just after in-person school reopening appears to drag the evolution of other age-groups according to Granger causality. This might be taken as an indirect indication of household transmission from offspring to parents with important societal implications for the aggregate dynamics of infections.

KEYWORDS

COVID-19, SARS-CoV-2, in-person school reopening, non-pharmaceutical intervention, interrupted time-series analysis

Highlights

- Interrupted time-series analyses show that in-person school reopening precedes and correlates with a posterior growth in contagion in almost all Spanish regions that reopened at different moments in time during September 2020 in Spain.
- A more granular analysis of the dynamics of age-groups in the Spanish region of Catalonia indicates that infections among individuals aged 10–19 grew earlier and faster than the rest just after school-reopening, driving the evolution of other age-groups in a Granger causal process.

Introduction

In many countries, several non-pharmacological interventions (NPIs) to mitigate the spread of SARS-CoV-2 in the community have been tested and proved effective, such as online schooling, mandatory mask wearing or the closure of bars and restaurants (1–8). Some published studies focus on specific factors, like age (9–11), while others define more general models to identify the effects of non-pharmacological interventions (6), or the effects of the vaccination process on the population (12). A study of the pandemic situation in Catalonia proposes the use of a *Digital Twin* (13). In that study, a combination of a simulation and an optimization model through a continuous validation process allows understanding the effects of the different NPIs on the population by analyzing the change points brought about by new cases. Similarly, a study applying Bayesian inference to a type of epidemiological SIR model (14) analyzes the change points and infers the effects of different interventions on the evolution of new cases.

During the first wave of the pandemic in Spain (47.3 M inhabitants), and as a part of the lockdown, in-person schooling was shut down. Nonetheless, both the effectiveness and social consequences of in-person school closure remain a controversial issue. Different studies employing various approaches, for school facilities (15–21) and specifically for child care facilities (22), have tried to discern if schools are a vector for the propagation of the infection and whether children have an impact on that spread (23). Two other studies with a similar focus to ours at trying to estimate the effect of in-person school reopening arrived at opposing conclusions when analyzing the second pandemic wave in Italy: one found a link between school reopening and the resurgence of the virus (24), while the other did not (25). For other similar diseases, like influenza outbreaks, closure of in-person schooling has been an effective non-pharmaceutical intervention (26, 27). Being the spread of SARS-CoV-2 mainly airborne (28–30), knowing that to talk increases the transmission risk (31), and that the risk raises in poorly ventilated environments (32), it seems plausible that online schooling will reduce community transmission as compared to in-person schooling. Besides that, children seem to have equivalent nasopharyngeal viral loads to adults (16, 33–35), even though the youngest (ages 0–10) may have had lower susceptibility (36) therefore some studies suggest that the transmission is mainly in households (37), although other suggest that although they have lower susceptibility, the youngest ones are more infectious than older individuals (38). Therefore, the spread on schools would remain high if limited measures are applied to mitigate transmission (35, 39). These different evidences lead to the definition of several official advices and reports with the purpose to lessen viral outbreaks in schools in the context of in-person schooling (40, 41) with special focus on the Accumulated Incidence (AI) in the community.

In this paper we analyze the role of in-person school reopening in Spain on the evolution of infections. Using an interrupted time-series perspective, we explore and model the dynamics followed by the different Spanish regions, Autonomous Communities, that reopened schools at varying moments in time during September 2020. The impact of school reopening is understood in the models as an external shock or interruption to the series. The evidence points to a correlation between school reopening and a posterior outbreak in contagion across most ACs. We further provide a plausible causal mechanism for that association by studying the Catalan situation in more detail. For this case, we analyze actual data on the evolution of infections among the different age-groups and their interconnected dynamics, identifying some key sociological patterns. A sudden burst in contagions among school-age individuals (10–19) takes place just after in-person school reopening and appears to drag the dynamics of other age-groups. We argue that actual data may contain age-dependent measurement error. Therefore, we replicate our interrupted time-series analysis using corrections for measurement bias as a robustness test. We weight the actual data by the levels of prevalence by age-group as obtained from large-scale probability sample surveys (42–45). This reanalysis confirms our main findings, what constitutes a strong robustness test, and offers further light into additional phenomena overlooked in the official incidence rate records. The data sources we use are provided by the Open Data service of the Catalonia regional government (46), accessed through the Socrata connector (47), and the National Statistics Institute INE (48).

School reopening across Spanish regions

In Figure 1 we present the number of daily COVID-19 cases detected in each Autonomous Community (AC, from now on) from the 1st of January until November 11th of 2020 (49). The date of school reopening is indicated with a vertical red line, and a dotted red line shows the 14th day after reopening. This range of time corresponds to the most likely incubation period for a child who contracted the virus on the 1st day of reopening and used as the official quarantine period. Schools were scheduled to open in different dates during September depending on what the government of each AC had arranged. In the cases in which pre-secondary and secondary education did not start the same day, we used the opening of secondary education as older children are assumed to have a stronger capacity to infect others. The figure includes the cases of Ceuta and Melilla which are Spanish autonomous cities in the North of Africa.

In all but two cases we observe an exponential growth in contagion 14 days after school reopened in September. The two exceptions are the Madrid Community and the Canary Islands. In these two cases, the peak of the second wave
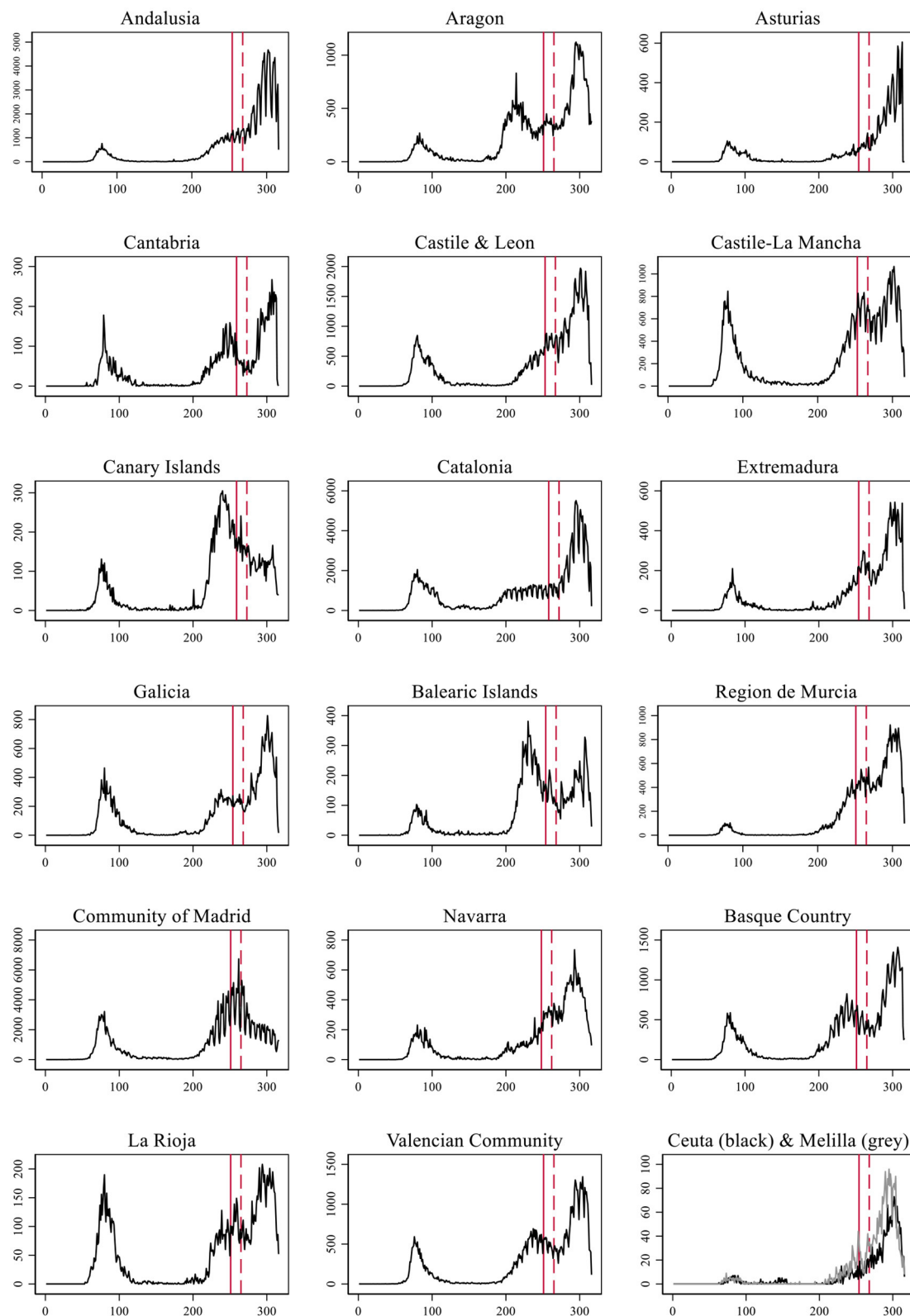
FIGURE 1
Evolution of the number of COVID-19 cases per day in each Autonomous Community of Spain.
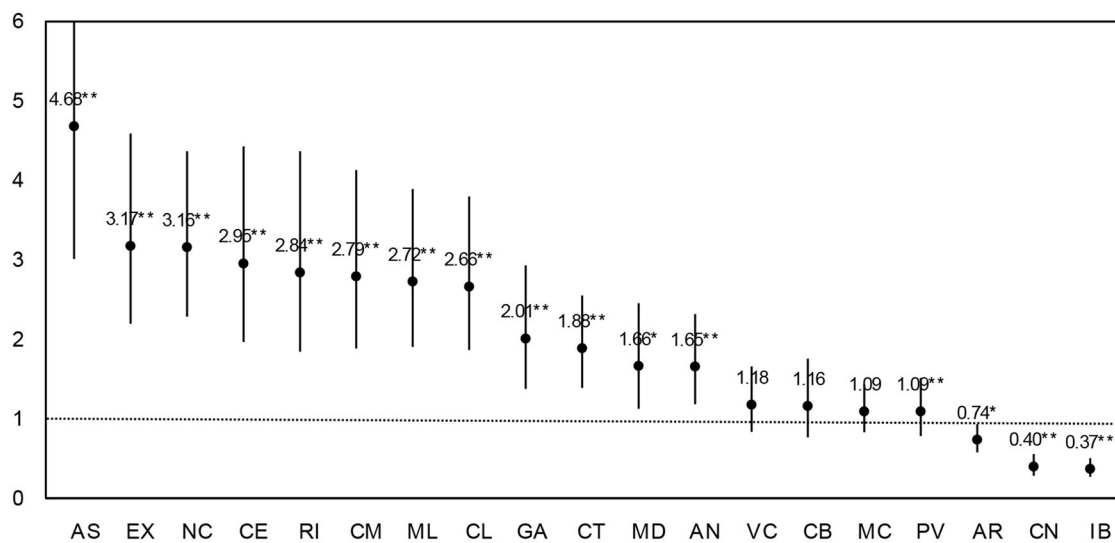
FIGURE 2
Effect of school reopening in each ACs. Estimates are incidence-rate ratios from Poisson regressions corrected for over dispersion. 95% CI.
**$p < 0.01$, *$p < 0.05$.

occurred before school reopening and containment measures were already applied previously. In the remaining 15 cases along with the two autonomous cities, the pattern is of an exponential growth. In six ACs the upsurge came after the second wave was being successfully contained, leading to a third wave: Cantabria, Castile-La Mancha, Extremadura, Balearic Islands, Basque Country, Region of Murcia, Navarra, Rioja, and Valencian Community. In four cases, the exponential growth came after a stationary situation: Andalusia, Aragon, Catalonia, and Galicia. In the four remaining cases, the number of daily contagions was already increasing before, but school reopening established the point where it definitively bursted.

Next, we perform a set of interrupted time-series Poisson regression models corrected for over dispersion, one for each AC, using the incidence rate as a dependent variable and having as predictors a linear trend (time) and a dummy variable representing an external shock to the series: in-person school reopening (the intervention), where 1 is the time-period with in-person classes and 0 otherwise. Therefore, in the models, the incidence rate (r) is assumed to be given by:

$$r = \log^{-1}(\alpha_0 + \beta_1\,(time) + \beta_2\,(intervention))$$
$$= e^{\alpha_0 + \beta_1\,(time) + \beta_2\,(intervention)} \qquad (1)$$

Full results of those regressions are presented in Supplementary Table 1. According to this modeling strategy and looking at the incidence-rate ratios (the change in the incidence rate due to the intervention) in Figure 2, school reopening implied a clear raise in the risk of contagion for the general population in all but three ACs (84% of ACs).

The exceptions were Aragon (AR), Cantabria (CN), and the Balearic Islands (IB). Asturias was the most affected AC. Reopening face-to-face classes increased 4.7 times the rate of infection in this region as compared to the period when schools were closed. In seven other ACs the rate of infection tripled (or almost) after the reopening: Extremadura, Navarra, Ceuta, Rioja, Castile-La Mancha, Melilla, and Castile and Leon. In other four cases the rate doubled or nearly: Galicia, Catalonia, Madrid, and Andalusia. In the remaining four ACs, the impact of reopening was still relevant implying an increase in cases in between 20 and 10% (Valencian Community, Cantabria, Murcia, and the Basque Country).

Next, we provide an estimate of the average effect of school reopening across ACs (r_{ij}) by using the pooled dataset of all ACs. We employ a panel data approach that conveniently accounts for the clustering of cases in geographical units (see Table 1). We run a random effects Poisson regression considering entity-specific intercepts for Autonomous Communities. The model contains a dummy variable for school reopening and a time trend, as shown in the following equation:

$$r_{it} = e^{\alpha_0 + \beta_1\,(time_{it}) + \beta_2\,(intervention_{it}) + u_i} \qquad (2)$$

For $i = 1, \ldots, 19$ ACs and $t = 1, \ldots, 320$ days observed. The random effects u_i are assumed to be normally distributed with mean 0 and variance σ_u^2.

On average, school reopening has a sizeable and significant statistical effect. Estimates for school reopening indicate a robust mean impact of school reopening on the incidence of daily

TABLE 1   Random effects Poisson regressions with entity-specific intercepts (ACs).

	IRRs
School reopening	1.760***
	(0.005)
Time (linear trend)	1.009***
	(0.000)
Constant	36.167***
	−8.699
Ln Alpha	0.15
Alpha	1.162
Log likelihood	−538,910.76
Observations	5,966

The estimates of Poisson models are incidence-rate ratios (IRRs).
Standard errors are in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

FIGURE 3
Ratio of the incidence-rate ratios of opening secondary education with respect to pre-secondary.

infections across ACs[1]. The incidence-rate ratio associated to school reopening implies a 76% average increase in risk for the population of becoming infected when comparing before and after reopening. Of course, with observational evidence, third variables temporally coinciding with school reopening, such as the return to work, could confound the association. Correlation does not imply causation, but neither precludes it.

Secondary and pre-secondary education (kindergarten and primary school) started at different moments in time in eight ACs. We can benefit from these naturally occurring phenomenon and use it to test whether the opening of any of the two educational stages had a stronger aggregate impact than the other. We run two separate Poisson regression models for each of these eight ACs, one using time and the date of pre-secondary education reopening as a predictor variable and the other employing the date of secondary education reopening instead. Detailed results are shown in Supplementary Table 3. In Figure 3, we present the ratio of the effects of opening secondary education with respect to pre-secondary (the ratio of the incidence rate ratio). In five out of eight cases (63%), there are almost no differences between coefficients. In the remaining three cases the opening of pre-secondary education had a stronger impact than the start of secondary education.

Indirect evidence of household transmission

To gain further insight as to which mechanisms may drive the outbreak of infections coinciding with in-person school

---

1   The estimates of an equivalent fixed effects model adjusting for time-invariant unobserved heterogeneity portrays the exact same estimates as the random effects model (see Supplementary Table 2).
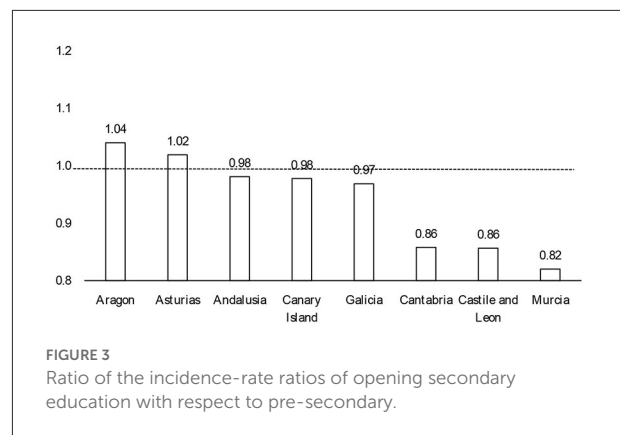
reopening, we study the Catalan case with more detailed data on age-groups (46). We explore the rate of infections per day within each age-group. Our hypothesis is that contagion inside family units with children might have been crucially boosted due to the school reopening. In aggregate terms, the return to in-person classes would have fostered a silent spread of the virus through the community with visible societal consequences 2 weeks later. Lacking direct measures on family units, we study the aggregate dynamics of infection in age-groups that might be involved. Individuals in their forties (40–49) are more likely to have children between the ages of 10 and 19 and live together with them (50). Using aggregate time series data, in the following analysis we show how these two age groups evolve similarly over time during the second wave of the pandemic, and that school reopening might be one main driver of the exponential growth in infections among children aged 10–19, dragging the evolution of older adults.

In Figure 4 we present the detailed development of the daily number of cases in Catalonia across the two pandemic waves observed. Again, we marked with a red line the moment of the reopening of schools, and with a dotted red line the passing of 14 days of the reopening.

In the first pandemic wave the number of cases observed per day was clearly less than the actual cases due to a lack of testing and plenty underreporting. A remarkable contention of the virus followed the strict lockdown that spanned from mid-March to June the 21st. Infections dropped to a minimum throughout July. During the end of June, cases started raising again but stabilized in a sort of plateau. A plausible explanation for this growth is a concurrent raise in testing during that period, instead of an actual increase in the number of cases (see Supplementary Figure 3).

Besides the growth in testing efforts, the surge in infections that lead to the summer plateau could also be connected to the reopening of bars and restaurants, a share of employees

FIGURE 4
Evolution of the number of COVID-19 cases per day in Catalonia.
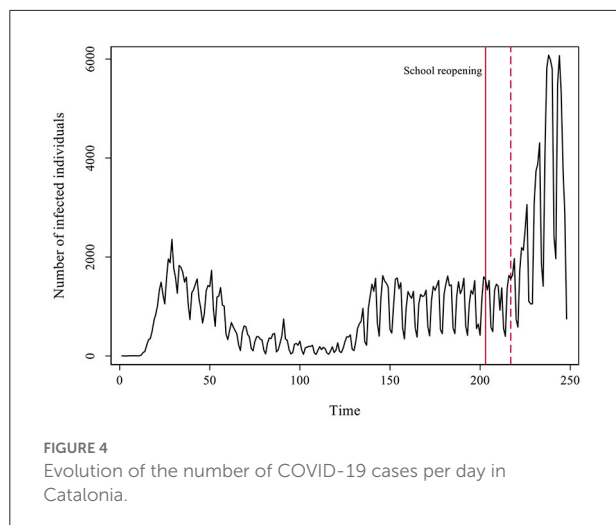
going back to work, and friends and family gatherings. In any case, during that plateau the dynamics was stationary (unit root DF test = −5.187 with a *p*-value = 0.000 for the period). The reopening of schools happened on September 15th and 14 days later a clear exponential growth in the number of cases took place. The co-occurrence in time of two phenomena does not prove causation. The increase in cases could have just temporally coincided with school reopening and be motivated by other factors instead. In any event, school reopening as a cause fulfills one of the rules of causality, that causes must temporally precede effects. Besides that, we already observed a similar pattern across many other ACs. In what follows, we analyze a plausible causal mechanism connecting in-person school reopening with the exponential surge in infections by studying the coevolution of age-groups involved in the process.

We implement a smoothing transformation of the time-series data for each age-group with a non-parametric procedure using locally weighted regressions[2], (see Supplementary Figure 1 to inspect the graphs with the original incidence count data). Figure 5 presents these estimates for all age-groups together, which help to visualize the patterns emerging from the data. The 10–19 age-group is the first experiencing an exponential growth just after school reopening following the plateau phase, and the one with a faster and larger increase in the rate of cases.

We focus on studying people in their forties (40–49 years old) as they are in a stage of the life cycle likely to have school children at home between the ages of 10 and 19[3]. After school reopening, in households where 40-year-olds and their offspring live together, contagion risk would be higher than in other family units. Ever since, not only parents could potentially infect their children but also vice versa. First, we compare the coevolution

_____

2   Using the lowess command in Stata.

3   According to official statistics (INE), in Catalonia the average age for a woman to have a child during the period from 2000 to 2019 was 31.13 years, the highest in Europe (48).

of youths between 10 and 19 years of age with people in their forties as well as with individuals in their thirties and fifties (Figure 6)[4]. Overall, these three older age-groups are somehow similar in terms of lifestyle and habits. They all loosely belong to the middle-aged category of the human life cycle, clearly differentiated from other life stages such as childhood, youth, or old age. They also portray a similar dynamic.

Our hypothesis regarding the mechanism that connects the evolution of the middle-aged with the 10- to 19-year-old individuals is that they live together in the same households, and the transmission from offspring to adults may have substantially increased due to in-person school reopening. If we compare the development of these two age segments over the period, we see that during the first wave of the pandemic both dynamics were uncorrelated. Middle-aged people got infected, but there were almost no cases (detected) among youths. Besides underreporting and a higher level of asymptomatic cases, schools were closed during the first wave. In the phase after the lockdown, cases among this younger group started to increase, but always remained at a lower level than middle-aged individuals. We consider people in their forties as clearly representative of the middle life stage of the life cycle and the age-group most likely to parent youngster in between 10 and 19 years of age. The plateau phase implied a stationary state for both groups (unit root DF test = −5.323 and −4.963, respectively, with a *p*-value = 0.000). At this stage, the higher level of infections among individuals in their forties could be related to going back to work, and other sort of gatherings. It could well be that, inside the household, contagion from parents to offspring was predominant at that moment. The opening of schools brings a stark increase in youth infections. Cases among younger people (10–19 years old) start rising before the growth among older adults (e.g., 40–49). If we compare a critical range of time, that between September 11th and the 7th of October, infections among youths were increasing faster than among their elders. OLS regressions with deterministic time trends yield a slope of 7.73 in the case of the young age-group and of 5.66 in the older one (full results not shown for simplicity). In fact, the steepest exponential growth of all age-groups takes place among individuals between 10 and 19 years (see Table 2).

The smoothing procedure helped us at visually appreciating how the increase in cases among young people aged 10–19 years preceded the subsequent increase in the group aged 40 to 49 and was larger in magnitude. With the aim of testing the role of the 10–19 age-group, after in-person school reopening, to drive the evolution of the pandemic during the second wave through their impact on people of other age-groups, especially middle-aged people, we now perform a set of time series tests using actual incidence rates. As a robustness check, we additionally

_____

4   Analyses including the other middle-aged groups are included in the Supplementary Figure 1.
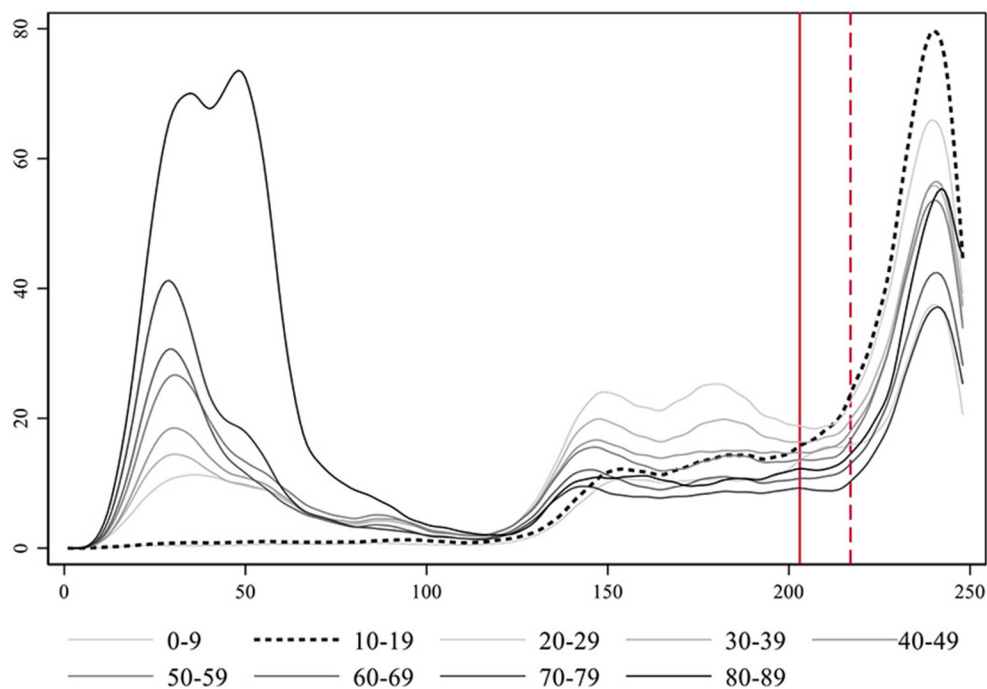
FIGURE 5
Smoothed estimates using locally weighted regression of the evolution of the rate of COVID-19 cases per day in Catalonia by age-groups.
Notice that the age-group 90+ is not represented for clarity purposes, due to its large incidence levels during the first wave.
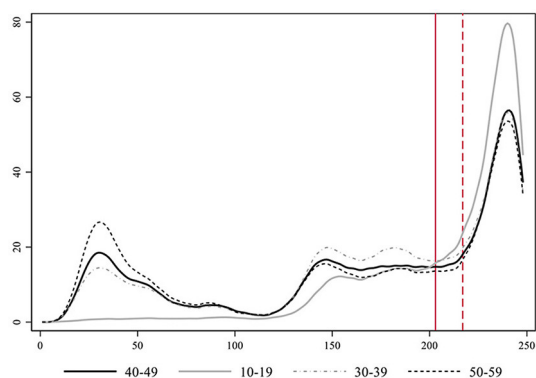
FIGURE 6
Smoothed estimates using locally weighted regression of the
evolution of the number of COVID-19 cases per day in
Catalonia by specific age-groups.

TABLE 2 Fitting deterministic linear and exponential time trends to
the evolution of the different age-groups over the second pandemic
wave in Catalonia.

	R-squared		
	Linear	**Exponential**	**Ratio**
0–9 years old	0.49	0.62	1.25
10–19 years old	0.45	0.76	1.69
20–29 years old	0.49	0.60	1.23
30–39 years old	0.49	0.59	1.21
40–49 years old	0.47	0.57	1.22
50–59 years old	0.46	0.54	1.18
60–69 years old	0.46	0.55	1.21
70–79 years old	0.44	0.54	1.23
80–89 years old	0.37	0.42	1.13

perform the same tests on a weighted version of the time-series
data adjusted for prevalence levels in each age-group.

Granger causality test

The Granger causality test (51) is a time-series procedure
to verify if the evolution of one time series is able to predict
another time series. Table 3 shows a group of Granger causality
tests to evaluate the effect of the 10–19 age-group series on the
40–49 age-group series. It presents a set of nested OLS regression
models with the 40–49 age-group series as the dependent
variable and the lagged dependent variable (with up to 10 lags)
and the 10–19 age-group variable (also with up to 10 lags) as
independent variables. This specification can be expressed using

the following equation:

$$y_t = \alpha_0 + (\beta_1 y_{t-1} + \cdots + \beta_{10} y_{t-10}) + (\beta_{11} x_{t-1} + \cdots + \beta_{20} x_{t-10}) + u_t \quad (3)$$

where the level of infections in the 40–49 age-group y at time t is a function of a constant α_0; the lagged dependent variable, in up to 10 consecutive lags ($\beta_1 y_{t-1}, \ldots, \beta_{10} y_{t-10}$); and the lagged independent variable: the level of infections in the 10–19 age-group in up to 10 consecutive lags ($\beta_{11} x_{t-1}, \ldots, \beta_{20} x_{t-10}$). The term u_t is the error term of the time-series regression.

We are interested in the F-statistic of the models that will eventually allow us to reject the null hypothesis. We reject the null in eight of the 10 models. Only with lags one and two the F-statistic is below the critical threshold. This has a substantive meaning: it takes longer than one or two single lags for the dynamics of the 10–19 age-group series to influence the 40–49 series. In the remaining models with more lags, the p-value associated to the F-statistic is always under 0.05 ($p < 0.000$) indicating that we can reject the null hypothesis that all coefficients of lag of the independent variable (10–19 age group series) are equal to 0. Therefore, we can state that the 10–19 age-group Granger causes the 40–49 age-group series.

Instead of this stream of causality from children to adults, could the level of parental infections be driving the level of infections of their offspring? To test it, we reverted the former Granger causality analysis so that the 10–19 age-group series is now the dependent variable (y_t) and the 40–49 series the independent variable (x_{t-n}). This would allow us checking whether there is a sort of reverse process by which the 40-year-olds are those who cause youths to get infected. As shown in Table 4, there is also evidence of this line of causation, but it is substantially weaker. In only three of the 10 models, we observe a Granger causal process. In any case, a bidirectional association among both series is consistent with the notion of a feedback relationship due to cohabitation of these age-groups in the same family units within households.

Chow test

In addition, we may want to verify when this relationship between the two time-series appears. We perform a test to check whether the opening of schools, as an external shock, implies a key disruption in the series under study here (Table 5). The Chow test is calculated after an OLS regression with the lagged dependent variable and the lagged independent variable as regressors together with the interaction of school reopening with both age-group series. The equation can be portrayed as follows:

$$y_t = \alpha_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \beta_3 z + \beta_4 (y_{t-1} \cdot z) + \beta_5 (x_{t-1} \cdot z) + u_t \quad (4)$$

where the level of infections in the 40–49 age-group y at time t is a function of a constant α_0, the lagged dependent variable

$\beta_1 y_{t-1}$), the lagged independent variable (the level of infection of the 10–19 age-group expressed by $\beta_2 x_{t-1}$), a dummy variable representing school reopening $\beta_3 z$, the interaction of school reopening with the lagged dependent variable $\beta_4 (y_{t-1} \cdot z)$ and the lagged independent variable $\beta_5 (x_{t-1} \cdot z)$. The term u_t is the error term of the time-series regression.

The null hypothesis for the Chow test means no break. If the p-value is <0.05, we can reject the null in favor of the alternative that there is a break. Our results indicate that the null hypothesis can be rejected, and we can conclude that school reopening caused a break in the regression coefficients.

Weighting by prevalence as a robustness check

The use of actual incidence records involves assuming that measurement error does not substantially distort our inferences. The proportion of asymptomatic cases is a key aspect to understand the pattern of the SARS-CoV-2 spread. Previous research (52) establishes that almost 60% of infected people report no symptomatology during an early stage of the disease, although symptoms can appear later as a result of being tested in the presymptomatic phase (53). This serves as a basis to discuss regarding the proportionality of the diagnostic effort done in all the age-groups, and if this can affect the analysis. Some studies suggested that the age range from 0 to 20 is highly asymptomatic (54). Moreover, other analyses seem to show that the prevalence on children is higher than previously thought (55), being prevalence a good estimator for capturing the true incidence on the population. Therefore, raw incidence data certainly contains statistical biases due to non-random factors such as the degree of asymptomatic individuals, which vary by age-group, or differences in diagnosis efforts on each age-segment of the population. In contrast to official incidence records, prevalence studies are implemented using random sampling, which allows obtaining more representative and realistic incidence estimates by age-groups. When data does not come from a random sample, as in the official records of infected individuals, it is susceptible of containing systematic error from the self-selection of symptomatic infected individuals that correlates with aging, or to over represent certain population segments for whom public diagnosis efforts are higher, such as younger individuals, but using less representative sampling procedures. The number of tests done for the age group from 10 to 19 is huge as compared to those performed on other age groups (see Supplementary Figure 3). However, this does not imply an improvement in detection, since it depends on the method used to perform the testing, and on whether the samples are correctly selected. As an example, the tests performed in a classroom typically composed of 25 students due to the detection of a positive index case will result in largely negative tests results. This is because only about 8% of infective individuals are

TABLE 3 Granger causality test for the 40−49 age-group series as dependent variable.

	(1) 40−49	(2) 40−49	(3) 40−49	(4) 40−49	(5) 40−49	(6) 40−49	(7) 40−49	(8) 40−49	(9) 40−49	(10) 40−49
L1.40−49	0.674***	0.579***	0.2	0.255*	0.342***	0.209*	0.152	0.307***	0.344***	0.213*
	(0.115)	(0.142)	(0.134)	(0.141)	(0.129)	(0.122)	(0.113)	(0.11)	(0.11)	(0.115)
L2.40−49		0.221	0.674***	0.717***	0.53***	0.603***	0.431***	0.361***	0.396***	0.34***
		(0.137)	(0.136)	(0.143)	(0.136)	(0.126)	(0.116)	(0.1)	(0.112)	(0.115)
L3.40−49			0.04	−0.029	−0.042	−0.203	−0.131	−0.022	−0.024	0.219*
			(0.123)	(0.147)	(0.139)	(0.131)	(0.12)	(0.104)	(0.098)	(0.115)
L4.40-49				−0.051	0.471***	0.289**	0.148	0.045	0.108	0.133
				(0.124)	(0.134)	(0.129)	(0.119)	(0.103)	(0.098)	(0.096)
L5.40−49					−0.514***	−0.142	−0.331***	−0.163	−0.223**	−0.215**
					(0.114)	(0.127)	(0.118)	(0.103)	(0.098)	(0.096)
L6.40−49						0.051	−0.02	0.173*	0.221**	0.217**
						(0.118)	(0.117)	(0.103)	(0.098)	(0.096)
L7.40−49							0.664***	0.513***	0.632***	0.645***
							(0.108)	(0.1)	(0.098)	(0.096)
L8.40−49								−0.349***	−0.448***	−0.274**
								(0.106)	(0.113)	(0.121)
L9.40−49									−0.16	−0.08
									(0.106)	(0.116)
L10.40−49										−0.407***
										(0.105)
L1.10−19	0.089	0.11	0.408***	0.397***	0.34***	0.348***	0.292***	0.314***	0.102	0.185*
	(0.075)	(0.111)	(0.104)	(0.105)	(0.096)	(0.095)	(0.093)	(0.086)	(0.096)	(0.096)
L2.10−19		−0.096	−0.822***	−0.876***	−0.858***	−0.825***	−0.58***	−0.435***	−0.236**	−0.256**
		(0.102)	(0.12)	(0.127)	(0.116)	(0.108)	(0.103)	(0.091)	(0.104)	(0.114)
L3.10−19			0.413***	0.525***	0.638***	0.563***	0.411***	0.279***	0.281***	0.183*
			(0.089)	(0.131)	(0.127)	(0.118)	(0.109)	(0.095)	(0.09)	(0.104)
L4.10−19				−0.038	−0.625***	−0.314**	−0.222**	−0.161*	−0.186**	−0.201**
				(0.094)	(0.123)	(0.123)	(0.112)	(0.096)	(0.091)	(0.089)
L5.10−19					0.603***	0.036	0.207*	0.199**	0.214**	0.198**
					(0.086)	(0.12)	(0.113)	(0.097)	(0.092)	(0.09)
L6.10−19						0.295***	0.157	−0.113	−0.083	−0.075
						(0.092)	(0.109)	(0.098)	(0.093)	(0.09)
L7.10−19							−0.224**	0.166*	0.021	0.021
							(0.091)	(0.095)	(0.093)	(0.091)
L8.10−19								−0.186**	0.166	0.048
								(0.085)	(0.113)	(0.116)
L9.10−19									−0.213**	−0.192
									(0.094)	(0.122)
L10.10−19										0.189**
										(0.094)
Constant	3.897***	3.037***	1.546	1.734	2.126**	1.903**	1.26	1.416*	1.558**	1.921***
	(1.057)	(1.169)	(1.051)	(1.082)	(1.009)	(0.949)	(0.871)	(0.754)	(0.721)	(0.713)
Observations	259	258	257	256	255	254	253	252	251	250
R-squared	0.648	0.652	0.743	0.744	0.79	0.822	0.856	0.895	0.907	0.913
Granger test										
F	1.41	0.56	16.50	12.13	21.93	15.33	6.89	4.94	3.67	3.47
Sig.	0.236	0.572	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Standard errors are in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

TABLE 4 Granger causality test for the 10−19 age−group series as dependent variable.

	(1) 10–19	(2) 10–19	(3) 10–19	(4) 10–19	(5) 10–19	(6) 10–19	(7) 10–19	(8) 10–19	(9) 10–19	(10) 10–19
L1.10–19	0.726***	0.808***	1.119***	1.097***	0.992***	0.943***	0.771***	0.943***	0.782***	0.845***
	(0.096)	(0.143)	(0.14)	(0.141)	(0.124)	(0.115)	(0.118)	(0.099)	(0.115)	(0.117)
L2.10–19		−0.041	−0.878***	−0.962***	−0.944***	−0.862***	−0.631***	−0.374***	−0.203	−0.178
		(0.132)	(0.16)	(0.17)	(0.149)	(0.131)	(0.131)	(0.104)	(0.125)	(0.139)
L3.10–19			0.528***	0.624***	0.691***	0.541***	0.41***	0.212*	0.225**	0.1
			(0.119)	(0.176)	(0.163)	(0.143)	(0.138)	(0.108)	(0.108)	(0.127)
L4.10–19				0.045	−0.727***	−0.231	−0.178	−0.086	−0.104	−0.125
				(0.125)	(0.159)	(0.149)	(0.142)	(0.11)	(0.109)	(0.109)
L5.10–19					0.966***	−0.022	0.192	0.195*	0.208*	0.201*
					(0.111)	(0.145)	(0.143)	(0.111)	(0.11)	(0.11)
L6.10–19						0.636***	0.291**	−0.111	−0.091	−0.091
						(0.112)	(0.138)	(0.112)	(0.111)	(0.11)
L7.10–19							0.12	0.847***	0.765***	0.764***
							(0.115)	(0.109)	(0.112)	(0.111)
L8.10–19								−0.636***	−0.375***	−0.458***
								(0.098)	(0.136)	(0.142)
L9.10–19									−0.223**	−0.272*
									(0.112)	(0.149)
L10.10–19										0.23**
										(0.115)
L1.40–49	0.225	0.205	−0.19	−0.117	0.027	−0.153	−0.126	−0.108	−0.044	−0.134
	(0.147)	(0.183)	(0.18)	(0.188)	(0.166)	(0.148)	(0.143)	(0.125)	(0.132)	(0.14)
L2.40–49		−0.061	0.488***	0.573***	0.368**	0.498***	0.322**	0.199*	0.149	0.072
		(0.176)	(0.182)	(0.191)	(0.175)	(0.152)	(0.148)	(0.115)	(0.134)	(0.141)
L3.40–49			−0.056	−0.08	0.002	−0.253	−0.146	0.051	0.038	0.263*
			(0.165)	(0.196)	(0.18)	(0.158)	(0.151)	(0.119)	(0.118)	(0.14)
L4.40–49				−0.193	0.568***	0.317**	0.145	−0.038	0.009	0.035
				(0.166)	(0.173)	(0.155)	(0.151)	(0.118)	(0.118)	(0.117)
L5.40–49					−1.001***	−0.308**	−0.451***	−0.176	−0.22*	−0.227*
					(0.147)	(0.154)	(0.149)	(0.118)	(0.118)	(0.118)
L6.40–49						−0.141	−0.032	0.193	0.225*	0.234**
						(0.142)	(0.148)	(0.118)	(0.118)	(0.118)
L7.40–49							0.309**	−0.063	−0.012	−0.01
							(0.137)	(0.115)	(0.118)	(0.117)
L8.40–49								−0.082	−0.197	−0.068
								(0.122)	(0.136)	(0.147)
L9.40–49									0.027	0.137
									(0.127)	(0.142)
L10.40–49										−0.377***
										(0.128)
Constant	0.564	1.252	−0.273	0.127	1.131	0.987	0.635	0.733	0.768	1.086
	(1.354)	(1.503)	(1.407)	(1.447)	(1.301)	(1.147)	(1.104)	(0.864)	(0.865)	(0.869)
Observations	259	258	257	256	255	254	253	252	251	250
R–squared	0.753	0.754	0.804	0.806	0.852	0.89	0.902	0.942	0.944	0.946
Granger tests										
F	2.34	0.67	2.59	2.41	10.89	4.04	2.91	1.10	1.20	1.93
Sig.	0.127	0.510	0.053	0.050	0.000	0.001	0.001	0.366	0.295	0.043

Standard errors are in parentheses.
***p < 0.01, **p < 0.05, *p < 0.1.

TABLE 5 Chow test of school reopening.

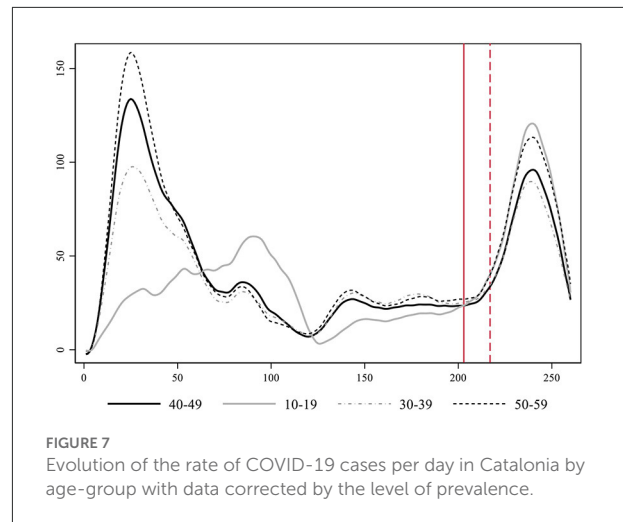|	Coef.	SE
Constant	2.71*	1.196
40–49 years old		
Lag 1	0.721**	0.151
10–19 years old		
Lag 1	0.028	0.151
School opening	10.42**	2.732
School * 40–49	0.104	0.261
School * 10–19	−0.161	0.213
Observations	259	
R-squared	0.676	
Chow test		
F	7.230	
P-value	0.000	

**p < 0.01, *p < 0.05.

responsible for 60% of the cases (56), and highly asymptomatic individuals are less infectious (57).

As a robustness test of our main analysis, we use data from large scale prevalence studies in Spain (42–45) to weight the actual incidence records and try compensating for the aforementioned biases. This robustness test implies weighting each age-group's time-series by their specific percentual level of detection before reanalyzing the data (see the explanation of the calculation procedure on the Supplementary material).

In Figure 7 we present the smoothed estimates of the incidence rates weighted by prevalence. Again, we can clearly identify how the 10–19 series grows exponentially faster and more intensely than the middle-aged series just after school reopening during the second wave. The increase among the 50-year-old individuals becomes now the second in importance. Furthermore, weighted data allows appreciating a more realistic estimate of the true overall magnitude of the first wave, which was far wider than the second. In any case, incidence among youths (10–19) during the first wave was rather low coinciding with a period when in-person school was closed.

The 10–19 age-group moves from being the second least infected group during the first wave (just after children between 0 and 9 years of age as shown in Supplementary Figure 5) to be leading the levels of contagion during the second wave, both in terms of the timing an intensity of its growth. The key exogenous contextual element that varies between these two waves and may be responsible for this difference is in-person schooling. Contagion among youths related to in-person classes does not necessarily translate into a life-threatening health risk for this age-group, however, it increases the danger in aggregate terms for the transmission to individuals of older age-groups across society at large, and especially within the multigenerational households where these young people live. Figure 7 further allows appreciating an overlooked increased in contagion of youths taking place after the first wave, when containment

FIGURE 7
Evolution of the rate of COVID-19 cases per day in Catalonia by age-group with data corrected by the level of prevalence.

measures were slightly relaxed. This growth finally went down during the summer months. If we rerun the Granger causality tests specified with the 40–49 series as determined by the 10–19 and up to 10 lags, we obtain evidence of Granger causality processes in three of its lags (see Supplementary Table 3). All in all, the replication of the analysis on the weighted time-series increases our confidence in the validity of our main results.

Discussion

In-person school reopening taking place at different dates during September 2020 precedes and correlates with a posterior growth in contagion in almost all Spanish regions. The time-series analysis of Catalan age-groups indicates that contagion among young individuals aged 10–19 after school-reopening grows earlier and faster than the rest, Granger causing the evolution of other age-groups. The lack of public awareness of this phenomenon might be due to a collective cognitive confusion regarding the actual role of school reopening. Different studies at the individual level show that children become less infected and are less infectious than older individuals (58). From this fact many conclude that it was rather safe to keep schools opened. This inference could actually be a sort of fallacy. Even if children are less infectious and get less infected (some estimates say by half), it has been proven they are infectious and infect. Moreover, even a smaller proportion of infected individuals can imply a large number of actual cases when the target population is very large: the whole school children population. As a result, the aggregate role of opening schools for the expansion of the pandemic can be collectively underestimated.

Like almost all the statistical tests, there is strong evidence for the correlation of the different effects presented here, but the causation if hidden. Granger causality is a statistical hypothesis test for determining whether one time series can forecast another one. Notice that it is only capable of testing the temporal

relation between the two time-series, since the true causation is a complex philosophical issue, here we can only assess if one time-series forecasts another time-series.

To test the robustness of our findings we implemented a weighting procedure based on prevalence studies to estimate the actual percent of detection. This allows us to generate a new time-series that represents the real cases. Reproducing our analysis with this corrected data yield equivalent results, enforcing our confidence in the findings.

Other studies have also analyzed the impact of school closures together with other non-pharmacological interventions. These studies employed large datasets that included multiple countries and various non-pharmacological interventions, and in all of them it was observed that the closure of schools provided a reduction in the R_t (6–8, 59). School reopening seems to have an impact on the R_t when this non-pharmaceutical intervention is lifted and applied, and coherently with the mortality (21), as is described on (60).

Furthermore, another study (61), using a methodology similar to that employed in our research, observed that parental exposure to open schools is associated with a somewhat higher rate of PCR-confirmed SARS-CoV-2 infection OR 1.17; CI 95% 1.03–1.32. It was also higher among teachers, PCR-confirmed SARS-CoV-2 infection OR 2.01; CI 95% 1.52–2.67.

In addition, a different research (62) robustly estimated that the closure of schools, like other interventions to reduce contacts in large groups, is one of the most effective interventions to contain the spread of COVID-19 by reducing the daily incidence.

While previous research has identified the overall impact of different non-pharmacological interventions in the reduction of SAR-CoV-2 spread, our study focuses more in depth on one of those interventions (school closure/reopening), in a specific context (Spain and Catalonia), at a particular moment in time (the second wave) and using an interrupted time-series approach. Our method can be easily reproduced in other countries to eventually find comparable patterns.

From our analysis we can contemplate the possibility that school reopening may generate a retro-feedback with parents' return to work and social activity, leading to an exponential growth, as observed in Catalonia and other Spanish ACs during September and October of 2020.

Despite its cost, online or hybrid schooling could have been a cost-effective option considering the potential role of schools as drivers of the virus in the community. The spread of the virus may imply higher expenses when medical, economic, and social costs of closing economic activities due to the arrival of a new viral wave are contemplated altogether. This understanding could help policy makers to find suitable solutions to limit the spread of the virus in the community such as using tele-education while keeping onsite schools for parents that need it, improving the ventilation of classes with HEPA filters, or reducing the ratios for onsite school.

Posterior virus variants, such as the B.1.1.7 detected in the UK, seem to increase the transmission rate among children. If this is confirmed, new analysis should be performed to assess how it will amplify the transmission rate in the community. The estimated effects of school reopening would constitute a downward estimation of the real impact in a context where new variants are widespread.

All in all, the findings presented here are consequential not only for the particular case of study, but more generally. Heated debates about the adequacy and safety of in-person school reopening have been held around the world. Different considerations regarding its costs and benefits have been casted, however, the full implications of its costs might not been weighted accurately enough. We believe our findings constitute a contribution in this direction.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: Generalitat de Catalunya, Registre de casos de COVID-19 realitzats a Catalunya. Segregació per sexe i edat. Dades Obertes (2020; February 21, 2021) Tyler, Socrata Open Data Server. Socrata (2020; June 2, 2020). CNE, Incidencias acumuladas e indicadores de transmisibilidad (2020; February 21, 2021).

Author contributions

RT: statistical modeling and analysis and scientific writing. PF: data preparation and literature review. JG-A: review and additional help. All authors contributed to the article and approved the submitted version.

Funding

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. 2022.990277/full#supplementary-material

References

1. Matzinger P, Skinner J. Strong impact of closing schools, closing bars and wearing masks during the Covid-19 pandemic: results from a simple and revealing analysis. *medRxiv*. (2020) 2020:20202457. doi: 10.1101/2020.09.26.20202457

2. Andrejko KL, Pry JM, Myers JF, Fukui N, DeGuzman JL, Openshaw J, et al. Effectiveness of face mask or respirator use in indoor public settings for prevention of SARS-CoV-2 infection — California, February–December 2021. *Morb Mortal Wkly Rep*. (2022) 71:212–6. doi: 10.15585/mmwr.mm7106e1

3. Huang J, Fisher BT, Tam V, Wang Z, Song L, Shi J, et al. *The Effectiveness of Government Masking Mandates On COVID-19 County-Level Case Incidence Across the United States, 2020*. (2022). Available online at: https://doi.org/10.1377/hlthaff. 2021.01072 (accessed February 28, 2022).

4. Leech G, Rogers-Smith C, Monrad JT, Sandbrink JB, Snodin B, Zinkov R, et al. Mask wearing in community settings reduces SARS-CoV-2 transmission. *Proc Natl Acad Sci USA*. (2022) 119:e2119266119. doi: 10.1073/pnas.2119266119

5. Gettings J, Czarnik M, Morris E, Haller E, Thompson-Paul AM, Rasberry C, et al. Mask use and ventilation improvements to reduce COVID-19 incidence in elementary schools — Georgia, November 16–December 11, 2020. *Morb Mortal Wkly Rep*. (2021) 70:779–84. doi: 10.15585/mmwr.mm7021e1

6. Li Y, Bpt K, Mbchb H, Mbbs N, Wang X, Nair H, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis*. (2020). doi: 10.1016/S1473-3099(20)30785-4

7. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. *Science*. (2020) 9338:eabd9338. doi: 10.1126/science.abd9338

8. Liu Y, Morgenstern C, Kelly J, Lowe R, Jit M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Med*. (2021) 19:40. doi: 10.1186/s12916-020-01872-8

9. Oshinubi K, Al-Awadhi F, Rachdi M, Demongeot J. Data analysis and forecasting of COVID-19 pandemic in Kuwait based on daily observation and basic reproduction number dynamics Kuwait. *J Sci*. (2021) 2021:1–30. doi: 10.1101/2021.07.24.21261059

10. Oshinubi K, Buhamra SS, Al-Kandari M, Waku J, Rachdi M, Demongeot J. Age dependent epidemic modeling of COVID-19 outbreak in Kuwait, France, and Cameroon. *Healthc*. (2022) 10:482. doi: 10.3390/healthcare10030482

11. Seligmann H, Vuillerme N, Demongeot J. Unpredictable, counter-intuitive geoclimatic and demographic correlations of covid-19 spread rates. *Biology*. (2021) 10:623. doi: 10.3390/biology10070623

12. Griette Q, Demongeot J, Magal P. What can we learn from COVID-19 data by using epidemic models with unidentified infectious cases? *Math Biosci Eng*. (2022) 19:537–94. doi: 10.3934/mbe.2022025

13. Fonseca P, Garcia J, García V, Pi X. Sars-cov-2 spread forecast dynamic model validation thorough digital twin approach, catalonia case study. *Mathematics*. (2021) 9:1–17. doi: 10.3390/math9141660

14. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*. (2020) 9789:eabb9789. doi: 10.1126/science.abb9789

15. Viner RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc Heal*. (2020) 4:397–404. doi: 10.1016/S2352-4642(20)30095-X

16. Davies NG, Klepac P, Liu Y, Prem K, Pearson AB, Quilty BJ, 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

17. Garcia-Alamino JM, Tobías A. Incidencia de infección por SARS-CoV-2 en la comunidad y su impacto en la primera semana de reapertura de colegios en Cataluña. *Gac. Sanit*. (2021) 35:302–3. doi: 10.1016/j.gaceta.2020.09.007

18. Kim S, Kim YJ, Peck KR, Jung E. School opening delay effect on transmission dynamics of Coronavirus disease 2019 in Korea: based on mathematical modeling and simulation study. *J Korean Med Sci*. (2020) 35:1–9. doi: 10.3346/jkms.2020.35.e143

19. Leeb RT, Price S, Sliwa S, Kimball A, Szucs L, Caruso E, et al. COVID-19 trends among school-aged children - United States, March 1-September 19, 2020. *Morb Mortal Wkly Rep*. (2020) 69:1–6. doi: 10.15585/mmwr.mm6939e2

20. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. (2020) 584:7. doi: 10.1038/s41586-020-2405-7

21. Auger KA, Shah SS, Richardson T, Hartley D, Hall M, Warniment A, et al. Association between statewide school closure and COVID-19 incidence and mortality in the US. *J Am Med Assoc*. (2020) 45229:859–70. doi: 10.1001/jama.2020.14348

22. Lopez AS, Hill M, Antezano J, Vilven D, Rutner T, Bogdanow L, et al. Transmission dynamics of COVID-19 outbreaks associated with child care facilities — Salt Lake City, Utah, April–July 2020. *Morb Mortal Wkly Rep*. (2020) 69:1319–23. doi: 10.15585/mmwr.mm6937e3

23. Hyde Z. COVID-19, children, and schools: overlooked and at risk. *Med J Aust*. (2020) 1:50934. doi: 10.5694/mja2.50934

24. Casini L, Roccetti M. Reopening Italy's schools in September 2020: a Bayesian estimation of the change in the growth rate of new SARS-CoV-2 cases. *BMJ Open*. (2021) 11:e051458. doi: 10.1136/bmjopen-2021-051458

25. Gandini S, Rainisio M, Iannuzzo ML, Bellerba F, Cecconi F, Scorrano L, et al. Cross-sectional and prospective cohort study of the role of schools in the SARS-CoV-2 second wave in Italy. *Lancet Reg Heal Eur*. (2021) 5:100092. doi: 10.1016/j.lanepe.2021.100092

26. Jackson C, Vynnycky E, Hawker J, Olowokure B, Mangtani P. School closures and influenza: systematic review of epidemiological studies. *BMJ Open*. (2013) 3:2149. doi: 10.1136/bmjopen-2012-002149

27. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, et al. Closure of schools during an influenza pandemic. *Lancet Infect Dis*. (2009) 9:473–81. doi: 10.1016/S1473-3099(09)70176-8

28. Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc Natl Acad Sci USA*. (2020) 117:14857–63. doi: 10.4324/9781003141402-3

29. Morawska L, Cao J. Airborne transmission of SARS-CoV-2: the world should face the reality. *Environ Int*. (2020) 139:105730. doi: 10.1016/j.envint.2020.1 05730

30. Tang S, Mao Y, Jones RM, Tan Q, Ji JS, Li N, et al. Aerosol transmission of SARS-CoV-2? Evidence, prevention and control. *Environ Int*. (2020) 144:106039. doi: 10.1016/j.envint.2020.106039

31. Asadi S, Wexler AS, Cappa CD, Barreda S, Bouvier NM, Ristenpart WD. Aerosol emission and superemission during human speech increase with voice loudness. *Sci Rep*. (2019) 9:1–10. doi: 10.1038/s41598-019-38808-z

32. Miller SL, Nazaroff WW, Jimenez JL, Boerstra A, Buonanno G, Dancer SJ, et al. Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor Air*. (2020) 2020:27. doi: 10.1101/2020.06.15.20132027

33. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*. (2021) 80:73. doi: 10.1126/science.abi5273

34. Heald-Sargent T, Muller WJ, Zheng X, Rippe J, Patel AB, Kociolek LK. Age-related differences in nasopharyngeal severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) levels in patients with mild to moderate coronavirus disease 2019 (COVID-19). *J Am Med Assoc Pediatr*. (2020) 174:902. doi: 10.1001/jamapediatrics.2020.3651

35. Fateh-Moghadam P, Battisti L, Molinaro S, Fontanari S, Dallago G, Binkin N, et al. Contact tracing during Phase I of the COVID-19 pandemic in the Province of Trento, Italy: key findings and recommendations. *medRxiv*. (2020) 2020:20127357. doi: 10.1101/2020.07.16.20127357

36. Viner RM, Mytton OT, Bonell C, Melendez-Torres GJ, Ward J, Hudson L, et al. Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis. *J Am Med Assoc Pediatr*. (2020) 2020:1–14. doi: 10.1101/2020.05.20.20108126

37. Lachassinne E, de Pontual L, Caseris M, Lorrot M, Guilluy C, Naud A, et al. SARS-CoV-2 transmission among children and staff in daycare centres during a nationwide lockdown in France: a cross-sectional, multicentre, seroprevalence study. *Lancet Child Adolesc Heal*. (2021) 4642:1–9. doi: 10.1016/S2352-4642(21)00024-9

38. Li F, Li YY, Liu MJ, Fang LQ, Dean NE, Wong GWK, et al. Household transmission of SARS-CoV-2 and risk factors for susceptibility and infectivity in Wuhan: a retrospective observational study. *Lancet Infect Dis*. (2021) 3099:1–11. doi: 10.1016/S1473-3099(20)30981-6

39. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. *J Infect Dis*. (2020) 2020:1–8. doi: 10.1101/2020.07.19.20157362

40. UNESCO. *Framework for Reopening Schools*. (2020).

41. Jones AJE, Young A, Clevenger K, Salimifard P, Wu E, Luna ML, et al. Healthy schools: risk reduction strategies for reopening schools. *Harvard TH Chan Sch Public Heal Heal Build Progr*. (2020).

42. Ministerio de Sanidad. *ESTUDIO ENE-COVID19: PRIMERA RONDA Estud Nac SERO-EPIDEMIOLOGÍA LA Infecc POR SARS-COV-2 EN ESPAÑA*. Madrid (2013). p. 3.

43. Ministerio de Sanidad. *ESTUDIO ENE-COVID19: SEGUNDA RONDA Estud Nac SERO-EPIDEMIOLOGÍA LA Infecc POR SARS-COV-2 EN ESPAÑA, 2–7*. Madrid (2020).

44. Ministerio de Sanidad. *ESTUDIO ENE-COVID: INFORME FINAL Estud Nac SERO-EPIDEMIOLOGÍA LA Infecc POR SARS-COV-2 EN ESPAÑA, 32*. Madrid (2020).

45. Ministerio de Sanidad. *ESTUDIO ENE-COVID: CUARTA RONDA Estud Nac SERO-EPIDEMIOLOGÍA LA Infecc POR SARS-COV-2 EN ESPAÑA, 2–7*. Madrid (2020).

46. Generalitat de Catalunya. *Registre de casos de COVID-19 realitzats a Catalunya Segregació per sexe i edat Dades Obertes*. Madrid (2020).

47. Tyler. *Socrata Open Data Server*. Seattle, WA: Socrata (2020).

48. INE. *Edad Media a la Maternidad por provincia, según orden del nacimiento*. Madrid: Indicadores de Fecundidad (2020).

49. CNE. *Incidencias acumuladas e indicadores de transmisibilidad* Madrid (2020).

50. INE. *Edad Media a la Maternidad por provincia, según orden del nacimiento*. Madrid: Indicadores de Fecundidad (2019).

51. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Essays Econom II Collect Pap Clive W J Granger*. (2008) 37:31–47. doi: 10.1017/CCOL052179207X.002

52. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung SM, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis*. (2020) 94:154–5. doi: 10.1016/j.ijid.2020.03.020

53. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: a living systematic review and meta-analysis. *PLoS Med*. (2020) 17:1–25. doi: 10.1371/journal.pmed.1003346

54. Jung CY, Park H, Kim DW, Choi YJ, Kim SW, Chang TI. Clinical characteristics of asymptomatic patients with COVID-19: a nationwide cohort study in South Korea. *Int J Infect Dis*. (2020) 99:266–8. doi: 10.1016/j.ijid.2020.08.001

55. Hippich M, Holthaus L, Assfalg R, Zapardiel-Gonzalo J, Kapfelsperger H, Heigermoser M, et al. A public health antibody screening indicates a 6-fold higher SARS-CoV-2 exposure rate than reported cases in children. *Med*. (2020) 2:149–63.e4. doi: 10.1016/j.medj.2020.10.003

56. Laxminarayan R, Wahl B, Dudala SR, Gopal K, Mohan C, Neelima SK, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science*. (2020) 28:1–43. doi: 10.1101/2020.07.14.20 153643

57. Sayampanathan AA, Heng CS, Pin PH, Pang J, Leong TY, Lee VJ. Infectivity of asymptomatic versus symptomatic COVID-19. *Lancet*. (2020) 6736:32651. doi: 10.1016/S0140-6736(20)32651-9

58. Fleischer M, Schumann L, Hartmann A, Walker RS, Ifrim L, von Zadow D, et al. Pre-adolescent children exhibit lower aerosol particle volume emissions than adults for breathing, speaking, singing and shouting. *J R Soc Interface*. (2022) 19:833. doi: 10.1098/rsif.2021.0833

59. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav*. (2020) 2020:7199. doi: 10.1101/2020.07.06.20147199

60. Poole S, Brendish NJ, Tanner AR, Clark TW. Physical distancing in schools for SARS-CoV-2 and the resurgence of rhinovirus. *Lancet Respir Med*. (2020) 2600:4–5. doi: 10.1016/S2213-2600(20)30502-6

61. Vlachos J, Hertegård E, Svaleryd HB. The effects of school closures on SARS-CoV-2 among parents and teachers. *Proc Natl Acad Sci USA*. (2021) 118:e2020834118. doi: 10.1073/pnas.2020834118

62. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. *Sci Rep*. (2021) 11:1–13. doi: 10.1038/s41598-021-81442-x

frontiers | Frontiers in Public Health

Fractional differential equation modeling of the HBV infection with time delay and logistic proliferation

Deshun Sun[1,2]*†, Jingxiang Liu[3]†, Xiuyun Su[1]† and Guoxian Pei[1]*

[1]Intelligent Medical Innovation Center, Southern University of Science and Technology Hospital, Shenzhen, China, [2]Shenzhen Key Laboratory of Tissue Engineering, Shenzhen Second People's Hospital (The First Hospital Affiliated to Shenzhen University, Health Science Center), Shenzhen, China, [3]School of Marine Electrical Engineering, Dalian Maritime University, Dalian, China

In this article, a fractional-order differential equation model of HBV infection was proposed with a Caputo derivative, delayed immune response, and logistic proliferation. Initially, infection-free and infection equilibriums and the basic reproduction number were computed. Thereafter, the stability of the two equilibriums was analyzed based on the fractional Routh–Hurwitz stability criterion, and the results indicated that the stability will change if the time delay or fractional order changes. In addition, the sensitivity of the basic reproduction number was analyzed to find out the most sensitive parameter. Lastly, the theoretical analysis was verified by numerical simulations. The results showed that the time delay of immune response and fractional order can significantly affect the dynamic behavior in the HBV infection process. Therefore, it is necessary to consider time delay and fractional order in modeling HBV infection and studying its dynamics.

KEYWORDS

HBV model, time delay, fractional order, stability, Hopf bifurcation

Introduction

Hepatitis B virus (HBV) can attack the liver and cause both acute and chronic diseases and further lead to fibrosis, cirrhosis, or even cancer. It is estimated that 296 million people have chronic hepatitis B, and 1.5 million new infections are reported each year; 820 000 people died of hepatitis B infections in 2019 (1). Therefore, HBV has become a major public health problem affecting human health (2).

Mathematical modeling and analysis of infectious viruses help understand the infection mechanism and realize the disease progression (3–5). Furthermore, mathematical modeling can also provide new insights to find the key factors to treat infectious diseases (6). In 1996, the basic ordinary differential equation (ODE) model of HBV infection was established with uninfected cells, infected cells, and free viruses (7). This is an early mathematical model for studying the spread of viruses. As research progresses, the mathematical modeling of virus transmission has become more and more complicated. For instance, Peter et al. (3) established a deterministic ODE model with six compartments to study the transmission dynamics of measles and obtained the best fit using available data,

which could help health workers in decision-making and policymakers to frame policies to eradicate the spread of measles in Nigeria. Mayowa et al. (5) divided the population into six classes and formulated a six-compartmental deterministic model to investigate the effect of vaccination on the dynamics of tuberculosis in a given population. All the aforementioned mathematical models are based on ordinary differential equations with bilinear incidence rate.

Subsequently, a large number of dynamic models were proposed to describe and analyze virus infection according to different biological mechanisms (8–13); for example, because hepatocytes have the ability to regenerate, the models are constrained by the number of healthy and infected hepatocytes. Li et al. (10) developed a logistic growth model of HBV. Moreover, in order to characterize the time of a body's immune response after the virus infection of target cells, time delay has been considered. Therefore, Zhang et al. (13) proposed a susceptible-vaccinated-exposed-infectious-removed (SVEIR) epidemic model with two time delays and constructed a Lyapunov function to discuss the asymptotic stability of the positive equilibrium point. Babasola et al. (14) modeled the spread of COVID-19 with a convex incidence rate incorporated with a time delay and proved that delay can destabilize the system and lead to periodic oscillation.

In recent years, a fractional derivative for describing memory, history, and heredity effects in modeling physical, chemical, financial, and biological systems has received increasing attention (15–28). For example, Diethelm (29) used a fractional-order model to simulate the dynamics of a dengue fever outbreak. The results showed that the simulation accuracy of the fractional-order model is much higher than that of the integer-order derivative. Gilberto et al. (30) proposed a fractional-order model to research the dynamics of influenza A (H1N1), and the results showed that the fractional-order model was in good agreement with real data. Similarly, Ogunrinde et al. (27) divided the population into five classes and proposed a fractional-order differential equation model to study COVID-19. The basic reproduction number was calculated by the spectral radius method, and the stability analysis of the model was carried out by constructing the Lyapunov function. Finally, the parameters were estimated by collected data, and the model can offer guidance to policymakers.

In addition to the mathematical modeling of fractional differential equations for the aforementioned infectious diseases, there are also many studies that use the fractional-order model to characterize the process of HBV infections (31–33). For example, Simelane and Dlamini (33) established a fractional-order HBV model with a saturated incidence rate by using the Caputo fractional derivatives. Then, the basic reproduction number was calculated, and the stability of the equilibriums was discussed. The simulation results demonstrated that the fractional-order model is more appropriate for modeling HBV transmission dynamics than the integer-order model. The time of HBV entry into the healthy liver cells and the production of

new virus particles should be taken into account; therefore, Gao et al. (32) established a three-dimensional delayed fractional-order HBV model, which included healthy hepatocytes, infected hepatocytes, and free viruses, as follows:

$$
\begin{cases}
{}^{CF}_{0}D_t^{\sigma_1} x(t) = \lambda_1{}^{\sigma_1} - \mu_1{}^{\sigma_1} x(t) - \beta_1{}^{\sigma_1} x(t)v(t) + \delta_1{}^{\sigma_1} y(t), \\
{}^{CF}_{0}D_t^{\sigma_1} y(t) = \beta_1{}^{\sigma_1} x(t)v(t) - (\alpha_1{}^{\sigma_1} + \delta_1{}^{\sigma_1})y(t), \\
{}^{CF}_{0}D_t^{\sigma_2} v(t) = c_1{}^{\sigma_2} y(t-\tau)e^{-\rho\tau} - \gamma_1{}^{\sigma_2} v(t).
\end{cases} \tag{1}
$$

This model has not considered the cytotoxic T lymphocyte (CTL) and alanine aminotransferase (ALT) levels, which reflect the extent of liver damage. Therefore, the items of CTL and ALT will be considered in our established model.

However, until now, no study has been designed to analyze the dynamics of HBV involving logistic proliferation, time delay, and items of CTL and ALT by fractional-order differential equations. Motivated by the aforementioned discussion, we proposed a fractional-order differential equation model with time delay and logistic proliferation in order to better understand the transmission mechanism of HBV in the human body.

The remaining part of this article is organized as follows: Section Mathematical model deals with the formulation of the model. Section Equilibriums and the basic reproduction number discusses the infection-free and infection equilibriums and the basic reproduction number. Section Equilibriums and the basic reproduction number discusses the stability analysis of the two equilibriums and analyzes the sensitivity of the basic reproduction number. Section Numerical simulation gives an account of the numerical simulations of equilibriums and the Hopf bifurcation. Finally, Section Conclusion and discussion comprises the conclusion and discussion.

Mathematical model

Therefore, based on our work (34), we proposed a fractional-order differential equation model with time delay and logistic proliferation as follows:

$$
\begin{cases}
\frac{d^\alpha x(t)}{dt} = \xi^\alpha + r^\alpha x(t)\left(1 - \frac{x(t)+y(t)}{T_{\max}^\alpha}\right) - d^\alpha x(t) - b^\alpha x(t)v(t), \\
\frac{d^\alpha y(t)}{dt} = b^\alpha x(t)v(t) - a^\alpha y(t) - k_1^\alpha y(t-\tau)z(t-\tau), \\
\frac{d^\alpha v(t)}{dt} = k^\alpha y(t) - \varepsilon^\alpha v(t) - k_2^\alpha y(t-\tau)z(t-\tau), \\
\frac{d^\alpha z(t)}{dt} = k_3^\alpha y(t-\tau)z(t-\tau) - k_4^\alpha z(t), \\
\frac{d^\alpha w(t)}{dt} = k_5^\alpha + k_6^\alpha y(t)z(t) - k_7^\alpha w(t).
\end{cases}
$$

$$\tag{2}$$

where the variables x, y, v, z, and w represent the uninfected cells, infected cells, viruses, CTL level, and ALT level, respectively; ξ and r are the production rate and proliferation rate of uninfected cells, respectively; $T_{\max}$ is the maximum hepatocyte count in the liver; d is the death rate of uninfected

cells; b is the infection rate of uninfected cells to become infected cells; a is the death rate of infected cells; k_1 represents the cure rate of infected cells by CTL; k and ε are the production rate and death rate of free viruses, respectively; k_2 represents the clearance rate of free viruses by CTL; k_3 and k_4 are the production rate and death rate of CTL, respectively; k_5 is the natural production rate of ALT; k_6 is the production rate of ALT from infected cells; k_7 is the death rate; τ is time delay with the order of $\alpha(0 < \alpha \le 1)$; and $\frac{d^\alpha x(t)}{dt}$, $\frac{d^\alpha y(t)}{dt}$, $\frac{d^\alpha v(t)}{dt}$, $\frac{d^\alpha z(t)}{dt}$, and $\frac{d^\alpha w(t)}{dt}$ denote the Caputo fractional derivatives. Hence, $\frac{d^\alpha x(t)}{dt}$ is defined as follows:

$$\frac{d^\alpha x_i}{dt} = I^{n-\alpha}\frac{d^n x}{dt^n} = \frac{1}{\Gamma(n-\alpha)}\int_0^t (t-s)^{(n-\alpha-1)} x^{(n)}(s)ds \quad (3)$$

where $n-1 < \alpha < n, n \in \mathbb{N}$ and $\Gamma(\textbf{.})$ is the gamma function. When $0 < \alpha < 1$,

$$\frac{d^\alpha x}{dt} = \frac{1}{\Gamma(1-\alpha)}\int_0^t \frac{x'(s)}{(t-s)^\alpha}ds \quad (4)$$

Based on the aforementioned model, the equilibriums and stability analysis are discussed in Section Equilibriums and the basic reproduction number.

Equilibriums and the basic reproduction number

In the following paragraphs, the equilibriums and the basic reproduction number are discussed.

Equilibriums

The method to compute the equilibrium is to set $\frac{d^\alpha x(t)}{dt} = 0$, $\frac{d^\alpha y(t)}{dt} = 0$, $\frac{d^\alpha v(t)}{dt} = 0$, $\frac{d^\alpha z(t)}{dt} = 0$, and $\frac{d^\alpha w(t)}{dt} = 0$. Hence, we get the following equations:

$$\begin{cases} \xi^\alpha - d^\alpha x(t) + r^\alpha x(t)\left(1 - \frac{x(t)+y(t)}{T_{\max}^\alpha}\right) - b^\alpha x(t)v(t) = 0, \\ b^\alpha x(t)v(t) - a^\alpha y(t) - k_1^\alpha y(t-\tau)z(t-\tau) = 0, \\ k^\alpha y(t) - \varepsilon^\alpha v(t) - k_2^\alpha y(t-\tau)z(t-\tau) = 0, \\ k_3^\alpha y(t-\tau)z(t-\tau) - k_4^\alpha z(t) = 0, \\ k_5^\alpha + k_6^\alpha y(t)z(t) - k_7^\alpha w(t) = 0. \end{cases} \quad (5)$$

The infection-free equilibrium $\boldsymbol{E_0}$ denotes $x \ne 0, w \ne 0, y = v = z = 0$; thus, the infection-free equilibrium is as follows:

$$E_0 = (x_0, y_0, v_0, z_0, w_0) = \left(\frac{T_{\max}^\alpha}{2r^\alpha}\left[-(d^\alpha - r^\alpha) + \sqrt{(d^\alpha - r^\alpha)^2 + \frac{4\xi^\alpha r^\alpha}{T_{\max}^\alpha}}\right], 0, 0, 0, \frac{k_5^\alpha}{k_7^\alpha}\right)$$

Similarly, the infection equilibrium E_1, which denotes $x \ne 0, y \ne 0, v \ne 0, z \ne 0, w \ne 0$, was computed by the following equations:

$$\begin{cases} \xi^\alpha - d^\alpha x^* + r^\alpha x^*\left(1 - \frac{x^*+y^*}{T_{\max}^\alpha}\right) - b^\alpha xv = 0, \\ b^\alpha x^* v^* - a^\alpha y^* - k_1^\alpha y^* z^* = 0, \\ k^\alpha y^* - \varepsilon^\alpha v^* - k_2^\alpha y^* z^* = 0, \\ k_3^\alpha y^* z^* - k_4^\alpha z^* = 0, \\ k_5^\alpha + k_6^\alpha y^* z^* - k_7^\alpha w^* = 0. \end{cases} \quad (6)$$

The previous equations were solved, and the infection equilibrium was obtained as follows:

$$x^* = -\frac{B}{3A} + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, y^* = \frac{k_4^\alpha}{k_3^\alpha}, v^* = \frac{a^\alpha k_2^\alpha k_4^\alpha + k^\alpha k_1^\alpha k_4^\alpha}{b^\alpha k_2^\alpha k_3^\alpha x^* + \varepsilon^\alpha k_1^\alpha k_3^\alpha}, z^* = \frac{k^\alpha}{k_2^\alpha} - \frac{\varepsilon^\alpha(a^\alpha k_2^\alpha - k^\alpha k_1^\alpha)}{k_2^\alpha(b^\alpha k_2^\alpha x^* + \varepsilon^\alpha k_1^\alpha)}, w^* = \frac{k_5^\alpha}{k_7^\alpha} + \frac{k^\alpha k_4^\alpha k_6^\alpha}{k_2^\alpha k_3^\alpha k_7^\alpha} - \frac{\varepsilon^\alpha k_4^\alpha k_6^\alpha(a^\alpha k_2^\alpha - k^\alpha k_1^\alpha)}{k_2^\alpha k_3^\alpha k_7^\alpha(b^\alpha k_2^\alpha x^* + \varepsilon^\alpha k_1^\alpha)}.$$

where

$$A = \frac{b^\alpha r^\alpha k_2^\alpha}{T_{\max}^\alpha}, B = -\left[b^\alpha k_2^\alpha(r^\alpha - d^\alpha - \frac{r^\alpha k_4^\alpha}{T_{\max}^\alpha k_3^\alpha}) - \frac{\varepsilon^\alpha r^\alpha k_1^\alpha}{T_{\max}^\alpha}\right],$$
$$C = -\left[b^\alpha \xi^\alpha k_2^\alpha + \varepsilon^\alpha k_1^\alpha(r^\alpha - d^\alpha - \frac{r^\alpha k_4^\alpha}{T_{\max}^\alpha k_3^\alpha}) - \frac{b^\alpha k_4^\alpha(a^\alpha k_2^\alpha + k^\alpha k_1^\alpha)}{k_3^\alpha}\right],$$
$$D = -\varepsilon^\alpha \xi^\alpha k_1^\alpha, p = \frac{3AC-B^2}{3A^2}, q = \frac{27A^2D-9ABC+2B^3}{27A^3}.$$

Thus, the infection equilibrium is as follows:

$$E_1 = (x^*, y^*, v^*, z^*, w^*)$$
$$= \left(x^*, \frac{k_4^\alpha}{k_3^\alpha}, \frac{a^\alpha k_2^\alpha k_4^\alpha + k^\alpha k_1^\alpha k_4^\alpha}{b^\alpha k_2^\alpha k_3^\alpha x^* + \varepsilon^\alpha k_1^\alpha k_3^\alpha}, \frac{k^\alpha}{k_2^\alpha} - \frac{\varepsilon^\alpha(a^\alpha k_2^\alpha - k^\alpha k_1^\alpha)}{k_2^\alpha(b^\alpha k_2^\alpha x^* + \varepsilon^\alpha k_1^\alpha)}, \frac{k_5^\alpha}{k_7^\alpha} + \frac{k^\alpha k_4^\alpha k_6^\alpha}{k_2^\alpha k_3^\alpha k_7^\alpha} - \frac{\varepsilon^\alpha k_4^\alpha k_6^\alpha(a^\alpha k_2^\alpha - k^\alpha k_1^\alpha)}{k_2^\alpha k_3^\alpha k_7^\alpha(b^\alpha k_2^\alpha x^* + \varepsilon^\alpha k_1^\alpha)}\right)$$

Basic reproduction number

The basic reproduction number can be calculated by the method of integral operator spectral radius given as follows:

$$R_0 = \rho(FV^{-1})$$

Thus, the basic reproduction number of E_0 is as follows:

$$R_0 = \frac{b^\alpha k^\alpha x_0}{a^\alpha \varepsilon^\alpha},$$

where

$$F = \begin{bmatrix} b^\alpha v & 0 & b^\alpha x \\ 0 & k^\alpha y & 0 \end{bmatrix}, V = \begin{bmatrix} 0 & a^\alpha + k_1^\alpha ze^{-\lambda\tau} & 0 \\ 0 & k_2^\alpha ze^{-\lambda\tau} & \varepsilon^\alpha \end{bmatrix}.$$

Similarly, the basic reproduction number of E_1 is as follows:

$$R_1 = \frac{b^\alpha k^\alpha x^*}{\varepsilon^\alpha\left(a^\alpha + \frac{k^\alpha k_1^\alpha}{k_2^\alpha} - \frac{\varepsilon^\alpha k_1^\alpha(a^\alpha k_2^\alpha - k^\alpha k_1^\alpha)}{k_2^\alpha(b^\alpha k_2^\alpha x^* + \varepsilon^\alpha k_1^\alpha)}\right)}$$

Stability and sensitivity analyses

The local asymptotic stability of E_0 and E_1 is discussed in this part.
First, the Jacobi matrix was computed as follows:

$$
Jac = \begin{bmatrix}
-d^\alpha + r^\alpha - \frac{2r^\alpha x + y}{T_{\max}^\alpha} & -\frac{r^\alpha x}{T_{\max}^\alpha} & -b^\alpha x & 0 & 0 \\
b^\alpha v & -a^\alpha - k_1^\alpha z e^{-S^\alpha \tau} & b^\alpha x & -k_1^\alpha y e^{-S^\alpha \tau} & 0 \\
0 & k^\alpha - k_2^\alpha z e^{-S^\alpha \tau} & -\varepsilon^\alpha & -k_2^\alpha y e^{-S^\alpha \tau} & 0 \\
0 & k_3^\alpha z e^{-S^\alpha \tau} & 0 & k_3^\alpha y e^{-S^\alpha \tau} - k_4^\alpha & 0 \\
0 & k_6^\alpha z & 0 & k_6^\alpha y & -k_7^\alpha
\end{bmatrix}
$$

(7)

Based on the previous Jacobi matrix, we got the characteristic determinant:

$$
\left| S^\alpha I - Jac \right|
$$

$$
= \begin{vmatrix}
S^\alpha + d^\alpha - r^\alpha + \frac{2r^\alpha x + y}{T_{\max}^\alpha} & \frac{r^\alpha x}{T_{\max}^\alpha} & b^\alpha x & 0 & 0 \\
-b^\alpha v & S^\alpha + a^\alpha + k_1^\alpha z e^{-S^\alpha \tau} & -b^\alpha x & k_1^\alpha y e^{-S^\alpha \tau} & 0 \\
0 & -k^\alpha + k_2^\alpha z e^{-S^\alpha \tau} & S^\alpha + \varepsilon^\alpha & k_2^\alpha y e^{-S^\alpha \tau} & 0 \\
0 & -k_3^\alpha z e^{-S^\alpha \tau} & 0 & S^\alpha + k_4^\alpha - k_3^\alpha y e^{-S^\alpha \tau} & 0 \\
0 & -k_6^\alpha z & 0 & -k_6^\alpha y & S^\alpha + k_7^\alpha
\end{vmatrix}
$$

Let $S^\alpha = \lambda$, then the simplified characteristic determinant is as follows:

$$
\left| \lambda I - Jac \right| = \begin{vmatrix}
\lambda + d^\alpha - r^\alpha + \frac{2r^\alpha x + y}{T_{\max}} & \frac{r^\alpha x}{T_{\max}} & b^\alpha x & 0 & 0 \\
-b^\alpha v & \lambda + a^\alpha + k_1^\alpha z e^{-\lambda \tau} & -b^\alpha x & k_1^\alpha y e^{-\lambda \tau} & 0 \\
0 & -k^\alpha + k_2^\alpha z e^{-\lambda \tau} & \lambda + \varepsilon^\alpha & k_2^\alpha y e^{-\lambda \tau} & 0 \\
0 & -k_3^\alpha z e^{-\lambda \tau} & 0 & \lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda \tau} & 0 \\
0 & -k_6^\alpha z & 0 & -k_6^\alpha y & \lambda + k_7^\alpha
\end{vmatrix}
$$

Local asymptotic stability of the infection-free equilibrium

The characteristic determinant at the infection-free equilibrium (E_0) is as follows:

$$
\left| \lambda I - Jac \right| = \begin{vmatrix}
\lambda + d^\alpha - r^\alpha + \frac{2r^\alpha x_0}{T_{\max}^\alpha} & \frac{r^\alpha x_0}{T_{\max}^\alpha} & b^\alpha x & 0 & 0 \\
0 & \lambda + a^\alpha & -b^\alpha x_0 & 0 & 0 \\
0 & -k^\alpha & \lambda + \varepsilon^\alpha & 0 & 0 \\
0 & 0 & 0 & \lambda + k_4^\alpha & 0 \\
0 & 0 & 0 & 0 & \lambda + k_7^\alpha
\end{vmatrix}
$$

$$
= (\lambda + d^\alpha - r^\alpha + \frac{2r^\alpha x_0}{T_{\max}^\alpha})(\lambda + k_4^\alpha)(\lambda + k_7^\alpha) \big[(\lambda + a^\alpha)(\lambda + \varepsilon^\alpha)
$$

$$
- b^\alpha k^\alpha x_0 \big]
$$

When $\left| \lambda I - Jac \right| = 0$, the eigenvalues are $\lambda_1 = -d^\alpha + r^\alpha - \frac{2r^\alpha x_0}{T_{\max}^\alpha}$, $\lambda_2 = k_4^\alpha$, $\lambda_3 = k_7^\alpha$, $\lambda_4 = \frac{-(a^\alpha + \varepsilon^\alpha) + \sqrt{(a^\alpha + \varepsilon^\alpha)^2 - 4(a^\alpha \varepsilon^\alpha - b^\alpha k^\alpha x_0)}}{2}$,

and $\lambda_5 = \frac{-(a^\alpha + \varepsilon^\alpha) - \sqrt{(a^\alpha + \varepsilon^\alpha)^2 - 4(a^\alpha \varepsilon^\alpha - b^\alpha k^\alpha x_0)}}{2}$.

Since $d > r$ and $R_0 = \frac{b^\alpha k^\alpha x_0}{a^\alpha \varepsilon^\alpha} < 1$, we have $\lambda_{1,2,3,4,5} < 0$. Thus, $\left| \arg(S_{1,2,3,4,5}) \right| > \frac{\alpha \pi}{2}$.

Thus, we get the conclusion that when $R_0 = \frac{b^\alpha k^\alpha x_0}{a^\alpha \varepsilon^\alpha} < 1$, E_0 is locally asymptotically stable.

Local asymptotic stability of the infection equilibrium

The characteristic determinant at the infection equilibrium (E_1) is as follows:

$$\begin{vmatrix} \lambda + d^\alpha - r^\alpha + \frac{2r^\alpha x + y}{T_{\max}^\alpha} & \frac{r^\alpha x}{T_{\max}^\alpha} & b^\alpha x & 0 & 0 \\ -b^\alpha v & \lambda + a^\alpha + k_1^\alpha z e^{-\lambda\tau} & -b^\alpha x & k_1^\alpha y e^{-\lambda\tau} & 0 \\ 0 & -k^\alpha + k_2^\alpha z e^{-\lambda\tau} & \lambda + \varepsilon^\alpha & k_2^\alpha y e^{-\lambda\tau} & 0 \\ 0 & -k_3^\alpha z e^{-\lambda\tau} & 0 & \lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau} & 0 \\ 0 & -k_6^\alpha z & 0 & -k_6^\alpha y & \lambda + k_7^\alpha \end{vmatrix}$$

$$= \left(\lambda + k_7^\alpha\right) \left\{ \begin{array}{l} (\lambda + A_0) \left[\begin{array}{l} \left(\lambda^2 + (a^\alpha + \varepsilon^\alpha)\lambda + a^\alpha\varepsilon^\alpha + k_1^\alpha z\lambda e^{-\lambda\tau} + \varepsilon^\alpha k_1^\alpha z e^{-\lambda\tau}\right)\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right) \\ + b^\alpha x \left(-k^\alpha + k_2^\alpha z e^{-\lambda\tau}\right)\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right) + k_1^\alpha k_3^\alpha yz \left(\lambda + \varepsilon^\alpha\right) e^{-2\lambda\tau} \\ + b^\alpha k_2^\alpha xy e^{-\lambda\tau} \end{array} \right] \\ + b^\alpha v \left[\frac{r\alpha x}{T_{\max}^\alpha}\left(\lambda + \varepsilon^\alpha\right)\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right) - b^\alpha x \left(\begin{array}{l}\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right)\left(k_2^\alpha z e^{-\lambda\tau} - k^\alpha\right) \\ + k_2^\alpha k_3^\alpha yz e^{-2\lambda\tau}\end{array}\right)\right] \end{array} \right\}$$

where $A_0 = d^\alpha - r^\alpha + \frac{2r^\alpha x + y}{T_{\max}^\alpha}$.

For convenience, we made the following simplifications:

$$\left(\lambda^2 + (a^\alpha + \varepsilon^\alpha)\lambda + a^\alpha\varepsilon^\alpha + k_1^\alpha z\lambda e^{-\lambda\tau} + \varepsilon^\alpha k_1^\alpha z e^{-\lambda\tau}\right)\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right)$$
$$= \lambda^3 + A_1\lambda^2 + A_2\lambda + A_3 + A_4\lambda^2 e^{-\lambda\tau} + A_5\lambda e^{-\lambda\tau} + A_6 e^{-\lambda\tau} + A_7\lambda e^{-2\lambda\tau} + A_8 e^{-2\lambda\tau}$$

where

$$A_1 = a^\alpha + \varepsilon^\alpha + k_4^\alpha, A_2 = a^\alpha\varepsilon^\alpha + (a^\alpha + \varepsilon^\alpha)k_4^\alpha, A_3 = a^\alpha\varepsilon^\alpha k_4^\alpha, A_4 = (k_1^\alpha z - k_3^\alpha y),$$
$$A_5 = \varepsilon^\alpha k_1^\alpha z + k_1^\alpha k_4^\alpha z - (a^\alpha + \varepsilon^\alpha)k_3^\alpha y, A_6 = \varepsilon^\alpha k_1^\alpha k_4^\alpha z - a^\alpha\varepsilon^\alpha k_3^\alpha y, A_7 = -k_1^\alpha k_3^\alpha yz,$$
$$A_8 = -\varepsilon^\alpha k_1^\alpha k_3^\alpha yz.$$

$$b^\alpha x \left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right)\left(-k^\alpha + k_2^\alpha z e^{-\lambda\tau}\right) + k_1^\alpha k_3^\alpha yz(\lambda + \varepsilon)^\alpha e^{-2\lambda\tau} + b^\alpha k_2^\alpha xy e^{-\lambda\tau}$$
$$= A_9\lambda + A_{10} + A_{11}\lambda e^{-\lambda\tau} + A_{12}e^{-\lambda\tau} + A_{13}\lambda e^{-2\lambda\tau} + A_{14}e^{-2\lambda\tau}$$

where

$$A_9 = -b^\alpha k^\alpha x, A_{10} = -b^\alpha k^\alpha k_4^\alpha x, A_{11} = b^\alpha k_2^\alpha xz, A_{12} = b^\alpha k^\alpha k_3^\alpha xy + b^\alpha k_2^\alpha k_4^\alpha xz + b^\alpha k_2^\alpha xy,$$
$$A_{13} = k_1^\alpha k_3^\alpha yz, A_{14} = \varepsilon^\alpha k_1^\alpha k_3^\alpha yz - bk_2^\alpha k_3^\alpha xyz$$

$$b^\alpha v \left[\frac{r^\alpha x}{T_{\max}^\alpha}\left(\lambda + \varepsilon^\alpha\right)\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right) - b^\alpha x \left(\begin{array}{l}\left(\lambda + k_4^\alpha - k_3^\alpha y e^{-\lambda\tau}\right)\left(k_2^\alpha z e^{-\lambda\tau} - k^\alpha\right) \\ + k_2^\alpha k_3^\alpha yz e^{-2\lambda\tau}\end{array}\right)\right]$$
$$= A_{15}\lambda^2 + A_{16}\lambda + A_{17} + A_{18}\lambda e^{-\lambda\tau} + A_{19}e^{-\lambda\tau} + A_{20}e^{-2\lambda\tau}$$

where

$$A_{15} = \frac{b^\alpha r^\alpha xv}{T_{\max}^\alpha}, A_{16} = b^\alpha v\left(\frac{r^\alpha k_4^\alpha x}{T_{\max}^\alpha} + \frac{\varepsilon^\alpha r^\alpha x}{T_{\max}} + b^\alpha k^\alpha x\right), A_{17} = b^\alpha v\left(\frac{\varepsilon^\alpha r^\alpha k_4^\alpha x}{T_{\max}^\alpha} + b^\alpha k^\alpha k_4^\alpha x\right),$$

$$A_{18} = -b^\alpha v\left(\frac{r^\alpha k_3^\alpha xy}{T_{\max}^\alpha} + b^\alpha k_2^\alpha xz\right), A_{19} = -b^\alpha v\left(\frac{\varepsilon^\alpha r^\alpha k_3^\alpha xy}{T_{\max}^\alpha} + b^\alpha k_2^\alpha k_4^\alpha xz + b^\alpha k^\alpha k_3^\alpha xy\right),$$
$$A_{20} = -b^{2\alpha}xv(k_2^\alpha k_3^\alpha yz - k_2^\alpha k_3^\alpha yz).$$

Thus, the characteristic determinant becomes as follows:

$$
(\lambda + k_7)\left\{
\begin{array}{l}
(\lambda + A_0)\left[
\begin{array}{l}
\lambda^3 + A_1\lambda^2 + A_2\lambda + A_3 + A_4\lambda^2 e^{-\lambda\tau} + A_5\lambda e^{-\lambda\tau} + A_6 e^{-\lambda\tau} + A_7\lambda e^{-2\lambda\tau} \\
+A_8 e^{-2\lambda\tau} + A_9\lambda + A_{10} + A_{11}\lambda e^{-\lambda\tau} + A_{12} e^{-\lambda\tau} + A_{13}\lambda e^{-2\lambda\tau} + A_{14} e^{-2\lambda\tau}
\end{array}
\right] \\
+A_{15}\lambda^2 + A_{16}\lambda + A_{17} + A_{18}\lambda e^{-\lambda\tau} + A_{19} e^{-\lambda\tau} + A_{20} e^{-2\lambda\tau}
\end{array}
\right\}
$$

$$
= \left(\lambda + k_7^\alpha\right)\left\{
\left[
\begin{array}{l}
\lambda^4 + B_1\lambda^3 + B_2\lambda^2 + B_3\lambda + B_4 + B_5\lambda^3 e^{-\lambda\tau} + B_6\lambda^2 e^{-\lambda\tau} + B_7\lambda e^{-\lambda\tau} + B_8 e^{-\lambda\tau} \\
+B_9\lambda^2 e^{-2\lambda\tau} + B_{10}\lambda e^{-2\lambda\tau} + B_{11} e^{-2\lambda\tau}
\end{array}
\right]
\right\}
$$

where

$$B_1 = A_1 + A_0, B_2 = A_2 + A_9 + A_0 A_1 + A_{15}, B_3 = A_3 + A_{10} + A_0 A_2 + A_0 A_9 + A_{16},$$
$$B_4 = A_0 A_3 + A_0 A_{10} + A_{17}, B_5 = A_4, B_6 = A_5 + A_{11} + A_0 A_4, B_7 = A_6 + A_{12} + A_0 A_5 + A_0 A_{11} + A_{18},$$
$$B_8 = A_0 A_6 + A_0 A_{12} + A_{19}, B_9 = A_7 + A_{13}, B_{10} = A_8 + A_{14} + A_0 A_7 + A_0 A_{13}, B_{11} = A_0 A_8 + A_0 A_{14} + A_{20}.$$

For further simplification, we derived the following assignment:

$$C_1 = B_1 + k_7, C_2 = B_2 + k_7 B_1, C_3 = B_3 + k_7 B_2, C_4 = B_4 + k_7 B_3, C_5 = k_7, C_6 = B_5, C_7 = B_6 + k_7 B_5,$$
$$C_8 = B_7 + k_7 B_6, C_9 = B_8 + k_7 B_7, C_{10} = k_7 B_8, C_{11} = B_9, C_{12} = B_{10} + k_7 B_9, C_{13} = B_{11} + k_7 B_{10}, C_{14} = k_7 B_{11}.$$

The characteristic determinant is as follows:

$$
\begin{aligned}
H(\lambda; \tau) = & \lambda^5 + C_1\lambda^4 + C_2\lambda^3 + C_3\lambda^2 + C_4\lambda + C_5 + (C_6\lambda^4 + C_7\lambda^3 + C_8\lambda^2 + C_9\lambda + C_{10})e^{-\lambda\tau} \\
& + (C_{11}\lambda^3 + C_{12}\lambda^2 + C_{13}\lambda + C_{14})e^{-2\lambda\tau} = 0
\end{aligned}
\tag{8}
$$

When $\tau = 0$, the previous equation becomes as follows:

$$\lambda^5 + D_1\lambda^4 + D_2\lambda^3 + D_3\lambda^2 + D_4\lambda + D_5 = 0 \tag{9}$$

where

$$D_1 = C_1 + C_6, D_2 = C_2 + C_7 + C_{11}, D_3 = C_3 + C_8 + C_{12}, D_4 = C_4 + C_9 + C_{13}, D_5 = C_5 + C_{10} + C_{14}.$$

Based on equation (9), we get the following lemma by applying the Routh–Hurwitz criterion.

Lemma If equation (9) satisfies $\Delta_1 \equiv D_1 > 0$, $\Delta_2 \equiv \begin{vmatrix} D_1 & 1 \\ D_3 & D_2 \end{vmatrix} > 0$, and $\Delta_3 \equiv \begin{vmatrix} D_1 & 1 & 0 \\ D_3 & D_2 & D_1 \\ D_5 & D_4 & D_3 \end{vmatrix} > 0$, E_1 is locally asymptotically stable

when $\tau = 0$.

Proof. The detailed proof can be referred to Peter et al. (26), Ogunrinde et al. (27).

The aforementioned lemma indicated that when $\tau = 0$, all roots of $H(\lambda; \tau)$ are to the left of the imaginary axis, and some roots may cross to the right from the imaginary axis as τ increases. Thus, E_1 is unstable because of its positive real parts.

Then, the stability of system (2) was investigated when $\tau > 0$.

Both sides of equation (8) were multiplied by $e^{\lambda\tau}$:

$$
\begin{aligned}
& (C_6\lambda^4 + C_7\lambda^3 + C_8\lambda^2 + C_9\lambda + C_{10}) + (\lambda^5 + C_1\lambda^4 + C_2\lambda^3 + C_3\lambda^2 + C_4\lambda + C_5)e^{\lambda\tau} \\
& + (C_{11}\lambda^3 + C_{12}\lambda^2 + C_{13}\lambda + C_{14})e^{-\lambda\tau} = 0
\end{aligned}
\tag{10}
$$

Suppose the aforementioned equation has a purely imaginary root $\lambda = i\omega$ $(\omega > 0)$, then we have $e^{i\omega} = \cos\omega + i\sin\omega$, $e^{-i\omega} = \cos\omega - i\sin\omega$. Substituting $\lambda = i\omega$ into equation (10), we have

$$C_6\lambda^4 + C_7\lambda^3 + C_8\lambda^2 + C_9\lambda + C_{10} = C_6\omega^4 - C_7\omega^3 i - C_8\omega^2 + C_9\omega i + C_{10} \tag{11}$$

$$(\lambda^5 + C_1\lambda^4 + C_2\lambda^3 + C_3\lambda^2 + C_4\lambda + C_5)e^{\lambda\tau} = (C_1\omega^4 - C_3\omega^2 + C_5)\cos\omega\tau + (\omega^5 - C_2\omega^3 + C_4\omega)\cos\omega\tau i + (-\omega^5 + C_2\omega^3 - C_4\omega)\sin\omega\tau + (C_1\omega^4 - C_3\omega^2 + C_5)\sin\omega\tau i \quad (12)$$

$$(C_{11}\lambda^3 + C_{12}\lambda^2 + C_{13}\lambda + C_{14})e^{-\lambda\tau} = (-C_{12}\omega^2 + C_{14})\cos\omega\tau + (-C_{11}\omega^3 + C_{13}\omega) \times \cos\omega\tau i + (-C_{11}\omega^3 + C_{13}\omega)\sin\omega\tau + (C_{12}\omega^2 - C_{14})\sin\omega\tau i \quad (13)$$

Therefore, equation (10) becomes as follows:

$$(C_1\omega^4 - C_3\omega^2 - C_{12}\omega^2 + C_5 + C_{14})\cos\omega\tau + (\omega^5 - C_2\omega^3 - C_{11}\omega^3 + C_4\omega + C_{13}\omega)\cos\omega\tau i$$
$$+(-\omega^5 + C_2\omega^3 - C_{11}\omega^3 - C_4\omega + C_{13}\omega)\sin\omega\tau + (C_1\omega^4 - C_3\omega^2 + C_{12}\omega^2 + C_5 - C_{14})\sin\omega\tau i$$
$$+C_6\omega^4 - C_7\omega^3 i - C_8\omega^2 + C_9\omega i + C_{10} = 0$$

For convenience, we assumed the following:

$$a_1 = C_1\omega^4 - C_3\omega^2 - C_{12}\omega^2 + C_5 + C_{14}, \quad a_2 = \omega^5 - C_2\omega^3 - C_{11}\omega^3 + C_4\omega + C_{13}\omega,$$
$$a_3 = -\omega^5 + C_2\omega^3 - C_{11}\omega^3 - C_4\omega + C_{13}\omega, \quad a_4 = C_1\omega^4 - C_3\omega^2 + C_{12}\omega^2 + C_5 - C_{14}.$$

Thus, we get

$$a_1\cos\omega\tau + a_2\cos\omega\tau i + a_3\sin\omega\tau + a_4\sin\omega\tau i + C_6\omega^4 - C_7\omega^3 i - C_8\omega^2$$
$$+C_9\omega i + C_{10} = 0 \quad (14)$$

The real part after separating the real and imaginary parts is as follows:

$$a_1\cos\omega\tau + a_3\sin\omega\tau = -C_6\omega^4 + C_8\omega^2 - C_{10} = D_1 \quad (15)$$

and the imaginary part is as follows:

$$a_2\cos\omega\tau + a_4\sin\omega\tau = C_7\omega^3 - C_9\omega = D_2 \quad (16)$$

It follows from the real part and imaginary part that

$$\cos\omega\tau = \frac{a_4D_1 - a_3D_2}{a_1a_4 - a_2a_3}; \sin\omega\tau = \frac{a_1D_2 - a_2D_1}{a_1a_4 - a_2a_3} \quad (17)$$

Suppose equation (10) has $\tilde{n}(1 \le \tilde{n} \le 5)$ positive real roots, denoted by $x_n(1 \le n \le \tilde{n})$. Let $\sqrt{x_n} = \omega$, we get

$$\cos(\sqrt{x_n}\tau) = Q_n$$
$$= \frac{(C_1\omega^4 - C_3\omega^2 + C_{12}\omega^2 + C_5 - C_{14})(-C_6\omega^4 + C_8\omega^2 - C_{10}) - (-\omega^5 + C_2\omega^3 - C_{11}\omega^3 - C_4\omega + C_{13}\omega)(C_7\omega^3 - C_9\omega)}{(C_1\omega^4 - C_3\omega^2 - C_{12}\omega^2 + C_5 + C_{14})(C_1\omega^4 - C_3\omega^2 + C_{12}\omega^2 + C_5 - C_{14}) - (\omega^5 - C_2\omega^3 - C_{11}\omega^3 + C_4\omega + C_{13}\omega)(-\omega^5 + C_2\omega^3 - C_{11}\omega^3 - C_4\omega + C_{13}\omega)}$$

$$\sin(\sqrt{x_n}\tau) = P_n$$
$$= \frac{(C_1\omega^4 - C_3\omega^2 - C_{12}\omega^2 + C_5 + C_{14})(C_7\omega^3 - C_9\omega) - (\omega^5 - C_2\omega^3 - C_{11}\omega^3 + C_4\omega + C_{13}\omega)(-C_6\omega^4 + C_8\omega^2 - C_{10})}{(C_1\omega^4 - C_3\omega^2 - C_{12}\omega^2 + C_5 + C_{14})(C_1\omega^4 - C_3\omega^2 + C_{12}\omega^2 + C_5 - C_{14}) - (\omega^5 - C_2\omega^3 - C_{11}\omega^3 + C_4\omega + C_{13}\omega)(-\omega^5 + C_2\omega^3 - C_{11}\omega^3 - C_4\omega + C_{13}\omega)}$$

Let

$$\tau_n^{(j)} = \begin{cases} \frac{1}{\sqrt{x_n}}\left[\arccos(Q_n) + 2j\pi\right], & if\ P_n \ge 0 \\ \frac{1}{\sqrt{x_n}}\left[2\pi - \arccos(Q_n) + 2j\pi\right], & if\ P_n < 0 \end{cases}$$

Here, the positive integer n satisfies $1 \le n \le \tilde{n}, j = 0, 1, 2, \ldots$

Thus, from the aforementioned equation, we know that the characteristic equation has a pair of purely imaginary roots $\pm i\sqrt{x_n}$. We define $\lambda_n^{(j)}(\tau) = \alpha_n^{(j)}(\tau) + i\omega_n^{(j)}(\tau)$ as the root of equation (10) near $\tau_n^{(j)}$ for every $1 \leq n \leq \tilde{n}$ and j, satisfying $\alpha_n^{(j)}(\tau_n^{(j)}) = 0$ and $\omega_n^{(j)}(\tau_n^{(j)}) = \sqrt{x_n}$. In summary, we arrived at the following theorem:

Theorem 1 When $\tau \in [0, \tau_{n_0}^{(0)})$ and there are positive real roots in equation (10), infection equilibrium E_1 is locally asymptotically stable, where

$$\tau_{n_0}^{(0)} = \min\left\{ \tau_n^{(j)} \Big| 1 \leq n \leq \tilde{n}, j = 0, 1, 2, \dots \right\}.$$

Proof. When $\tau \in [0, \tau_{n_0}^{(0)})$ and equation (10) have no positive real roots, where $\tau_{n_0}^{(0)} = \min\left\{ \tau_n^{(j)} \Big| 1 \leq n \leq \tilde{n}, j = 0, 1, 2, \dots \right\}$, all the roots have strictly negative real parts. Thus, E_1 is locally asymptotically stable for $\tau \in [0, \tau_{n_0}^{(0)})$.

Sensitivity of the basic reproduction number

In this part, the sensitivity index of the basic reproduction number is explored in order to find out the most sensitive parameter that can significantly affect the basic reproduction number and give proper treatment strategies (3).

The sensitivity index can be computed by using the following equation:

$$K_q^{R_0} = \frac{\partial R_0}{\partial q} \times \frac{q}{R_0} \tag{18}$$

The basic reproduction number of E_0 is as follows:

$$R_0 = \frac{b^\alpha k^\alpha x_0}{a^\alpha \varepsilon^\alpha}, \text{ where } x_0 = \frac{T_{\max}^\alpha}{2r^\alpha} \left[-(d^\alpha - r^\alpha) + \sqrt{(d^\alpha - r^\alpha)^2 + \frac{4\xi^\alpha r^\alpha}{T_{\max}^\alpha}} \right].$$

The results of sensitivity indexes (Table 1) demonstrated that the infection rate of uninfected cells to become infected cells (b), production rate of free viruses (k), maximum hepatocyte counts in the liver ($T_{\max}$), and production rate of uninfected cells (ξ) have the highest positive index. Therefore, decreasing the infection rate, the production rate of free viruses, and the production rate of uninfected cells can help treat patients with hepatitis B. On the contrary, the death rate of infected cells (a), the death rate of free viruses (ε), and the death rate of uninfected cells (d) have the highest negative index. This also suggests that increasing the death rate of infected cells, the death rate of free viruses, and the death rate of uninfected cells can also keep $R_0 < 1$ and help the treatment of patients with hepatitis B.

Numerical simulation

In this section, a simulation is carried out to prove the accuracy of the aforementioned theoretical analysis.

Algorithm

Before the simulation, first, we provide the algorithm to solve the fractional-order differential equation (35, 36):

$$\begin{cases} x(t_k) = \left[\xi^\alpha - d^\alpha x(t_{k-1}) + r^\alpha x(t_{k-1}) \left(1 - \frac{x(t_{k-1}) + y(t_{k-1})}{T_{\max}^\alpha}\right) - b^\alpha x(t_{k-1})v(t_{k-1}) \right] h^{q_1} - \sum_{j=v}^{k} c_j^{(q_1)} x(t_{k-j}), \\[2mm] y(t_k) = \left[b^\alpha x(t_{k-1})v(t_{k-1}) - ay(t_{k-1}) - k_1^\alpha y(t_{k-m-1})z(t_{k-m-1}) \right] h^{q_1} - \sum_{j=v}^{k} c_j^{(q_1)} y(t_{k-j}), \\[2mm] v(t_k) = \left[k^\alpha y(t_{k-1}) - \varepsilon^\alpha v(t_{k-1}) - k_2^\alpha y(t_{k-m-1})z(t_{k-m-1}) \right] h^{q_1} - \sum_{j=v}^{k} c_j^{(q_1)} v(t_{k-j}), \\[2mm] z(t_k) = \left[k_3^\alpha y(t_{k-m-1})z(t_{k-m-1}) - k_4^\alpha z(t_{k-1}) \right] h^{q_1} - \sum_{j=v}^{k} c_j^{(q_1)} z(t_{k-j}), \\[2mm] w(t_k) = \left[k_5^\alpha + k_6^\alpha y(t_{k-1})z(t_{k-1}) - k_7^\alpha w(t_{k-1}) \right] h^{q_1} - \sum_{j=v}^{k} c_j^{(q_1)} w(t_{k-j}), \end{cases} \tag{19}$$

TABLE 1 Sensitivity indexes of R_0 to model parameters.

Parameter	Sensitivity index
b	0.9000
k	0.9000
a	-0.8100
ε	-0.8100
$T_{\max}$	0.9298
d	-0.8951
r	-0.4518
ξ	0.8716

TABLE 2 Description of parameters and values when $R_0 < 1$.

Parameter	Description	Value	Source
ξ	Production rate of uninfected cells	4.6664	(37)
d	Death rate of uninfected cells	2.1897	Estimated
r	Proliferation rate of uninfected cells	0.0924	Estimated
$T_{\max}$	Maximum hepatocyte counts in the liver	4.2843	Estimated
b	Infection rate of uninfected cells to become infected cells	1.4042	Estimated
a	Death rate of infected cells	3.8707	Estimated
k_1	Cure rate of infected cells by CTL	1.8838	(37)
k	Production rate of free viruses	1.3655	(37)
ε	Death rate of free viruses	1.48663	Estimated
k_2	Clearance rate of free viruses by CTL	1.2661	Estimated
k_3	Production rate of CTL	3.8549	(37)
k_4	Death rate of CTL	1.1395	Estimated
k_5	Natural production rate of ALT	1.8789	(37)
k_6	Production rate of ALT from infected cells	0.12002	Estimated
k_7	Death rate of ALT	1.2557	Estimated

where T_{sim} is time length, $k = 1, 2, 3, ..., N$, $N = [T_{sim}/h]$, $m = [\tau/h]$, and $x(0) = x_0, v(0) = v_0, w(0) = w_0, y(t) = y_0, z(t) = z_0, t \in [-\tau, 0]$ are the initial conditions. $c_0^{(q)} = 1, c_j^{(q)} = \left(1 - \frac{1+q}{j}\right) c_{j-1}^{(q)}$.

Simulation of asymptotically stable infection-free equilibrium

First, we simulate the case of infection-free. The parameters are shown in Table 2.

The time length is 400, and the initial conditions are $x(0) = 1$, $v(0) = 1$, $w(0) = 1$, $y(t) = 1$, $z(t) = 1$, $t \in [-\tau, 0]$. The

order $\alpha = 0.9$ and the time delay $\tau = 0.7$. Therefore, we have $E_0 = (x_0, y_0, v_0, z_0, w_0) = (2.0289, 0, 0, 0, 1.4372)$, and the basic reproduction number $R_0 = 0.7546$.

The behaviors of the uninfected cells (x), infected cells (y), free viruses (v), CTLs (z), and ALT (w) are shown in Figure 1. In Figure 1, all individuals converge to the infection-free equilibrium E_0, and the basic reproduction number R_0 is 0.7546, which is smaller than 1. This coincides with our theoretical analysis, which showed that when $R_0 < 1$, the infection-free equilibrium E_0 is asymptotically stable.

Simulation of asymptotically stable infection equilibrium

The theoretical analysis of the infection equilibrium is verified in this section. Similarly, the parameters are shown in Table 3.

The initial conditions are the same as in the previous section. The time length is 400. The order $\alpha = 0.9$, and the time delay $\tau = 1.2$. Therefore, the infection equilibrium is as follows: $E_1 = (x_1, y_1, v_1, z_1, w_1) = (1.2762, 0.3341, 0.5338, 1.0787, 1.4802)$, and the basic reproduction number is $R_1 = 2.0132$.

Figure 2 is the behavior of the uninfected cells (x), infected cells (y), free viruses (v), CTLs (z), and ALT (w) with $R_1 > 1$ and $\tau = 1.2$. From Figure 2, we know that although all the individuals oscillate at the beginning, they converge to infection equilibrium E_1 shortly. Figure 3 shows the phase portraits of the uninfected cell–infected cell–free virus space; the arrow indicates the direction of convergence of the phase portraits, and it converges to the infection equilibrium E_1 (red dot).

Simulation of the Hopf bifurcation of the infection equilibrium

In this subsection, the Hopf bifurcation of the infection equilibrium is simulated. All parameters are the same as those in Section Simulation of asymptotically stable infection equilibrium, except $\tau = 3.2$.

Figure 4 shows that when $\tau = 3.2$, the uninfected cells (x), infected cells (y), free viruses (v), CTLs (z), and ALT (w) oscillate periodically around the infection equilibrium E_1. Figure 5 shows the phase portraits of the uninfected cell–infected cell–free virus space, and when $\tau = 3.2$, the phase portraits are a stable limit cycle which is around the infection equilibrium E_1. The bifurcation diagram (Figure 6) shows that the stability of infection equilibrium E_1 changes at $\tau = 1.2$.
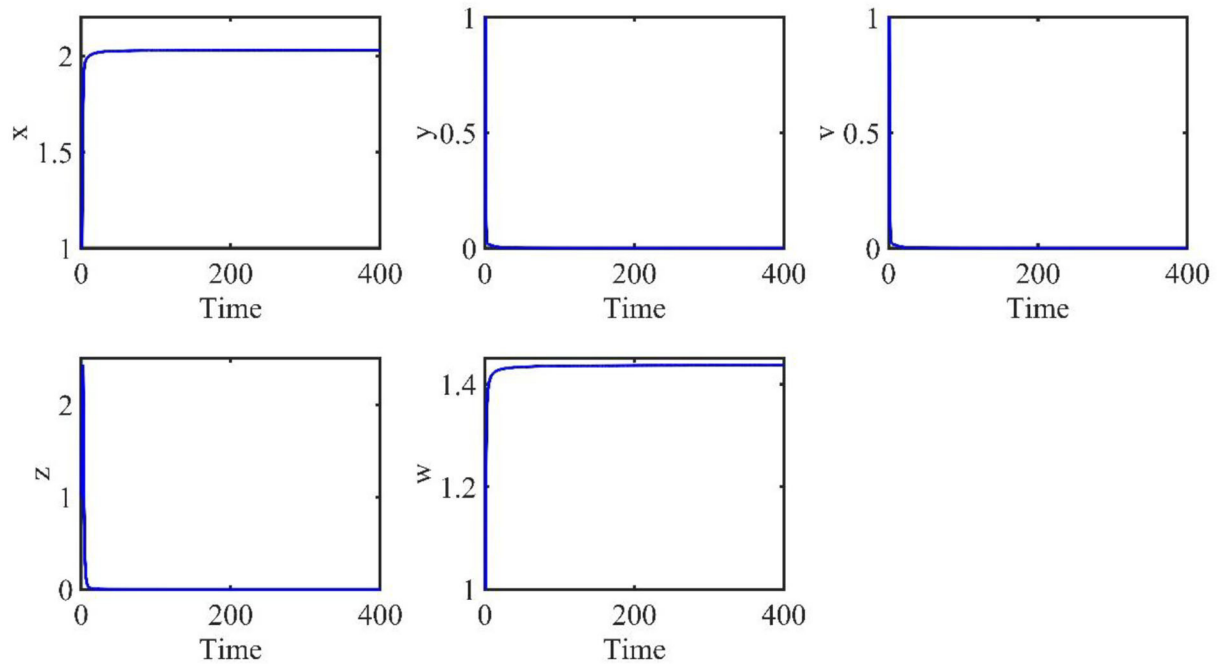
FIGURE 1
Dynamic change trend of uninfected cells (x), infected cells (y), free viruses (v), CTLs (z), and ALT (w) with $R_0 < 1$ and $\tau = 0.7$.

TABLE 3 Description of parameters and values when $R_1 > 1$.

Parameter	Description	Value	Source
ξ	Production rate of uninfected cells	4.6664	(37)
d	Death rate of uninfected cells	2.1897	Estimated
r	Proliferation rate of uninfected cells	0.0924	Estimated
$T_{\max}$	Maximum hepatocyte counts in the liver	4.2843	Estimated
b	Infection rate of uninfected cells to become infected cells	2.4042	Estimated
a	Death rate of infected cells	2.8707	Estimated
k_1	Cure rate of infected cells by CTL	1.8838	(37)
k	Production rate of free viruses	2.3655	(37)
ε	Death rate of free viruses	0.48663	(37)
k_2	Clearance rate of free viruses by CTL	1.2661	Estimated
k_3	Production rate of CTL	3.8549	(37)
k_4	Death rate of CTL	1.1395	Estimated
k_5	Natural production rate of ALT	1.8789	(37)
k_6	Production rate of ALT from infected cells	0.12002	Estimated
k_7	Death rate of ALT	1.2557	Estimated

Simulation of phase portraits with different orders

In this section, the phase portraits with different orders are studied by using the method of numerical simulation.

The initial order is $\alpha = 0.75$, and with step of 0.05, the order increases to 0.99. The phase portraits also used the uninfected cell–infected cell–free virus space. As shown in Figure 7, when $\tau = 1.2$ and the order (α) increases from 0.75 to 0.99, the volume of the phase portraits becomes

FIGURE 2
Dynamic change trend of uninfected cells (x), infected cells (y), free viruses (v), CTLs (z), ALT (w), and phase portraits of the xyz-space with $R_1 > 1$ and $\tau = 1.2$.

FIGURE 3
Phase portraits of the xyz-space with $R_1 > 1$ and $\tau = 1.2$.

bigger and the phase portraits become more complicated. Furthermore, the numerical simulations indicated that when the order increases from 0.75 to 0.95, the uninfected cell–infected cell–free virus space converges to the infection equilibrium E_1. However, when $\alpha = 0.99$, the phase portrait is a stable limit cycle, which is around the infection equilibrium E_1. This indicated that the order can significantly affect the stability of the system.

Conclusion and discussion

In this study, a fractional differential model of HBV infection with time delay and logistic proliferation was proposed in order to better understand the infection mechanism and realize the infection progression. First, the infection-free equilibrium, infection equilibrium, and the basic reproduction number were computed. In epidemiology, R_0 is considered the most important parameter, which provides an insight into how the disease spreads and helps us understand how to control the disease. Therefore, we proved that if the basic reproduction number $R_0 = \frac{b^\alpha k^\alpha x_0}{a^\alpha \varepsilon^\alpha} < 1$, the infection-free equilibrium (E_0) is locally asymptotically stable, which indicated that if the basic reproduction number $R_0 < 1$ can be controlled in patients, hepatitis B will disappear. Similarly, the stability analysis of the infection-free equilibrium (E_1) was discussed. In addition, the Hopf bifurcation of the infection equilibrium was studied at the theoretical level. Furthermore, sensitivity was analyzed to screen out the parameters that can significantly affect the basic reproduction number in our model. The results indicated that decreasing the infection rate (b), production rate of free viruses (k) and production rate of uninfected cells (ξ) can significantly decrease the basic reproduction number (R_0). Similarly, increasing the death rate of infected cells (a), the
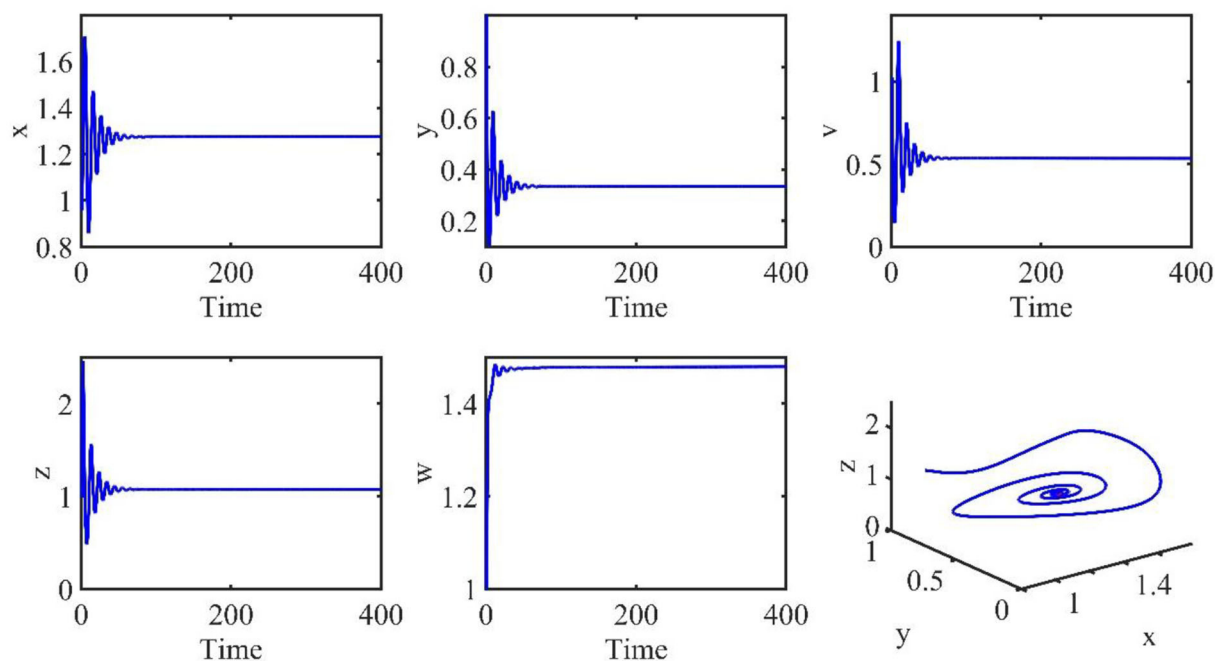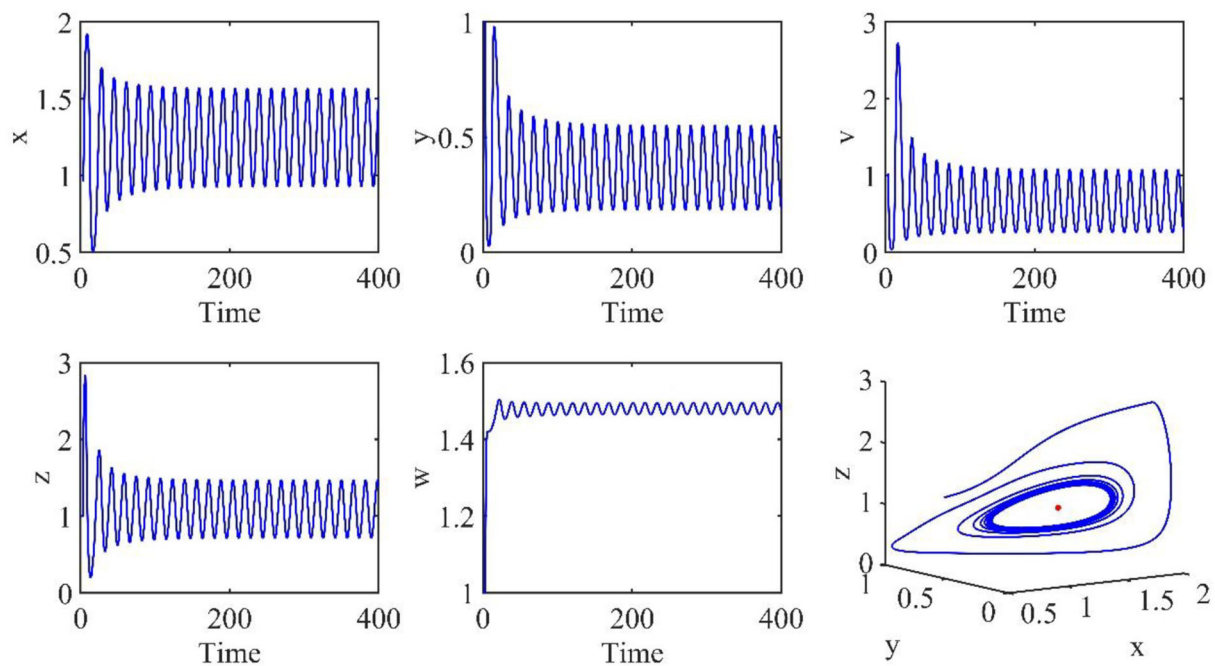
FIGURE 4

Dynamic change trend of uninfected cells (x), infected cells (y), free viruses (v), CTLs (z), ALT (w), and phase portraits of the xyz-space with $\tau = 3.2$.
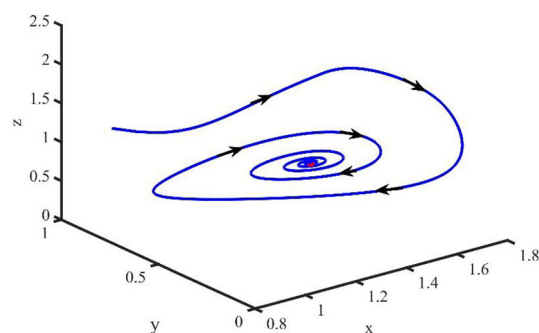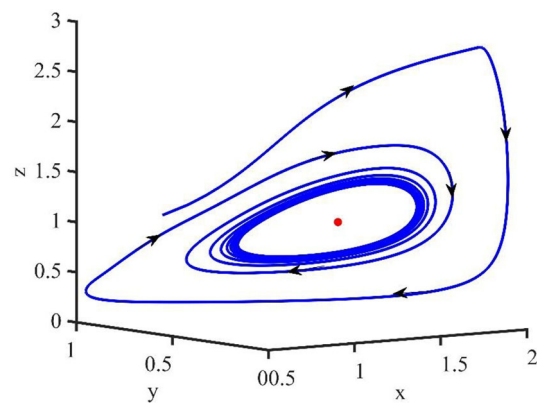

FIGURE 5

Phase portraits of the xyz-space with $\tau = 3.2$.
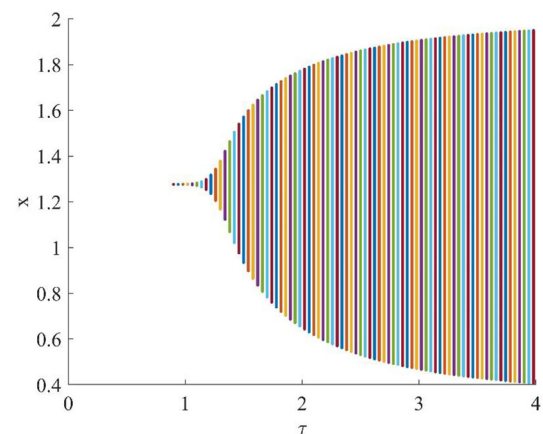
FIGURE 6

One parameter bifurcation diagram with respect to τ.

death rate of free viruses (ε) and he death rate of uninfected cells (d) can also decrease the basic reproduction number (R_0). Therefore, in order to keep $R_0 < 1$, the patient can decrease parameters b, k, and ξ or increase a, ε, and d to achieve the purpose of treatment.

In order to verify the accuracy of the aforementioned theoretical analysis, the numerical simulations were carried out. The simulation results showed that when $R_0 < 1$ and $\tau < 1.2$,

the infection-free equilibrium E_0 is asymptotically stable, which indicates that the disease will disappear. When $R_1 > 1$ and $\tau < 1.2$, the infection equilibrium E_1 is asymptotically stable, which indicates that the disease could be mitigated and will lead to a lower infectious class over a period. However, with the increase in τ, the uninfected cells, infected cells, free viruses, CTL levels, and ALT levels oscillate periodically around the infection
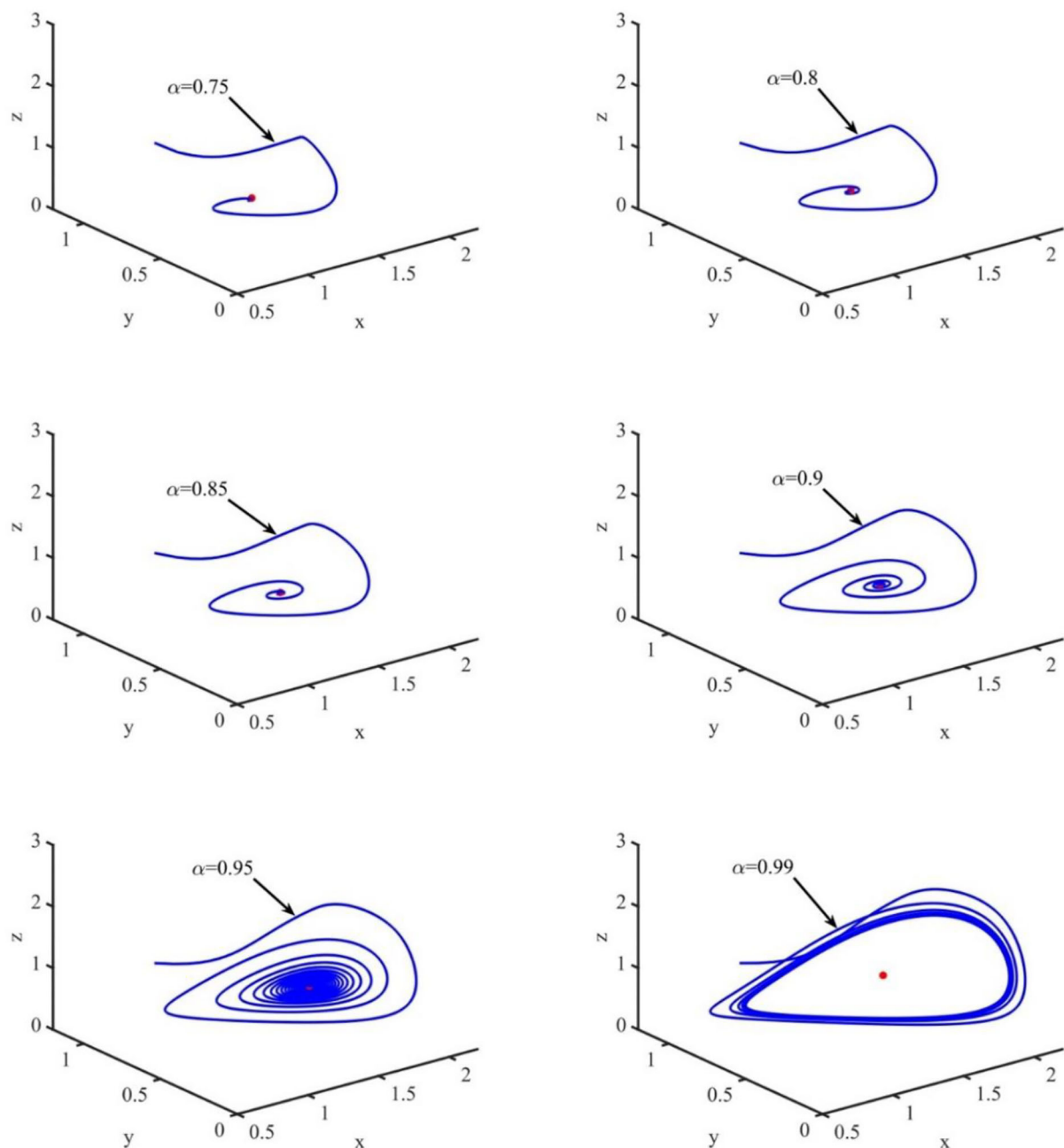
FIGURE 7
Phase portraits of the xyz-space with $\tau = 1.2$ and $\alpha = 0.75$, $\alpha = 0.80$, $\alpha = 0.85$, $\alpha = 0.90$, and $\alpha = 0.99$.

equilibrium E_1, and the phase portrait is a stable limit cycle, which around the infection equilibrium E_1 indicate that the disease would be out of control. Furthermore, the simulations also indicated that the order can significantly affect the stability of the system. For example, if the order is in the range of 0.75–0.95, the phase portraits converge to the infection equilibrium E_1, and when $\alpha = 0.99$, the phase portrait is a stable limit cycle.

Therefore, time delay and fractional order are necessary factors that should be considered in modeling HBV infection and for researching dynamic characteristics. Although the process of HBV infection is more complicated than is established in this study, we believe that the model and analysis can play an important role in improving the HBV treatment regimen.

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 authors.

Author contributions

DS: conceptualization, methodology, software, and writing—reviewing and editing. JL: software, methodology, and writing. XS: supervision, writing—review editing. GP: writing—review editing. All authors contributed to the article and approved the submitted version.

Funding

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. WHO. *Hepatitis B.* (2021). Available online at: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b (accessed June 24, 2022).

2. Li M, Zu J. The review of differential equation models of HBV infection dynamics. *J Virol Methods.* (2019) 266:103–13. doi: 10.1016/j.jviromet.2019.01.014

3. James Peter O, Ojo MM, Viriyapong R, Abiodun Oguntolu F. Mathematical model of measles transmission dynamics using real data from Nigeria. *J Diff Equ Appl.* (2022) 28:753–70. doi: 10.1080/10236198.2022.2079411

4. Peter OJ, Qureshi S, Yusuf A, Al-Shomrani M, Idowu AA. A new mathematical model of COVID-19 using real data from Pakistan. *Result Phys.* (2021) 24:104098. doi: 10.1016/j.rinp.2021.104098

5. Ojo MM, Peter OJ, Goufo EFD, Panigoro HS, Oguntolu FA. Mathematical model for control of tuberculosis epidemiology. *J Appl Math Comput.* (2022). doi: 10.1007/s12190-022-01734-x

6. Elaiw AM, AlShamrani NH. Global stability of humoral immunity virus dynamics models with nonlinear infection rate and removal. *Nonlinear Anal Real World Appl.* (2015) 26:161–90. doi: 10.1016/j.nonrwa.2015.05.007

7. Nowak MA BS, Hill AM. Viral dynamics in hepatitis B virus infection. *Proc Natl Acad Sci U S A.* (1996) 93:4398–402. doi: 10.1073/pnas.93.9.4398

8. Su Y, Sun D, Zhao L. Global analysis of a humoral and cellular immunity virus dynamics model with the Beddington–DeAngelis incidence rate. *Math Methods Appl Sci.* (2015) 38:2984–93. doi: 10.1002/mma.3274

9. Manna K, Chakrabarty SP. Chronic hepatitis B infection and HBV DNA-containing capsids: modeling and analysis. *Commun Nonlinear Sci Numer Simul.* (2015) 22:383–95. doi: 10.1016/j.cnsns.2014.08.036

10. Li J, Wang K, Yang Y. Dynamical behaviors of an HBV infection model with logistic hepatocyte growth. *Math Comput Model.* (2011) 54:704–11. doi: 10.1016/j.mcm.2011.03.013

11. Prakash M, Balasubramaniam P. Bifurcation analysis of macrophages infection model with delayed immune response. *Commun Nonlinear Sci Numer Simul.* (2016) 35:1–16. doi: 10.1016/j.cnsns.2015.10.012

12. Sun C, Cao Z, Lin Y. Analysis of stability and Hopf bifurcation for a viral infectious model with delay. *Chaos, Soliton, and Fractals.* (2007) 33:234–45. doi: 10.1016/j.chaos.2005.12.029

13. Zhang Z, Kundu S, Tripathi JP, Bugalia S. Stability and Hopf bifurcation analysis of an SVEIR epidemic model with vaccination and multiple time delays. *Chaos, Solitons, and Fractals.* (2020) 131:109483. doi: 10.1016/j.chaos.2019.109483

14. Babasola O, Oshinubi K, Peter OJ, Onwuegbuche FC, Khan I. Time-delayed modelling of the COVID-19 dynamics with a convex incidence rate. *Res Sq.* (2022). doi: 10.21203/rs.3.rs-1814397/v1

15. Chinnathambi R, Rihan FA. Stability of fractional-order prey–predator system with time-delay and Monod–Haldane functional response. *Nonlinear Dyn.* (2018) 92:1637–48. doi: 10.1007/s11071-018-4151-z

16. Sun D, Lu L, Liu F, Duan L, Wang D, Xiong J. Analysis of an improved fractional-order model of boundary formation in the *Drosophila* large intestine dependent on Delta–Notch pathway. *Adv Differ Equ.* (2020) 2020:377. doi: 10.1186/s13662-020-02836-1

17. Mondal S, Lahiri A, Bairagi N. Analysis of a fractional order eco-epidemiological model with prey infection and type 2 functional response. *Math Methods Appl Sci.* (2017) 40:6776–89. doi: 10.1002/mma.4490

18. Almeida R. Analysis of a fractional SEIR model with treatment. *Appl Math Lett.* (2018) 84:56–62. doi: 10.1016/j.aml.2018.04.015

19. George Maria Selvam A, Abraham Vianny D. Bifurcation and dynamical behaviour of a fractional order Lorenz-Chen-Lu like chaotic system with discretization. *J Phys Conf Ser.* (2019) 1377:012002. doi: 10.1088/1742-6596/1377/1/012002

20. Wang X, Wang Z, Huang X, Li Y. Dynamic analysis of a delayed fractional-order SIR model with saturated incidence and treatment functions. *Int J Bifurc Chaos.* (2019) 28:1850180. doi: 10.1142/S0218127418501808

21. Li H, Huang C, Li T. Dynamic complexity of a fractional-order predator–prey system with double delays. *Phys A Stat Mech Appl.* (2019) 526:120852. doi: 10.1016/j.physa.2019.04.088

22. Balci E, Öztürk I, Kartal S. Dynamical behaviour of fractional order tumor model with Caputo and conformable fractional derivative. *Chaos, Solitons, and Fractals.* (2019) 123:43–51. doi: 10.1016/j.chaos.2019.03.032

23. Ye Y, Kou C. Stability analysis for a fractional HIV-1 model with time delay. In *4th International Conference on Biomedical Engineering and Informatics.* IEEE, New York, NY (2011).

24. Nwajeri UK, Fadugba D, Ohaeri EO, Oshinubi K, Ogunrinde RR, Ogunrinde PB. *Co-Dynamic Model of Drug Trafficking and Money Laundering Coupled with Atangana-Baleanu Derivative.* (2022). doi: 10.2139/ssrn.4165832

25. Peter OJ. Transmission Dynamics of Fractional Order Brucellosis Model Using Caputo–Fabrizio Operator. *Int. J. Diff. Equ.* (2020) 2020:2791380. doi: 10.1155/2020/2791380

26. Peter OJ, Shaikh AS, Ibrahim MO, Nisar KS, Baleanu D, Khan I, et al. Analysis and dynamics of fractional order mathematical model of COVID-19 in Nigeria using Atangana–Baleanu operator. *Comput Mater Contin.* (2021) 66:1823–48. doi: 10.32604/cmc.2020.012314

27. Ogunrinde RB, Nwajeri UK, Fadugba SE, Ogunrinde RR, Oshinubi KI. Dynamic model of COVID-19 and citizens reaction using fractional derivative. *Alex Eng J.* (2021) 60:2001–12. doi: 10.1016/j.aej.2020.09.016

28. Peter OJ, Yusuf A, Oshinubi K, Oguntolu FA, Lawal JO, Abioye AI, et al. Fractional order of pneumococcal pneumonia infection model with Caputo Fabrizio operator. *Result Phys.* (2021) 29:104581. doi: 10.1016/j.rinp.2021.104581

29. Diethelm K, A. fractional calculus based model for the simulation of an outbreak of dengue fever. *Nonlinear Dyn.* (2012) 71:613–9. doi: 10.1007/s11071-012-0475-2

30. González-Parra G, Arenas AJ, Chen-Charpentier BM. A fractional order epidemic model for the simulation of outbreaks of influenza A(H1N1). *Math Methods Appl Sci.* (2014) 37:2218–26. doi: 10.1002/mma.2968

31. Danane J, Allali K, Hammouch Z. Mathematical analysis of a fractional differential model of HBV infection with antibody immune response. *Chaos, Solitons, and Fractals.* (2020) 136:109787. doi: 10.1016/j.chaos.2020.109787

32. Gao F, Li X, Li W, Zhou X. Stability analysis of a fractional-order novel hepatitis B virus model with immune delay based on Caputo–Fabrizio derivative. *Chaos, Solitons, and Fractals.* (2021) 142:110436. doi: 10.1016/j.chaos.2020.110436

33. Simelane SM, Dlamini PG, A. fractional order differential equation model for Hepatitis B virus with saturated incidence. *Result Phys.* (2021) 24:104114. doi: 10.1016/j.rinp.2021.104114

34. Su Y, Sun D. Optimal control of anti-HBV treatment based on combination of traditional Chinese medicine and western medicine. *Biomed Signal Process Control.* (2015) 15:41–8. doi: 10.1016/j.bspc.2014.09.007

35. Adebisi AF, Ojurongbe TA, Okunlola KA, Peter OJ. Application of Chebyshev polynomial basis function on the solution of Volterra integro-differential equations using Galerkin method. *Math Comput Sci.* (2021) 2:41–51. doi: 10.30511/mcs.2021.540133.1047

36. Christie I, Adebayo T, Peter O, Folasade A. Numerical solution of two-dimensional Fredholm integro-differential equations by Chebyshev integral operational matrix method. *J Appl Math Comput Mech.* (2022) 21:29–40. doi: 10.17512/jamcm.2022.1.03

37. Sun D, Liu F. Analysis of a new delayed HBV model with exposed state and immune response to infected cells and viruses. *Biomed Res Int.* (2017) 2017:7805675. doi: 10.1155/2017/7805675

Check for updates

Modeling vaccination and control strategies for outbreaks of monkeypox at gatherings

Pei Yuan[1,2†], Yi Tan[1,2†], Liu Yang[1,2,3†], Elena Aruffo[1,2†],
Nicholas H. Ogden[2,4], Jacques Bélair[2,5], Julien Arino[2,6],
Jane Heffernan[2,7], James Watmough[2,8], Hélène Carabin[9,10,11,12]
and Huaiping Zhu[1,2]*

[1]Laboratory of Mathematical Parallel Systems (LAMPS), Department of Mathematics and Statistics, York University, Toronto, ON, Canada, [2]Canadian Centre for Diseases Modeling (CCDM), York University, Toronto, ON, Canada, [3]School of Mathematics and Statistics, Northeast Normal University, Changchun, China, [4]Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, Saint-Hyacinthe, QC, Canada, [5]Département de Mathématiques et de Statistique, Université de Montréal, Montréal, QC, Canada, [6]Department of Mathematics, University of Manitoba, Winnipeg, MB, Canada, [7]Department of Mathematics and Statistics, York University, Toronto, ON, Canada, [8]Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada, [9]Département de Pathologie et Microbiologie, Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, QC, Canada, [10]Département de médecine sociale et préventive, École de santé publique de l'Université de Montréal, Montréal, QC, Canada, [11]Centre de Recherche en Santé Publique (CReSP) de l'université de Montréal et du CIUSS du Centre Sud de Montréal, Montréal, QC, Canada, [12]Groupe de Recherche en Épidémiologie des Zoonoses et Santé Publique (GREZOSP), Université de Montréal, Saint-Hyacinthe, QC, Canada

Background: The monkeypox outbreak in non-endemic countries in recent months has led the World Health Organization (WHO) to declare a public health emergency of international concern (PHEIC). It is thought that festivals, parties, and other gatherings may have contributed to the outbreak.

Methods: We considered a hypothetical metropolitan city and modeled the transmission of the monkeypox virus in humans in a high-risk group (HRG) and a low-risk group (LRG) using a Susceptible-Exposed-Infectious-Recovered (SEIR) model and incorporated gathering events. Model simulations assessed how the vaccination strategies combined with other public health measures can contribute to mitigating or halting outbreaks from mass gathering events.

Results: The risk of a monkeypox outbreak was high when mass gathering events occurred in the absence of public health control measures. However, the outbreaks were controlled by isolating cases and vaccinating their close contacts. Furthermore, contact tracing, vaccinating, and isolating close contacts, if they can be implemented, were more effective for the containment of monkeypox transmission during summer gatherings than a broad vaccination campaign among HRG, when accounting for the low vaccination coverage in the overall population, and the time needed for the development of the immune responses. Reducing the number of attendees and effective contacts during the gathering could also prevent a burgeoning outbreak, as could restricting attendance through vaccination requirements.

Conclusion: Monkeypox outbreaks following mass gatherings can be made less likely with some restrictions on either the number and density of attendees in the gathering or vaccination requirements. The ring vaccination strategy

inoculating close contacts of confirmed cases may not be enough to prevent potential outbreaks; however, mass gatherings can be rendered less risky if that strategy is combined with public health measures, including identifying and isolating cases and contact tracing. Compliance with the community and promotion of awareness are also indispensable to containing the outbreak.

Introduction

Monkeypox, a zoonosis, has been recorded since early May 2022 in at least 30 non-endemic countries including Spain, the United States, Germany, the United Kingdom, France, and Canada (1). As of 21 July 2022, the cumulative number of confirmed cases exceeded 15,000 globally (2). On 23 July 2022, the World Health Organization (WHO) declared monkeypox a Public Health Emergency of International Concern (PHEIC) due to outbreaks in multiple countries and continents (3). The Public Health Agency of Canada (PHAC) reported 1,410 cases of monkeypox as of 14 October 2022, mostly occurring in Quebec, Ontario, and British Columbia (4). The unusual outbreak emerged in non-endemic areas of the world associated with transmission among gay, bisexual, and other men who have sex with men (gbMSM) (5). Although at the time of writing this paper, the epidemic was declining, there remains a pressing need to understand the epidemic and potential control methods (5–7).

Monkeypox virus, which is closely related to smallpox, is an enveloped double-stranded DNA virus, with two clades, the Central African clade and the West African clade (7, 8). The former is more virulent with reported fatality rates in Africa of 10% for the Central Africa clade and 3.6% for the West African clade (7). The incubation period ranges from 5 to 21 days, after which infected individuals may initially have flu-like symptoms, then, 1–3 days later, a characteristic skin rash develops. The recovery period may take 2–4 weeks (8). In the recent outbreak, there are atypical clinical observations. The majority of the patients are gbMSM who reported genital lesions which subsequently develop into skin lesions on other body sites, although with more limited distribution than reported in the previous outbreaks (9).

In Canada, control is based on vaccines and non-pharmaceutical interventions including recommendations for testing and isolation of cases, and, where possible, tracing of contacts (10). In June 2022, the National Advisory Committee on Immunization (NACI) released a guideline on using an orthopoxvirus (Imvamune®) vaccine with potential efficacy against monkeypox (11). The guideline recommends pre-exposure prophylaxis (PrEP) vaccination for adults at high risk of exposure (occupational or otherwise) and also post-exposure prophylaxis (PEP).

After 2 years of restrictions on gatherings due to the control and prevention of the COVID-19 pandemic, mass gatherings related to festivals and ceremonies are now allowed with no attendance limitations (12). In many Canadian provinces, local festivals recorded attendance close to the pre-pandemic level (12–15), which has led to concerns about the spread and possible outbreaks of monkeypox. WHO also expressed concerns that more infections could arise in Europe and elsewhere (16) due to private and social gatherings during festivals, parties, and holidays. In fact, in the United States, many cases were reported linked to large social gatherings, such as pride events, pool parties, and bathhouses (17, 18). Consequently, it is essential to assess the effect of gathering events on monkeypox transmission to inform public health on the most effective control measures.

Transmission risk at a gathering is mainly associated with the gathering size (19) and is proportional to the population density at the gathering place (20). Using an individual-based model, Moritz et al. (21) showed that for a mass gathering event (MGE) with 200,000 participants, there is a 23.6% increase in positive cases attributed to MGE for the transmission of COVID-19. Also, the effect of increased density of contacts during Hajj was estimated to generate a 78-fold increase in meningococcal infection that impacts not only pilgrims but also the local population (22).

To investigate the dynamics of monkeypox and provide information to public health for prevention and control, especially at gatherings, we established a SEIR modeling framework to assess the effect of the vaccination and other control methods. The vaccination in a high-risk group and ring vaccination strategy along with testing and isolation of cases and contact tracing, as well as the possible interventions during gathering events, are also considered. We mainly focused on assessing the effectiveness of public health control measures, including preventive vaccination or vaccination post-exposure, to simulate the scenarios of gathering with different numbers of attendees and different levels of interventions to inform public health decision-making. Our findings suggested that reactive

ring vaccination may itself not be enough; however, if close contacts of cases can be identified, vaccinated, and isolated, an outbreak after MGEs may be prevented.

Methods

Modeling overview

The vast majority of reported cases occurring in the recent outbreaks in non-endemic areas have been linked to specific high-risk locations and populations. Hence, to better capture the infection dynamics within different risk settings, we considered the population to be divided into two subgroups: a low-risk population (LRG), which is defined as individuals who behave in such a way that their possibility of becoming infected is reduced, and a high-risk population (HRG), which is defined as individuals whose behavior makes them at higher risk of acquiring the infection. For simplicity, henceforth, we use subscript 1 for the LRG and subscript 2 for the HRG. Those two groups interact between and within groups as represented by a contact matrix (c_{ij}, $i, j = 1, 2$), defined using the assumptions in Yuan et al. (23). We assumed that there is no movement of population between the risk groups unless there is a gathering event such that a proportion of LRG people may become part of the HRG.

The infection dynamic follows the SEIR framework, which is extended to include the prodromal stage, vaccinated (partially, fully), and quarantined (tested and confirmed, vaccinated, and susceptible). Susceptible (S) individuals become infected, and move to the exposed compartment (E), after encountering an infectious individual from either LRG or HRG, assuming that the latter group is with higher susceptibility than the first group. After a latent period, the prodromal stage (P) begins, and during this phase no symptoms are apparent, but the individual can shed the virus (24). This period is then followed by the symptomatic infectious stage (I) and then recovery (R) occurs. Infectious individuals with symptoms may be tested, then quarantined (Q), while their contacts, which might be susceptible, exposed, or pre-symptomatic, can be vaccinated and quarantined (Q_s or Q_v) to prevent any further spread of the infection. Although the model does not include demographics, we assumed that infection-related death might occur among infectious individuals. Isolated individuals who develop infection will remain isolated until recovery. Individuals in both prodromal and symptomatic infectious stages can transmit the infection; however, infections in the prodromal phase are assumed to be less infectious than those in the symptomatic phase. The population structure and flow diagram of the disease are shown in Figure 1.

We included the current vaccination process in Canada (11) to explore its effectiveness by assuming that only high-risk people and close contacts of confirmed cases will be vaccinated.

We also examined the public health measures of testing and isolating cases. Model assumptions, variables, and parameters are summarized in Tables 1–5, and the model equations are presented in Supplementary material S1.

Transmission

We used the assumptions of Yuan et al. (23) on the transmission of the monkeypox virus. The probability of transmission per contact was assumed to vary between 12.2 and 24.5% among the HRG and between 0.37 and 0.74%, among LRG, as calculated from the basic reproduction number R_0 derived from our simplified model without public health control measures (see Supplementary material S2 for details) and all the other parameters being fixed (Tables 3, 4).

Public health interventions

On 10 June 2022, NACI issued interim guidance on the use of Imvamune® in the context of monkeypox outbreaks in Canada (11). Imvamune, initially developed for the prevention of smallpox, is a two-dose vaccine with the second dose administered 28 days after the first one. The immune response is detectable by week 2 after the first dose and peaked at week 6 after dose 2 in a randomized, open-label trial designed to compare the effectiveness of Imvamune with the second-generation replicating smallpox vaccine (11). Given the recent emergence of cases and the use of vaccines, there are no available data indicating the effectiveness of Imvamune vaccination against monkeypox infection; however, studies of vaccine effectiveness (VE) of smallpox vaccine may provide a general estimation. In the context of PEP, the median effectiveness in preventing smallpox disease with vaccination at 1–3 days after exposure was estimated at 80% (30). There are a lot of uncertainties about the effectiveness of Imvamune against monkeypox infection, although some observational studies suggested an efficacy of about 85% among fully vaccinated individuals (11). Hence, in our model, we assumed that the efficacy of the first and second doses ranges between 40 and 60%, and 70 and 85%, respectively.

For simplification, given the extensive vaccine campaign against smallpox until 1972, we assumed that all individuals born before that year are fully vaccinated and therefore protected in the model. We defined here that partially and fully vaccinated represent the individuals who received one or two doses of vaccine, respectively.

PEP with contact tracing (PEPCT)

Post-Exposure Prophylaxis with Contact Tracing (PEPCT) in our model means that individuals (without symptoms) with
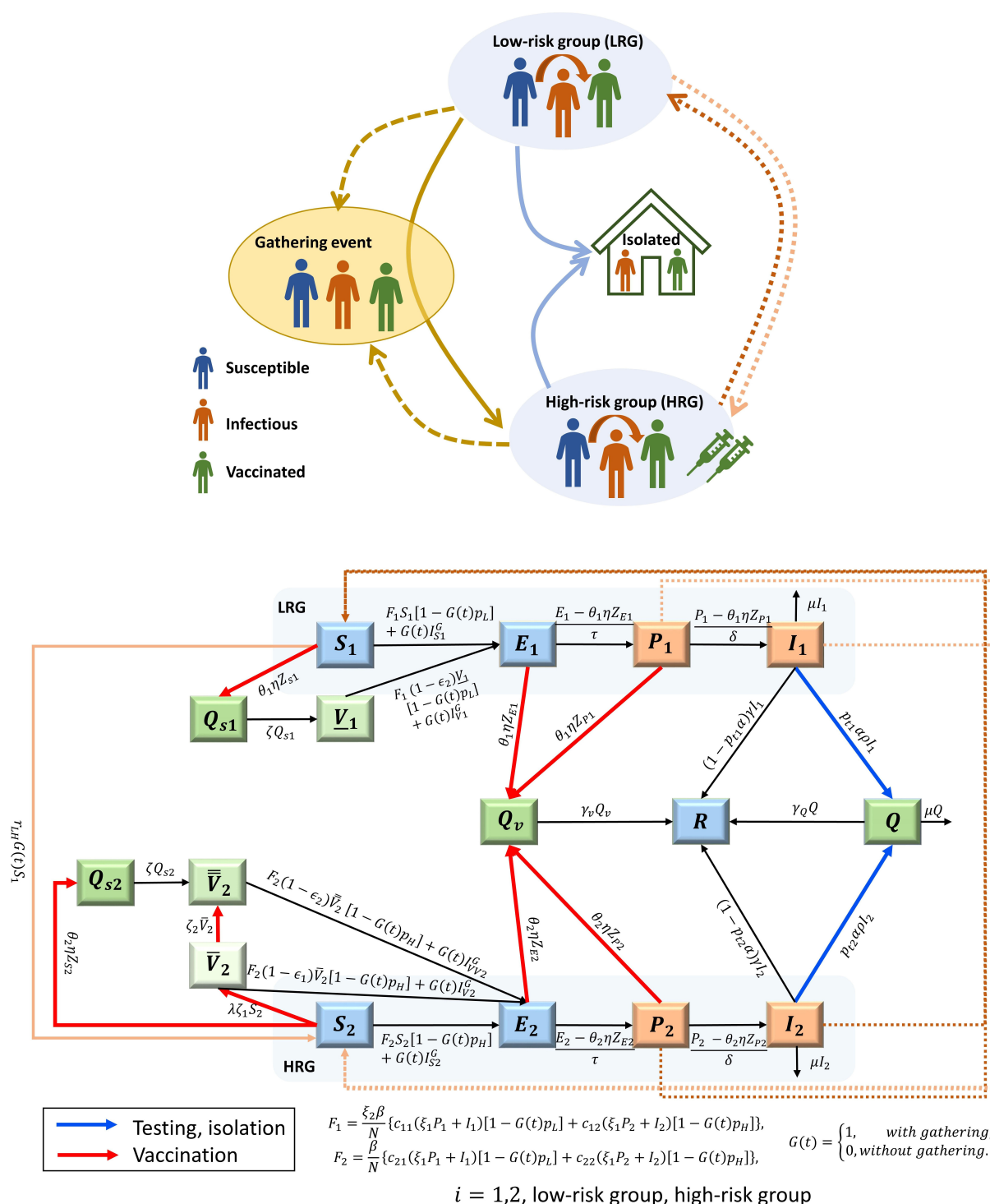
FIGURE 1
Schematic diagram **(A)** and flow chart **(B)** of the MPX transmission among the population classified with low-risk group (LRG, $i = 1$) and high-risk group (HRG, $i = 2$), considering the gathering event. In groups i: Susceptible (S_i), Exposed (E_i), infectious, prodromal phase (P_i), infectious, acute phase, with rash symptom (I_i), Recovered R; Quarantined (Q), Quarantined and vaccinated (Q_v); Quarantined and Susceptible (Q_{si}); partially vaccinated HRG individuals, $\overline{V}_2$, fully vaccinated HRG individuals $\overline{\overline{V}}_2$, fully vaccinated LRG individuals (born before 1972), $\underline{V}_1$. The individuals in the orange compartments are infectious, while those in dark green compartments are quarantined and hence not involved in the transmission. The compartments of partially and fully vaccinated are presented with light and middle green colors respectively. The arrow with red color represents the process of vaccinating the HRG individuals and the exposures from contact tracing, while the arrow with blue color represents the testing and isolation of symptomatic infections. The dash lines show the transmission routes between the LRG and HRG. The description of the parameters shown in the flow diagram can be found in Tables 3, 4.

TABLE 1 Model assumptions.

Model assumptions

Population classification	• The human population is divided into two groups, low-risk group (LRG, i = 1) and high-risk group (HRG, i = 2).
Demographic	• Immigration, birth, and natural death of the population are ignored. The mortality from infection is considered.
Monkeypox transmission	• Infectious individuals in both prodromal phase (with lower infectiousness ξ_1, assumed to be 75% of that during the rash stage) and acute phase can transmit the virus.
	• The tested positive, confirmed and isolated individuals and the vaccinated and quarantined individuals are not involved in the transmission.
	• The susceptibility of LRG individuals is lower than that of HRG individuals, with a scaling factor ξ_2 (assumed to be 0.05).
	• The contacts among LRG individuals mainly come from the LRG individuals, with the proportion k_1 (assumed to be 0.6) of the contact rate baseline in the human population.
	• The contact rate in the HRG is higher than that in the LRG, by a factor k_2 of the contact rate baseline (assumed to be 1.3).
	• The contact between HRG individuals and LRG individuals is low, accounting for $(1 - k_1)$ of the contact rate baseline.
	• The vertical transmission is ignored.
Testing and isolation	• Individuals with rash symptoms will go to hospital seeking medical help and then be tested, confirmed and isolated and this process may take $1/\rho$ days.
	• Proportion of confirmed infections (α) will comply with the isolation strategy.
Vaccination	• The two-dose vaccine considered is Imvamune®, and a second dose is offered 28 days after the first dose.
	• Individuals receiving the first dose after 2 weeks when the immune response is detectable, and the second dose of Imvamune after 6 weeks when the immune response peaked, are considered as partially vaccinated and fully vaccinated, respectively. Given the vaccination campaign against smallpox, individuals born before 1972 are considered to be fully vaccinated.
	• Post-Exposure Prophylaxis with contact tracing (PEPCT): Individuals (without symptom) with high-risk exposures to a confirmed case of monkeypox may be vaccinated and then quarantined.
	• Pre-Exposure Prophylaxis in HRG (PrEPH): Susceptible individuals in HRG may be vaccinated.
	• The effectiveness of the first and second dose is assumed to be ϵ_1 and ϵ_2, respectively.
Gathering event	• During the period of gathering event, the contact rate between LRG and HRG and within HRG may increase κ_1 and κ_2 times, respectively.
	• Some susceptible individuals in the LRG may transit to the HRG.

exposure to any confirmed case of monkeypox may be traced, vaccinated, and then quarantined. The effect of the vaccine in PEPCT is not significant as the individuals with exposure will be required to isolate and quarantine until fully recovered.

The proportion of PEPCT is represented by $\theta_i, (i = 1, 2)$, which should be the product of the vaccination proportion of the traced close contact and the proportion of contacts that could be traced. Also, it takes time to trace and isolate which in fact can be modeled by the parameter $1/\eta$, the average time from exposure to becoming vaccinated and isolated.

We denote by $M_{I_1}(t)$ and $M_{I_2}(t)$ the number of new daily confirmed monkeypox cases in the LRG and HRG at time t, respectively, and are defined as follows:

$$M_{I_1}(t) = \rho I_1(t)$$
$$M_{I_2}(t) = \rho I_2(t)$$

where $1/\rho$ is the average number of days infectious individuals spend between showing rash symptoms and being tested and confirmed.

Following Yuan et al. (23), the number of close contacts of newly confirmed cases in the LRG, who are in the exposed state $Z_{E1}(t)$ and the prodromal phase $Z_{P1}(t)$ at time t can be calculated as

$$Z_{E1}(t) = [\xi_2\beta\sum_{i=1}^{2} c_{2i}M_{I_1}(t)\frac{c_{11}}{c_{11} + c_{12}}$$
$$+ \beta\sum_{i=1}^{2} c_{2i}M_{I_2}(t)\frac{c_{21}}{c_{21} + c_{22}}]\frac{\tau}{\tau + \delta},$$

$$Z_{P1}(t) = [\xi_2\beta\sum_{i=1}^{2} c_{2i}M_{I_1}(t)\frac{c_{11}}{c_{11} + c_{12}}$$
$$+ \beta\sum_{i=1}^{2} c_{2i}M_{I_2}(t)\frac{c_{21}}{c_{21} + c_{22}}]\frac{\delta}{\tau + \delta}.$$

Similarly, we obtained the number of close contacts of newly confirmed cases in the HRG, who are in the exposed state $Z_{E2}(t)$ and the prodromal phase $Z_{P2}(t)$, as

$$Z_{E2}(t) = [\xi_2\beta\sum_{i=1}^{2} c_{2i}M_{I_1}(t)\frac{c_{12}}{c_{11} + c_{12}}$$
$$+ \beta\sum_{i=1}^{2} c_{2i}M_{I_2}(t)\frac{c_{22}}{c_{21} + c_{22}}]\frac{\tau}{\tau + \delta},$$

$$Z_{P2}(t) = [\xi_2\beta\sum_{i=1}^{2} c_{2i}M_{I_1}(t)\frac{c_{12}}{c_{11} + c_{12}}$$
$$+ \beta\sum_{i=1}^{2} c_{2i}M_{I_2}(t)\frac{c_{22}}{c_{21} + c_{22}}]\frac{\delta}{\tau + \delta}.$$

TABLE 2 Variables used in the modeling of monkeypox transmission and their assumed initial values.

Variables	Description	Initial value		Ref.
$S_i(t)$	Number of susceptible individuals in group i at day t	$i = 1$	$i = 2$	Assumed
		$(1 - p_{HRG})N_p$	$p_{HRG}N_p - 3$	
$E_i(t)$	Number of exposed individuals in group i at day t	$i = 1$	$i = 2$	Assumed
		0	1	
$P_i(t)$	Number of infectious individuals in the prodromal phase in group i at day t	$i = 1$	$i = 2$	Assumed
		0	1	
$I_i(t)$	Number of infectious individuals with rash symptoms in the acute phase in group i at day t	$i = 1$	$i = 2$	Assumed
		0	1	
$Q(t)$	Number of tested, confirmed and isolated symptomatic individuals at day t	0		Assumed
$R(t)$	Number of recovered individuals at day t	0		Assumed
$\underline{V_1}(t)$	Number of fully vaccinated individuals in LRG at day t	$(1 - p_{HRG}) \times 38.4\% \times N_p$		(25)
$\underline{V_2}(t)$	Number of partially vaccinated individuals in HRG at day t	0		Assumed
$\overline{\overline{V}}_2(t)$	Number of fully vaccinated individuals in HRG at day t	$p_{HRG} \times 38.4\% \times N_p$		(25)
$Q_v(t)$	Number of vaccinated and quarantine individuals who has been exposed to the confirmed cases at day t	0		Assumed
$Q_{S_i}(t)$	Number of susceptible individuals in population i, i =1,2 who has been exposed to the confirmed cases at day t, and consequently vaccinated and isolated.	0		Assumed

Also, the number of close contacts of newly confirmed cases that are susceptible in the LRG and HRG is

$$Z_{S1}(t) = (1 - \xi_2\beta) \sum_{i=1}^{2} c_{2i}M_{I_1}(t) \frac{c_{11}}{c_{11} + c_{12}}$$
$$+ (1 - \beta) \sum_{i=1}^{2} c_{2i}M_{I_2}(t) \frac{c_{21}}{c_{21} + c_{22}},$$
$$Z_{S2}(t) = (1 - \xi_2\beta) \sum_{i=1}^{2} c_{2i}M_{I_1}(t) \frac{c_{12}}{c_{11} + c_{12}}$$
$$+ (1 - \beta) \sum_{i=1}^{2} c_{2i}M_{I_2}(t) \frac{c_{22}}{c_{21} + c_{22}}.$$

PrEPH in HRG (PrEPH)

Pre-exposure prophylaxis in HRG (PrEPH) refers to administering the vaccine to individuals at high risk of exposure to the virus. Hence, in our model, vaccination is only administered to those in HRG. Since people born before 1972 have been vaccinated with the smallpox vaccine, therefore, in the initial state in the model, a proportion of the LRG and HRG populations is considered fully vaccinated for simplicity.

Testing and isolation

Testing and isolation are crucial steps to detect monkeypox infections and stop the virus from spreading. Individuals with clinical illnesses where monkeypox is suspected should be tested and proceeded for self-isolation before the negative test result is received, while individuals with positive test results should isolate at home until they recover. However, it can take several days from when infected individuals develop symptoms to seek medical help and then get tested, with large variations depending on the individuals' behaviors. In addition, the recovery period takes 2–4 weeks, and with respect to the isolation strategy may not be total, hence compliance with the isolation strategy for those tested and confirmed monkeypox cases are also included in our model.

Gatherings

Gathering events, here refer to mass gatherings, are defined by the WHO as "more than a specified number of persons at a specific location for a specific purpose for a defined period of time" (31). Gatherings may contribute significantly to the

TABLE 3 Parameters used in the modeling of monkeypox transmission (assumed).

Parameter	Definition	Value (range)	Ref.
Demographic related			
p_{HRG}	Proportion of the population that is in the HRG	0.035	Assumed
N_p	The total number of populations in the hypothetical city	5,000,000	Assumed
Transmission related			
$1/\gamma_Q$	Average number of days of recovery needed for isolated individuals, days	28	Assumed
$1/\gamma_v$	Average number of days of recovery needed for exposed individuals vaccinated and quarantined, days	21	Assumed
ξ_1	Scaling factor of infectiousness of infected in the prodromal phase compared to infections with rash symptoms	0.75	Assumed
ξ_2	Scaling factor of susceptibility of LRG individuals compared to HRG individuals	0.03	Assumed
β	Probability of transmission per contact among HRG	(0.122, 0.245)	Assumed
k_1	Proportion of contacts within the LRG in overall contacts	0.6	Assumed
k_2	Scaling factor of contact rate among HRG compared to baseline contact	1.3	Assumed
Vaccine related			
ϵ_1	The effectiveness of first dose of Imvamune against monkeypox infection	40% (40–60%)	Assumed
ϵ_2	The effectiveness of second dose of Imvamune against monkeypox infection	80% (70–85%)	Assumed
Public health control measures related			
$1/\lambda$	The average time to achieve the vaccination coverage in HRG, days	40 (30–90)	Assumed
p_v	The vaccination coverage in HRG	0.6 (0.1–0.9)	Assumed
θ_1	The vaccination proportion of individuals in LRG who are the close contact with the confirmed cases	0.35 (0.1–0.6)	Assumed
θ_2	The vaccination proportion of individuals in HRG who are the close contact with the confirmed cases	0.6 (0.3–0.8)	Assumed
r_{LH}	The daily transition rate of the low-risk susceptible individuals to the high-risk, 1/days	0.0005 (0.0001–0.001)	Assumed
c_G	Number of effective contacts during the gathering, persons	30 (10–50)	Assumed
p_{t1}	Proportion of symptomatic individuals in LRG who has been tested and isolated	0.3 (0.1–0.5)	Assumed
p_{t2}	Proportion of symptomatic individuals in HRG who has been tested and isolated	0.5 (0.3–0.7)	Assumed
α	Proportion of individuals who comply with the isolation strategy	0.4 (0.3–0.6)	Assumed

spread of infectious diseases, as was extensively studied during the SARS-CoV-2 pandemic (19–21).

Modeling of gathering event in the absence of a specific intervention

Not only will individuals from HRG attend the gathering event, but also individuals in the LRG will join and some of them may transit to the HRG; therefore, facilitating virus spreading and posing the risk of a possible outbreak. In our model, we denote $G(t)$ as an indicator parameter if there is a gathering event.

$$G(t) = \begin{cases} 1, & \text{with gathering event,} \\ 0, & \text{without gathering event.} \end{cases}$$

The daily transition rate from the low-risk susceptible individuals to the high-risk ones is r_{LH} days^{-1}. During the period of gathering events, we assume that the total number of attendees is N_G and the proportion of attendees from LRG and HRG is p_{GL} and p_{GH}, respectively. The proportion of LRG and HRG individuals attending the gathering event on day t is calculated as

$$p_L(t) = \frac{p_{GL}N_G}{N_1(t)}, \quad p_H(t) = \frac{p_{GH}N_G}{N_2(t)},$$

respectively, and the number of infectious attendees during the gathering on day t is given by

$$N_{GI}(t) = p_L(t)\left[P_1(t) + I_1(t)\right] + p_H(t)\left[P_2(t) + I_2(t)\right],$$

TABLE 4 Parameters used in the modeling of monkeypox transmission (from literature).

Parameter	Definition	Value (range)	Ref.
Demographic related			
p_{50+}	The proportion of individuals who were born before 1972 and has been fully vaccinated to the total populations	38.4%	(25)
Transmission related			
τ	Average incubation period of MPX, days	13	(8)
δ	Average number of days from prodromal phase to acute phase, days	3	(8)
μ	Daily disease induced death rate, 1/days	3.6%/21	(7)
$1/\gamma$	Average number of days of recovery needed for infectious individuals with rash symptoms, days	21	(8)
c_0	Baseline contact rate among the overall population, per day	10.8	(26)
Vaccine related			
$1/\zeta_1$	The time from the first dose administered to the immune response start, days	14	(11)
$1/\zeta_2$	The time from the immune response started after first dose to immune response peak after second dose, days	56	(11)
Gathering event related			
T_1	The starting time of the gathering event	Jul. 20, 2022	(27)
T_2	The end time of the gathering event	Aug. 1, 2022	(27)
Public health control measures related			
T_v	The starting time of the vaccination strategy conducted	June 10, 2022	(28)
$1/\eta$	The average time of the close contact of confirmed cases from the exposed to be traced and then vaccinated, days	7 (1–14)	(11)
$1/\rho$	The average days from when infected individuals develop symptoms to seek medical help and then get tested, days	7 (1–13)	(29)

where $N_1(t) = S_1(t) + E_1(t) + P_1(t) + I_1(t) + \underline{V}_1(t)$ and $N_2(t) = S_2(t) + E_2(t) + P_2(t) + I_2(t) + \overline{V}_2(t) + \overline{\overline{V}}_2(t)$. Note that the number of recovered individuals is small and not included for simplicity.

Following Champredon et al. (19), we calculated the expected minimum probability of transmissions per attendee that will occur during the gathering, considering the vaccination of attendees, yielding

$$r_G(t) = 1 - \left(\left(1 - \frac{N_{GI}(t)}{N_G - 1} \beta \right)^{c_G} \right)$$

where c_G is the number of effective contacts with an infectious individual during the gathering. An effective contact is defined as a contact where there is physical contact and enough exposure time between individuals which may result in the transmission of monkeypox. Hence, the number of LRG-susceptible individuals infected during the gathering is,

$$I_{S1}^G = p_L S_1 r_G,$$

and the number of fully vaccinated LRG individuals infected during the gathering is,

$$I_{V1}^G = p_L \underline{V}_1 r_G (1 - \epsilon_2).$$

Similarly, we obtained the number of susceptible, partially vaccinated, and fully vaccinated individuals from HRG infected during the gathering as,

$$I_{S2}^G = p_H S_2 r_G, \, I_{V2}^G = p_H \overline{V}_2 r_G (1 - \epsilon_1), \, I_{VV2}^G = p_H \overline{\overline{V}}_2 r_G (1 - \epsilon_2),$$

respectively.

The expected minimum number of transmissions that will occur during the gathering is thus,

$$I^G = I_{S1}^G + I_{V1}^G + I_{S2}^G + I_{V2}^G + I_{VV2}^G.$$

Modeling vaccination intervention specific to gatherings

If public health interventions, including vaccination strategy, are applied at the gathering event to prevent the transmission, we assumed that only the fully vaccinated individuals are allowed to attend the gathering, which includes both vaccinated individuals and the individuals vaccinated but not effectively protected (infected, and in the prodromal state). Hence, the proportion of attendance in the qualified LRG and

TABLE 5 Parameters defined in the modeling of monkeypox transmission.

Parameter	Definition	Value (range)	Ref.
c_{ij}	Matrix of contacts among the groups	$c_{11} = c_0 k_1, c_{21} = c_{12},$ $c_{12} = c_0 (1 - k_1),$ $c_{22} = c_0 k_2.$	
Gathering event related			
$G(t)$	Indicator parameter if gathering event take place or not	-	Defined
N_G	The number of attendees of the gathering event	-	Defined
T	The duration of gathering event, days	$T_2 - T_1$	Defined
p_{GH}	Proportion of individuals attending the gathering event who is from HRG	$1 - p_{GL}$	Defined
p_L	Proportion of LRG individuals attending the gathering event at day t	-	Defined
p_H	Proportion of HRG individuals attending the gathering event at day t	-	Defined

HRG on day t is

$$\bar{p}_L(t) = \frac{p_{GL} N_G}{\underline{V}_1(t) + P_{V1}(t)}, \bar{p}_H(t) = \frac{p_{GH} N_G}{\overline{\overline{V}}_2(t) + P_{V2}(t)},$$

where $P_{V1} = \frac{(1-\epsilon_2)V_1(t)}{S_1(t)+(1-\epsilon_2)\underline{V}_1(t)}\left[P_1(t) + E_1(t)\right]$ and

$$P_{V2} = \frac{(1-\epsilon_2)\overline{\overline{V}}_2(t)}{S_2(t) + (1-\epsilon_1)\overline{V}_2(t) + (1-\epsilon_2)\overline{\overline{V}}_2(t)}\left[P_2(t) + E_2(t)\right]$$

are the number of individuals in the prodromal state who are fully vaccinated but not effectively protected in the LRG and HRG, respectively.

Hence, the number of infectious attendees during the gathering on day t, which can only include the fully vaccinated individuals and those infected but in the prodromal stage, is

$$\overline{N_{GI}}(t) = \overline{p_L}\frac{(1-\epsilon_2)\underline{V}_1(t)}{S_1(t) + (1-\epsilon_2)\underline{V}_1(t)}P_1(t)$$
$$+ \overline{p_H}\frac{(1-\epsilon_2)\overline{\overline{V}}_2(t)}{S_2(t) + (1-\epsilon_1)\overline{V}_2(t) + (1-\epsilon_2)\overline{\overline{V}}_2(t)}P_2(t).$$

Since only fully vaccinated individuals are allowed to attend the gathering, the minimum probability of transmissions per

attendee will become

$$\overline{r_G}(t) = 1 - \left(\left(1 - \frac{\overline{N_{GI}}(t)}{N_G - 1}\beta(1-\epsilon_2)\right)^{c_G}\right).$$

Hence, the expected minimum number of transmissions that will occur during the gathering is

$$\overline{I_{V1}^G} = \overline{p_L}(t)\,\underline{V}_1(t)\,\overline{r_G}(t), \overline{I_{VV2}^G} = \overline{p_H}(t)\,\overline{\overline{V}}_2\overline{r_G}(t)$$

respectively.

Scenario analysis

We conducted numerical simulations with the setting of the hypothetical metropolitan city, starting on 1 May 2022, and the model was run for 2 years (730 days). The vaccination started to be administered in HRG individuals on 10 June 2022 (28). The initial values, parameters with fixed values, and the range of some parameters used for simulations are presented in Tables 2–4. We investigated five different scenarios listed in Table 6 by presenting the projection of daily new infections in LRG and HRG (per 100,000 individuals). In scenarios 1–3, the projection of mean and 95% confidence interval of daily new infections are obtained from 5,000 parameter sets sampling from the prior distribution (uniform) of parameters, by the Latin hypercube sampling (LHS) method (32, 33). We only present the mean of all the simulations in scenarios 2 and 3 for an intuitive interpretation of the results. We explored the proportion of vaccination coverage needed to prevent transmission under different assumptions in scenarios 4 and 5, where other parameters are fixed at the values presented in Tables 3, 4. The analyses were conducted using MATLAB (R2020a) (34).

Sensitivity analysis

Sensitivity analyses were conducted to address the uncertainty of the parameters through the LHS and the partial rank correlation coefficient (PRCC) method (32). We generated 5,000 samples of parameters related to vaccination strategy, including the efficacy of the vaccine, the vaccination coverage in HRG, the days needed to achieve the vaccination coverage, and the vaccination proportion of close contacts of confirmed cases in HRG and LRG, and the parameters associated with gathering events, including the number of attendees of gathering, the effective contacts in the gathering, the proportion of attendees from LRG individuals. The ranges of parameters used in the sensitivity analysis are reported in Supplementary Table S1. We calculated the value of PRCC to investigate the relationship between the parameters and the model outputs of cumulative cases, which above 0.5 were considered to be significant.

TABLE 6 Lists of settings of scenarios to projections of monkeypox infection.

Scenarios	Settings
Scenario 1	Gathering events and no public health (PH) control measures are conducted. • 10, 000 daily attendees, $R_0 = 1.5$ (Baseline). • 10, 000 daily attendees, $R_0 = 3$. • 100,000 daily attendees, $R_0 = 3$.
Scenario 2	Gathering events (no interventions on the attendees) with 100,000 daily attendees, 10, 30 or 50 effective contacts in the gathering and $R_0 = 3$. • With testing, isolation, and PEPCT strategy, but without PrEPH. • With testing, isolation, and PrEPH (90% HRG received 1 dose vaccine as of July 20), but without PEPCT strategy.
Scenario 3	Gathering events with 100,000 daily attendees (fully vaccinated), 10, 30 or 50 effective contacts in the gathering and $R_0 = 3$. • With testing, isolation, and PrEPH (60% HRG received 1 dose vaccine as of Jul 20) and PEPCT strategy. • With testing, isolation, and PrEPH (90% HRG received 1 dose vaccine as of Jul 20), testing and isolation, but without PEPCT strategy.
Scenario 4	Gathering events (no interventions on the attendees) with 100,000 daily attendees, and $R_0 = 3$. With testing, isolation, and without PEPCT strategy, and other settings.
Scenario 5	Gathering events (no interventions on the attendees) with 100,000 daily attendees, 50 effective contacts in the gathering and $R_0 = 3$. With testing, isolation, vaccinating 30% close contact of confirmed cases in LRG, and other settings.

Some general setting and descriptions

Gathering event	Starting from July 20 to August 1, 2022.
No PH control measures	No interventions on monkeypox transmission, like testing, isolation, tracing or vaccination.
$R_0 = 1.5$ or 3	The probability of per contact among HRG is 12.2 % or 24.5%.
Testing, isolation	The testing and isolation strategy are implemented and the testing proportion of symptomatic infection in HRG and LRG is 50% and 30% respectively.
PEPCT strategy	Vaccinating the 60% close contact of confirmed cases in HRG and 30% close contact of confirmed cases in LRG.
Other settings	The time from individuals developing symptoms, to seeking medical help and then being tested and isolated is 7 days, and the time for tracing and vaccinating the close contacts of confirmed cases is 7 days, the daily transition rate from LRG to HRG during the gathering is 0.0005, and the proportion of individuals adhering to the isolation strategy is 40%, and the 10% of attendees is LRG individuals.

Results

Impact of no public health measures implementation after gatherings

We projected the daily new infections in LRG and HRG when the MGE occurred between 20 July and 1 August 2022 (Figure 2, Scenario 1 described in Table 6). We varied the number of participants and transmissibility levels of the monkeypox virus, under the assumption that public health control measures are not in place. A large outbreak of cases follows MGEs, with the number of cases increasing as the number of attendees and probability of transmission increase, raising the transmission. However, with the lowest number of participants and lower value of R_0, the outbreak shows the beginning of an increasing trend 400 days after the gathering. For a gathering with 10,000 daily attendees, the risk of a monkeypox outbreak is low if the transmission probability per contact among the HRG individuals of the monkeypox virus is 12.2% ($R_0 = 1.5$). On the other hand, we could observe a large outbreak if the number of participants increases to 100,000, with an average peak of daily new infections in HRG of around 150 per 100,000 people. If $R_0 = 3$ and attendance increases to 100,000, the outbreak will immediately follow the gathering event, with a larger peak size of infections (500 per 100,000).

Effects of PEPCT strategies

Figures 3, 4 (Scenario 2 and 3 described in Table 6) show the impact of PEPCT and PrEPH strategies under the possibility of a mass gathering with 100,000 attendees and relatively high transmission efficiency of the monkeypox virus ($R_0 = 3$). Overall, if testing and isolation of symptomatic cases are in place, and contact tracing is effective, the PEPCT strategy is more beneficial to the control of monkeypox outbreaks, compared to the PrEPH strategy, and this is expected given the time needed to develop an immune response and the low vaccine coverage in the overall population in the simulations. With the implementation of the PEPCT strategy, tracing, vaccinating, and isolating 60 and 30% of close contacts of confirmed monkeypox cases in HRG and LRG, respectively, and maximum effective contacts in the gathering, the average peak of daily new infection in HRG was below 4 per 100,000 attendees (Figure 3, second panel). However, this number exceeds 90 per 100,000 if 90% of HRG individuals received at least 1 dose of vaccine before the gathering event started, but without PEPCT (Figure 3, fourth panel).

Similar results were obtained if all attendees of the gathering event are fully vaccinated (Figure 4). If the PEPCT strategy is implemented and the effective contacts during the gathering are 50, the mean peak of daily infection in HRG is below 1 per 100,000, although only 60% of individuals were administered 1

FIGURE 2
Projections of daily new infections (per 100 k) in the LRG and HRG without public health control measures if there is a gathering event from 20 July to 1 August 2022, with 10,000 or 100,000 daily attendees. The reproduction number of monkeypox transmission is 1.5 or 3. The gray shaded bar represents the gathering event occurring. The detailed setting can be found in Table 6, scenario 1.

dose of vaccine before the gathering. Conversely, with only 10 effective contacts in the gathering, the mean peak of daily new infection in HRG exceeds 35 per 100, 000, if vaccinating 90% of HRG individuals with 1 dose before the gathering but without a PEPCT strategy. Regardless of whether there was a vaccination requirement for the attendance of gathering events, PEPCT strategies are critical to containing monkeypox transmission arising from gatherings.

Results for LRG follow the same trends of HRG, but with a smaller magnitude.

Measures to prevent outbreaks after gatherings: Restricting effective contacts or vaccination

The peak size of the monkeypox outbreak is significantly associated with the number of effective contacts in the gathering if there are no restrictions on the gathering activities (Figure 3). Thus, the public health control measures aiming at constraining

effective contacts during the gathering are essential to prevent the possible outbreak under this circumstance.

However, contact during the gathering would have a slight effect on the progression of the monkeypox transmission if only fully vaccinated individuals are allowed to attend the gathering (Figure 4). Moreover, whether there is a vaccination requirement for the attendees of gathering activities or not, public health control measures, such as PEPCT, are needed to prevent an outbreak resulting from MGEs.

Identification of the best combination of vaccination coverage and gathering intervention

Figure 5 shows a contour plot of peak infection in LRG and HRG under Scenario 4 (described in Table 6), when effective contacts and vaccine coverage (one dose by July 20) are varied. These results permit the determination of the vaccination coverage in HRG and constraints of contacts in gathering needed to prevent monkeypox transmission if the gathering

FIGURE 3
Projections of daily new infections (per 100 k) in the LRG and HRG with different public health control strategies under the different effective contacts in the gathering if there is a gathering event occurring from 20 July to 1 August 2022, with daily 100,000 attendees. **(A)** With PEPCT, testing and isolation, but without PrEPH; **(B)** with PrEPH, testing and isolation, but without PEPCT. Note that the reproduction number of monkeypox transmission is 3. The testing proportion of symptomatic infection in HRG and LRG is 50 and 30%, respectively. The PEPCT strategy represents that we vaccinated the 60% close contacts in HRG and 30% close contact in LRG. The gray shaded bar represents the gathering event occurring. The detailed setting can be found in Table 6, scenario 2.

event occurs. In this scenario, testing and isolation (30 and 50% of symptomatic infections in LRG and HRG) are included. As illustrated in Figure 5, public health interventions on constraining contacts during the gathering are more beneficial to contain the transmission, compared to the strategy of increasing vaccination coverage in HRG shortly before the event. The infection in the LRG could be kept below 1 per 100,000 if the effective contacts in the gathering are 46. However, the contacts should be < 15 to maintain a low prevalence (10 per 100,000) in HRG.

Identification of the best combination of PEPCT and PrEPH strategy

The contour plot of peak infection in LRG and HRG with varying vaccination proportions of PEPCT and PrEPH is shown in Figure 6, if all the individuals were allowed to attend the mass gathering events (100,000 attendees), along with assumptions listed in Table 6 (Scenario 5). If vaccination coverage of 1 dose in HRG as of 20 July is 10%, to maintain the infection in LRG below 1 per 100,000, at least 38% of close contact of confirmed cases in
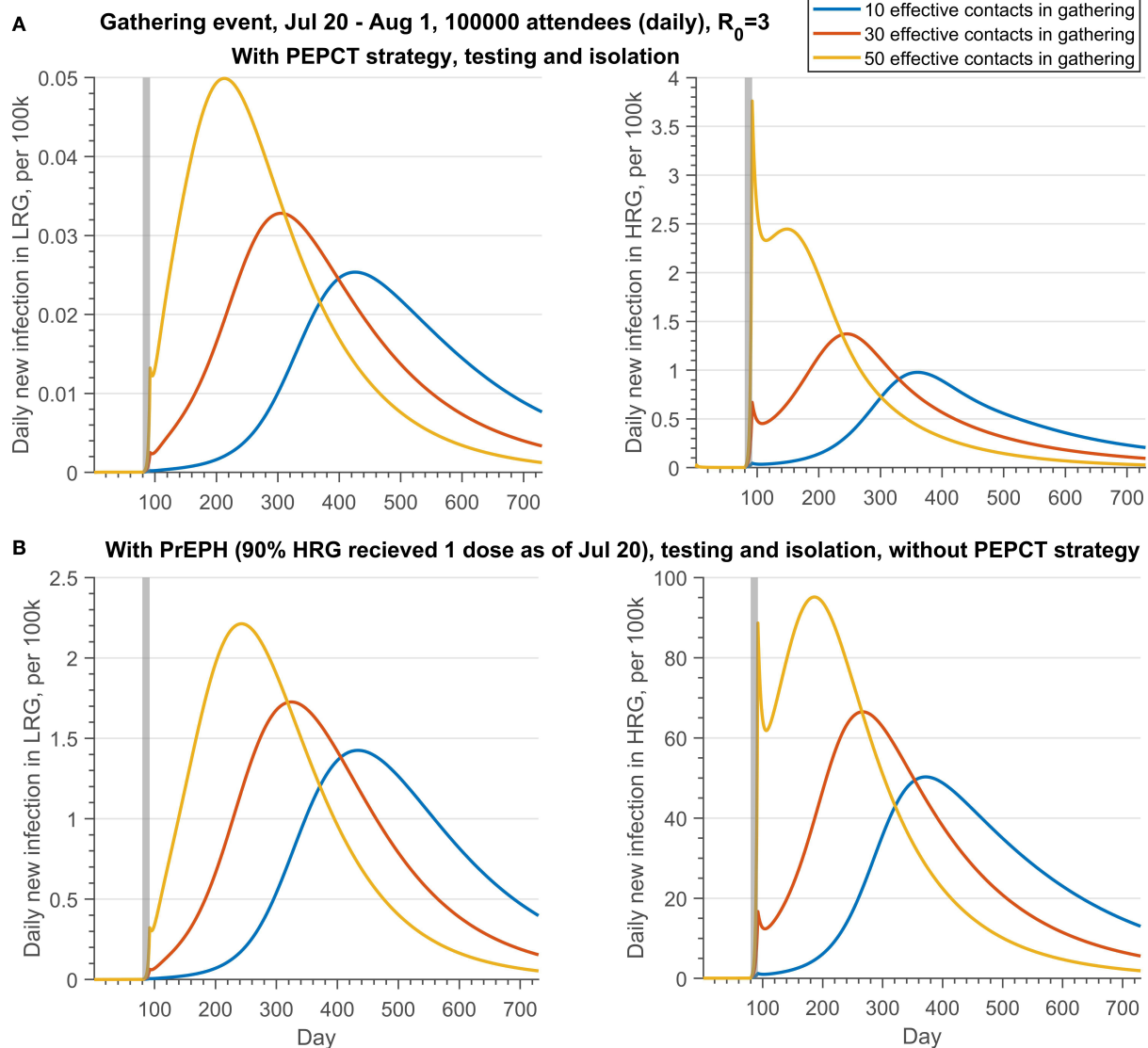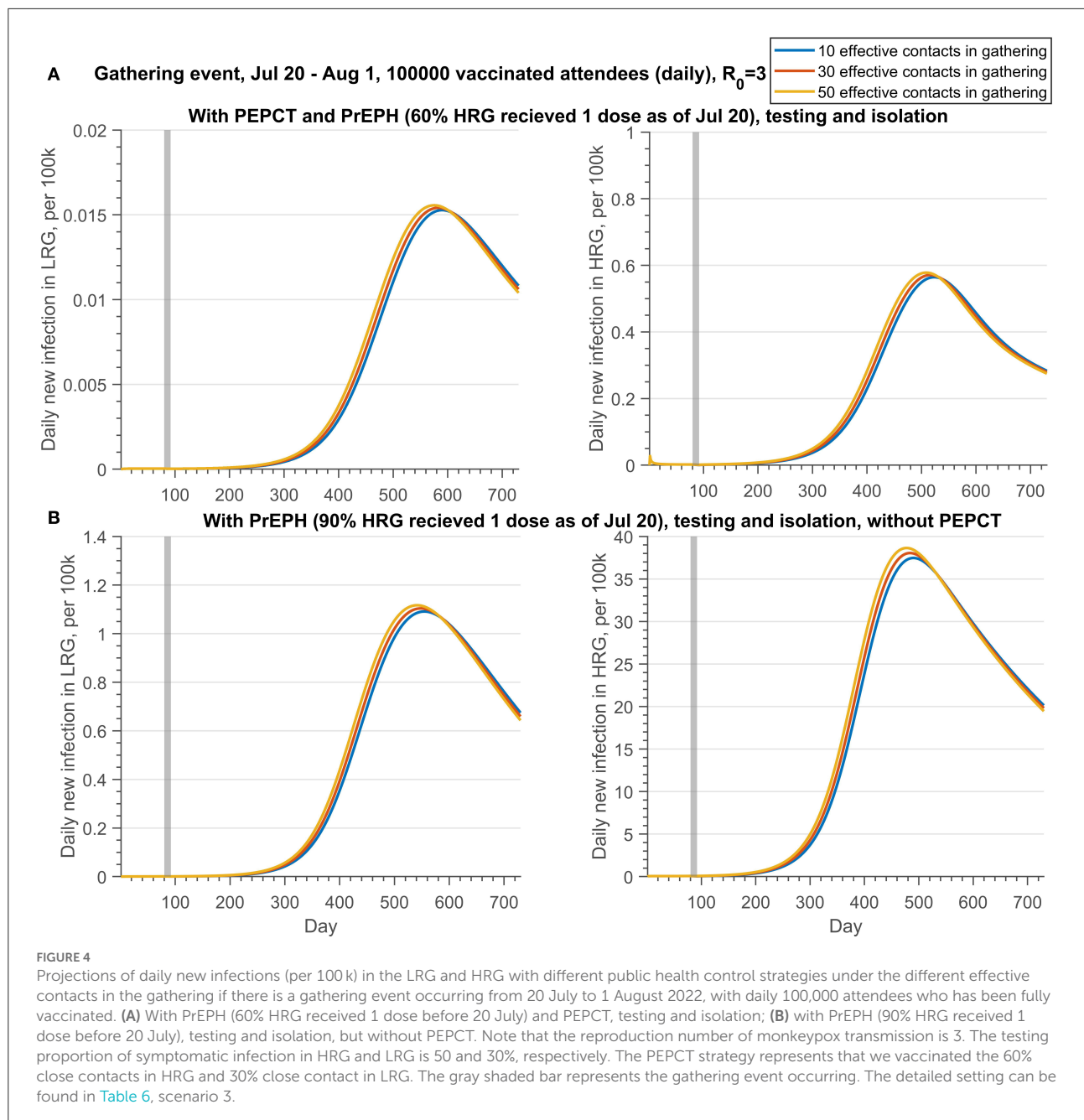
FIGURE 4
Projections of daily new infections (per 100 k) in the LRG and HRG with different public health control strategies under the different effective contacts in the gathering if there is a gathering event occurring from 20 July to 1 August 2022, with daily 100,000 attendees who has been fully vaccinated. **(A)** With PrEPH (60% HRG received 1 dose before 20 July) and PEPCT, testing and isolation; **(B)** with PrEPH (90% HRG received 1 dose before 20 July), testing and isolation, but without PEPCT. Note that the reproduction number of monkeypox transmission is 3. The testing proportion of symptomatic infection in HRG and LRG is 50 and 30%, respectively. The PEPCT strategy represents that we vaccinated the 60% close contacts in HRG and 30% close contact in LRG. The gray shaded bar represents the gathering event occurring. The detailed setting can be found in Table 6, scenario 3.

HRG should be traced, vaccinated, and isolated. However, this proportion needs to increase to 75% to keep the infection in HRG below 1 per 100,000.

Sensitivity analysis

The observed number of attendees, the number of effective contacts, and the transition rate from LRG to HRG during the gathering are significantly positively correlated with the cumulative cases (Figure 7A). On the other hand, the proportion of attendees from the LRG is negatively correlated to the cumulative cases, but not significant, which indicates a possibility of the dilution effect of the LRG attendees on the gathering events in consideration of the fact that they have a lower possibility to transmit the virus to the LRG community.

The sensitivity analysis also showed that the efficacy of the second dose and the vaccine coverage of close contact with confirmed cases in HRG and LRG present a negative correlation with cumulative cases (Figure 7B). Nevertheless, the first dose
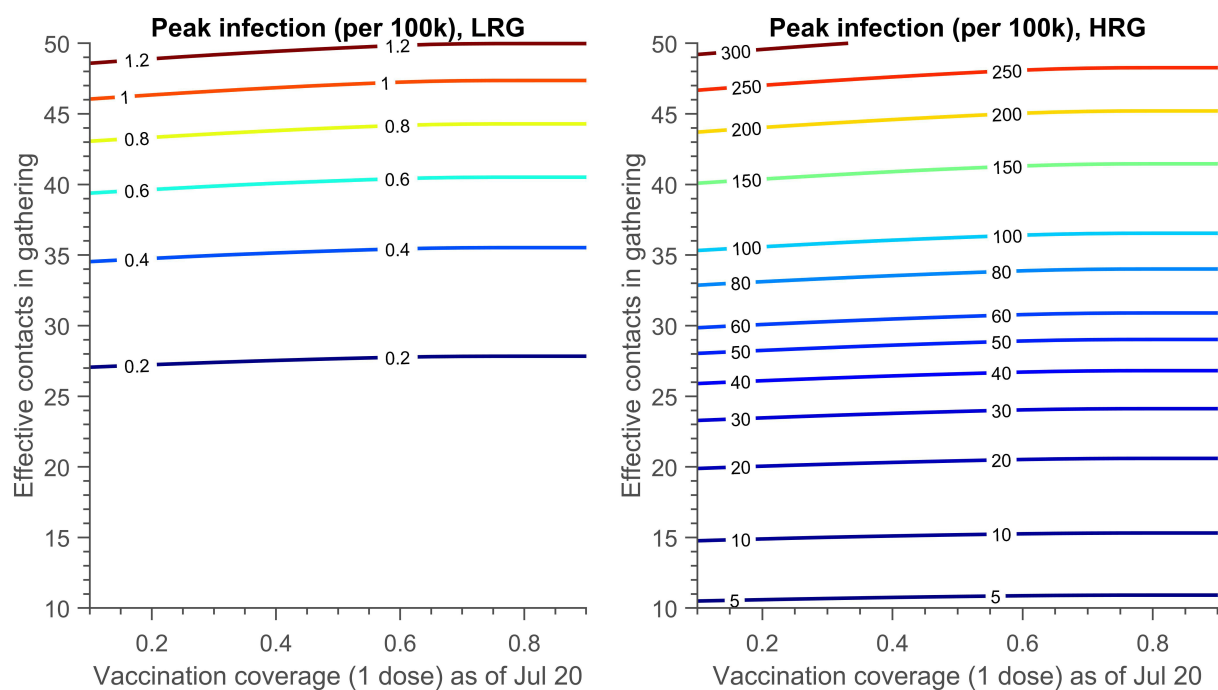
FIGURE 5
The contour plot of peak infection in LRG and HRG (per 100 k) with different effective contacts in the gathering and the vaccination coverage achieved (1 dose) on 20 July. There is no PEPCT strategy and the testing proportion of symptomatic infection in HRG and LRG is 50 and 30% respectively. The detailed setting can be found in Table 6, scenario 4.
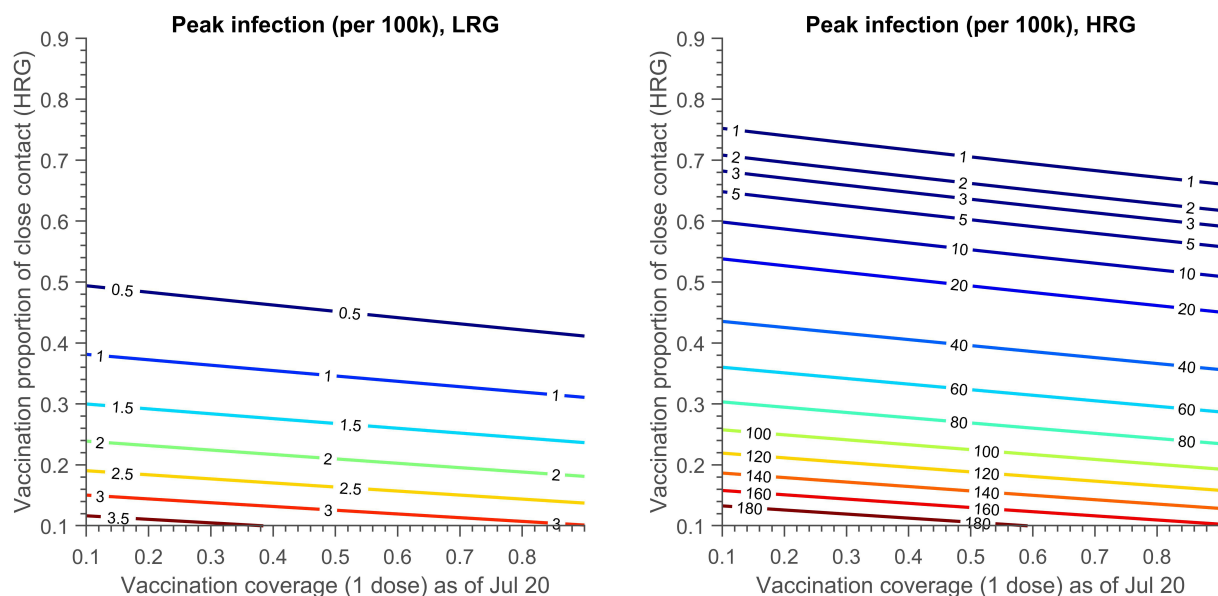
FIGURE 6
The contour plot of peak infection in LRG and HRG (per 100 k) with vaccination proportion of close contact in HRG and the vaccination coverage achieved (1 dose) on 20 July. The testing proportion of symptomatic infection in HRG and LRG is 50 and 30%, respectively. A total of 30% close contact of confirmed cases in LRG are traced, vaccinated and quarantined. The detailed setting can be found in Table 6, scenario 5.
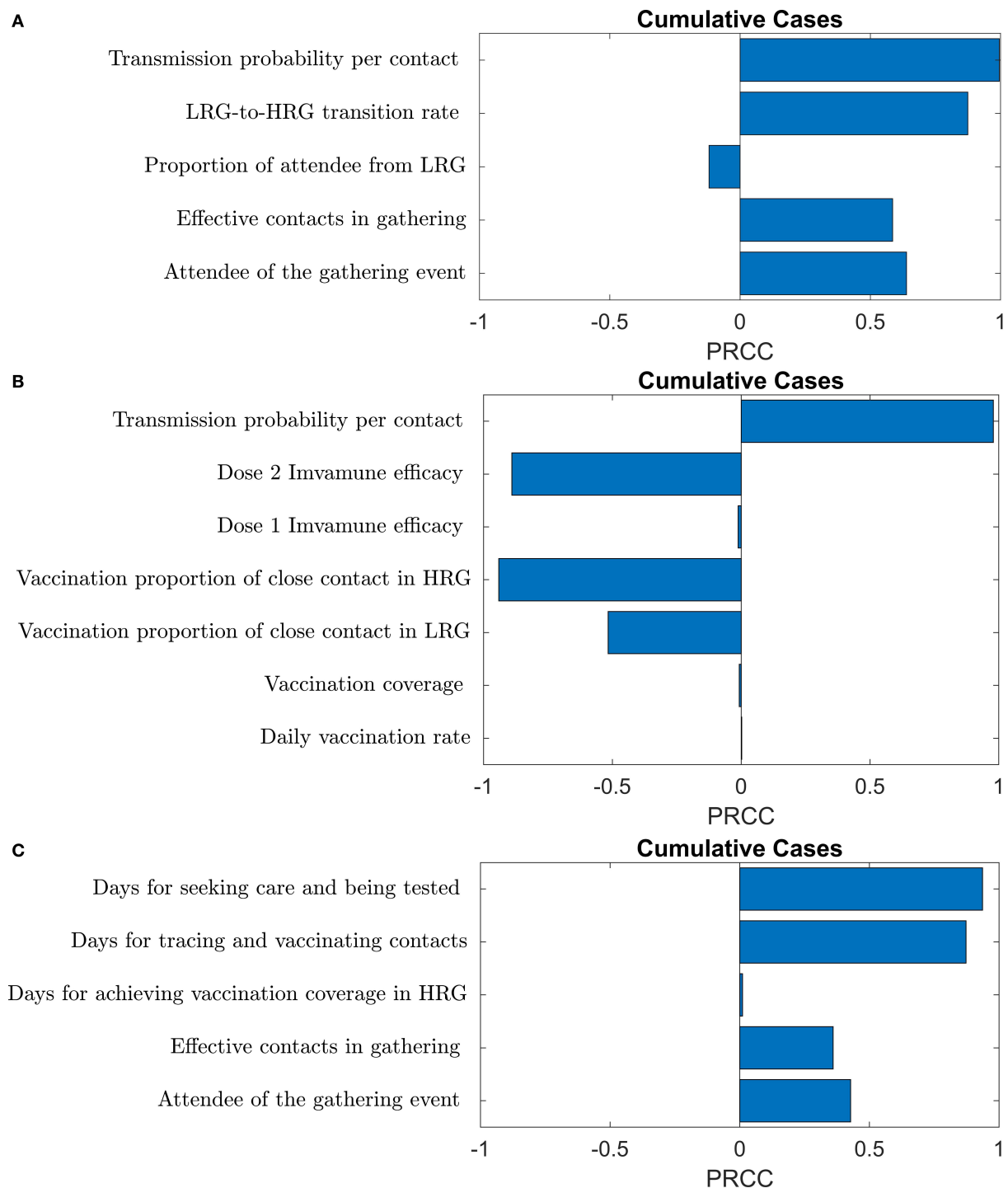
FIGURE 7
The PRCC plots of **(A)** gathering event-related **(B)** transmission and vaccine-related **(C)** public health control strategy-related parameters on cumulative cases.

efficacy and its coverage in HRG, and the daily vaccination rate in HRG are not significant to the cumulative cases, due to the small proportion of HRG individuals in the total population (Figure 7C).

The sensitivity analysis confirmed that the PEPCT strategy and the interventions in the gathering activities are crucial to curb monkeypox spreading at gatherings. Moreover, the efforts of case detection, testing and isolation, and contact tracing,

of symptomatic monkeypox cases affect the progression of the disease significantly (Figure 7C).

Discussion

The classification of the monkeypox outbreak as a PHEIC by WHO underlines its global seriousness (3). In this study, we employed a mathematical modeling approach to explore how vaccinations and other public health measures can be implemented, in a hypothetical metropolitan city, to prevent the spread of monkeypox in human populations after mass gatherings events. Our results suggested that the risk of a monkeypox outbreak after gatherings is high, especially if the number of attendees was large and public health measures were not in place. However, effective public health measures can support the containment of monkeypox transmission at mass gatherings by a combination of constraints on effective contact at the gatherings, implementation of PrEPH, testing and isolation of symptomatic cases, and contact tracing and PEP (here studied as one process PEPCT). Vaccination requirements for participants in mass gathering events could play a crucial role in curbing the spreading of viruses. In addition, tracing, vaccinating, and isolating the close contacts who are exposed to cases was more beneficial to contain monkeypox transmission compared to a PrEP vaccination campaign in HRG individuals that begins shortly before the event.

Our novel model structure, with consideration of saturation of contacts at gatherings, allowed us to assess the monkeypox transmission risk on the occasion that gathering activities occur. Our results suggested that ring vaccination, along with efficient contact tracing and isolation, can be a powerful tool to halt the spread of the monkeypox virus linked to mass gathering events. Either limiting the gathering size and density, requiring the vaccination of attendees, or both may be essential for safe social gathering events. Additional measures were also required, such as a high level of effort to test and isolate confirmed cases and PEPCT, which required rigorous contact tracing and compliance with isolation policies.

In these simulations, the containment effect of PEPCT was determined by the vaccine efficacy, the proportion of contacts traced, the days needs to trace the exposures and vaccination, and the compliance with the quarantine after vaccination. Hence, PEPCT may not work as expected due to the uncertainty of efficacy and availability of the vaccine, as well as the difficulty in identifying the people who are most at risk from infection (35). Moreover, the significant containment effect of the PEPCT strategy mainly results from the transmission cut-off due to the contact tracing and isolation of the exposures. In reality, there is considerable difficulty in identifying contacts with cases, thus limiting the PEPCT strategy. Containing the spreading of monkeypox by ring vaccination protocol is also

greatly dependent on the willingness of inoculation of the individual with exposure and compliance with the quarantine after vaccination.

In addition, the precise effectiveness of the smallpox vaccine against monkeypox infection in the human population remains unclear, although initial studies begin to suggest high effectiveness (36). The rollout of the vaccine campaign in HRG, or extending the eligibility to individuals with a moderate risk, requires the support of enough stockpiles of the vaccine. Given the limited vaccine capacity, the public health sector should prioritize vaccines for the communities at risk (37). However, targeting specific communities or groups of people may deepen stigma and hinder tracing, vaccination, and identification of cases (5, 35, 38). Also, only vaccinating the individuals with high risk was not enough to prevent the outbreak in a scenario with an increased transmission efficiency of the virus (possibly due to accelerated evolution) and there may be increased spillover from HRG to LRG (Supplementary Figure S1).

Although our modeling scenario simulations are conducted in a hypothetical metropolitan city with parameters from literature regarding Canadian cities, our model can be easily applied to any jurisdictions and areas where data are available, with the refinement of key parameters like the efficacy of vaccines to inform public health decision making. But the specific numbers required to halt the transmission need to be re-examined in a given region with local data. Besides, it is known that animals are a reservoir and part of the transmission to the human population, but this is not included in our work as we are focusing on the effect of vaccination and gathering events, and it can underestimate the transmission risk (23). We will include it in further study.

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

HZ, PY, YT, LY, EA, and NO: conceptualization. YT, LY, and PY: data curation. LY, PY, YT, and HZ: formal analysis. PY, YT, LY, EA, and HZ: methodology and writing—original draft. PY and YT: software, validation, and visualization. JB, JA, JH, JW, NO, HC, EA, PY, YT, LY, and HZ: writing—review and editing. HZ, JA, JB, JH, JW, and HC: funding acquisition. HZ: supervision. All authors contributed to the article and approved the submitted version.

Funding

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. 2022.1026489/full#supplementary-material

References

1. World Health Organization. *Multi-Country Monkeypox Outbreak in Non-Endemic Countries: Update*. (2022). Available online at: https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON388 (accessed July 21, 2022).

2. Our World in Data. *Monkeypox*. (2022). Available online at: https://ourworldindata.org/monkeypox (accessed July 26, 2022).

3. World Health Organization. *Second Meeting of the International Health Regulations (2005) (IHR) Emergency Committee Regarding the Multi-country Outbreak of Monkeypox*. (2022). Available online at: https://www.who.int/news/item/23-07-2022-second-meeting-of-the-international-health-regulations-(2005)-(ihr)-emergency-committee-regarding-the-multi-country-outbreak-of-monkeypox (accessed July 26, 2022).

4. Government of Canada. *Monkeypox: Outbreak Update*. (2022). Available online at: https://www.canada.ca/en/public-health/services/diseases/monkeypox/outbreak-update.html (accessed October 18, 2022).

5. The New York Times. *Monkeypox and the Gay Community*. (2022). Available online at: https://www.nytimes.com/2022/06/24/briefing/monkeypox-gay-community.html (accessed July 31, 2022).

6. Kozlov M. Monkeypox vaccination begins-can the global outbreaks be contained? *Nature*. (2022) 606:444–5. doi: 10.1038/d41586-022-01587-1

7. Bunge EM, Hoet B, Chen L, Lienert F, Weidenthaler H, Baer LR, et al. The changing epidemiology of human monkeypox—a potential threat? a systematic review. *PLOS Negl Trop Dis*. (2022) 16:e0010141. doi: 10.1371/journal.pntd.0010141

8. World Health Organization. *Monkeypox. World Health Organization*. (2022). Available online at: https://www.who.int/news-room/fact-sheets/detail/monkeypox (accessed June 2, 2022).

9. Lum FM, Torres-Ruesta A, Tay MZ, Lin RTP, Lye DC, Rénia L, et al. Monkeypox: disease epidemiology, host immunity and clinical interventions. *Nature Reviews Immunology*. (2022) 22:597–613. doi: 10.1038/s41577-022-00775-4

10. Government of Canada. *Interim Guidance on Infection Prevention and Control for Suspect, Probable or Confirmed Monkeypox Within Healthcare Settings*. (2022). Available online at: https://www.canada.ca/en/public-health/services/diseases/monkeypox/health-professionals/interim-guidance-infection-prevention-control-healthcare-settings.html (accessed July 21, 2022).

11. Government of Canada. *NACI Rapid Response: Interim Guidance on the Use of Imvamune® in the Context of Monkeypox Outbreaks in Canada*. (2022). Available online at: https://www.canada.ca/en/public-health/services/immunization/national-advisory-committee-on-immunization-naci/guidance-imvamune-monkeypox.html (accessed July 15, 2022).

12. Mccarthy Tetrault. *COVID-19: Recovery and Re-opening Tracker*. (2022). Available online at: https://www.mccarthy.ca/en/insights/articles/covid-19-recovery-and-re-opening-tracker (accessed July 21, 2022).

13. CBC News. *'Back in the saddle': 2022 Calgary Stampede Winds Down with Near Pre-Pandemic Numbers*. (2022). Available online at: https://www.cbc.ca/news/canada/calgary/calgary-stampede-2022-near-pre-pandemic-numbers-1.6523589 (accessed August 01, 2022).

14. The Globe and Mail. *Summer Festival Sales to Hit New Heights in 2022*. (2022). Available online at: https://www.theglobeandmail.com/life/article-summer-festival-sales-to-hit-new-heights-in-2022/ (accessed August 01, 2022).

15. CTV News. Hundreds of thousands celebrate return of Toronto Pride parade to downtown streets. (2022). Available online at: https://toronto.ctvnews.ca/hundreds-of-thousands-celebrate-return-of-toronto-pride-parade-to-downtown-streets-1.5963641 (accessed July 21, 2022)

16. CTV News. *How Concerned Should We Be About Monkeypox?*. (2022). Available online at: https://www.ctvnews.ca/health/how-concerned-should-we-be-about-monkeypox-1.5914385 (accessed July 21, 2022).

17. NBC Bay Area. *SF's LGBTQ Community on Alert After Monkeypox Cases Linked to Pride Parties*. (2022). Available online at: https://www.nbcbayarea.com/news/local/san-francisco/san-francisco-pride-monkeypox/2937348/ (accessed July 31, 2022).

18. CNN Health. *Monkeypox Spreading in 'Cluster Events,' But Vaccines Can Help Stop it, Local Health Officials Say*. (2022). Available online at: https://www.cnn.com/2022/07/21/health/monkeypox-clusters-local-officials/index.html (accessed July 31, 2022).

19. Champredon D, Fazil A, Ogden N H. Simple mathematical modelling approaches to assessing the transmission risk of SARS-CoV-2 at gatherings. *CCDR*. (2021) 47:184–94. doi: 10.14745/ccdr.v47i04a02

20. Liu C, Huang J, Chen S, Wang D, Zhang L, Liu X, et al. The impact of crowd gatherings on the spread of COVID-19. *Environ Res*. (2022) 213:113604. doi: 10.1016/j.envres.2022.113604

21. Moritz S, Gottschick C, Horn J, Popp M, Langer S, Klee B, et al. The risk of indoor sports and culture events for the transmission of COVID-19. *Nat Commun*. (2021) 12:1–9. doi: 10.1038/s41467-021-25317-9

22. Coudeville L, Amiche A, Rahman A, Arino J, Tang B, Jollivet O, et al. Disease transmission and mass gatherings: a case study on meningococcal infection during Hajj. *BMC Infect Dis*. (2022) 22:1–10. doi: 10.1186/s12879-022-07234-4

23. Yuan P, Tan Y, Yang L, Aruffo E, Ogden NH, Belair J, et al. Assessing transmission risks and control strategy for monkeypox as an emerging zoonosis in a metropolitan area. *J Med Virol*. (2022) 1–15. doi: 10.1101/2022.06.28.22277038. [Epub ahead of print].

24. European Center for Disease Prevention and Control. *Considerations for Contact Tracing During the Monkeypox Outbreak in Europe, 2022*. (2022). Available online at: https://www.ecdc.europa.eu/sites/default/files/documents/Considerations-for-Contact-Tracing-MPX_June%202022.pdf (accessed July 31, 2022).

25. Statista. *Resident Population of Canada in 2021, by Age.* (2022). Available online at: https://www.statista.com/statistics/444868/canada-resident-population-by-age-group/ (accessed August 01, 2022).

26. Brankston G, Merkley E, Fisman DN, Tuite AR, Poljak Z, Loewen PJ, et al. Quantifying contact patterns in response to COVID-19 public health measures in Canada. *BMC Public Health.* (2021) 21:1–10. doi: 10.1186/s12889-021-12080-1

27. Fashion Magazine. *All the Pride Events Happening Across Canada this Summer.* (2022). Available online at: https://fashionmagazine.com/flare/pride-events-calendar-canada-2022/ (accessed August 01, 2022).

28. GovernmentofCanada. *Summary of National Advisory Committee on Immunization (NACI) Rapid Response of June 10, 2022.* (2022). Available online at: https://www.canada.ca/content/dam/phac-aspc/documents/services/immunization/national-advisory-committee-on-immunization-naci/guidance-imvamune-monkeypox/summary-june-10-2022/summary-june-10-2022-en.pdf (accessed July 31, 2022).

29. Adler H, Gould S, Hine P, Snell LB, Wong W, Houlihan CF, et al. Clinical features and management of human monkeypox: a retrospective observational study in the UK. *Lancet Infect Dis.* (2022) 22:1153–62. doi: 10.1016/S1473-3099(22)00228-6

30. Massoudi MS, Barker L, Schwartz B. Effectiveness of postexposure vaccination for the prevention of smallpox: results of a delphi analysis. *J Infect Dis.* (2003) 188:973–6. doi: 10.1086/378357

31. World Health Organization: *Public Health for Mass Gatherings: Key Considerations.* (2022). Available online at: https://www.who.int/publications/i/item/public-health-for-mass-gatherings-key-considerations (accessed October 16, 2022).

32. Mckay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics.* (2000) 21:239–45. doi: 10.1080/00401706.2000.10485979

33. Yuan P, Aruffo E, Tan Y, Yang L, Ogden NH, Fazil A, et al. Projections of the transmission of the Omicron variant for Toronto, Ontario, and Canada using surveillance data following recent changes in testing policies. *Infect Dis Model.* (2022) 7:83–93. doi: 10.1016/j.idm.2022.03.004

34. Math Works. *Matlab for Artificial Intelligence.* (2022). Available online at: https://www.mathworks.com (accessed July 10, 2022).

35. Kupferschmidt K. Monkeypox vaccination plans take shape amid questions. *Science.* (2022) 376:1142–3. doi: 10.1126/science.add3743

36. Payne AB, Ray LC, Kugeler KJ, Fothergill A, White EB, Canning M, et al. Incidence of monkeypox among unvaccinated persons compared with persons receiving ≥ 1 JYNNEOS vaccine dose—32 US jurisdictions, July 31–September 3, 2022. *MMWR Morb Mortal Wkly Rep.* (2022) 71:1278–82. doi: 10.15585/mmwr.mm7140e3

37. Knight J, Tan D, Mishra S. Maximizing the impact of limited vaccine supply under different epidemic conditions: a two-city monkeypox modelling analysis. *medRxiv. [Preprint].* (2022). doi: 10.1101/2022.08.18.22278949

38. Kozlov M Monkeypox declared a global emergency: will it help contain the outbreaks? *Nature.* (2022). doi: 10.1038/d41586-022-02054-7

Check for updates

Spatio-temporal evolution of the COVID-19 across African countries

Bechir Naffeti[1]*, Sebastien Bourdin[2], Walid Ben Aribi[1],
Amira Kebir[1,3] and Slimane Ben Miled[1]

[1]Laboratory of BioInformatics, bioMathematics and bioStatistics, Institute Pasteur of Tunis, Tunis,
Tunisia, [2]Métis Lab, EM Normandie Business School, Le Havre, France, [3]Preparatory Institute for
Engineering Studies of Tunis, University of Tunis, Tunis, Tunisia

The aim of this study is to make a comparative study on the reproduction number R_0 computed at the beginning of each wave for African countries and to understand the reasons for the disparities between them. The study covers the two first years of the COVID-19 pandemic and for 30 African countries. It links pandemic variables, reproduction number R_0, demographic variable, median age of the population, economic variables, *GDP* and *CHE* per capita, and climatic variables, mean temperature at the beginning of each waves. The results show that the diffusion of COVID-19 in Africa was heterogeneous even between geographical proximal countries. The difference of the basic reproduction number R_0 values is very large between countries and is significantly correlated with economic and climatic variables *GDP* and temperature and to a less extent with the mean age of the population.

1. Introduction

On January 30, 2020, the World Health Organization (WHO) declared COVID-19 as a Public Health Emergency of International Concern[1] and by March 11, 2020, declared the first pandemic caused by the coronavirus. Up to July 2021, COVID-19 has affected over 187 million people with more than 4 million associated deaths and in addition, has induced catastrophic public health and socio-economic affliction globally (1).

The first cases in Africa to be reported by WHO were respectively, on February 14, 18, and 25, 2020, in Egypt, Algeria, and Nigeria. These first cases have nearly coincided with those in Europe, which is likely the original source of pathogen introduction in Africa.[2] Since then, the virus has spread quite quickly (see Figure 1) (2). Up to June 6, 2020, most African countries have crossed the threshold of 1,000 cases and the whole

---

1 https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov)

2 https://www.afro.who.int/news/covid-19-cases-top-10-000-africa, https://covid19.ncdc.gov.ng/

FIGURE 1
Spread of COVID-19 in Africa.
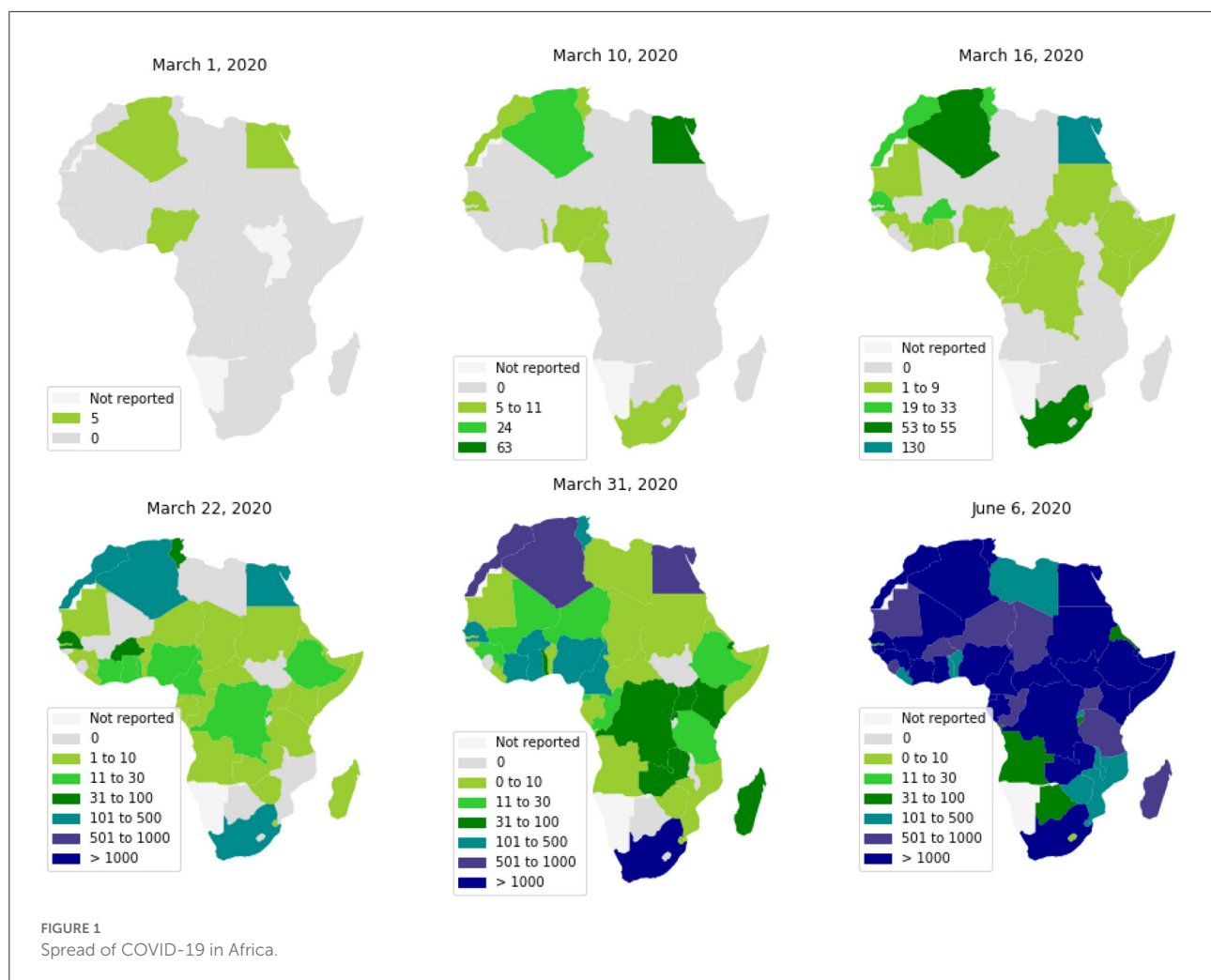
continent had 175, 423 cumulative cases and 4, 862 reported deaths. The WHO had predicted that 29 to 44 million Africans would be infected with COVID-19 during the first year of the pandemic, and 83 to 190 thousand Africans would had die if they don't uphold containment measures.[3]

The high levels of poverty, weak health systems, and a large number of crowded urban areas, make the virus particularly devastating in African countries.[4] However, the warmer climate, the population youth, and the boosted immunity by long exposure to previous endemic pathogens, would allow the continent to mitigate the risk of the pandemic (3). In this context, the diversity of COVID-19's dynamics throughout Africa and its relationship to socioeconomic and

environmental factors can help us better understand the epidemic's determinism.

Like European countries (4), at the beginning of the COVID-19 epidemic, most African countries implemented strict Non-Pharmaceutical intervention (NPI) to limit the spread of this pandemic (5–7). This has included: the obligation to mask wearing and social distancing measures at the individual level, frontier closure, the closure of schools, universities, and public places, the closure of mosques and churches, and the prohibition of movement between cities and provinces. These measures have contributed in reducing the spread of the pandemic (8, 9). However, considering the socio-economical heterogeneity of the African countries (10), the response to these measures differed from one country to another as evidenced by the disparities between regions in infected cases and wave numbers (11).

To date, few studies have analyzed how the pandemic spread in Africa and how its intensity varied over time (12–16). Moreover, to our knowledge, no study has been conducted

---
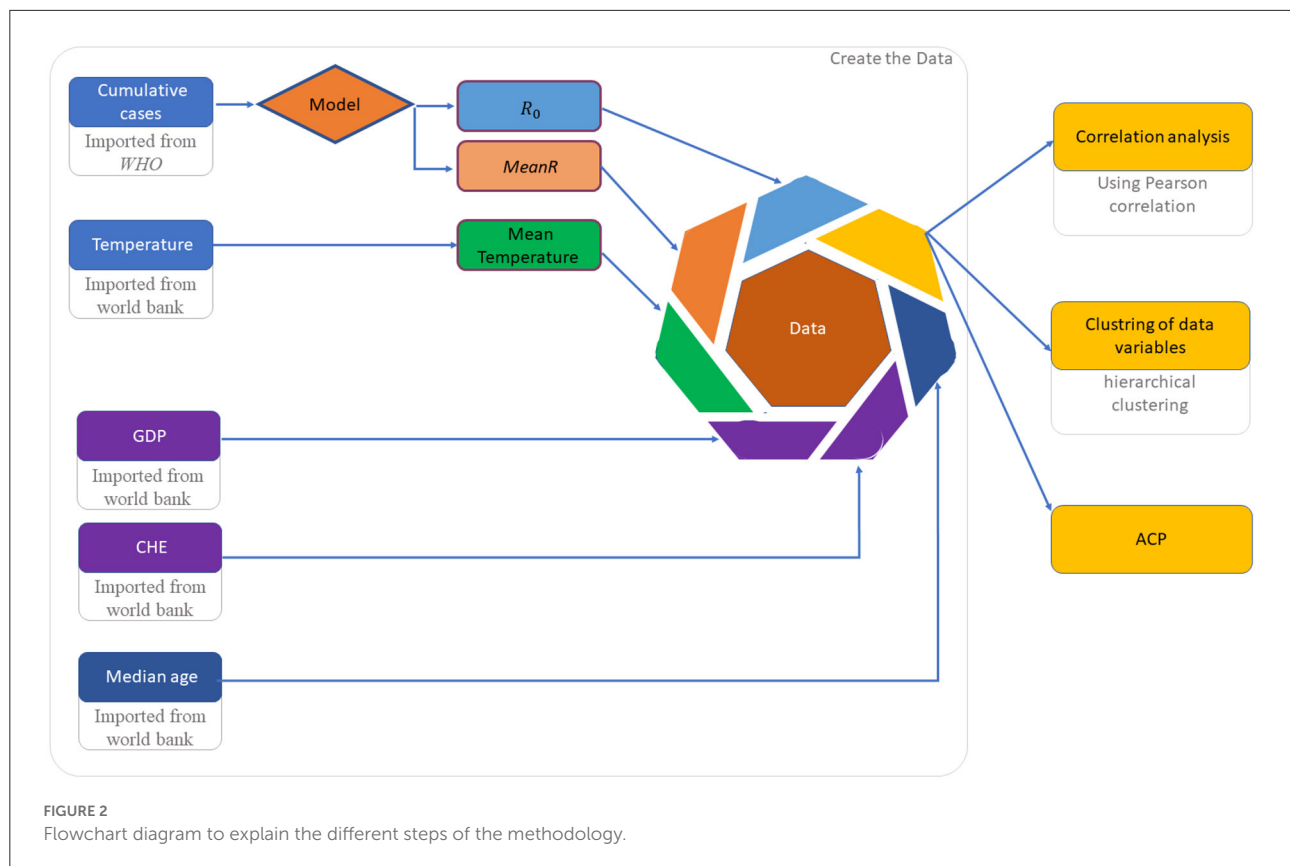
FIGURE 2
Flowchart diagram to explain the different steps of the methodology.

to analyze what are the determinants that could explain the geography of the pandemic.

This study aims at analyzing the Spatio-temporal evolution of the COVID-19 infection across 30 African countries and for each wave until March, 2022. And to provide demo-economical and environmental factors that can better explain the regional heterogeneity of the basic reproduction rate, R_0. To this end, we calculate R_0 at the early beginning of each wave, in order to avoid taking into account the NPI measure. We then make a correlation analysis between R_0 and collected demographic, economic, and climatic data so as to assess how these factors may account for the regional variations of the pandemic.

The document is organized as follows: In Section 2, the material and method are presented. In Section 3, results and discussion are given. Finally, the conclusion is given in Section 4.

2. Materials and methods

In order to comprehend the differences between African countries, we collected epidemiological data from 45 African countries. Due to the quality of the data, this list was reduced to 30 countries distributed between North, South, East and West Africa. These countries are: Algeria, Angola, Burkina Faso, Cameroon, Chad, Ivory Coast, Egypt, Ethiopia, Guinea, Guinea-Bissau, Kenya, Libya, Madagascar, Mali, Mauritania, Morocco, Mozambique, Namibia, Niger, Nigeria, RDC, Rwanda, Senegal, Somalia, South Africa, Sudan, Tanzania, Tunisia, Zambia, and Zimbabwe.

Up to March, 2022, With the exception of Tanzania, Madagascar, Chad, and Burkina Faso, which had three waves, and Kenya, Algeria, Tunisia, and Zambia, which had five waves, nearly all of the thirty African countries analyzed had four waves. For all countries, the Omicron variant generated the most recent wave.

We took into account six epidemiological, demo-economical and climate factors for each country:

- Epidemiological variables are: The basic reproduction numbers, R_0, of each wave, used to analyze the temporal evolution of the COVID-19 wave by wave at each country. The second one is the mean value of R_0 over waves, denoted by $MeanR$ where $MeanR = \frac{1}{n}\sum_{i=1}^{n} R_0^i$, R_0^i is the R_0 of the wave i and n is the number of waves. the $MeanR$ is used to for an inter-countries comparison.
- Economic variables: The current health expenditure (CHE), and the gross domestic product (GDP) were

FIGURE 3
Frequency distribution and histogram of *MeanR* of all waves.
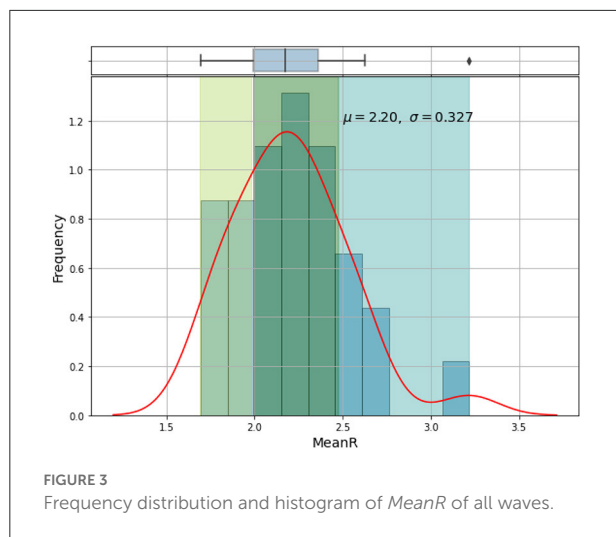
TABLE 1 Distribution of *MeanR* by Country.

	MeanR	**Country**
Group 1	[2.49, 3.22]	Senegal, Zimbabwe, South Africa,
		Angola, Zambia, Ethiopia
Group 2	[1.99, 2.43]	Mali, RDC, Guinea, Sudan, Algeria
		Kenya, Nigeria, Mauritania, Libya
		Guinea-Bissau, Namibia, Morocco
		Rwanda, Côte d'Ivoire, Tunisia
		Mozambique, Tanzania
Group 3	[1.69, 1.907]	Madagascar, Niger, Egypt, Somalia,
		Chad, Burkina-Faso, and Cameroon

collected from World Bank data.[5] It has been shown that these variables have an impact on the propagation of the pandemic in several countries (17–19).

- Climate variable: Mean of the country's temperature at the periods of the beginning waves.[6,7]
- Demographic variable: The median of ages of the population population (see text footnote 5) as older patients are at higher risk of developing severity (20).

For each wave and each African country, basic reproduction rate, R_0, was computed using the method developed in (21). This method is based on a SIR model, which is an Ordinary Differential Equations (ODE) that describes a structured population through three classes: S (susceptible), I (infected: reported and unreported), and R (removed: recovered or die). For more detail about the system of ordinary differential equations, parameters identification and how to determine the R_0 values see Appendix. As it is difficult to estimate the impact of control policies in the calculation of R_0, we chose to calculate R_0 with data from the first days of each wave. Indeed, we assume that at the beginning of each wave the control policies are very little applied or non-existent, so the growth of the pandemic is exponential.

To measure the degree of the relationship between variables, we use the Pearson correlation defined by (22).

For the clustering countries with similar data variables, we use an "unsupervised learning" method, the hierarchical clustering (23).

In this method, it is not necessary to specify an initial number of clusters to run the algorithm. Dendrogram was used to visualize the partitioning of the data.

Impact data variables were summarized and visualized using Principal Component Analysis (PCA) (24).

Data set implementation and analysis is described in Figure 2.

3. Results and discussion

In this section, we perform inter-country and intra-country analyses integrating economic, climatic, and demographic factors.

3.1. Inter-country analysis

Based on the mean of R_0, *MeanR*, distribution across waves (see Figure 3), we divided countries into three groups (see Table 1 and Figure 4).

We observe that more than 50% of the countries have *MeanR* values in [1.99, 2.37], mainly located in north Africa. Moreover, 25% of the other countries show higher values of *MeanR* and are located especially in South Africa.

3.2. Intra-country distribution

When comparing the first three waves, we can see that the R_0 values for waves 4 and 5 have significantly grown (see Table 2). This makes sense given how quickly the Delta and Omicron variants have spread.

Based on the values of R_0 and for each wave, we clustered the countries into three groups (see Table 3 and Figure 5). We observe that countries with the highest *MeanR*, corresponding to Group 1, had experienced a strong first or second wave (countries in Group 1 for the first or second wave).

---

5   https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS

6   http://www.climatemps.com/

7   https://climateknowledgeportal.worldbank.org/country/central-african-republic/climate-data-historical

FIGURE 4
Distribution of *MeanR* among 30 countries in Africa.
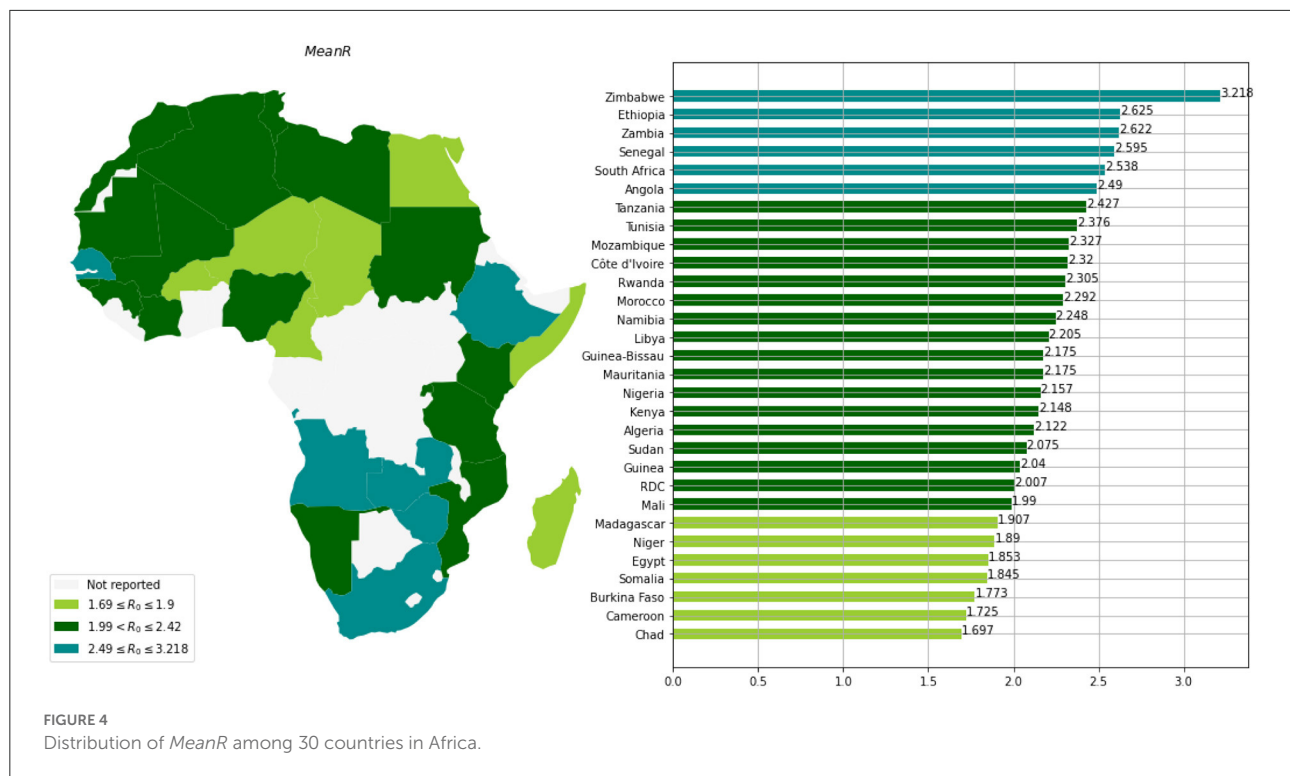
TABLE 2   Mean R_0 over countries by waves.

	First wave	Second wave	Third wave	Fourth wave	Fifth wave
R_0	1.87	2.04	1.91	2.95	3.53

Low R_0 waves were experienced by the countries with the lowest *MeanR* Group (Group 3 for the *MeanR*). Indeed, some of the least affected countries (belonging to Group 2 or 3), have experienced three weak waves, such as Kenya and Guinea (Group 3, for the three waves) or a medium wave as Libya (Group 3 for waves 1 and group 2 for wave 2) RDC (Group 3 for waves 1 and 2 and Group 2 for wave 3) and Mauritania and Cameroon (Group 3 for waves 1 and 3).

We note that, in general, countries in the first Group for the first wave (except for Senegal, Tanzania, and Sudan) experienced a weaker second and third wave. Conversely, countries that experienced a weaker first wave (Groups 2 and 3), experienced a stronger second or third wave (Group 1). Indeed, in Tunisia, the first and third waves (belonging to Group 1), were significant, but the second wave was less so (belonging to Group 3). Senegal experienced three major waves (belonging to Group 1). Finally, South Africa, Chad, Morocco, and Algeria had a powerful first wave (belonging to Group 1), a moderate second wave (belonging to Group 2), and a weak third wave (belonging to Group 3).

3.3. Impact of economic factor

Next, we looked at the relationship between the mean R_0 values, *MeanR*, and the Gross Domestic Product (*GDP*), and the Current Health Expenditure (*CHE*) (see Table 4 and data in the Appendix).

It is revealed that the *MeanR* is highly positively correlated to *GDP* and is moderately positively correlated to *CHE*. Indeed, countries with the highest *GDP*s in Africa (*GDP*s above US$3000 per capita), especially South Africa and some North African countries like Tunisia, Morocco, and Algeria, experienced a significant first wave (see Table 5). These countries were the first to be impacted by the epidemic because of their degree of development, which makes them more accessible to international trade (see Figure 6).

The lower relationship between R_0 and *CHE* may be explained by two facts: Firstly, we measured R_0 at the start of the wave when public health interventions were either not yet in place or were poorly in place. Secondly, the

TABLE 3 Distribution of the country between a Group of R_0 for each wave.

	First wave		Second wave		Third wave		Fourth wave		Fifth wave	
	$[R_0]$	Country	$[R_0]$	Country	$[R_0]$	Country	$[R_0]$	Country	$[R_0]$	Country
Group 1	[1.95, 3.25]	Guinea-Bissau	[2.19, 3.72]	Madagascar	[2.26, 2.57]	Mozambique	[4.05, 5.4]	Angola, Zimbabwe	[4.23]	Zambia
		Mali		Burkina Faso		Guinea		Ethiopia, Côte d'Ivoire		
		Senegal		Nigeria		Tunisia				
		Chad		Mauritania		Zimbabwe				
		Sudan		Sudan		Rwanda				
		Namibia		Mozambique		Senegal				
		Algeria		Senegal		Libya				
		Tanzania		Egypt						
		Tunisia		Zambia						
		Morocco		Tanzania						
		South Africa		Zimbabwe						
Group 2	[1.54, 1.8]	Rwanda	[1.74, 2.06]	Algeria	[1.96, 2.13]	Côte d'Ivoire	[2.64, 3.47]	Guinea-Bissau, Mauritania	[3.34, 3.72]	Tunisia
		Nigeria		Chad		Namibia		Senegal, Rwanda		Kenya
		Somalia		Ethiopia		RDC		Mozambique, Nigeria		
		Côte d'Ivoire		Mali		Somalia		South Africa, Niger		
		Ethiopia		Rwanda		Zambia		Libya, RDC, Namibia		
		Madagascar		Libya		Mali		Morocco, Zambia, Guinea		
		Niger		Namibia		Ethiopia				
		Zimbabwe		Morocco		Tanzania				
		Burkina Faso		Cameroon						
				South Africa						
Group 3	[1.34, 1.46]	Angola	[1.41, 1.7]	Somalia	[1.14, 1.86]	Mauritania	[1.51, 1.46]	Somalia, Kenya, Mali	[2.85]	Algeria
		Kenya		Niger		Nigeria		Algeria, Tunisia		
		Libya		Guinea-Bissau		Egypt		Cameroon, Sudan, Egypt		
		Mauritania		Côte d'Ivoire		Sudan				
		Zambia		Tunisia		Algeria				
		Mozambique		RDC		Guinea-Bissau				
		Guinea		Kenya		South Africa				
		RDC		Angola		Kenya				
		Cameroon		Guinea		Morocco				
						Madagascar				
						Cameroon				
						Angola				
						Burkina-Faso				
						Niger				
						Chad				

Groups 1, 2, and 3: Are the respectively the classes of the highest, the average and the lowest values. **[R_0]:** Is the interval of the R_0 values at each class.
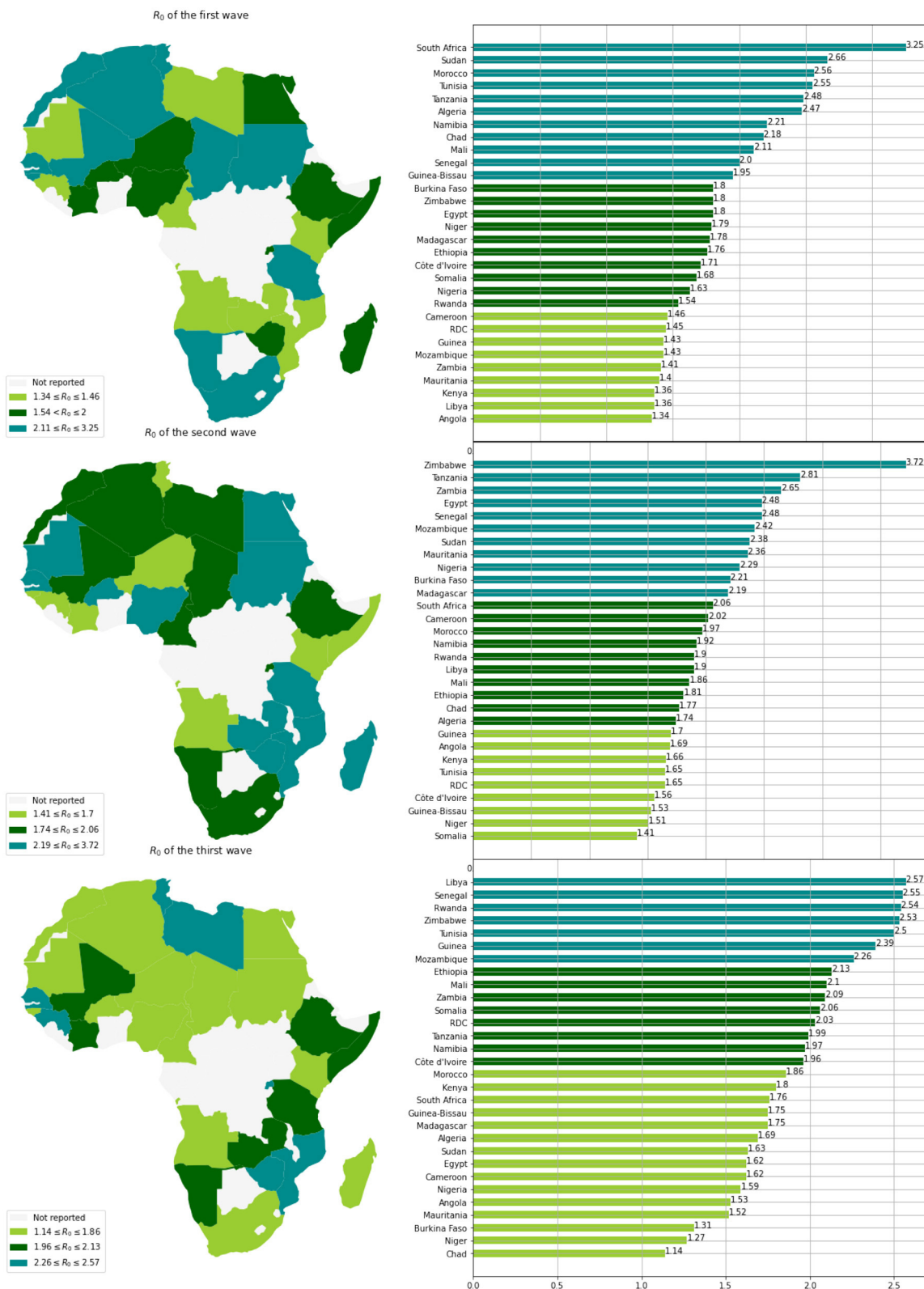
FIGURE 5
(Continued)
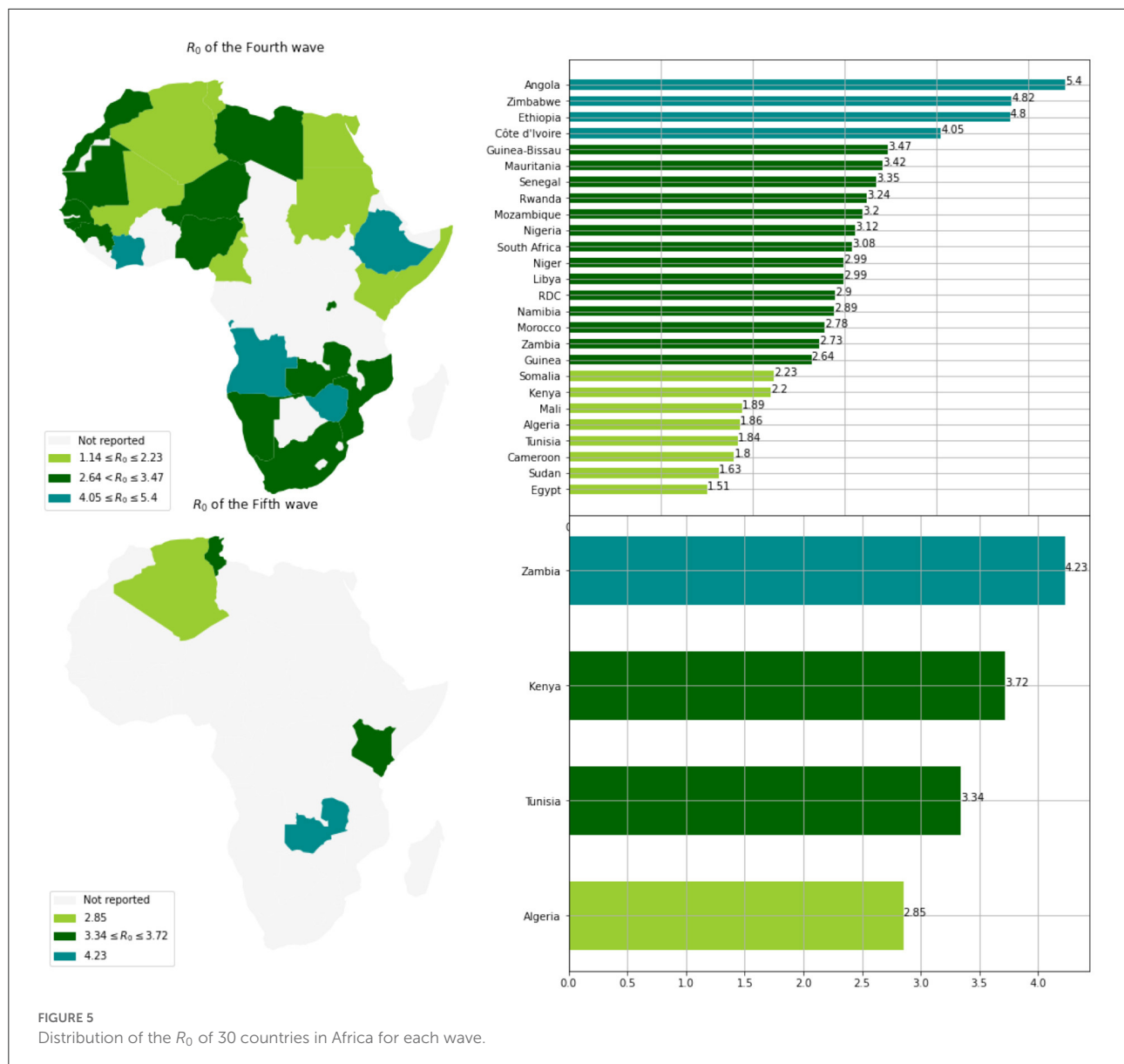
FIGURE 5
Distribution of the R_0 of 30 countries in Africa for each wave.

CHE plays a role in terms of preparedness and impact to improve the public health policy between waves in terms of screening capacity. A country with a high *CHE* has the material, human and technological resources to perform the volume of diagnostic tests and thus has the capacity to rapidly identify confirmed cases. This implies a strong dependence between the number of tests and *CHE* (17). It was noted that countries with low health system investment, *CHE*, often have a low testing capacity which makes it difficult to assess the true extent of COVID-19. For example, as of mid-April 2020, the Democratic Republic of the Congo was only performing about 200 tests per day (25), Senegal about 300 tests per day, and Ethiopia about 400 tests per day while the number

of tests was 3493 in South Africa.[8] For these countries, the question of the quality of the data and the reality of the virus circulation arises.

3.4. Impact of demographic factors

According to Table 4, there is a correlation between the demographic factors, i.e., the median age and *MeanR*. We observe (see Figure 5), that most countries with a median age

---

TABLE 4 Results of correlation analysis.

Country	Correlation coefficients
MeanR vs. _GDP_	
Mozambique, Mauritania, Nigeria Libya, Madagascar, Kenya, South Africa, RDC, Chad, Côte d'Ivoire, Sudan, Mali, Niger, Guinea-Bissau, Morocco, Tunisia, Somalia, Namibia, Burkina Faso, Guinea, South Africa, Algeria	0.722
MeanR vs. _CHE_	
Mozambique, Mauritania, Nigeria, Libya, Cameroon, Rwanda, Madagascar, South Africa, Guinea, RDC, Chad, Côte d'Ivoire, Sudan, Mali, Morocco, Tunisia, Zimbabwe,Kenya, Angola,	0.563
MeanR vs. Median age	
Mozambique, Mauritania, Nigeria, Libya, Cameroon, Rwanda, Madagascar, South Africa, Guinea, RDC, Chad, Côte d'Ivoire, Sudan, Mali, Niger, Morocco, Tunisia, Somalia, Namibia, Burkina Faso,Kenya, Guinea-Bissau,	0.626
MeanR vs. Mean Temperature	
Mozambique, Mauritania, Nigeria, Libya, Rwanda, Madagascar, Kenya, Guinea, RDC, Chad, Côte d'Ivoire, Sudan, Mali, Niger, Guinea-Bissau, Algeria, Tunisia, Angola, Somalia, Tanzania, Zambia, Namibia, Burkina Faso, Morocco, South Africa	−0.729

Country: The country used for the correlation analysis. **Correlation coefficient:** Is the value of the Pearson coefficient. **MeanR vs. Y:** The analysis of the correlation between the variable _MeanR_ and _Y_. where $Y \in \{GDP, CHE, Median_{age}, Mean\ Temperature\}$. We Show that if economic and the age are increasing, the _MeanR_ increasing while it is decreasing if the temperature increasing.

under 18 have a _MeanR_ less than 2, including Niger, Mali, Chad, Somalia, and Burkina Faso. While South Africa and other countries with a median age greater than 27 have a _MeanR_ greater than 2.

This result may be explained by the fact that older people are over-represented in the COVID-19 data since they are more likely to be tested and have more serious infections (26). In contrast, younger people tend to be in better health than older ones, making them more immune to infection. This has been observed in the influenza pandemic in Africa where children and adolescents had a negligible epidemiological impact (27).

3.5. Impact of climatic factors

The annual temperature and _MeanR_ are negatively highly correlated, as seen in Table 4 and Figure 5. Indeed, we observe that almost all of the countries with a lower value of _meanR_ have a dry climate and a high annual temperature (annual temperature greater than 27°C), in contrast to the northern countries, which have a lower annual temperature (annual temperature less than 23°C; see Figure 4).

Note that, negative correlation had already been observed in China (28), in several Latin American countries (29, 30), in the U.S.A. (31) and in Japan (32). From a biological point of view, low humidity dries out the nasal mucosa and impairs the stability of the aerosol droplets and therefore virus particles (33). Hence, the virus replication is limited by temperatures (20 and 30°C) (34).

3.6. Clustering of African country from epidemic, economic, demographic, and climatic variables

We performed a hierarchical clustering and a principal component analysis (PCA) (Figures 7–10). The PCA depicts two primary axes (PC1, PC2) that together account for 75% of country variation, 54.2% for the PC1 and 20.6% for the PC2. The primary parameters in PC2 are _MeanR_, median age, and _GDP_. For PC1, the primary parameter is the temperature.

We then performed a Hierarchical Clustering of the countries (see Figure 7). We were able to divide the countries into two distinct clusters. Morocco, Algeria, Tunisia, Libya, Egypt, Namibia, and South Africa make up the first cluster, which spans north and south Africa. This cluster is characterized by high median age and a high _GDP > US_$3000) (in [21.8, 32.7] years old). Except for Egypt (_MeanR_ = 1.8), this cluster has witnessed an average of _MeanR_ > 2.2. We have identified five countries that had a significant first wave.

The second cluster is in turn divided into sub-clusters, the first sub-cluster includes Tanzania, Ethiopia, Angola, Zambia, Rwanda, and Zimbabwe. This cluster is characterized by high values of _MeanR_. _MeanR_ in [2.4, 3.21] in Zimbabwe, a middle _GDP_ in US$ [797, 1800], and a median age < 20 years.

The latest cluster includes the rest of the countries. This cluster is characterized by a high annual temperature of more than 25°C and a _GDP_ of between US$2350 and US$438. These countries are distinguished by _MeanR_ values in [1.69, 2.59]. Twelve of the 18 countries of the second Group are included in this cluster.

TABLE 5 Distribution of the countries between classes waves of R_0 and GDP (in US\$$).

	GDP > 3000	1000 < GDP ≤ 3000	GDP ≤ 1000
First wave			
$C_1(1)$	Morocco, South Africa, Tunisia, Namibia, Algeria,	Tanzania, Senegal	Sudan, Chad, Guinea-Bissau Mali,
$C_2(1)$	Egypt,	Zimbabwe, Côte d'Ivoire	Somalia, Madagascar, Rwanda, Burkina-Faso, Ethiopia, Niger,
$C_3(1)$	Libya	Guinea, Cameroon Angola, Kenya, Mauritania,	Mozambique, RDC, Zambia
Second wave			
$C_1(2)$	Egypt	Tanzania, Senegal, Zimbabwe, Mauritania	Sudan, Madagascar, Mozambique, Zambia Burkina-Faso,
$C_2(2)$	Morocco, South Africa, Algeria, Namibia, Libya	Cameroon	Chad, Mali, Rwanda, Ethiopia
$C_3(2)$	Tunisia	Guinea, Angola, Côte d'Ivoire, Kenya,	Guinea-Bissau, Somalia, Niger, RDC
Third wave			
$C_1(3)$	Libya, Tunisia	Zimbabwe, Guinea	Mozambique, Rwanda, Ethiopia
$C_2(3)$	Namibia	Tanzania, Senegal, Côte d'Ivoire	Zambia, Mali, RDC, Somalia,
$C_3(3)$	Egypt, Morocco, Algeria, South Africa,	Mauritania, Cameroon, Angola, Kenya	Sudan, Madagascar, Guinea-Bissau, Niger, Burkina-Faso, Chad
Fourth wave			
$C_1(4)$		Zimbabwe, Côte d'Ivoire, Angola,	Ethiopia, Guinea-Bissau
$C_2(4)$	Libya, Namibia, South Africa, Morocco,	Guinea, Senegal Mauritania	Mozambique, Rwanda, RDC, Niger Zambia,
$C_3(4)$	Tunisia, Egypt, Algeria	Cameroon, Kenya	Mali, Somalia, Sudan
Fifth wave			
$C_1(5)$			Zambia
$C_2(5)$	Tunisia	Kenya	
$C_3(5)$	Algeria		
MeanR			
C_1	South Africa	Zimbabwe, Senegal, Angola	Ethiopia, Zambia
C_2	Morocco, Algeria, Tunisia, Libya	Tanzania, Guinea, Kenya, Mauritania, Nigeria, Côte d'Ivoire	Mozambique, Sudan, Guinea-Bissau, Rwanda, Mali,RDC,
C_3	Egypt	Cameroon	Somalia, Madagascar, Niger, Chad, Burkina-Faso

$C_i^{(j)}$: is the group $i, i = 1, 2, 3$ at wave $j, j = 1, 2, 3, 4, 5$. We show that countries with the highest GDPs in Africa (GDPs above US3000\$ per capita), especially South Africa and some North African countries like Tunisia, Morocco, and Algeria are experienced a significant first wave.

Country	Date
Morocco	2020-02-14
Egypt	2020-02-14
Kenya	2020-02-24
Tunisia	2020-02-28
Cameroon	2020-02-29
Senegal	2020-03-03
Algeria	2020-03-04
Cote d'Ivoire	2020-03-05
South Africa	2020-03-09
Burkina Faso	2020-03-12
Mali	2020-03-16
RDC	2020-03-18
Madagascar	2020-03-20
Libya	2020-03-24
Niger	2020-03-30
Guinea	2020-03-31
Sudan	2020-04-06
Tanzania	2020-04-06
Nigeria	2020-04-09
Somalia	2020-04-10
Mozambique	2020-04-11
Chad	2020-04-15
Guinea-Bissau	2020-04-20
Ethiopia	2020-05-02
Zambia	2020-05-04
Angola	2020-05-13
Mauritania	2020-05-22
Namibia	2020-05-26
Rwanda	2020-05-31
Zimbabwe	2020-07-02

FIGURE 6

Distribution of the t_0 for the first wave.

4. Conclusion

The objective of our paper was to document the Spatio-temporal variations in the baseline reproduction rate R_0 and to understand the reasons for the different disparities between them. We highlight that more developed countries experienced a higher incidence in the first wave, which can be explained by their higher international exposure. We also show that the quality of health systems played a key role in limiting virus-related mortality. Consistent with the literature, we also show that countries with younger populations were less affected by the pandemic. Finally, we show that climate also plays a determining role in explaining the reproduction rate R_0. At the end of the analysis of the determinants, we have made a clustering of the countries in order to identify which ones have been the most suffering during this pandemic or on the contrary which ones have been the most resistant.

Our results show that the geography of the pandemic in Africa largely overlaps with the geography of the wealth of the states. Consequently, the fight against poverty and the development of health infrastructures are sine-qua-non conditions for an effective fight against future epidemics or pandemic crises that could occur.

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

BN, AK, and SBe contributed to conception and design of the study. BN and WB organized the database. BN and SBe performed the statistical analysis. BN, SBo, and WB wrote the first draft of the manuscript. BN, SBo, AK, and SBe wrote the final draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FIGURE 7
Hierarchical clustering (Dendrogram) for the pandemic, economic, demographic, and climatic variables allows countries to be grouped into 3 separate clusters.

FIGURE 8
Principal components (PC) plot from the principal component analysis (PCA) on the pandemic, economic, demographic, and climatic variables.

Funding

This work was supported in part by the French Ministry for Europe and Foreign Affairs *via* the project REPAIR COVID-19-Africa coordinated by the Pasteur International Network association, by European Union's Horizon 2020 research and innovation program under grant agreement No. 883441 (STAMINA) and the Tunisian Ministry for High Education *via* PRF project PRFCOV19-D5P1 Evaluer, prédire, agir: stratégie fédérée pour la lutte contre la 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

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. 2022.1039925/full#supplementary-material

FIGURE 9
Correlation circles for the two first PC planes.

FIGURE 10
Projection of the 30 African countries of the PCA's plot on the first PC plane.

References

1. Jeanne L, Bourdin S, Nadou F, Noiret G. Economic globalization and the COVID-19 pandemic: global spread and inequalities. *GeoJournal*. (2022). doi: 10.1007/s10708-022-10607-6. [Epub ahead of print].

2. Mehtar S, Preiser W, Lakhe NA, Bousso A, TamFum JJM, Kallay O, et al. Limiting the spread of COVID-19 in Africa: one size mitigation strategies do not fit all countries. *Lancet Global Health*. (2020) 8:e881–3. doi: 10.1016/S2214-109X(20)30212-6

3. El-Sadr WM, Justman J. Africa in the path of COVID-19. *N Engl J Med*. (2020) 383:e11. doi: 10.1056/NEJMp2008193

4. Bourdin S, Ben Miled S, Salhi J. The drivers of policies to limit the spread of COVID-19 in Europe. *J Risk Fin Manage*. (2022) 15:67. doi: 10.3390/jrfm15020067

5. Haider N, Osman AY, Gadzekpo A, Akipede GO, Asogun D, Ansumana R, et al. Lockdown measures in response to COVID-19 in nine sub-Saharan African countries. *BMJ Global Health*. (2020) 5:e003319. doi: 10.1136/bmjgh-2020-003319

6. Maeda JM, Nkengasong JN. The puzzle of the COVID-19 pandemic in Africa. *Science*. (2021) 371:27–28. doi: 10.1126/science.abf8832

7. BenMiled S, Borgi C, Hsairi M, Somrani N, Kebir A. Hospital bed capacity across in Tunisia hospital during the first four waves of the COVID-19 pandemic: a descriptive analysis. *medRxiv [Preprint]*. (2022). doi: 10.1101/2022.08.23.22279122

8. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*. (2020) 585:410–3. doi: 10.1038/s41586-020-2293-x

9. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. (2020) 584:257–61. doi: 10.1038/s41586-020-2405-7

10. Lwasa S. Appreciating the heterogeneity in the unity of Africa: a socio-ecological perspective on Africa's geographies. *Can Geogr*. (2019) 63:594–602. doi: 10.1111/cag.12576

11. Zhang F, Karamagi H, Nsenga N, Nanyunja M, Karinja M, Amanfo S, et al. Predictors of COVID-19 epidemics in countries of the World Health Organization African Region. *Nat Med*. (2021) 27:2041–7. doi: 10.1038/s41591-021-01491-7

12. Martinez-Alvarez M, Jarde A, Usuf E, Brotherton H, Bittaye M, Samateh AL, et al. COVID-19 pandemic in west Africa. *Lancet Global Health*. (2020) 8:e631–2. doi: 10.1016/S2214-109X(20)30123-6

13. Salyer SJ, Maeda J, Sembuche S, Kebede Y, Tshangela A, Moussif M, et al. The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study. *Lancet*. (2021) 397:1265–75. doi: 10.1016/S0140-6736(21)00632-2

14. Bailey D, Clark J, Colombelli A, Corradini C, De Propris L, Derudder B, et al. Regions in a time of pandemic. *Region Stud*. (2020) 54:1163–74. doi: 10.1080/00343404.2020.1798611

15. Bourdin S, Jeanne L, Nadou F, Noiret G. Does lockdown work? A spatial analysis of the spread and concentration of COVID-19 in Italy. *Region Stud*. (2021) 55:1182–93. doi: 10.1080/00343404.2021.1887471

16. McCann P, Ortega-Argilés R, Yuan PY. The COVID-19 shock in European regions. *Region Stud*. (2022) 56:1142–60. doi: 10.1080/00343404.2021.1983164

17. Oshinubi K, Rachdi M, Demongeot J. Analysis of reproduction number R0 of COVID-19 using current health expenditure as gross domestic product percentage (CHE/GDP) across countries. In: Oshinubi K, Rachdi M, Demongeot J, editors. *Healthcare*. vol. 9. MDPI (2021). p. 1247. doi: 10.3390/healthcare9101247

18. Asfahan S, Shahul A, Chawla G, Dutt N, Niwas R, Gupta N. Early trends of socio-economic and health indicators influencing case fatality rate of COVID-19 pandemic. *Monaldi Arch Chest Dis*. (2020) 90. doi: 10.4081/monaldi.2020.1388

19. Kompas T, Grafton RQ, Che TN, Chu L, Camac J. Health and economic costs of early and delayed suppression and the unmitigated spread of COVID-19: the case of Australia. *PLoS ONE*. (2021) 16:e0252400. doi: 10.1371/journal.pone.0252400

20. Barek MA, Aziz MA, Islam MS. Impact of age, sex, comorbidities and clinical symptoms on the severity of COVID-19 cases: a meta-analysis with 55 studies and 10014 cases. *Heliyon*. (2020) 6:e05684. doi: 10.1016/j.heliyon.2020.e05684

21. Liu Z, Magal P, Seydi O, Webb G. Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions. *Biology*. (2020) 9:50. doi: 10.3390/biology9030050

22. Ree MJ. Correlation and regression: applications for industrial organizational psychology and management. *Organ Res Methods*. (2002) 5:200. doi: 10.1177/1094428102005002005

23. Nielsen F (ed.). Hierarchical clustering. In: *Introduction to HPC with MPI for Data Science*. Springer (2016). p. 195–211. doi: 10.1007/978-3-319-21903-5_8

24. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscip Rev Comput Stat*. (2010) 2:433–59. doi: 10.1002/wics.101

25. Wells CR, Stearns JK, Lutumba P, Galvani AP. COVID-19 on the African continent. *Lancet Infect Dis*. (2020) 20:1368–70. doi: 10.1016/S1473-3099(20)30374-1

26. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HH, Mercer SW, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Global Health*. (2020) 8:e1003–17. doi: 10.1016/S2214-109X(20)30264-3

27. Munro AP, Faust SN. Children are not COVID-19 super spreaders: time to go back to school. *Arch Dis Childhood*. (2020) 105:618–9. doi: 10.1136/archdischild-2020-319474

28. Li H, Xu XL, Dai DW, Huang ZY, Ma Z, Guan YJ. Air pollution and temperature are associated with increased COVID-19 incidence: a time series study. *Int J Infect Dis*. (2020) 97:278–82. doi: 10.1016/j.ijid.2020.05.076

29. Bolano-Ortiz TR, Camargo-Caicedo Y, Puliafito SE, Ruggeri MF, Bolano-Diaz S, Pascual-Flores R, et al. Spread of SARS-CoV-2 through Latin America and the Caribbean region: a look from its economic conditions, climate and air pollution indicators. *Environ Res*. (2020) 191:109938. doi: 10.1016/j.envres.2020.109938

30. Prata DN, Rodrigues W, Bermejo PH. Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. *Sci Tot Environ*. (2020) 729:138862. doi: 10.1016/j.scitotenv.2020.138862

31. Li AY, Hannah TC, Durbin JR, Dreher N, McAuley FM, Marayati NF, et al. Multivariate analysis of black race and environmental temperature on COVID-19 in the US. *Am J Med Sci*. (2020) 360:348–56. doi: 10.1016/j.amjms.2020.06.015

32. Ujiie M, Tsuzuki S, Ohmagari N. Effect of temperature on the infectivity of COVID-19. *Int J Infect Dis*. (2020) 95:301–3. doi: 10.1016/j.ijid.2020.04.068

33. Schaffer F, Soergel M, Straube D. Survival of airborne influenza virus: effects of propagating host, relative humidity, and composition of spray fluids. *Arch Virol*. (1976) 51:263–73. doi: 10.1007/BF01317930

34. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog*. (2007) 3:e151. doi: 10.1371/journal.ppat.0030151

35. Bozorg-Haddad O, Solgi M, Loáiciga HA. *Meta-Heuristic and Evolutionary Algorithms for Engineering Optimization*. John Wiley & Sons (2017). doi: 10.1002/9781119387053

Appendix: Model description

This model consists of the following system of ordinary differential equations:

$$\begin{cases} S'(t) = -\tau S(t)[I(t) + U(t)] \\ I'(t) = \tau S(t)[I(t) + U(t)] - \nu I(t) \\ R'(t) = \nu_1 I(t) - \eta R(t) \\ U'(t) = \nu_2 I(t) - \eta U(t) \end{cases} \tag{1}$$

Where $t \geq t_0$ the time in days, t_0 is the beginning date of each wave, $S(t)$ is the number of individuals susceptible to infection at time t, $I(t)$ is the number of infectious individuals at time t, $R(t)$ is the number of reported infectious individuals at time t and $U(t)$ is the number of unreported infectious individuals at time t. This system is supplemented by initial condition at time $t = t_0$, (S_0, I_0, R_0, U_0).

We assume that the cumulative number of reported symptomatic cases at time t is proportional to the cumulative number of symptomatic cases for each time t. Let's denote the proportion coefficient by f. Therefore, the rate of asymptomatic infectious becoming reported symptomatic is $\nu_1 = f\nu$ and the rate of asymptomatic infectious becoming unreported symptomatic is $\nu_2 = (1 - f)\nu$.

Table A1 represents the set of parameters that are fixed by the hypothesis and those evaluated by the country model.

We assume that $\eta = \frac{1}{7}$ and $\nu = \frac{1}{7}$, are fixed for all African countries, which means that the average period of infectiousness of both unreported symptomatic infectious individuals and reported symptomatic infectious individuals and that the average period of infectiousness is 7 days. The fraction of total infectious cases that are reported f is unknown and varies from region to region.

The cumulative number of the reported symptomatic infectious cases at time t is obtained by using the following equation (21):

$$CR(t) = \nu_1 \int_{t_0}^{t} I(s) ds \tag{2}$$

Since in the early stage of the epidemic, all the infected components of the system grow exponentially and the number of

Table A1 Parameters and initial conditions of the model 1.

Symbol	Interpretation	Method
t_0	Time at which the epidemic started	Fitted
S_0	Number of susceptible at time t_0	Fixed
I_0	Number of asymptomatic infectious at time t_0	Fitted
U_0	Number of unreported symptomatic infectious at time t_0	Fitted
τ	Transmission rate	Fitted
$\frac{1}{\nu}$	Average time during which asymptomatic infectious are asymptomatic	Fitted
f	Fraction of asymptomatic infectious that become reported symptomatic infectious	Estimated
$\nu_1 = f\nu$	Rate at which asymptomatic infectious become reported symptomatic	Fitted
$\nu_2 = (1-f)\nu$	Rate at which asymptomatic infectious become unreported symptomatic	Fitted
$\frac{1}{\eta}$	Average time symptomatic infectious have symptoms	Fixed

susceptible remains unchanged during a relatively short period of time t, we can fit an exponentially growing curve $CR(t)$ to the cumulative reported cases data defined by the following special form :

$$CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3 \tag{3}$$

with χ_1, χ_2 and χ_3 three positive numbers that we estimate using log-linear regression and the Genetic algorithm optimization method (35).

Following (21), we have:

$$\begin{cases} t_0 = \frac{1}{\chi_2}(\ln(\chi_3) - \ln(\chi_1)) \\ I_0 = \frac{\chi_1 \chi_2 \exp(\chi_2 t_0)}{f\nu} \\ U_0 = \frac{\nu_2}{\eta + \chi_2} I_0 = \frac{(1-f)\nu}{\eta + \chi_2} I_0 \\ \tau = \frac{\chi_2 + \nu}{S_0} \frac{\eta + \chi_2}{\nu_2 + \eta + \chi_2} \\ R_0 = \frac{\tau S_0}{\nu}(1 + \frac{\nu_2}{\eta}) \end{cases} \tag{4}$$

Check for updates

An NLP tool for data extraction from electronic health records: COVID-19 mortalities and comorbidities

Sana S. BuHamra[1]*, Abdullah N. Almutairi[1],
Abdullah K. Buhamrah[2], Sabah H. Almadani[1] and
Yusuf A. Alibrahim[3]

[1]Department of Information Science, Kuwait University, Kuwait City, Kuwait, [2]Surgery Department,
Al-Adan Hospital, Al Ahmadi, Kuwait, [3]Department of Medical Imaging, Faculty of Medicine,
University of Toronto, Toronto, ON, Canada

Background: The high infection rate, severe symptoms, and evolving aspects of the COVID-19 pandemic provide challenges for a variety of medical systems around the world. Automatic information retrieval from unstructured text is greatly aided by Natural Language Processing (NLP), the primary approach taken in this field. This study addresses COVID-19 mortality data from the intensive care unit (ICU) in Kuwait during the first 18 months of the pandemic. A key goal is to extract and classify the primary and intermediate causes of death from electronic health records (EHRs) in a timely way. In addition, comorbid conditions or concurrent diseases were retrieved and analyzed in relation to a variety of causes of mortality.

Method: An NLP system using the Python programming language is constructed to automate the process of extracting primary and secondary causes of death, as well as comorbidities. The system is capable of handling inaccurate and messy data, this includes inadequate formats, spelling mistakes and mispositioned information. A machine learning decision trees method is used to classify the causes of death.

Results: For 54.8% of the 1691 ICU patients we studied, septic shock or sepsis-related multiorgan failure was the leading cause of mortality. About three-quarters of patients die from acute respiratory distress syndrome (ARDS), a common intermediate cause of death. An arrhythmia (AF) disorder was determined to be the strongest predictor of intermediate cause of death, whether caused by ARDS or other causes.

Conclusion: We created an NLP system to automate the extraction of causes of death and comorbidities from EHRs. Our method processes messy and erroneous data and classifies the primary and intermediate causes of death of COVID-19 patients. We advocate arranging the EHR with well-defined sections and menu-driven options to reduce incorrect forms.

KEYWORDS

natural language processing, text mining, information extraction, SARS-CoV-2, mortality, decision tree, prediction

Introduction

The COVID-19 pandemic has had a significant impact on how and where healthcare is delivered effectively and efficiently. During the pandemic, the need for novel and current technologies arise to assist in predicting clinical outcomes in critical time with the high overflow of patients. Clinical (text) notes constitute a major source of medical data and are rarely used to their full capacity, even though they include a wealth of subjective information. Prior to electronic health records (EHRs), practitioners had to manually collect data from clinical notes, which was costly and difficult to scale up. Despite the expanding volumes of healthcare data, Kong (1) claims that over 80% of text, image, signal, and other medical data collections remain unstructured and unused. One main goal in medical research is to use EHRs to extract and analyze well-structured data. Many methods were devised and evaluated using EHRs for detecting patients with known risk factors for consequences such as stroke and significant bleeding (2), as well as investigating the difficulties of decoding and comprehending clinical narratives (3). Natural language processing (NLP) can expedite diagnosis and care to patients who are most vulnerable during pandemics by using textual data from medical records. According to Zhou et al. (4), only NLP can extract information about a patient's family history from free-text clinical papers. The researchers employed word embeddings and a Convolutional Neural Network (CNN) to recognize International Classification of Diseases (ICD-10) diagnostic codes in discharge notes and outperformed current methods with little data preparation (5).

Artificial Intelligence (AI) and Machine Learning (ML) technologies including NLP can be used to aid in the diagnosis and treatment of individuals suffering from acute and chronic diseases during the COVID-19 pandemic. DeCapprio et al. (6) used medical records that had already been made public as COVID-19 proxies (pneumonia, influenza, acute bronchitis, and upper respiratory illnesses). Zoabi et al. (7) came up with a machine learning decision tree model that predicts a positive COVID-19 infection in an RT-PCR test during the first month of the pandemic. Izquierdo et al. (8) used a mix of traditional epidemiological methods, NLP, and ML predictive modeling to find out what symptoms COVID-19 patients have that make them likely to be admitted to the ICU. Guan et al. (9) employed simple-tree XGBoost to identify high-risk COVID-19 cases and assessed how much faster causes of death may be identified using minimally preprocessed notes.

This study intends to construct an NLP system to automate the extraction of primary and secondary causes of death, as well as comorbidities, from the mortality EHRs of COVID-19 patients admitted to the ICU in Kuwait during the pandemic. Since many of the free-text notes were inadequately formatted, contained spelling mistakes and were placed in the wrong field, acquiring sufficient and reliable data was the largest hurdle. In fact, the causes of death in most records in our data were not expressed precisely nor was in the correct field although the EHRs file is mortality specific.

Other work in the literature used available clean EHRs for their analysis. However, EHRs may sometimes be inaccurate and noisy due to them being compiled under extreme pressures of time and manpower due to the large influx of patients with critical cases, such as the case during the pandemic. EHRs need to be first corrected and cleaned to be used for proper analysis or be used in medical systems such the Unified Medical Language System (UMLS) and SNOMED CT. Otherwise, a significant amount of information will be lost.

To correct the EHRs we used physicians as the domain knowledge experts to understand and extract the common mistakes in the EHRs that were done by their fellow physicians. Their knowledge and findings were converted to a Python language code to automate cleaning and fixing the data in the EHRs. Also, the Python code used the domain expert knowledge to distinguish between acute diseases and causes of death in some circumstances. In addition, the causes of death were classified to a direct cause or a related one. Comorbidities were used as an important factor in analyzing the cause of death. This will offer precise information on the casuality and spectrum of comorbidities in fatal instances, allowing for an accurate evaluation of COVID-19's hazardous nature. Finally, we have utilized a decision tree-based model to predict death due to ARDS or other complications. These findings can assist healthcare systems to plan for the spread of future pandemics and identify groups at risk.

Methods

The data

Data on COVID-19 mortalities were retrieved from Jaber Hospital's mortality Electronic Health Records (EHR) for all patients admitted to the ICU between March 7, 2020, and August 19, 2021, and death reported between March 7, 2020, and August 27, 2021. The data set contains 1691 cases after excluding 12 children (<17 years old) and 46 with no data entries. The monthly total death rate in Kuwait is depicted in Worldometer cite (10). On the final day of data collection for this study, the total number of COVID-19 deaths was reported to be 2415; thus, our sample size covers 70% (1691/2415) of the COVID-19 mortality population. We also covered all death peaks and pandemic main waves during this time.

Initially, the data was extracted as a pdf file and then converted to an Excel spreadsheet. Patients' demographics (age, gender, and residency), date of ICU admission, date of death, reasons for admission, admission diagnosis, final diagnosis, cause of death, brief history, brief summary, and contributing factors are all included in each record. To ensure confidentiality, all data was anonymized and all patient identifiers were

FIGURE 1
Word cloud plot of death causes and contributing factors.

removed. Additional data cleansing were also performed to ensure data accuracy.

Creating corpus of terminologies

The data sheet obtained from EHR has mainly eleven columns: date of admission to ICU, date of death, age at death, admission diagnosis, reason for admission, final diagnosis, cause of death (COD), brief history, brief summary, and contributing factors. Except for the first three, all remaining columns are text features.

The underlying cause of death, such as "COVID-19" or "COVID 19 pneumonia," was listed in the COD column in many records, whereas the primary/intermediate causes were found explicitly or indirectly in the brief summary or brief history columns. Furthermore, there were two major flaws in the free-text notes in the mortality EHR. The first issue is that many terminologies have misspellings or improper forms. For example, multiorgan failure is referred to as "Multi-Organs" or "Multiorgan Failure." The second issue is inconsistency in the reporting of text notes. The causes of death are not always listed in the data columns that you would expect. Comorbidities, on the other hand, are not consistently included in the list of contributing factors. As a result, we are unable to use the existing NLP tools or Unified Medical Language System (UMLS). We had to develop our own system to extract concepts, knowledge, and relationships from the mortality EHR at hand.

TABLE 1 Primary causes of death.

Primary COD (Abbrev.)	Alternative terms
Cardiopulmonary arrest (CPA)	Cardiopulmonary collapse, cardiorespiratory arrest, cardiorespiratory failure, cardiorespiratory collapse, circulatory collapse, asystole
Cardiac arrest (CA)	Cardiogenic shock, cardiovascular collapse, cardiac event, bradycardic arrest, STEMI
Respiratory failure (HRF)	Pulmonary failure, pulmonary arrest, pulmonary dysfunction, hypoxia, hypoxic, hypoxemia, hypoxemic, desaturate
Multiorgan failure (MOF)	MODS, multiple organ dysfunction syndrome, multi organ failure, multiple organ failure, multisystem failure
Hepatic failure (LF)	Liver failure, worsening liver function, hepatic failure
Renal failure (RF)	kidney failure, dialysis, CRRT
Septic shock (SS)	

CODs and comorbidity glossary tables

Our strategy is to extract the causes of death (COD) and comorbidities/diseases by using NLP techniques such as a bag-of-word (BoW) model. The BoW model will be applied on each column to extract all terms and phrases that represent the CODs and comorbidities for each patient. The model achieves this by tokenizing all text columns in the data sheet and creating a case/term occurrence matrix where each row represents a

TABLE 2 Intermediate causes of death.

Intermediate COD (Abbrev.)	Alternative terms
Acute respiratory distress syndrome (ARDS)	Mechanical ventilation, acute respiratory failure, hypoxic respiratory failure, HRF
Acute kidney failure (AKI)	Acute kidney injury, renal impairment, anuric, hyperkalemia, dialysis
Pulmonary embolism (PE)	DVT collapse, thrombosis
Heart failure (HF)	Rescue PCI, cardiomyopathy, myocarditis
Stroke (ST)	CVA, cerebrovascular accident, failed thrombolysis, hemorrhagic cerebral, subdural, subarachnoid hemorrhage, hge
Pneumothorax (PN)	Tension pneumothorax, hemothorax, hemopneumothorax, hydropneumothorax, pneumoperitoneum, bilateral chest tubes, chest tube
Myocardial infarction (MI)	STEMI, PCI, CCU, ischemic changes, cardiac strain, st elevation, troponin elevated, NStemi
Arrhythmia (AR)	Ventricular fibrillation, VFib, ventricular tachycardia, vtach, rhythm, atrial fibrillation, AF, PAF
Bleeding (BL)	ICH, hematoma, AVM, intracerebral hemorrhage, epistaxis, PRBC, transfusion, melena, upper GI bleeds
Disseminated intravascular coagulation (DI)	DIC
Urinary tract infection (UT)	UTI, urinary tract infection, urosepsis, E.col

TABLE 3 General disease categories (GDC), comorbidities and other risk factors.

GDC (Abbrev.)	Comorbidity/risk factor (Abbrev.)
Endocrine, nutritional, and metabolic diseases (ENMs)	Diabetes mellitus (DM), thyroid disease (THY), dyslipidemia (DLP), obesity (OB), Addison disease (ADs)
Diseases of the nervous system (DNS)	Stroke (CVA), Parkinson's disease (PD), dementia (DEM), multiple sclerosis (MS), epilepsy (EP), psychiatric disorders (OCD)
Diseases of the circulatory system (DCS)	Hypertension (HTN), anemia (IDA), pulmonary embolism (PE), peripheral vascular disease (PVD), bleeding disorders (BDs)
Cardiovascular system diseases (CVD)	Coronary artery disease (CAD), cardiomyopathy (HCM), valvular heart disease (AVR), heart failure (HF), arrhythmia (AF)
Respiratory diseases (RDs)	Asthma (BA), chronic obstructive pulmonary disease (COPD), lung disease (LD)
GI disorders (GIDs)	Inflammatory bowel disease (IBD), gastroesophageal reflux (GERD), liver disease (LD)
Diseases of the genitourinary (DGS)	Chronic kidney disease (CKD), benign prostatic hyperplasia (BPH)
Autoimmune disorders (ADs)	Rheumatoid arthritis (RA), Immunecompromised (IC) (a risk factor)
Ortho disorders (ODs)	Bone disorders (OA)
Infectious diseases (IDs)	HIV-infection (HIV)
Neoplasms (CRC)	Cancer (CA) of any kind
Congenital disorders (CDs)	Down syndrome (DS)

patient's case and each column represents each medical term relating to a cause of death or comorbidity, all other word tokens will be omitted. The cells of the matrix will contain a 0 or 1 representing the occurrence or absence of the term from the case. The terms related to cause of death will be categorized to three stages similar to the fashion of death certificates. These stages are the primary, intermediate and the underlying cause (which led to the intermediate).

The list of primary causes of death, according to WHO guidelines, denotes the condition (injury, complication, or disease) that directly preceded death. WHO issued an updated International Classification of Diseases (ICD) and health-related problems to accommodate COVID-19-related death complications (11). The condition(s) that led to the primary COD are reflected in the intermediate COD. Multiple

complications contributing to the intermediate COD were identified in the majority of COVID-19 decedents in the ICU in this study. Additionally, COVID-19 pneumonia was the most frequently encountered underlying cause in those ICU cases, resulting in an intermediate stage of complication.

In order to create BoW, the COD and comorbidity terms were extracted from the EHR in several steps. Starting with a preliminary text analysis using the text mining package (tm) and the word cloud generator package (wordcloud) in R to extract the most common terms (Figure 1). To create glossary tables, our medical experts validated the extracted terms by reviewing 50–100 EHRs at random. The process was repeated four times to ensure that the majority of the terminologies were covered. This helped identify alternative terminologies and misspelled terms. Tables 1, 2 show the refined
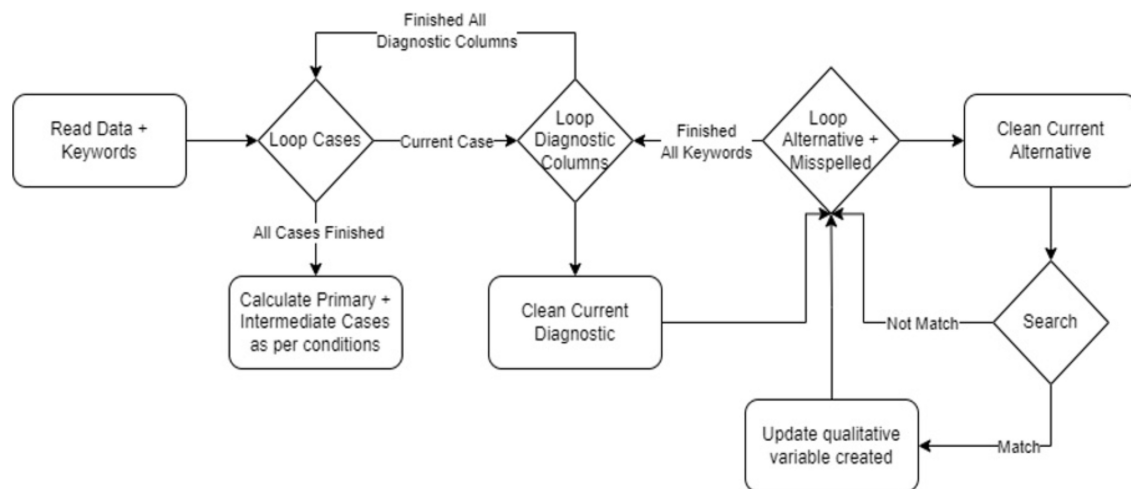
FIGURE 2
Primary and Intermediate CODs encoding flowchart.

FIGURE 3
Pseudocode for Primary COD.

list of primary COD and intermediate COD. In accordance with the International Classification of Diseases (11), Table 3 provides twelve general disease categories (GDC), 34 distinct comorbidities, and a risk factor associated with our data. Detailed versions of Tables 1–3, including all potential alternate terms and/or incorrect forms may be requested from the corresponding author.

Developing and applying NLP methods

We created an NLP method to identify, extract, and automatically encode natural language from mortality EHRs into structured clinical data. Tables 1, 2 are used as keywords to extract primary and intermediate CODs, while Table 3 presents

keywords to extract comorbidities. Method created in Python. Figure 2 shows our algorithm.

In this method, text is stripped of punctuation, special characters, capitalization, stop words, and tokenization. Used EHR variables include cause of death, final diagnosis, brief history, and brief summary. To create a case/COD term occurrence matrix, binary variables must be created for each primary/intermediate COD listed in Tables 1, 2. Initial occurrence matrix setting is zero. CODs or equivalents are compatible with tokens. The case/term occurrence matrix cell is set to 1 upon a match. Every case applies (rows). A COD abbreviation was not mistaken for a term, as PE is not present in hypertensive or hyperthyroid. Negation was also carefully handled; if a term is preceded by a negative or conditional word, it will not match. Exclusion words consist of (no, not, no sign of, non, no history of, no active, no previous medical, not known to have, no indications of, previous condition, old condition). Text format is used to list the final primary and intermediate CODs. The pseudocode used to extract the final primary COD is depicted in Figure 3.

Determining the actual intermediate CODs are handled differently. Multiple intermediate CODs are reported as a group. Our clinicians manually validated and separated the correct outcome to determine which disorders were terminal. A counter matching the extracted causes is also computed to help identify the terminal cause based on the most common causes to cross-check the accuracy of the findings.

The comorbidities for each case are identified using Table 3 in the same manner that CODs are identified. Preprocessed word tokens are extracted from the EHR reason for admission, contributing factors, admission diagnosis and brief summary.

Data manipulation and analysis

Original EHR mortality data had two sets of variables. First set included seven categorical and quantitative variables. Second set included eight free-text variables. The pdf data sheet was converted to an Excel sheet for data manipulation and cleaning. The second set of data was used to generate 70 variables using Python to determine death causes and comorbidities. During exploratory data analysis, we generated appropriate graphs (bar, pie, boxplots) and summary statistics (mean, median, SD, IQR). Hypothesis tests included Chi-square, TURF, ANOVA, and Kruskal Wallis. Finally, we built our prediction model with a decision tree. SPSS V23 and R were used for the statistical analysis.
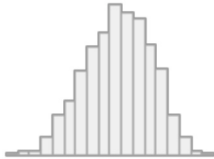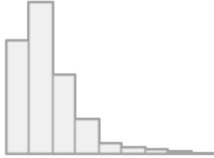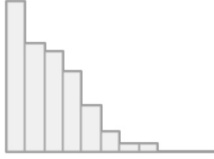
Results

Overall findings

The majority of the 1,691 anonymous COVID-19 decedents were male 963 (56.9%). The age at death ranges between 19.8 and 103.2 years with 63.8 years (SD 14.4). On the average the duration stay in ICU prior to death was 18.5

TABLE 4 Demographic and clinical characteristics.

Variable	Summary	Count (%)	Graph
Age	Mean (sd): 63.8 (14.4) min ≤ med ≤ max: 19.8 ≤ 64.5 ≤ 103.2 IQR (CV): 20.7 (0.2)	662 distinct values	
Age group	1. (< 50) years 2. (50–64) years 3. (65+) years	303 (17.9%) 573 (33.9%) 815 (48.2%)	
Gender	1. Female 2. Male	728 (43.1%) 963 (56.9%)	
ICU (days)	Mean (sd): 18.5 (12.8) min ≤ med ≤ max: 0 ≤ 16 ≤ 86 IQR (CV): 14 (0.7)	74 distinct values	
Total comorbidities	Mean (sd): 2.7 (1.9) min ≤ med ≤ max: 0 ≤ 3 ≤ 11 IQR (CV): 3 (0.7)	12 distinct values	
Total comorbidities group	Mean (sd): 2.6 (1.7) min ≤ med ≤ max: 0 ≤ 3 ≤ 6 IQR (CV): 3 (0.6)	0 : 172 (10.8%) 1 : 288 (18.1%) 2 : 333 (20.9%) 3 : 304 (19.1%) 4 : 245 (15.4%) 5 : 143 (9.0%) 6 : 110 (6.9%)	

days (SD 12.8). Two or more comorbidities were present (mean 2.5, SD 1.9) with hypertension and diabetes mellitus shared among more than half of them (Table 4). Since these patients died in the intensive care unit, COVID-19

pneumonia was mainly the underlying cause of death that resulted in intermediate and thus primary causes of death. COVID-19 pneumonia was detected in 94 percent of cases (1592/1691).

TABLE 5 Demographic and clinical characteristics by age group.

| Variable | N | Overall N = 1,691[a] | Age group (yrs.) | | | p-value[b] |
			Age < 50 N = 303[a]	Age [50-64] N = 573[a]	Age =65+ N = 815[a]	
Age	1,691	64 (54, 74)	43 (39, 47)	58 (54, 62)	75 (70, 81)	<0.001
Gender	1,691					0.003
Female		728 (43%)	114 (38%)	229 (40%)	385 (47%)	
Male		963 (57%)	189 (62%)	344 (60%)	430 (53%)	
ICU (days)	1,691	16 (10, 24)	14 (9, 23)	16 (10, 24)	16 (10, 24)	0.19
Total comorbidities	1,596	3 (1, 4)	1 (0, 2)	2 (1, 3)	3 (2, 5)	<0.001
Unknown		95	0	33	62	
Total group comorbidities	1,595					<0.001
0		172 (11%)	111 (37%)	61 (11%)	0 (0%)	
1		288 (18%)	85 (28%)	122 (23%)	81 (11%)	
2		333 (21%)	59 (19%)	136 (25%)	138 (18%)	
3		304 (19%)	30 (9.9%)	99 (18%)	175 (23%)	
4		245 (15%)	12 (4.0%)	72 (13%)	161 (21%)	
5		143 (9.0%)	2 (0.7%)	32 (5.9%)	109 (14%)	
6		110 (6.9%)	4 (1.3%)	18 (3.3%)	88 (12%)	
Unknown		96	0	33	63	

[a]Median (IQR) or Frequency (%).
[b]Kruskal-Wallis rank sum test; Pearson's Chi-squared test.

FIGURE 4
Combined intermediate complications and terminal Intermediate COD.

TABLE 6 Primary by intermediate causes of death.

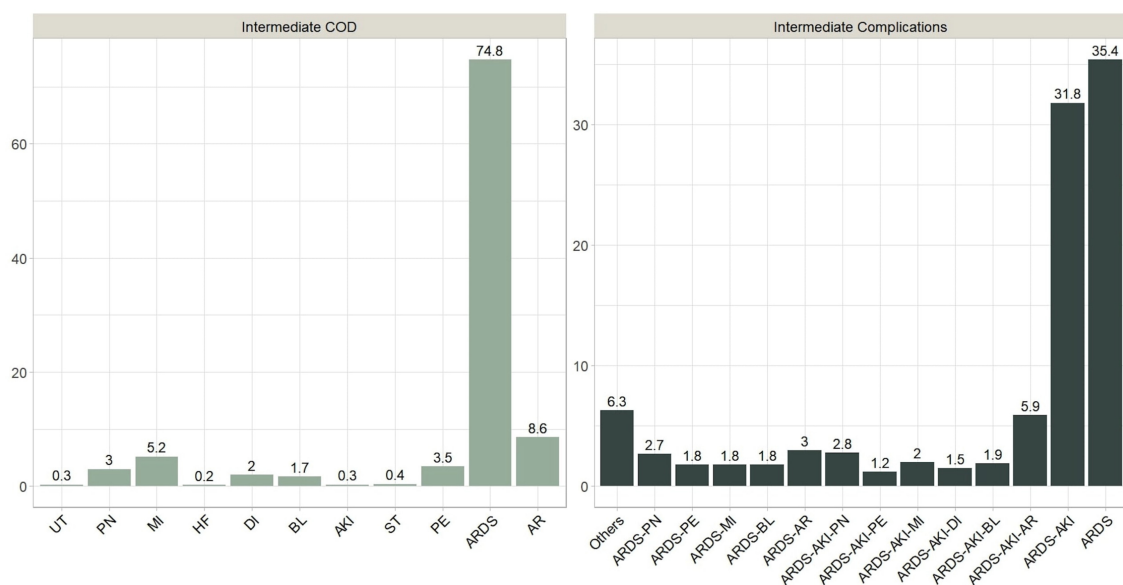| Inter-mediate COD	Primary COD count (%)							
	SS	CPA	HRF	CA	SS + MOF	MOF	RF	Total
AR	53 (7.9)	26 (8.6)	11 (4.7)	26 (14.4)	14 (10.4)	12 (9.6)	3 (6.8)	145 (8.6)
ARDS	523 (78.4)	220 (72.4)	194 (82.6)	117 (65)	94 (69.6)	8 (64.8)	35 (79.5)	1,265 (74.8)
BL	12 (1.8)	3 (1)	5 (2.1)	3 (1.7)	1 (0.7)	5 (4)	0 (0)	29 (1.7)
DI	14 (2.1)	2 (0.7)	0 (0)	1 (0.6)	11 (8.1)	4 (3.2)	1 (2.3)	33 (2)
HF	0 (0)	2 (0.7)	0 (0)	1 (0.6)	0 (0)	1 (0.8)	0 (0)	4 (0.2)
MI	23 (3.4)	26 (8.6)	4 (1.7)	21 (11.7)	5 (3.7)	9 (7.2)	0 (0)	88 (5.2)
PE	14 (2.1)	16 (5.3)	12 (5.1)	7 (3.9)	5 (3.7)	5 (4)	1 (2.3)	60 (3.5)
PN	22 (3.3)	8 (2.6)	7 (3)	3 (1.7)	2 (1.5)	6 (4.8)	3 (6.8)	51 (3)
ST	2 (0.3)	1 (0.3)	1 (0.4)	1 (0.6)	1 (0.7)	0 (0)	0 (0)	6 (0.4)
UT	2 (0.3)	0 (0)	1 (0.4)	0(0)	1 (0.7)	1(0.8)	0 (0)	5 (0.3)
Total	667 (100)	304 (100)	235 (100)	180 (100)	135 (100)	125 (100)	44 (100)	1,691 (100)

When the mean ICU stay was compared across the three age groups of <50, 50–64, and 65 or more, no significant difference (Table 5) using the ANOVA F-test (p-value = 0.903). On the other hand, testing for mean total comorbidities across these three age groups was significant (p-value <0.0001), and the Tukey B multiple comparison test reveals significance with three means for groups in homogenous subsets of mean total comorbidities of 1.25, 2.19, and 3.28, respectively.

Clinical characteristics and common causes of death among COVID-19 patients

We identified primary and secondary causes of death. Septic shock was the primary COD in 667 patients (39.4%), followed by cardiopulmonary arrest 304 (18.0%), respiratory failure 235 (13.9%), and cardiac arrest 180 (10.6%). The percentages of cases with (septic shock & MOF), MOF, and renal failure were 135 (8.0%), 125 (7.4%), and 44 (2.6%), respectively. Hepatic failure occurred in only one case and thus ignored from further analysis. On the other hand, ARDS was one of the main reasons for ICU admissions and was reported in all deaths. Numerous cases were reported in which a combination of intermediate death complications occurred. These cases were thoroughly examined by our physicians to determine which terminal complication is more likely to be classified as the intermediate COD. It was found that around 75% of these decedents had ARDS as an intermediate COD, while the remaining 25% had intermediate COD other than ARDS. Among the other causes are AKI, AR, BL, DI, HF, MI, PE, PN, ST, and UT. The frequency distribution of intermediate combined complications along with the frequency distribution of the terminal complication leading to intermediate COD are shown in Figure 4. Table 6 shows the count and percentage of counts for primary and intermediate causes, as well as the column percentages for primary causes. While ARDS is the most prevalent intermediate COD regardless of primary cause, AR and MI disorders were significantly (7.2–14.4%) linked with cardiac arrest and MOF.

Age distribution appears to be similar by primary COD, with a median age at death of 64.5 years and an interquartile range (IQR = 20.7). However, a few young patients, approximately the age of 20, died because of MOF or renal failure (Figure 5). The median length of stay in the ICU prior to death was approximately 16 days overall but was significantly longer (∼ 20 days) for those who died of septic shock or (septic shock + MOF). Patients who died because of MOF had an average of three or more comorbidities. Those who died of renal failure and (septic shock + MOF) died in a manner like that described above.

Those who died because of AR, DI, or MI had the highest average age (70 years) and total comorbidities (3 or more), as well as the shortest average stay in the ICU. Patients who died of HF were younger (average age 50 years) and had more than two comorbidities, with an ICU stay of < 20 days. We also noticed the sequences of (MI → septic shock) and (PE → respiratory failure) were associate with 4 or above comorbidities on the average.

Exploring the relationship between comorbidities and causes of death

The following is a list of the comorbidities of dead patients in this study. Hypertension (57%) is the most common condition, followed by diabetes (52%), coronary artery disease (23%), and chronic renal disease (14%). 12% for each arrhythmia and dyslipidemia, cancer (11%), Rheumatoid arthritis (10%),
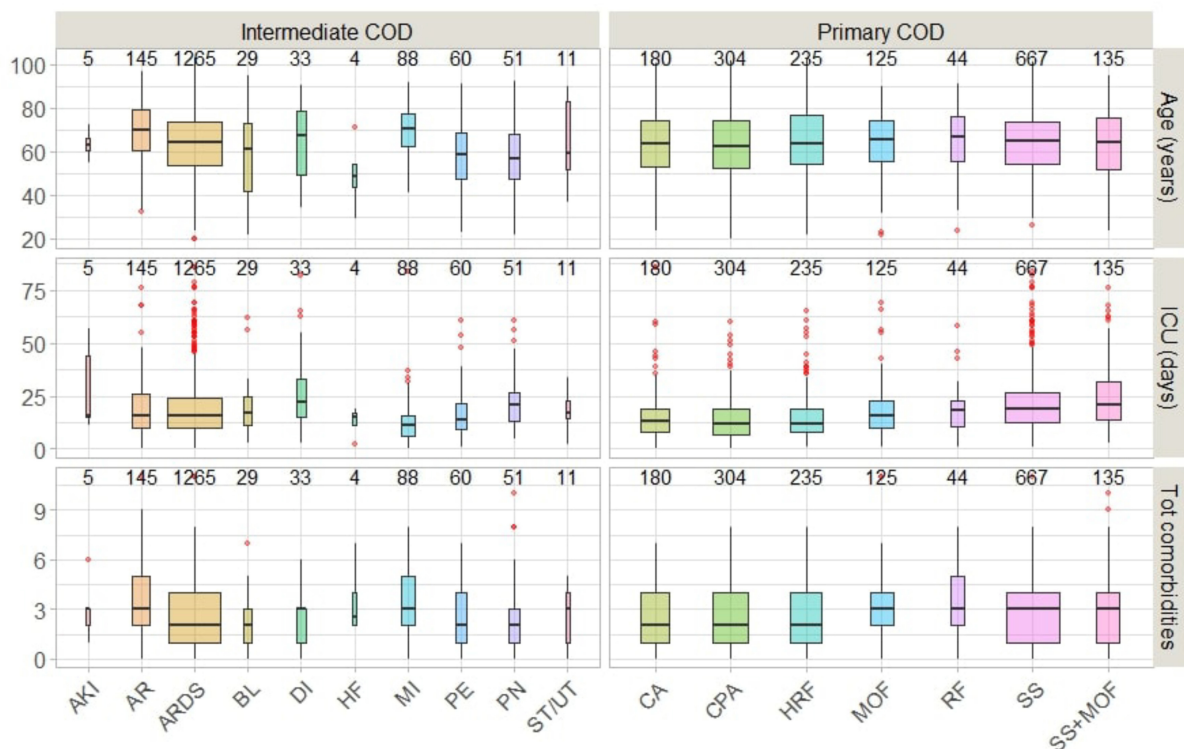
FIGURE 5
Primary and Intermediate CODs distribution by age and clinical characteristics.

TABLE 7 Best reach and frequency by group size.

	Size	Reach	Cases %	Count	Responses %
ADDED: HTN	1	962	56.9	962	27.7
ADDED: DM	2	1,139	67.4	1,834	52.9
KEPT: HTN					
ADDED: Cancer	3	1,195	70.7	2,026	58.4
KEPT: DM, HTN					
ADDED: RA	4	1,234	73.0	2,196	63.3
KEPT: Cancer, DM, HTN					
ADDED: AF	5	1,266	74.9	2,402	69.3
KEPT: Cancer, DM, HTN, RA					
ADDED: Obesity	6	1,296	76.6	2,550	73.5
KEPT: AF, Cancer, DM, HTN, RA					

obesity (9%), thyroid disease (8%), stroke (7%), pulmonary embolism (5%), asthma (4%), valvular heart disease (4%), bleeding disorders (4%), and 3% for each COPD and dementia. The remaining comorbidities with < 3% reported incidence include Anemia, heart failure, prostate hyperplasia, liver disease, epilepsy, cardiomyopathy, peripheral vascular disease, lung disease, psychiatric disorders, osteoporosis, multiple sclerosis, down syndrome, Parkinson's disease, inflammatory bowel disease, gastroesophageal reflux disease, Addison disease, and HIV infection.

Next, we present the results of the Total Unduplicated Reach and Frequency (TURF) method. TURF is a popular statistical technique in market research that ranks product combinations according to the number of customers who favor them (12). In this study, we applied the method in a clinical setting, treating comorbidities and patients as products and people.

The goal is to determine the most likely disease combinations that these patients share. The analysis traverses all possible combinations of comorbidities and records two statistics for each: reach and frequency. The reach is the percentage of individuals who exhibit at least one comorbidity in a given combination, and the frequency is the total number of times comorbidities are exhibited in a given combination. We tested the method for all comorbidities listed in Table 3 and a range of reach values. Table 7 provides a summary of the ideal choices according to the number of diseases (Size). For instance, the optimal combination of four comorbidities has a 73 percent

success rate with RA, cancer, DM, and HTN. This indicates that seventy-three percent of the patients had at least one of the conditions (rheumatological disorders, cancer, DM, HTN). If Diabetes and High Blood Pressure were eliminated from the analysis due to their high prevalence and we wanted to evaluate other possible combinations of diseases, the one with the highest prevalence was (obesity, CAD, Cancer, RA) with 43.6%.

When we looked at the general disease classification frequencies, we found that over 60% of the patients had circulatory (DCS) and endocrine (ENMS) disorders, one-third had cardiovascular diseases (CVD), and the remaining categories (RDs, CRC, DNS, ADs, DGS) varied from 8 to 15%. In compared to patients who died of ARDS/PE/Other, approximately 65 percent of patients who died of MI or AR had cardiovascular illnesses (Figure 6). Those who die from ARDS, on the other hand, usually have endocrine or circulatory system problems. Nervous system diseases were the least common among the PE dead. With chi-square test findings of (175.5, p-value 0.001) and (12.2, p-value $= 0.016$), the circulatory and nervous systems had the most significant association with intermediate COD.

FIGURE 6
Proportions of general disease categories by Intermediate COD.
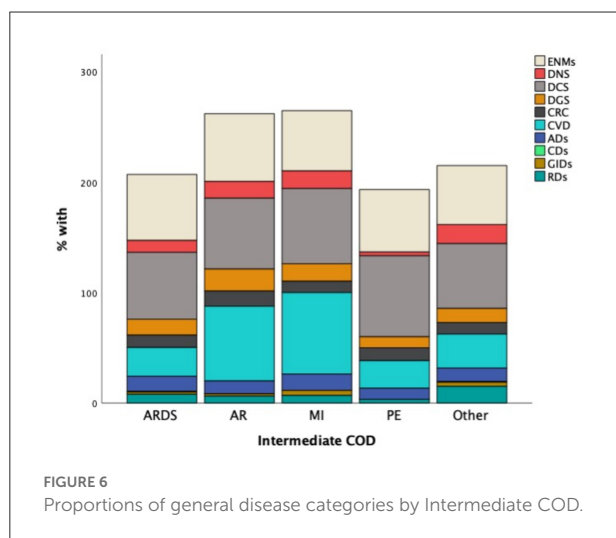
Predicting death due to ARDS or other causes

The total comorbidities distribution by age group of COVID-19 deaths due to ARDS or other cause is displayed in

FIGURE 7
Total comorbidities by age group due to ARDS or other causes of death.

FIGURE 8
Decision Tree prediction model of ARDS/Other causes of death.
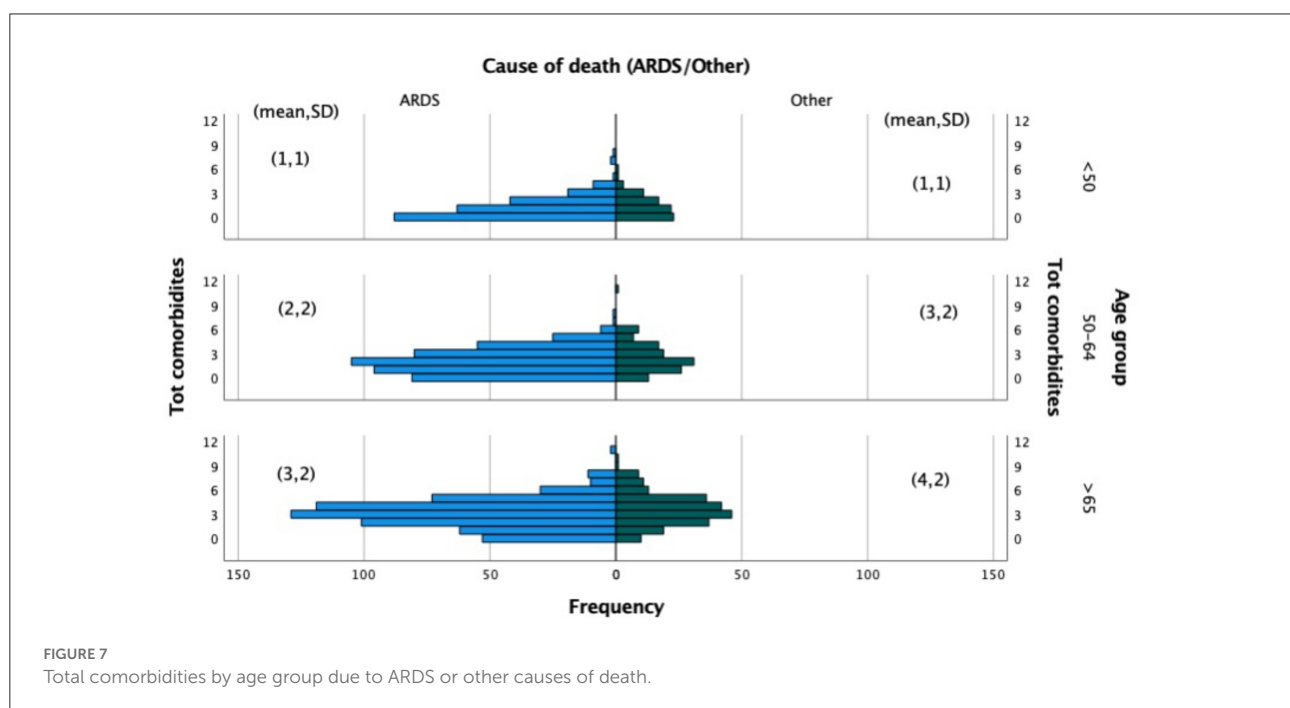
Figure 7. Patients under the age of 50 have a similar comorbidity distribution, with an average of one disease. Two comorbidities were found on average per age group (50–64) with more variation among those who died from causes other than ARDS. In contrast, older patients (age >65) who died from causes other than ARDS have an average of four comorbidities, compared to three for the other group who died mainly from ARDS.

In this section, we used decision tree (DT) to determine the most parsimonious predictors of intermediate COD among COVID-19 patients in intensive care units. Decision trees learn to divide data into smaller and smaller categories to forecast the goal. The test is represented by a node, while the numerous outcomes are represented by edges. The dividing process is repeated until no further gains can be obtained or a preset rule is reached. Three common decision tree techn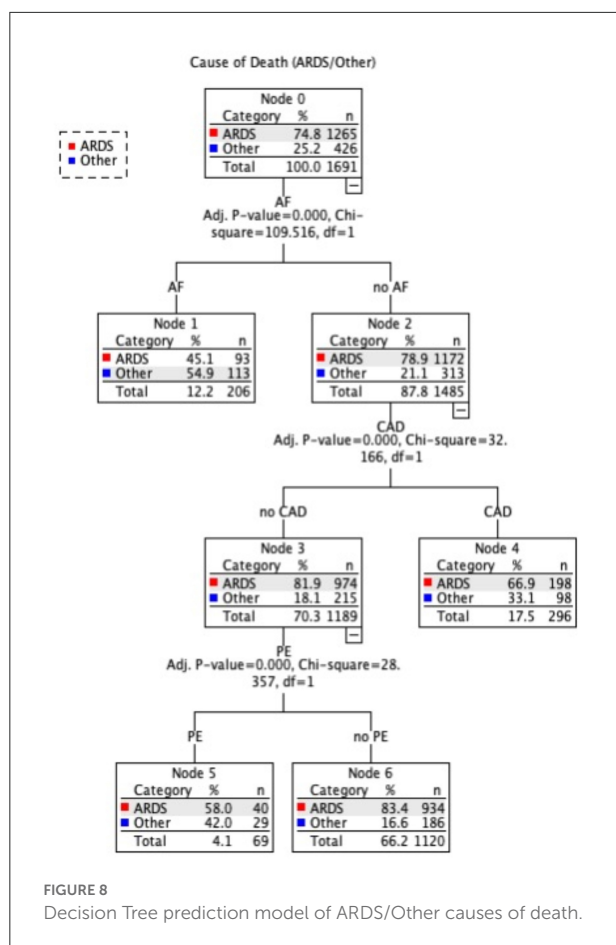iques include classification and regression tree (CART), chi-squared automatic interaction detection (CHAID), and quick unbiased efficient statistical tree (QUIEST). For mathematical explanations and performance comparisons of these DT approaches, see Lin et al. (13). Figure 8 illustrates the results of the QUEST model, which demonstrate that the existence of an arrhythmia (AF) was the best indicator of the intermediate cause (ARDS/Other). Patients

with AF are more likely to have a cause other than ARDS (54.9%). Node 1 is considered a terminal node for predicting a cause of death other than ARDS since no child nodes was found below it. In patients without AF, on the other hand, CAD was the second-best predictor of (ARDS/Other). In patients without AF but with CAD, the terminal Node 3 predicted 66.9 ARDS vs. 33.1% for other causes. PE is an additional predictor in the model for patients who do not have AF or CAD. ARDS is the main intermediate COD in this group, accounting for over 83% of patients without PE and 58% of patients with PE who died from ARDS. The risk and classification tables allow for a quick evaluation of the model's performance. The risk of misclassifying the cause of death is estimated to be 0.272 (or 27.2%), which is consistent with the results of the classification table, which show that 76% of causes of death are correctly classified.

Discussion

We used Machine learning NLP to extract clinical data and causes of death from EHRs for COVID-19 patients at Jaber Hospital in Kuwait. Consistency and completeness issues with the text data in these records made extraction difficult. During the pandemic, Jaber hospital was restricted to COVID-19 admissions, with most critical cases transferred from other hospitals. Many patient records were incomplete due to patients being transferred from district hospitals where their original medical records were kept. Machine learning and big data analytics have been used to investigate disease-related prognostic factors (14).

Several clinical characteristics have been linked to COVID-19 mortality. Age, gender, comorbidities, ICU stay, and disease severity are all factors. Increased proportions of 65-year-olds or older led to a significant age-mortality association (15, 16). Males were more likely to die from COVID-19 (17, 18). More than double the number of death patients had two or more comorbidities, according to Ayed et al. (17). Combining old age and comorbidities was also a factor in death (19) and survival time (20). On the other hand, Zhou et al. (21) reported a median (IQR) time of 18.5 (15–22) days from onset of symptoms to death. In our study, 815 (48%) of 1691 deceased ICU COVID-19 patients were over 65, men were more prevalent (56.9 vs. 43.1%), patients with two or more comorbidities accounted for 52% of cases, and the mean (SD) survival time to death was 18.5 (12.8) days. Hypertension and diabetes accounted for more than half of all cases in this study. This confirms prior research (17, 22–24). In COVID-19 patients, cardiovascular disease and secondary infections increase disease severity and mortality (15, 25, 26). Circulatory and cardiovascular diseases account for 61.6 and 32.5% of these patients, respectively; HIV-infections are rare. COVID-19 patients had a higher incidence of kidney and heart disease, and myocardium damage reduced survival (16, 27, 28).

Previous research on comorbidities and death causes has linked dysfunction to mortality (17, 29). In this study, decedents with MOF and renal failure averaged three or more comorbidities. Septic shock was the leading primary cause, accounting for 667 deaths (39.4%), followed by cardiopulmonary arrest (304 deaths, 18%), respiratory failure (235 deaths, 13.9%), and cardiac arrest (180 deaths, 10.6%). The most common intermediate COD, on the other hand, was ARDS (1265, 74.8%). We also found 849 (50.2%) cases of sepsis. Other findings (21) revealed that sepsis was the leading cause of death (59%) among the 54 pandemic deaths, followed by respiratory failure (54%), ARDS (31%), heart failure (23%), and septic shock (20%).

Acute respiratory distress syndrome (ARDS) is a severe COVID-19 consequence. Patients with moderate-to-severe ARDS require invasive mechanical ventilation and intensive medical therapy (30, 31). ARDS was one of the most common reasons for ICU hospitalizations, as it was recorded in 81.8% of ICU survivors and all fatalities (32). This is also demonstrated in our data, as all patients were admitted to the intensive care unit, and ARDS was a common morbid consequence. However, complications other than ARDS were deemed the predominant intermediate COD in 25% of the cases (Figure 4). As a result, we employed decision trees to forecast the most significant contributing factors to intermediate COD, namely ARDS or Other cause. "Other" denotes a complication associated with AKI, AR, BL, DI, HF, MI, PE, PN, ST, or UT. We encountered only three significant predictors, namely arrhythmia (AF), coronary artery disease (CAD), and pulmonary embolism (PE). Patients with AF were more likely to have an etiology other than ARDS. According to Elezkurtaj et al. (33), the majority of decedents died from COVID-19, with preexisting health conditions and comorbidities only contributing to the mechanism of death. We agree because, among the many variables examined in this study, only a few contributing factors were found to be significant with intermediate COD.

Strengths, limitations, and future work

The dynamic nature of the method, its usability, and its potential to maintain self-control all contribute to its strength. In addition, the sampled data span both significant pandemic waves and death peaks, accounting for 70% of the total reported COVID-19 fatality cases in Kuwait. The death rate drastically decreased after then. Therefore, our sample represents the population under consideration to a high degree of accuracy. Nevertheless, our study has several limitations. First, there is a chance of selection or referral bias as the research was conducted at a single location, i.e., Jaber Hospital. Second, the lack of information extracted from the inadequate documentation of the patient records. The absence of a symptom (such as obesity, smoking, etc.) does not necessarily suggest that a patient is

symptom-free. Thirdly, patients were typically transferred late in the course of their disease, and their medical records lacked vital medical history information. Such discrepancies in clinical data may result in information bias that contributes to a decrease in model precision.

Future studies could potentially investigate the impact of vaccines on the time to death, provide survival time estimates by cause of death, and perform spatiotemporal analyses of transferable patients. Knowing the COVID-19 death rate and patient survival rate can help risk management experts. COVID-19 or its evolving variants can be avoided, and strategies can be used to slow their spread.

Conclusion

We employ self-developed natural language processing (NLP) to automate the extraction of causes of death and comorbidities from the EHRs of COVID-19 decedents from the beginning of the pandemic through all major pandemic waves in this study. We structured the acquired text data and used it to conduct additional research.

We analyzed the demographic, clinical, and causes of death data for 1,691 ICU patients and discovered that the most common primary causes of death, which were documented in 54.8% of cases, were infection-related and included septic shock or sepsis-related multi-organ failure. The second most common cause of death was respiratory failure or cardiopulmonary arrest, which were documented in 32.2% of cases. Furthermore, cardiac arrest and renal failure account for 10.6 and 2.6% of all deaths, respectively. ARDS, on the other hand, was the most common cause of mortality in the intermediate stage. Arrhythmia (AF) was revealed to be the strongest predictor of intermediate cause (ARDS/Other) using machine learning decision tree analysis.

We recommend structuring the EHR with well-defined sections and providing menu-driven options for reporting causes of death and comorbidities to minimize misspellings or incorrect forms. Comprehensive assessment and user guidance are required for standards to be effectively integrated into EHR systems.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethical Review Committee (ERC) at Kuwait Ministry of Health (No. 1529/2020). Written informed consent for participation was not required for this study in

accordance with the national legislation and the institutional requirements.

Author contributions

SB conceived and gained ethical approval for the project. SB, AB, and YA participated in the retrieval, processing, and purification of data. AB and YA developed the clinical concepts, played a key role in establishing the data extraction clinical criteria, and validations. SB and AA created both the method and the programming. SB and SA carried out statistical analysis and produce visuals. SB, AA, and SA contributed to the paper's drafting. All authors have reviewed, offered comments, and approved the submission of the work.

Acknowledgments

We would like to express our gratitude to the administration of Jaber Al-Ahmad Hospital for their cooperation and support.

We would like to thank Eng. Naser Alibrahim for his support with Python coding.

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. Kong HJ. Managing unstructured big data in healthcare system. *Healthc Inform Res.* (2019) 25:1. doi: 10.4258/hir.2019.25.1.1

2. Wang SV, Rogers JR, Jin Y, Bates DW, Fischer MA. Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. *J Am Med Inform Assoc JAMIA.* (2017) 24:339–44. doi: 10.1093/jamia/ocw082

3. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform.* (2019) 7:e12239. doi: 10.2196/12239

4. Zhou L, Lu Y, Vitale CJ, Mar PL, Chang F, Dhopeshwarkar N, et al. Representation of information about family relatives as structured data in electronic health records. *Appl Clin Inform.* (2014) 5:349–67. doi: 10.4338/ACI-2013-10-RA-0080

5. John Lin CC Yu K, Hatcher A, Huang TW, Lee HK, Carlson J, et al. Identification of diverse astrocyte populations and their malignant analogs. *Nat Neurosci.* (2017) 20:396–405. doi: 10.1038/nn.4493

6. DeCapprio D, Gartner J, McCall CJ, Burgess T, Kothari S, Sayed S. Building a COVID-19 Vulnerability Index. *MedRxiv.* (2020). doi: 10.1101/2020.03.16.20036723

7. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *NPJ Digit Med.* (2021) 4:3. doi: 10.1038/s41746-020-00372-6

8. Izquierdo JL, Ancochea J, Soriano JB. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: retrospective study using machine learning and natural language processing. *J Med Internet Res.* (2020) 22:e21801. doi: 10.2196/21801

9. Guan X, Zhang B, Fu M, Li M, Yuan X, Zhu Y, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study. *Ann Med.* (2021) 53:257–66. doi: 10.1080/07853890.2020.1868564

10. Coronavirus. *Worldometer.* Available online at: https://www.worldometers.info/coronavirus/country/kuwait/ (accessed April 12, 2022).

11. (ICD-10). *International Classification of Diseases, Tenth Revision (ICD-10).* (2021). Available online at: https://www.cdc.gov/nchs/icd/icd10.htm (accessed April 11, 2022).

12. Data Scientist. *Reflections of a Data Scientist.* (2018). Available online at: https://www.reflectionsofadatascientist.com/2018/05/r-turf-analysis-spss.html (accessed November 11, 2022).

13. Lin CL, Fan CL. Evaluation of CART, CHAID, and QUEST algorithms: a case study of construction defects in Taiwan. *J Asian Archit Build Eng.* 18:539–53. doi: 10.1080/13467581.2019.1696203

14. Darabi H, Tsinis D, Zecchini K, Whitcomb W, Liss A. "Forecasting mortality risk for patients admitted to intensive care units using machine learning," In: *Procedia Computer Science, vol. 140.* Chicago, IL: Elsevier (2018). p. 306–313. doi: 10.1016/J.PROCS.2018.10.313

15. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* (2020) 46:846–8. doi: 10.1007/s00134-020-05991-x

16. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *J Med Virol.* (2020) 92:441–7. doi: 10.1002/jmv.25689

17. Ayed M, Borahmah AA, Yazdani A, Sultan A, Mossad A, Rawdhan H. Assessment of clinical characteristics and mortality-associated factors in COVID-19 critical cases in Kuwait. *Med Princ Pract.* (2021) 30:185–92. doi: 10.1159/000513047

18. Galbadage T, Peterson BM, Awada J, Buck AS, Ramirez DA, Wilson J, et al. Systematic review and meta-analysis of sex-specific COVID-19 clinical outcomes. *Front Med.* (2020) 7:348. doi: 10.3389/fmed.2020.00348

19. Moon SS, Lee K, Park J, Yun S, Lee YS, Lee DS. Clinical characteristics and mortality predictors of COVID-19 patients hospitalized at nationally-designated treatment hospitals. *J Korean Med Sci.* (2020) 35:e328. doi: 10.3346/jkms.2020.35.e328

20. Sousa GJB, Garces TS, Cestari VRF, Florêncio RS, Moreira TMM, Pereira MLD. Mortality and survival of COVID-19. *Epidemiol Infect.* (2020) 148:e123. doi: 10.1017/S0950268820001405

21. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet Lond Engl.* (2020) 395:1054–62. doi: 10.1016/S0140-6736(20)30566-3

22. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA*. (2020) 323:1574–81. doi: 10.1001/jama.2020.5394

23. Nada KM, Hsu E shuo, Seashore J, Zaidan M, Nishi SP, Duarte A, et al. Determining cause of death during Coronavirus Disease 2019 pandemic. *Crit Care Explor*. (2021) 3:e0419. doi: 10.1097/CCE.0000000000000419

24. Yan Y, Yang Y, Wang F, Ren H, Zhang S, Shi X, et al. Clinical characteristics and outcomes of patients with severe COVID-19 with diabetes. *BMJ Open Diabetes Res Care*. (2020) 8:e001343. doi: 10.1136/bmjdrc-2020-001343

25. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol Off J Ger Card Soc*. (2020) 109:531–8. doi: 10.1007/s00392-020-01626-9

26. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. (2020) 94:91–5. doi: 10.1016/j.ijid.2020.03.017

27. Arentz M, Yim E, Klaff L, Lokhandwala S, Riedo FX, Chong M, et al. Characteristics and outcomes of 21 critically ill patients with COVID-19 in Washington State. *JAMA*. (2020) 323:1612–4. doi: 10.1001/jama.2020.4326

28. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al. Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. *Travel Med Infect Dis*. (2020) 34:101623. doi: 10.1016/j.tmaid.2020.101623

29. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*. (2001) 286:1754–8. doi: 10.1001/jama.286.14.1754

30. Gibson PG, Qin L, Puah SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS. *Med J Aust*. (2020) 213:54–6.e1. doi: 10.5694/mja2.50674

31. Tzotzos SJ, Fischer B, Fischer H, Zeitlinger M. Incidence of ARDS and outcomes in hospitalized patients with COVID-19: a global literature survey. *Crit Care Lond Engl*. (2020) 24:516. doi: 10.1186/s13054-020-03240-7

32. Alshukry A, Ali H, Ali Y, Al-Taweel T, Abu-Farha M, AbuBaker J, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) patients in Kuwait. *PLoS ONE*. (2020) 15:e0242768. doi: 10.1371/journal.pone.0242768

33. Elezkurtaj S, Greuel S, Ihlow J, Michaelis EG, Bischoff P, Kunze CA, et al. Causes of death and comorbidities in hospitalized patients with COVID-19. *Sci Rep*. (2021) 11:4263. doi: 10.1038/s41598-021-82862-5

Effect of time-varying adherence to non-pharmaceutical interventions on the occurrence of multiple epidemic waves: A modeling study

Joel-Pascal Ntwali N'konzi[1,2]*, Chidozie Williams Chukwu[3] and Farai Nyabadza[3]

[1]African Institute for Mathematical Sciences, Kigali, Rwanda, [2]Maxwell Institute for Mathematical Sciences, University of Edinburgh, Heriot-Watt University, Edinburgh, United Kingdom, [3]Department of Mathematics and Applied Mathematics, University of Johannesburg, Johannesburg, South Africa

Non-pharmaceutical interventions (NPIs) play a central role in infectious disease outbreak response and control. Their usefulness cannot be overstated, especially during the early phases of a new epidemic when vaccines and effective treatments are not available yet. These interventions can be very effective in curtailing the spread of infectious diseases when adequately implemented and sufficiently adopted by the public. However, NPIs can be very disruptive, and the socioeconomic and cultural hardships that come with their implementation interfere with both the ability and willingness of affected populations to adopt such interventions. This can lead to reduced and unsteady adherence to NPIs, making disease control more challenging to achieve. Deciphering this complex interaction between disease dynamics, NPI stringency, and NPI adoption would play a critical role in informing disease control strategies. In this work, we formulate a general-purpose model that integrates government-imposed control measures and public adherence into a deterministic compartmental epidemic model and study its properties. By combining imitation dynamics and the health belief model to encode the unsteady nature of NPI adherence, we investigate how temporal variations in NPI adherence levels affect the dynamics and control of infectious diseases. Among the results, we note the occurrence of multiple epidemic waves as a result of temporal variations in NPI adherence and a trade-off between the stringency of control measures and adherence. Additionally, our results suggest that interventions that aim at increasing public adherence to NPIs are more beneficial than implementing more stringent measures. Our findings highlight the necessity of taking the socioeconomic and cultural realities of affected populations into account when devising public health interventions.

1. Introduction

An important aspect of the control of infectious diseases is the use of non-pharmaceutical interventions (NPIs) such as social distancing, mask-wearing, school closures, mass gathering bans, and lockdowns that aim to change key aspects of human behavior to reduce transmission rates. Their usefulness cannot be overstated, especially during the early phases of a new epidemic when vaccines are not available yet. The still ongoing COVID-19 pandemic has underscored the value of these interventions even beyond the early phases of an epidemic. Various studies have shown, for instance, the necessity of keeping NPIs in place during the vaccine rollout in 2021, a year after the start of the pandemic (1, 2). The literature on the usefulness of NPIs and their effects on the transmission dynamics of COVID-19 is vast (3–9).

However, as the COVID-19 pandemic has shown, disease control measures can come with unprecedented damages to economies and overall well-being that usually interfere with the willingness and capabilities of the public to adopt recommended health behaviors. This can lead to reduced and unsteady adherence to NPIs, making disease control more challenging to achieve. Additionally, the COVID-19 pandemic has highlighted the fact that faced with the same public health threat, countries will implement interventions of varied stringency and the level of public adherence to these interventions will vary widely between regions and over time (10–15). This has consequences on the effectiveness of NPIs over time. In particular, the study by Ge et al. (16) has highlighted the fluctuating nature of NPI performance over time at different geographical scales. These observations are not particular to the COVID-19 pandemic. The works by Matthews Pillemer et al. (17) and Wang et al. (18) showed, for instance, that similar patterns had been observed during the 2002-03 SARS outbreak in four different regions around the world and the 2013 H7N9 outbreak in China, respectively.

Many studies have investigated the predictors of adherence to NPIs during public health emergencies and causes of variations in adherence levels over time, especially during the COVID-19 pandemic (13, 17, 19–22). The most commonly identified factors that affect compliance with NPIs are risk perception, trust or lack of trust in public health authorities, misinformation, economic hardship, and the sociocultural realities of affected populations. NPI adherence patterns can significantly impact the effectiveness of disease control strategies and interact largely with disease dynamics over time. Factors like risk perception and perceived economic hardship due to NPI adherence are subject to temporal change as the disease unfolds resulting in a complex interaction between disease dynamics, the stringency of NPIs, and NPI adherence.

Untangling the feedback loop between human behavior and the spread of infectious diseases has caught the attention of many over the last two decades due to its importance for epidemic preparedness and control and this has led to what is now called *behavioral epidemiology of infectious diseases* (23–26). Though the pre-COVID-19 literature is vast, we have not come across many studies related to modeling the interplay between NPI stringency, NPI adherence, and the dynamics of infectious diseases. Much progress has been made since the beginning of the COVID-19 pandemic, however. The studies by Acuña-Zegarra et al. (27) and Iyaniwura et al. (28) both investigated the impact of adherence to NPIs on the dynamics of COVID-19 using an SEIR-like model. In particular, they both divide the population into 2 groups: individuals who adhere to all the NPIs are put in one group, and those who do not adhere in the other group. To capture the time-varying nature of NPI adherence, the two models allow flows between the 2 groups, though the model by Acuña-Zegarra et al. (27) incorporates only the movement from the adherents to the non-adherents group. However, though their modeling framework allows highlighting the effect of NPI adherence on disease dynamics, it is agnostic on the drivers of changes in NPI adherence status. Additionally, since most individuals only partially adhere to NPIs, it does not seem sensible to divide the population into adherents and non-adherents. Finally, their models do not capture the interaction between NPI intensity and NPI adherence, nor do they allow to answer the question of how disease dynamics interact with either of the two.

Other modeling studies on NPIs focused on assessing the effectiveness of NPIs or their usefulness in specific contexts (29–31). Though these models have played a major role in shedding light on the need for sustained adherence to NPIs during the COVID-19 pandemic, the authors studied the effect of NPI adherence on disease dynamics by simulating the models under different *fixed* NPI adherence scenarios, and thus failing to account for the impacts of temporal variations in NPI adherence on disease dynamics.

In this study, we aim to fill in this gap. We modify the classical deterministic Susceptible-Exposed-Infected-Recovered (SEIR) model to capture both the effect of disease control measures and the feedback loop between NPI adherence and disease dynamics. We use the model to investigate the effect of temporal variations in NPI adherence levels on the spread of infectious diseases. We use imitation dynamics and build upon the health belief model to encode the dynamics of the public level of adherence to NPIs. We do not take into account the intricacies of specific diseases such as the spread of different variants or strains and the availability of vaccines for sake of generality. The approach we use for modeling temporal variations in NPI adherence is similar to the one adopted in Jentsch et al. (32).

The rest of this paper is organized as follows. In Section 2 we present the mathematical model used in this study. Section 3 contains our results. In particular, we derive an expression for the basic reproduction number ($\mathcal{R}_0$) and study

the stability properties of the equilibrium solutions of the model. Furthermore, we perform a scenario analysis to investigate the effect of both the stringency of NPIs and temporal variations in the level of adherence on the disease dynamics. Lastly, in Section 4 we provide a discussion of our results.

2. Methods

2.1. Baseline model

We use a deterministic SEIR model that divides the population into four distinct compartments based on the respective disease state of individuals. The *Susceptible* (S) compartment contains individuals who can but have not yet been infected by the disease. Individuals in the *Exposed* (E) have been infected by the disease but cannot spread it, whereas those in the *Infected* (I) compartment can transmit the disease. The *Recovered* (R) compartment contains individuals who have recovered from the disease. Our model takes into account births and both natural and disease-induced deaths. For simplicity, we do not take reinfections into account. That is, we assume that individuals who recover from the disease become immune over the course of the epidemic. This is the case for diseases like measles, for example. This leads to the following system of differential equations:

$$\frac{dS}{dt} = \pi - \big(\lambda(t) + \mu\big)S, \tag{1a}$$

$$\frac{dE}{dt} = \lambda(t)S - (\kappa + \mu)E, \tag{1b}$$

$$\frac{dI}{dt} = \kappa E - (\gamma + \delta + \mu)I, \tag{1c}$$

$$\frac{dR}{dt} = \gamma I - \mu R, \tag{1d}$$

with nonnegative initial conditions $S(0) = S_0, E(0) = E_0, I(0) = I_0, R(0) = R_0$, and $x(0) = x(0)$. The disease state variables S, E, I, and R represent fractions of the population in the corresponding compartment, and model parameters are as described in Table 1 below.

In the classical SEIR model, non-pharmaceutical interventions are not explicitly taken into consideration, thus the *force of infection*, $\lambda(t)$, is given by

$$\lambda(t) = \beta_0 I(t), \tag{2}$$

where β_0 is the *effective contact rate* or *transmission rate* in the absence of control measures and $I(t)$ is the prevalence rate of the disease in the population at time t, i.e., the ratio of the number of infected individuals to the total population count.

TABLE 1 Description of state variables and model parameters.

Variables	Description
S	Fraction of susceptible individuals
E	Fraction of exposed individuals
I	Fraction of infectious individuals (prevalence rate)
R	Fraction of recovered individuals
x	Level of adherence to recommended control measures

Parameters	Description
π	Birth/recruitment rate
β_0	Effective contact rate in the absence of disease control measures
μ	Natural death rate
$1/\kappa$	Mean latent period
γ	Recovery rate of infected individuals
δ	Disease-induced per capita death rate
α	Disease control strength
r_i	Overall perceived susceptibility
m	Overall perceived severity of the disease
r_c	Overall perceived frustration with control measures
k	Overall sensitivity to changes in disease prevalence
ω	$r_i m / r_c$
K	k/r_c

2.2. Incorporating control measures

To incorporate the effect of NPIs, we assume that the stringency of interventions is proportional to the prevalence rate of the disease and that they are geared toward reducing the transmission rate of the disease. Hence, we use a prevalence-dependent contact rate and modify the force of infection in Equation (2) to obtain:

$$\lambda(t) = \frac{\beta_0 I(t)}{1 + \alpha I(t)}, \tag{3}$$

where α is a positive, constant parameter that measures the stringency of government policy responses to curb the epidemic. A probably more insightful way of thinking about the parameter α is that the fraction $1/\alpha$ defines the prevalence rate that has to be reached within the population for the policymakers to implement NPIs that aim to reduce the transmission rate by half. Hence, high values of α mean that decision-makers have a stringent policy approach in responding to the epidemic whereas lower values correspond to instances of a more hands-off policy approach. It is worth noting that the use of prevalence-dependent contact rates to model behavioral responses to disease outbreaks as in Equation (3) can be traced back to the work of Capasso and Serio (33) and was used by Gros et al. (34) to model short-term control strategies for the SARS-CoV-2 virus.

An important limitation of the force of infection in Equation (3) is that it implicitly assumes that individuals' compliance with disease control measures is constant over time. However, as shown earlier, this is not realistic in many instances.

2.3. Modeling the dynamics of NPI adherence level

To model temporal variations in adherence levels to NPIs, we introduce an additional variable, $x(t)$, that measures the *level of adherence* to disease control measures at time t. We modify the force of infection in Equation (3) to obtain:

$$\lambda(t) = \frac{\beta_0 I}{1 + \alpha I x(t)}. \tag{4}$$

We assume that $x(t)$ is an average over the population. Building upon the health belief model (35) and using imitation dynamics, as in the work by Bauch (36) on vaccinating behavior, we model the dynamics of the adherence level by

$$\frac{dx}{dt} = kx(1-x)(-r_c + r_i mI), \tag{5}$$

where r_c represents the *perceived frustration* with NPIs due to the socioeconomic, cultural, and emotional damages and inconveniences associated with NPI adoption; r_i and m represent the *perceived susceptibility* to the disease and the *perceived severity* of the disease, respectively; and k measures the *public responsiveness* to changes in disease prevalence. Defining $K = k/r_c$ and $\omega = r_i m/r_c$ as in Bauch (36), we then rewrite Equation (5) as:

$$\frac{dx}{dt} = Kx(1-x)(-1 + \omega I), \tag{6}$$

where $I = \dfrac{1}{\omega}$ is a threshold prevalence rate below which the adherence level would be decreasing. Hence, the two threshold prevalence rates $I = 1/\alpha$ and $I = 1/\omega$ are loose measures of tolerance for the disease at the decision-makers level and public level, respectively.

3. Results

In this section, we derive a mathematical expression for the basic reproduction number, $\mathcal{R}_0$, and study the stability properties of both disease-free and endemic equilibrium solutions in Section 3.1. In particular, we show that the model has two disease-free equilibria, of which one is always unstable and the other is globally asymptotically stable when $\mathcal{R}_0 < 1$ and unstable otherwise. When $\mathcal{R}_0 > 1$, the model has three endemic equilibrium solutions that exchange stability as $\mathcal{R}_0$ is varied. These analytical results are followed by extensive numerical simulations in Section 3.2.

3.1. Equilibria and basic reproduction number

Our model has two disease-free equilibria given by

$$\mathcal{E}_1 = (S_1, E_1, I_1, R_1, x_1) = \left(\frac{\pi}{\mu}, 0, 0, 0, 0\right), \tag{7}$$

and

$$\mathcal{E}_2 = (S_2, E_2, I_2, R_2, x_2) = \left(\frac{\pi}{\mu}, 0, 0, 0, 1\right), \tag{8}$$

corresponding to the case where there is no disease and no adherence to NPIs for $\mathcal{E}_1$, and the situation where the disease is absent but the population fully complies with prevention measures $\mathcal{E}_2$.

The basic reproduction number, $\mathcal{R}_0$, for this model is given by

$$\mathcal{R}_0 = \frac{\beta_0 \kappa \pi}{\chi \mu}, \tag{9}$$

with $\chi = (\kappa + \mu)(\gamma + \delta + \mu)$.

The disease-free equilibrium point $\mathcal{E}_1$ is globally asymptotically stable when $\mathcal{R}_0 < 1$ and unstable when $\mathcal{R}_0 > 1$, whereas $\mathcal{E}_2$ is always unstable (proof in Section 1 of the Supplementary material). This encapsulates the fact that when there is no disease, individuals have no incentive to adhere to NPIs, especially because the cost of adopting preventive measures, no matter how small it might be, would still be higher than the perceived risk of acquiring the disease when the disease is absent.

Additionally, the model has rich dynamics with three endemic equilibria given by

(a) $\mathcal{E}_3 \equiv (S^*, E^*, I^*, R^*, x^*)$, with

$$
\begin{aligned}
S^* &= \frac{\chi}{\beta \kappa}, \\
E^* &= \frac{\mu(\gamma + \delta + \mu)}{\kappa \beta}(\mathcal{R}_0 - 1), \\
I^* &= \frac{\mu}{\beta}(\mathcal{R}_0 - 1), \\
R^* &= \frac{\gamma}{\beta}(\mathcal{R}_0 - 1), \\
x^* &= 0,
\end{aligned}
\tag{10}
$$

exists when $\mathcal{R}_0 > 1$.

(b) $\mathcal{E}_4 \equiv (S^*, E^*, I^*, R^*, x^*)$, with

$$S^* = \frac{1}{\mu}\left(\pi - \frac{\chi}{\kappa\omega}\right) = \frac{\pi}{\mathcal{R}_0}\left(\mathcal{R}_0 - \frac{\beta}{\mu\omega}\right),$$

$$E^* = \frac{\gamma + \delta + \mu}{\kappa\omega},$$

$$I^* = \frac{1}{\omega}, \tag{11}$$

$$R^* = \frac{\gamma}{\mu\omega},$$

$$x^* = \frac{1}{\alpha}\left[\mathcal{R}_0 - \left(1 + \frac{\beta}{\mu\omega}\right)\right],$$

exists when

$$\mathcal{R}_0 > 1 + \frac{\beta}{\mu\omega}, \tag{12}$$

and

(c) $\mathcal{E}_5 \equiv (S^*, E^*, I^*, R^*, x^*)$, with

$$S^* = \frac{\alpha\pi\kappa + \chi}{\kappa(\beta + \alpha\mu)},$$

$$E^* = \frac{\mu(\gamma + \delta + \mu)}{\kappa(\beta + \alpha\mu)}(\mathcal{R}_0 - 1),$$

$$I^* = \frac{\mu}{\beta + \alpha\mu}(\mathcal{R}_0 - 1), \tag{13}$$

$$R^* = \frac{\gamma}{\beta + \alpha\mu}(\mathcal{R}_0 - 1),$$

$$x^* = 1,$$

exists whenever $\mathcal{R}_0 > 1$.

Theorem 1 (Stability of $\mathcal{E}_3$). The endemic equilibrium point $\mathcal{E}_3$ is locally asymptotically stable when

$$1 < \mathcal{R}_0 < 1 + \frac{\beta}{\mu\omega}, \tag{14}$$

and unstable otherwise.

Theorem 2 (Stability of $\mathcal{E}_5$). The endemic equilibrium point $\mathcal{E}_5$ is locally asymptotically stable when

$$\mathcal{R}_0 > 1 + \frac{\beta}{\mu\omega} + \frac{\alpha}{\omega} \quad (15) \quad \text{and} \quad \mathcal{R}_0 > \frac{1}{\mu}\left(\frac{\chi}{\xi_0} - 1\right) (16)$$

where $\chi = (\gamma + \delta + \mu)(\kappa + \mu)$ and $\xi_0 = \gamma + \delta + \kappa + 2\mu$. Moreover, if we assume the latency period is at least 1 day so that $\kappa \leq 1$, then (Equation 15) subsumes (Equation 16). The endemic equilibrium $\mathcal{E}_5$ is unstable when $\mathcal{R}_0 < 1 + \frac{\beta}{\mu\omega} + \frac{\alpha}{\omega}$.

Theorem 1 shows that for diseases that eventually spread within a population, the higher ω is, the more likely the public will keep some level of adherence to NPIs when the disease becomes endemic. On the other hand, Theorem 2 shows that the

TABLE 2 Parameter values used in simulations.

Parameter	Figures 1–5	Figure 6
π	0.000002	0.0017
μ	0.000002	0.0017
β_0	1.1	0.6
κ	1/6	0.2
γ	0.125	0.2
δ	0.001	0.004
K	0.4	0.3

likelihood of having full adherence to NPIs during the endemic stage is both proportional to ω and inversely proportional to the stringency of control measures, α. This highlights a trade-off between NPI stringency and adherence. (Proofs can be found in Section 2 of the Supplementary material).

3.2. Scenario analysis

We simulate the modified SEIR model given by Equations (1), (3), and (6) in different scenarios to better understand the effect of temporal variations in NPI adherence levels on disease dynamics. We keep all the epidemiological parameters fixed and vary only the values of α, the stringency of control measures, and ω, with $1/\omega$ measuring the tolerance of the disease in the public. Parameter values used in simulations are given in Table 2.

3.2.1. Baseline scenario

First, we consider the hypothetical situation where the NPI adherence level is fixed over time, i.e., $x(t) = x$. When $x = 0$ or $\alpha = 0$, the model reduces to the classical SEIR model without the effect of disease control interventions. It is easy to see that when the adherence level is constant over time, one can write $x(t) = 1$ without any loss of generality. We investigate the dynamics of the disease when the adherence level is fixed at $x = 1$ for different values of α. Figure 1 shows that increasing the stringency of NPIs can result in a substantial decrease in the prevalence at the peak of the epidemic. It is important to note that the model predicts that the epidemic curve can be flattened for high values of α, and the ultimate conclusion would be that the higher the stringency of NPIs, the better. However, this might not hold when temporal variations in NPI adherence are considered, as we show below.

3.2.2. Time-varying adherence level
3.2.2.1. Fixed NPI stringency

First, we fix the stringency of NPIs to $\alpha = 100$ and investigate the effect of varying ω. Figure 2 shows that if $\omega = $

FIGURE 1
Effect of increasing NPI stringency when the adherence level to NPIs is constant over time. It is clear that under this assumption, substantial increases in NPI stringency result in a considerable decrease in the prevalence at the peak of the epidemic.

FIGURE 2
Effect of increasing ω at fixed $\alpha = 100$. This figure shows that when the level of public tolerance for the disease is high, the prevalence at the peak of the epidemic can be much higher than what might be expected under the fixed NPI adherence level assumption.

100, i.e., the risk perceived by the public exceeds the barriers to NPI adherence only when disease prevalence is at least 1%, then the peak prevalence of the epidemic would exceed 20%, whereas it was lower than 10% even for mildly stringent NPIs ($\alpha = 50$) when the NPI adherence level was assumed constant throughout the epidemic. However, if the public perceives a much higher risk from the disease ($\omega \geq 6,000$), the level of adherence to control measures would be much higher and the peak prevalence can be reduced to around 6%, which is much closer to what was observed in Figure 1. It is noteworthy that even under perfect adherence to NPIs, the prevalence at peak cannot be further reduced without increasing the stringency of control measures.

3.2.2.2. Fixed public tolerance for the disease

We observed that under the fixed adherence assumption, i.e., $x(t) = 1$ for all t, a substantial increase in NPI stringency would result in a substantial decrease in the peak prevalence for relatively low values of α. However, this might not be the case when the NPI adherence level, $x(t)$, is time-varying. To see what might happen in this case, we fix $\omega = 100$ and simulate the model for different values of α. Figure 3 shows that increasing the NPI stringency, α, does not result in a substantial decrease in the peak prevalence when the public tolerance for the disease is not sufficiently low. This is in contrast with what is observed in Figure 1 when the NPI adherence level is constant over time. Moreover, for considerably high values of α, the level of adherence to NPIs fluctuates over time due to trade-offs between NPI stringency and adherence, leading to the occurrence of multiple epidemic waves. This seemingly counter-intuitive observation might be justified by the quite high public tolerance of the disease when $\omega = 100$ as the NPI adherence level increases only when the disease prevalence is at least 1%. This might happen either because the perceived risk for the disease is low or because the perceived frustration with NPIs is much higher than the perceived risk. This shows that not much can be achieved by imposing very stringent control measures if the capacity and willingness of the public to adopt such measures are not increased.

3.2.2.3. Striking the right balance between NPI stringency and adherence

To understand what might happen when NPI stringency is high but not too high, we fix NPI stringency at $\alpha = 500$ and simulate the model for different values of ω. Figures 4A, B show that for small values of ω, multiple epidemic waves can occur due to the trade-offs between NPI stringency and adherence. However, as ω is increased, the effectiveness of NPIs increases and the epidemic curve can be flattened (see Figure 4C). It is important to note that the peak prevalence for the flattened epidemic curve ($\omega = 6,000$) is around 1.4%, which is much lower compared to the 31.1% when $\omega = 50$ and lower than the 1.7% prevalence peak for the second wave when $\omega = 100$. This suggests that disease control can be much more effective with reasonably stringent NPIs if the willingness and capacity of the public to adopt such measures is increased.

3.2.2.4. What matters most: Increasing NPI stringency or adherence?

We compare the relative reduction in the prevalence at the first peak when α is increased at fixed ω and when ω is increased at fixed α. Figure 5 shows that increasing ω, and thus the adherence level to NPIs, results in a much more substantial reduction in the peak prevalence than increasing NPI stringency, α, at fixed ω.
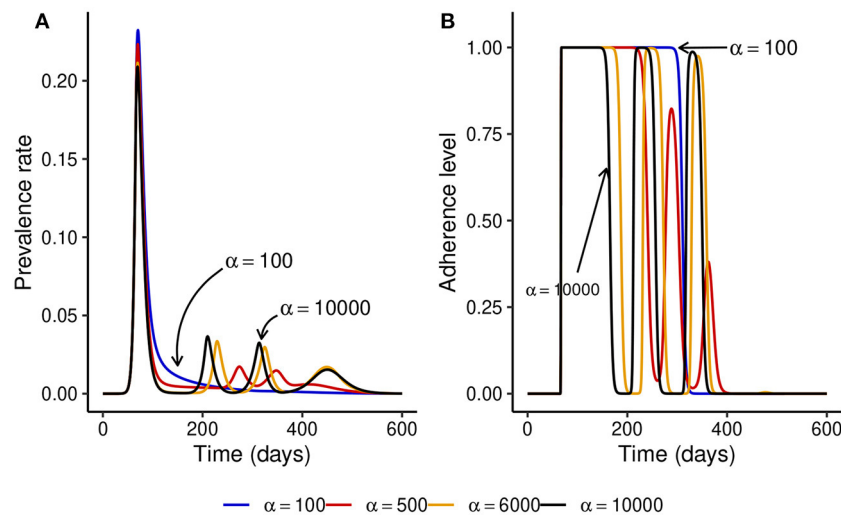
FIGURE 3
Effect of increasing the stringency of NPIs, α, when the level of adherence to NPIs is time-varying and $\omega = 100$ is fixed. This shows that if the public tolerance for the disease is not sufficiently low, increasing NPI stringency results only in a marginal decrease in the peak prevalence and can induce the occurrence of multiple epidemic waves. This is in contrast with what is observed in Figure 1 when the NPI adherence level is constant over time. **(A)** The prevalence rate of the disease. **(B)** Time-varying adherence level to NPIs.
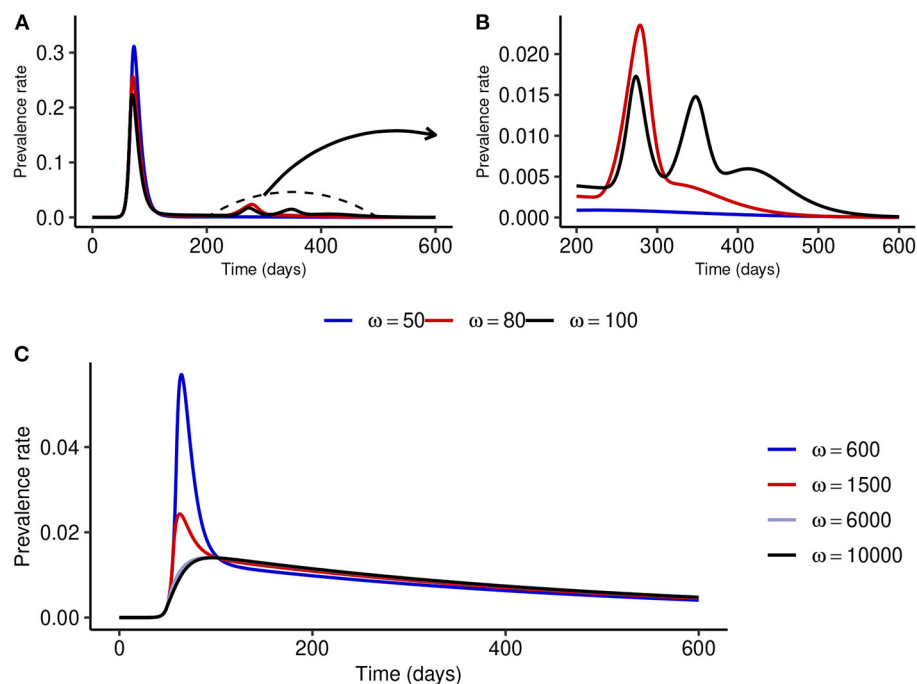
FIGURE 4
Effect of increasing ω at fixed medium-level NPI stringency ($\alpha = 500$). **(A, B)** Show that multiple epidemic waves can occur when ω is not high enough. Both the number of secondary waves and the prevalence at each secondary peak depend on the stringency of NPIs, α, and the public tolerance for the disease, $1/\omega$. **(C)** Shows that the epidemic curve can be flattened for considerably high values of ω.

3.2.2.5. Equilibrium solutions

Lastly, we simulate the modified SEIR model with a different set of parameter values to investigate the behavior of the system at equilibrium. In particular, we fix the NPI stringency at $\alpha = 2000$ and use ω as the bifurcation parameter. Figure 6 shows that when ω is low, the NPI adherence level decays to zero as

FIGURE 5

(A) Reduction in the peak prevalence when α is increased at fixed ω contrasted with the reduction when ω is increased at fixed ω. **(B)** Increasing ω, and thus the NPI adherence level, therefore, results in much higher marginal benefits than increasing the stringency of NPIs at fixed ω.
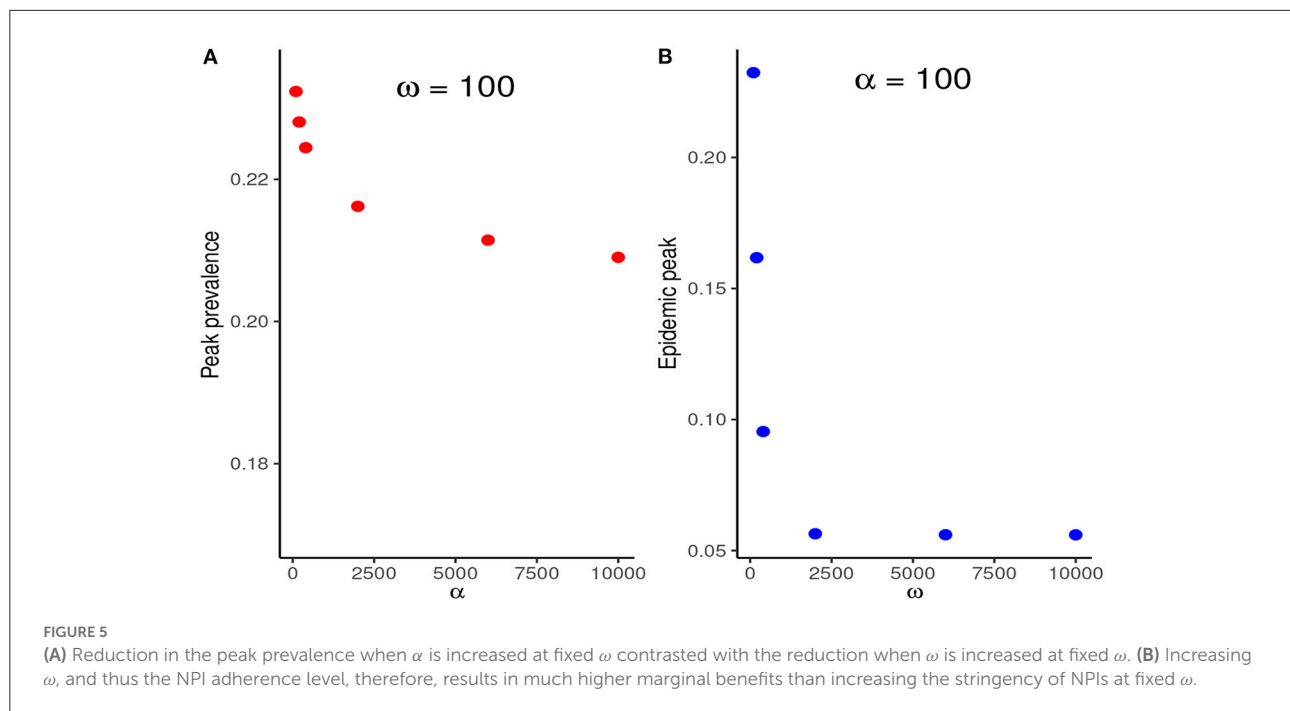
the disease becomes endemic. However, for larger values of ω, observe either partial or full adherence to NPIs at the endemic state depending on how large the value of ω is. Therefore, the system undergoes two transcritical bifurcations as the value of ω is varied. It is important to note that the exact values where these exchanges of stability occur depend on the values of other parameters. In the particular case of partial NPI adherence at equilibrium, increasing the value of ω results in both a decrease in the endemic prevalence rate and an increase in the NPI adherence level at equilibrium. Additionally, increasing the NPI stringency, α, at fixed ω can destabilize the system from full to partial adherence and even no adherence to NPIs at equilibrium.

4. Discussion

Non-pharmaceutical interventions play a central role in infectious disease outbreak response and control. These interventions can be very effective in curtailing the spread of infectious diseases when adequately implemented and sufficiently adopted by the population. However, NPIs can be very disruptive, and the socioeconomic and cultural challenges associated with their implementation interfere with both the ability and willingness of affected populations to adopt such interventions (25). This can lead to not only geographical but also temporal variations in adherence levels to NPIs during disease outbreaks and impede the control of infectious diseases (12). In this study, we investigated the interplay between NPI

stringency, temporal variations in NPI adherence levels, and disease dynamics using mathematical modeling. We showed that when the NPI adherence level is assumed constant over time, a substantial increase in the stringency of NPIs would result in a substantial decrease in the peak prevalence and the epidemic curve can be flattened by imposing stringent disease control measures. The ultimate conclusion, in this case, is that the higher the stringency of NPIs, the more effective disease control is. Though this agrees with the results by Gros et al. (34), we showed that this conclusion might not hold when temporal variations in NPI adherence levels are taken into account. For time-varying adherence levels to NPIs, the dynamics of the disease are no longer influenced only by the stringency of control measures, but also by risk perception and the perceived frustration associated with NPIs. Our results suggest that when the public tolerance for the disease is relatively high, which might occur when either the perceived risk is low or the perceived frustration with NPIs is high, the level of adherence to NPIs might fluctuate over time, consistent with the results by Crane et al. (14). We showed that these fluctuations in NPI adherence levels over time might induce the occurrence of multiple epidemic waves, which is consistent with the results by Ochab et al. (37) and Ngonghala et al. (38). Furthermore, our results suggest a trade-off between NPI stringency and adherence in the sense that the public would find it more difficult to sustain high levels of NPI adherence when the stringency level is high, which is again consistent with the results by Acuña-Zegarra et al. (27). This conclusion is supported
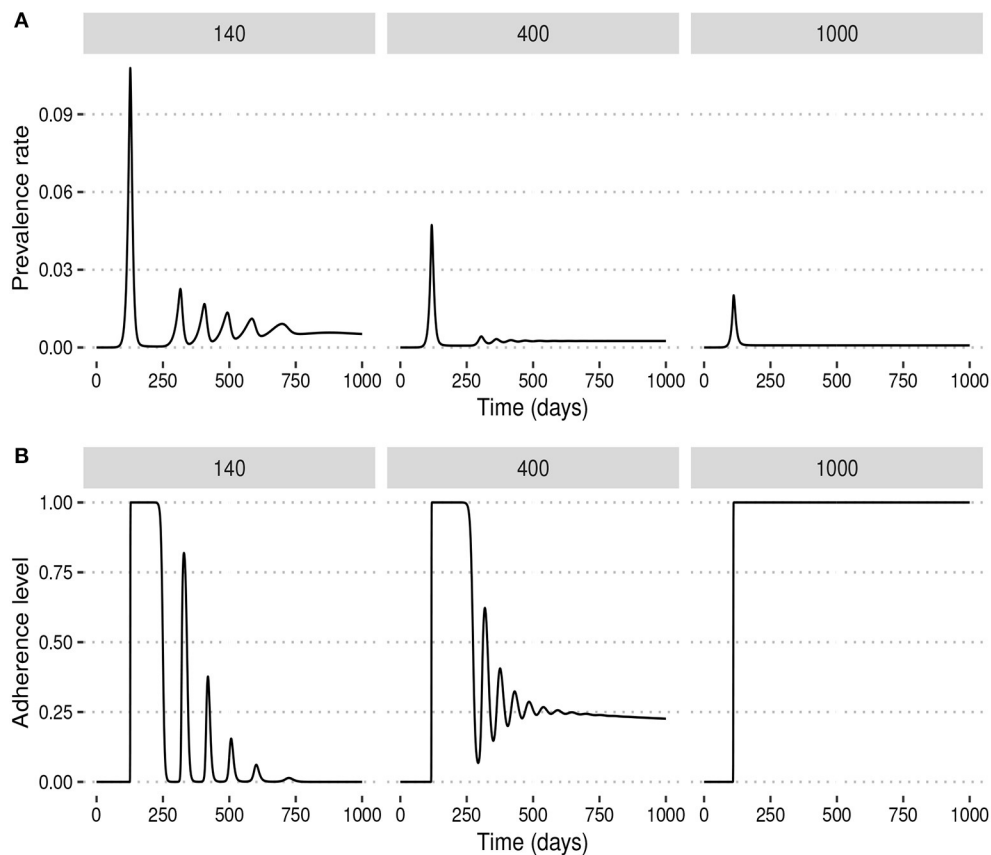
FIGURE 6
Behavior at equilibrium (endemic) for different values of the ω. For low values of ω, public tolerance for the disease is high and the system converges to a state where there is no adherence to NPIs during the endemic stage of the disease. For higher values of ω, however, NPI adherence during the endemic stage can be either partial or total depending on how large the value of ω is. **(A)** Prevalence rates. **(B)** Adherence levels.

both by our scenario analysis and analytical derivations of the stability properties of the endemic equilibrium points of our model.

Our model also suggests that interventions that aim at increasing the adherence level to NPIs might be much more valuable than increasing the overall NPI stringency. In particular, we showed that an epidemic can be effectively controlled even with not very stringent NPIs if the ability and willingness of the public to adopt such interventions is sufficiently increased. We, therefore, argue that for the effective control of infectious disease outbreaks, public health authorities should not only focus on implementing the right policies but also, more importantly, on devising strategies to both increase risk perception and decrease the frustration associated with NPIs. This agrees with the conclusions by Avusuglo et al. (31). Importantly, our results suggest that much consideration should be given to the socioeconomic and cultural realities of affected populations when devising public health policies for infectious disease control.

Our work has important limitations that present opportunities for future studies. First, our model has not been calibrated to epidemic data. Also, we have kept the model simple for generality. However, to reflect real viral epidemics, it might be useful to take factors such as re-infections, vaccine availability, and the spread of multiple variants of the same virus into account. Furthermore, we assumed that the level of frustration with NPIs is constant over time. This might not be realistic. Accounting for temporal variations in frustration levels associated with NPIs might yield more insight. Lastly, we assumed that the implementation of NPIs was triggered by the prevalence of the disease. However, there might be cases where NPIs are implemented based on other factors such as the availability of intensive care units or the number of disease-induced deaths. It would be insightful to check whether

the conclusions of this study are robust with respect to these other modeling approaches.

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

J-PN'k, CC, and FN: conception of the work, model formulation, critical revision of the important intellectual content, and provide approval for publication of the content. J-PN'k: model analysis and numerical simulations, drafting the work, and agreement to be accountable for all aspects of the work. CC and FN: supervision. All authors contributed to the article and approved the submitted version.

Acknowledgments

J-PN'k submitted this work as part of his master's thesis at the African Institute for Mathematical Sciences, South Africa. We thank Dr. H. Juliette T. Unwin (Imperial College London,

UK) for her feedback, which has substantially improved the quality of this draft.

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.2022.1087683/full#supplementary-material

References

1. Gozzi N, Bajardi P, Perra N. The importance of non-pharmaceutical interventions during the COVID-19 vaccine rollout. *PLoS Comput Biol.* (2021) 17:e1009346.

2. Yang J, Marziano V, Deng X, Guzzetta G, Zhang J, Trentini F, et al. Despite vaccination, China needs non-pharmaceutical interventions to prevent widespread outbreaks of COVID-19 in 2021. *Nat Hum Behav.* (2021) 5:1009–20. doi: 10.1038/s41562-021-01155-z

3. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* (2020) 584:257–61. doi: 10.1038/s41586-020-2405-7

4. Flaxman S, Mishra S, Gandy A, Unwin H, Coupland H, Mellan T, et al. Report 13: estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. *Nature.* (2020) 584:257–61.

5. He Y, Chen Y, Yang L, Zhou Y, Ye R, Wang X. The impact of multi-level interventions on the second-wave SARS-CoV-2 transmission in China. *PLoS ONE.* (2022) 17:e0274590. doi: 10.1371/journal.pone.0274590

6. Banholzer N, Van Weenen E, Lison A, Cenedese A, Seeliger A, Kratzwald B, et al. Estimating the effects of non-pharmaceutical interventions on the number of new infections with COVID-19 during the first epidemic wave. *PLoS ONE.* (2021) 16:e0252827. doi: 10.1371/journal.pone.0252827

7. Pozo-Martin F, Weishaar H, Cristea F, Hanefeld J, Bahr T, Schaade L, et al. The impact of non-pharmaceutical interventions on COVID-19 epidemic growth in the 37 OECD member states. *Eur J Epidemiol.* (2021) 36:629–40. doi: 10.1007/s10654-021-00766-0

8. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. *Sci Rep.* (2021) 11:1–13. doi: 10.1038/s41598-021-81442-x

9. Fair KR, Karatayev VA, Anand M, Bauch CT. Estimating COVID-19 cases and deaths prevented by non-pharmaceutical interventions, and the impact

of individual actions: A retrospective model-based analysis. *Epidemics.* (2022) 39:100557.

10. Yan B, Zhang X, Wu L, Zhu H, Chen B. Why do countries respond differently to COVID-19? A comparative study of Sweden, China, France, and Japan. *Am Rev Public Administ.* (2020) 50:762–9. doi: 10.1177/0275074020942445

11. Quigley AL, Trent M, Seale H, Chughtai AA, MacIntyre CR. Cross-sectional survey of changes in knowledge, attitudes and practice of mask use in Sydney and Melbourne during the 2020 COVID-19 pandemic. *BMJ Open.* (2022) 12:e057860. doi: 10.1136/bmjopen-2021-057860

12. Du Z, Wang L, Shan S, Lam D, Tsang TK, Xiao J, et al. Pandemic fatigue impedes mitigation of COVID-19 in Hong Kong. *Proc Nat Acad Sci.* (2022) 48:e2213313119. doi: 10.1073/pnas.2213313119

13. Petherick A, Goldszmidt R, Andrade EB, Furst R, Hale T, Pott A, et al. A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. *Nat Hum Behav.* (2021) 5:1145–60. doi: 10.1038/s41562-021-01181-x

14. Crane MA, Shermock KM, Omer SB, Romley JA. Change in reported adherence to nonpharmaceutical interventions during the COVID-19 pandemic, April-November 2020. *JAMA.* (2021) 325:883–5. doi: 10.1001/jama.2021.0286

15. Clinton M, Sankar J, Ramesh V, Madhusudan M. Changes in pattern of adherence to NPIs during the COVID-19 pandemic. *Indian J Pediatr.* (2021) 88:837. doi: 10.1007/s12098-021-03768-8

16. Ge Y, Zhang WB, Liu H, Ruktanonchai CW, Hu M, Wu X, et al. Impacts of worldwide individual non-pharmaceutical interventions on COVID-19 transmission across waves and space. *Int J Appl Earth Observat Geoinf.* (2022) 106:102649. doi: 10.1016/j.jag.2021.102649

17. Matthews Pillemer F, Blendon RJ, Zaslavsky AM, Lee BY. Predicting support for non-pharmaceutical interventions during infectious outbreaks: a four region analysis. *Disasters.* (2015) 39:125–45. doi: 10.1111/disa.12089

18. Wang F, Wei J, Shi X. Compliance with recommended protective actions during an H7N9 emergency: a risk perception perspective. *Disasters*. (2018) 42:207–32. doi: 10.1111/disa.12240

19. Nzaji MK, Mwamba GN, Miema JM, Umba EKN, Kangulu IB, Ndala DBB, et al. Predictors of non-adherence to public health instructions during the COVID-19 pandemic in the Democratic Republic of the Congo. *J Multidiscip Healthc*. (2020) 13:1215. doi: 10.2147/JMDH.S274944

20. Wright L, Fancourt D. Do predictors of adherence to pandemic guidelines change over time? A panel study of 22,000 UK adults during the COVID-19 pandemic. *Prevent Med*. (2021) 153:106713. doi: 10.1016/j.ypmed.2021.106713

21. Ding X, Brazel DM, Mills MC. Factors affecting adherence to non-pharmaceutical interventions for COVID-19 infections in the first year of the pandemic in the UK. *BMJ Open*. (2021) 11:e054200. doi: 10.1136/bmjopen-2021-054200

22. Santos JV, da Costa JG, Costa E, Almeida S, Cima J, Pita-Barros P. Factors associated with non-pharmaceutical interventions compliance during COVID-19 pandemic: a Portuguese cross-sectional survey. *J Public Health*. (2022) 26:fdac001. doi: 10.1093/pubmed/fdac001

23. Manfredi P, D'Onofrio A. *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases. SpringerLink: Bücher*. New York, NY: Springer (2013). Available online at: https://books.google.co.uk/books?id=IXhEAAAAQBAJ

24. Verelst F, Willem L, Beutels P. Behavioural change models for infectious disease transmission: a systematic review (2010-2015). *J R Soc Interface*. (2016) 12:13. doi: 10.1098/rsif.2016.0820

25. Bedson J, Skrip LA, Pedi D, Abramowitz S, Carter S, Jalloh MF, et al. A review and agenda for integrated disease models including social and behavioural factors. *Nat Hum Behav*. (2021) 5:834–46. doi: 10.1038/s41562-021-01136-2

26. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys Rep*. (2021) 913:1–52. doi: 10.1016/j.physrep.2021.02.001

27. Acu na-Zegarra MA, Santana-Cibrian M, Velasco-Hernandez JX. Modeling behavioral change and COVID-19 containment in Mexico: a trade-off between lockdown and compliance. *Math Biosci*. (2020) 325:108370. doi: 10.1016/j.mbs.2020.108370

28. Iyaniwura SA, Rabiu M, David JF, Kong JD. Assessing the impact of adherence to Non-pharmaceutical interventions and indirect transmission on the

dynamics of COVID-19: a mathematical modelling study. *Math Biosci Eng*. (2021) 18:8905–32. doi: 10.3934/mbe.2021439

29. Bisanzio D, Reithinger R, Alqunaibet A, Almudarra S, Alsukait RF, Dong D, et al. Estimating the effect of non-pharmaceutical interventions to mitigate COVID-19 spread in Saudi Arabia. *BMC Med*. (2022) 20:1–14. doi: 10.1186/s12916-022-02232-4

30. Yin L, Zhang H, Li Y, Liu K, Chen T, Luo W, et al. A data-driven agent-based model that recommends non-pharmaceutical interventions to suppress Coronavirus disease 2019 resurgence in megacities. *J R Soc Interface*. (2021) 18:20210112. doi: 10.1098/rsif.2021.0112

31. Avusuglo W, Bragazzi NL, Asgary A, Orbinski J, Wu J, Kong JD. *The Role of Behavioral Compliance to Non-Pharmaceutical and Pharmaceutical Interventions in the Fight Against COVID-19: Insights From a Behavior-Disease Economic Epidemic Model Coupled With Optimal Control Theory*. Available at SSRN 3922118 (2021).

32. Jentsch PC, Anand M, Bauch CT. Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *Lancet Infect Dis*. (2021) 21:1097–106. doi: 10.1016/S1473-3099(21)00057-8

33. Capasso V, Serio G. A generalization of the Kermack-McKendrick deterministic epidemic model. *Math Biosci*. (1978) 42:43–61. doi: 10.1016/0025-5564(78)90006-8

34. Gros C, Valenti R, Schneider L, Valenti K, Gros D. Containment efficiency and control strategies for the Corona pandemic costs. *Sci Rep*. (2021) 11:1–13. doi: 10.1038/s41598-021-86072-x

35. Janz NK, Becker MH. The health belief model: a decade later. *Health Educ Q*. (1984) 11:1–47. doi: 10.1177/109019818401100101

36. Bauch CT. Imitation dynamics predict vaccinating behaviour. *Proc R Soc B Biol Sci*. (2005) 272:1669–75. doi: 10.1098/rspb.2005.3153

37. Ochab M, Manfredi P, Puszynski K, d'Onofrio A. Multiple epidemic waves as the outcome of stochastic SIR epidemics with behavioral responses: a hybrid modeling approach. *Nonlinear Dyn*. (2022) 1–40. doi: 10.1007/s11071-022-07317-6

38. Ngonghala CN, Iboi E, Eikenberry S, Scotch M, MacIntyre CR, Bonds MH, et al. Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus. *Math Biosci*. (2020) 325:108364. doi: 10.1016/j.mbs.2020.108364

Mitigating co-circulation of seasonal influenza and COVID-19 pandemic in the presence of vaccination: A mathematical modeling approach

Bushra Majeed[1], Jummy Funke David[1], Nicola Luigi Bragazzi[1], Zack McCarthy[1], Martin David Grunnill[1], Jane Heffernan[2,3], Jianhong Wu[1]* and Woldegebriel Assefa Woldegerima[1]*

[1]Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada, [2]Centre for Disease Modeling, Department of Mathematics and Statistics, York University, Toronto, ON, Canada, [3]Modelling Infection and Immunity Lab, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

The co-circulation of two respiratory infections with similar symptoms in a population can significantly overburden a healthcare system by slowing the testing and treatment. The persistent emergence of contagious variants of SARS-CoV-2, along with imperfect vaccines and their waning protections, have increased the likelihood of new COVID-19 outbreaks taking place during a typical flu season. Here, we developed a mathematical model for the co-circulation dynamics of COVID-19 and influenza, under different scenarios of influenza vaccine coverage, COVID-19 vaccine booster coverage and efficacy, and testing capacity. We investigated the required minimal and optimal coverage of COVID-19 booster (third) and fourth doses, in conjunction with the influenza vaccine, to avoid the coincidence of infection peaks for both diseases in a single season. We show that the testing delay brought on by the high number of influenza cases impacts the dynamics of influenza and COVID-19 transmission. The earlier the peak of the flu season and the greater the number of infections with flu-like symptoms, the greater the risk of flu transmission, which slows down COVID-19 testing, resulting in the delay of complete isolation of patients with COVID-19 who have not been isolated before the clinical presentation of symptoms and have been continuing their normal daily activities. Furthermore, our simulations stress the importance of vaccine uptake for preventing infection, severe illness, and hospitalization at the individual level and for disease outbreak control at the population level to avoid putting strain on already weak and overwhelmed healthcare systems. As such, ensuring optimal vaccine coverage for COVID-19 and influenza to reduce the burden of these infections is paramount. We showed that by keeping the influenza vaccine coverage about 35% and increasing the coverage of booster or fourth dose of COVID-19 not only reduces the infections with COVID-19 but also can delay its peak time. If the influenza vaccine coverage is increased to 55%, unexpectedly, it increases

the peak size of influenza infections slightly, while it reduces the peak size of COVID-19 as well as significantly delays the peaks of both of these diseases. Mask-wearing coupled with a moderate increase in the vaccine uptake may mitigate COVID-19 and prevent an influenza outbreak.

1. Introduction

Despite the implementation of non-pharmaceutical interventions (NPIs) (1, 2) and the existence of highly effective vaccines (3), the "coronavirus disease 2019" (COVID-19) pandemic continues to plague the globe (4). Due to the emergence of multiple highly contagious strains (5) that can evade the immune response and make the existing vaccines less effective (6, 7), it can be expected that new waves of COVID-19 will arise (8), with COVID-19 becoming an endemic disease (9). If these waves occur during a typical influenza season in many regions of the world, then this would create a situation of co-circulation of multiple respiratory viruses, including influenza and respiratory syncytial virus (RSV) (10), among others. Since respiratory pathogens share similar symptoms, this poses a serious challenge to the global public health system (11). During the first 2 years of the still ongoing COVID-19 pandemic, seasonal influenza infections have been mitigated, likely due to the mandatory use of personal protective equipment (PPE) and the implementation of stringent packages of NPIs to contain the spread of COVID-19 (12). On the other hand, the lack of exposure to the influenza virus may also have decreased the population's immunity levels and increased susceptibility to influenza because of its low circulation in the two previous seasons (13). All this, taken together, may potentially lead to a larger seasonal influenza outbreak when COVID-19-induced social distancing and other restrictions are relaxed, creating the ideal situation for influenza–COVID co-circulating in the population.

The combined risk of the concurrent influenza epidemic and the COVID-19 pandemic is a serious global public health concern since it can be extremely difficult to anticipate influenza circulation in the upcoming winter with COVID-19. Some epidemiological observational studies have investigated the impact of SARS-CoV-2 and influenza viruses co-circulation in terms of prevalence rate of co-circulation, clinical outcomes, and imposed burden (11, 14).

Mathematical modeling can play a key role in accounting for interactions of a given pathogen with other infectious agents and in quantifying the real burden of each pathogen and the full impact and effectiveness of public health interventions targeting each infectious agent (15, 16). This is particularly relevant given that, in the current situation, where the SARS-CoV-2 virus is still transmitting continuously in the world despite the availability of many effective vaccines, the emergence of new variants of concern (VOCs) is inevitable.

Considering the additional burden of COVID-19 during the influenza season on population health and healthcare systems, including emergency departments (EDs), it is of paramount importance to investigate the effects of co-circulations when vaccines are available for both diseases. Therefore, the present mathematical model was developed with the aim of studying the impact of SARS-CoV-2 during an influenza season, quantitatively assessing the effects of the co-circulation of the two respiratory pathogens. In particular, in the present study, we are interested in finding optimal strategies to manage and control both influenza and COVID-19 outbreaks during the same season. Among many scenarios and interventions, we consider optimal strategies to delay and separate the peaks of the influenza outbreak and COVID-19 wave.

2. The co-circulation model

To investigate the co-circulation of influenza and SARS-CoV-2, a deterministic compartmental model formulated in terms of ordinary differential equations was employed. The objective of the study is to identify potential control strategies to mitigate the burden caused by both viruses in the near term, that is, during a single respiratory illness season, several simplifying assumptions were made to focus on essential elements relevant to the present study. Specifically, the following assumptions were made to enable such a focus. Changes in population demographics were not considered, that is, births and deaths were not modeled, and the population size remained constant. A closed population was considered and, therefore, no inbound or outbound travel occurred. It was assumed that the population mixed homogeneously. Age-related heterogeneities (e.g., susceptibility to infection, social contact mixing, and vaccination coverage) and spatial heterogeneities (e.g., testing and case reporting, social contact mixing, and level of pathogen circulation) were not considered. Given that

symptomatic influenza and COVID-19 share similar symptoms, the RT-PCR testing capacity was considered to be a shared resource between infections occurring by both influenza and SARS-CoV-2.

Due to the scarcity of data and for model simplification, we assume that infections are exclusive and neither pathogen can be supplanted within an infected host. In other words, we assume no co-infection and super-infection of both diseases. However, secondary infection is possible, that is, an individual, after recovery from one disease, can get infected by other diseases. Focusing on a single respiratory illness reason, it was assumed that upon infection by both influenza virus and SARS-CoV-2 (temporally distinct infections) complete immunity against infection by both pathogens was conferred. All individuals were assumed to be vaccinated before the considered influenza season started, and no vaccination occurred during the season. This means that, rather than modeling vaccination as a time-dependent process, administration of vaccines was modeled as a time-independent process and was embedded in the initial conditions.

The population was stratified into susceptible, isolated, infected, diagnosed, recovered, and hospitalized states, with further stratification based on epidemiological history (e.g., prior infection) of the two circulating viruses (Figure 1). Vaccination of individuals against both influenza and COVID-19 was considered in the model; therefore, dividing the population susceptible to infection by both pathogens into four classes: those individuals not vaccinated against either influenza or COVID-19 (denoted by S_u); those individuals vaccinated against COVID-19 and not influenza (V_c); those individuals vaccinated against influenza and not COVID-19 (V_f); and those individuals vaccinated against influenza and COVID-19 (V_{cf}). Each of these four groups could potentially be infected by SARS-CoV-2 and influenza, with vaccination modulating the infection risk in each class.

In the model, unvaccinated individuals (S_u) were infected by influenza at rate λ_f and with SARS-CoV-2 at rate λ_c. Vaccination status modulated the infection process as follows: Individuals vaccinated against influenza were assumed to be infected by influenza at a reduced rate $(1 - \phi_f)\lambda_f$, where ϕ_f is the vaccine effectiveness against infection by influenza. It has been suggested that influenza vaccination may reduce susceptibility to COVID-19 (17, 18); therefore, those individuals vaccinated against influenza were assumed to be infected by SARS-CoV-2 at the reduced rate $(1 - \eta)\lambda_c$, where $0 \leq \eta \leq 1$. Individuals vaccinated against both COVID-19 and influenza were therefore infected by SARS-CoV-2 at rate $\lambda_c(1 - \phi_c)(1 - \eta)$ and by influenza virus at rate $(1 - \phi_f)\lambda_f$.

Upon infection, and in light of the fact that influenza and COVID-19 share common symptoms, it was assumed that a fraction q of symptomatic individuals isolate themselves until their test results become available (thus entering the Q class); meanwhile, the remaining proportion $(1 - q)$

continues socio-economic activity without disruption (i.e., their contact patterns were assumed to remain unaltered). Those individuals in the isolated (Q) class that tested positive for COVID-19 were assumed to remain isolated and not infectious; meanwhile, those individuals who tested negative were assumed to end their isolation and resume normal mixing.

Those individuals in I_c are diagnosed and isolated, recovered naturally, or hospitalized from their primary infection. This is in contrast to individuals in I_f who either recover naturally or are hospitalized from their primary infection. Given that the study is concerned with a single respiratory disease season, we considered the case where recovery from this primary infection yields immunity to re-infection by the same pathogen. Meanwhile, those who recovered from primary infection (R^f and R^c) become susceptible to secondary infection. When an individual experiences primary and secondary infection, they are considered immune to infection by both influenza and SARS-CoV-2.

Infectious individuals, either influenza (I_f) or COVID-19 (I_c), may be diagnosed, and subsequently, COVID-19-tested positive individuals are isolated (moving to (D_c) compartment), while COVID-19-tested negative will move out from the Q class to (I_f) compartment. Later on, we will specify how the incidence of both diseases impacts the diagnosis speed since they share common symptoms.

Isolated individuals were assumed to transmit the disease longer. Furthermore, diagnosed (D_c) and infectious individuals (I_f) or (I_c) can recover from influenza or COVID-19, respectively, at the rate of γ_{cD} and γ_{fI} or γ_{cI} (moving to class R^f or R^c, respectively), or they will be admitted to hospitals with disease-induced severity, at the rate of τ_f from I_f class and at the rate of θ_c and τ_c from D_c and I_c classes, respectively.

We assume that after recovery from either of these diseases, an individual is completely immune (recall this research focuses on a flu season only) to that disease but susceptible to the other diseases. Individuals recovered from COVID-19 but infected with influenza are denoted by I_f^c and hospitalized as H_f^c. And those individuals recovered from influenza but infected with COVID-19 are denoted by I_c^f, with the diagnosed and isolated denoted by D_c^f and hospitalized H_c^f. Individuals experiencing such secondary infections become immune to both diseases after recovery from the second disease.

2.1. Standing assumptions

We assume that a fraction of the population receives the flu vaccine, and another fraction of the population receives the COVID-19 vaccine. A fraction of the population may receive vaccines for both influenza and COVID-19. We assume that they

FIGURE 1
A flow diagram of the transmission dynamics model, considering the co-circulation of influenza and SARS-CoV-2 viruses. On the right side of the diagram, classes I_c, D_c, and H_c represent that there are further sub-classes in each of these compartments. Similarly, flu classes have sub classes.
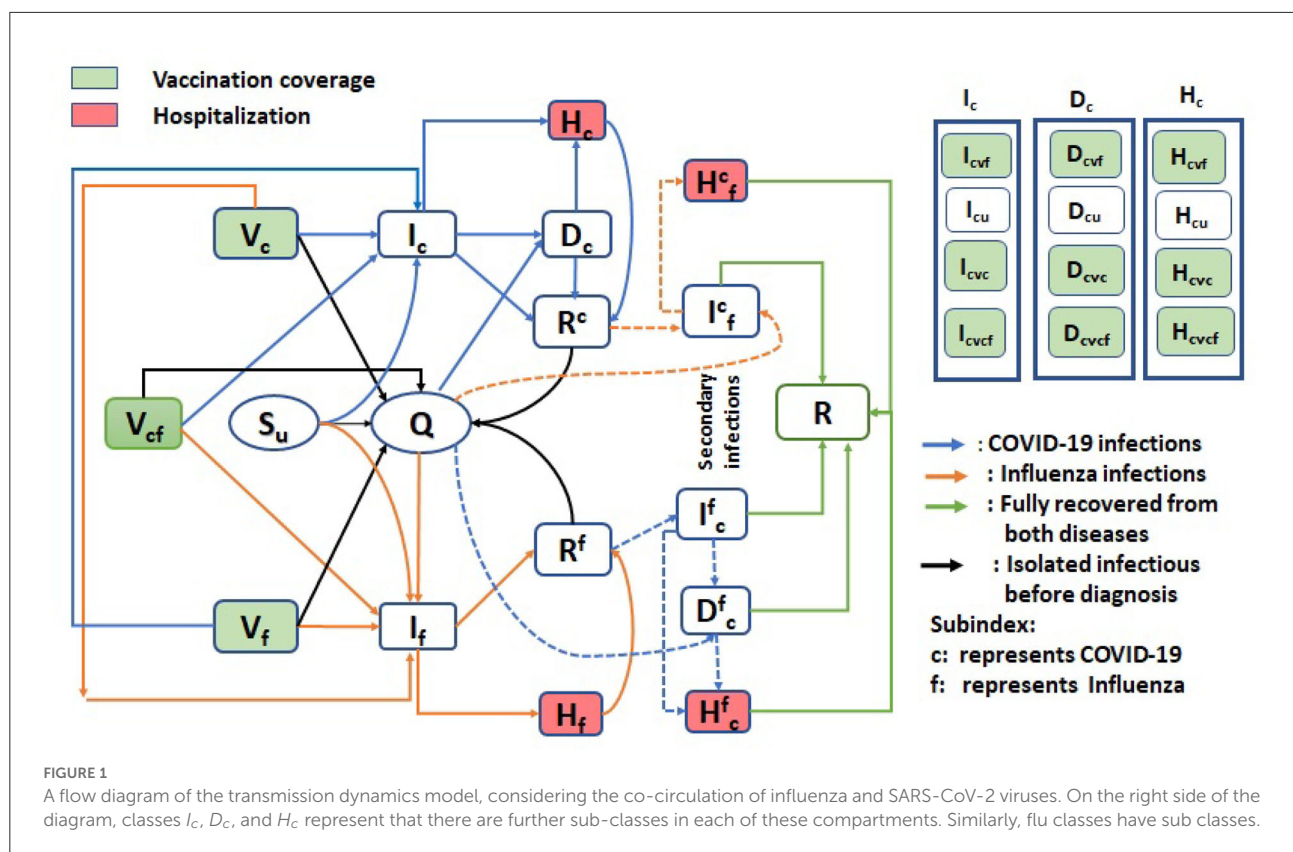
receive the vaccine(s) before our considered influenza season starts. We do not assume a perfect vaccine, so individuals who received a vaccine can still be infected by the disease intended. Furthermore, we assume that a vaccine against influenza or COVID-19 reduces the risk of being infected (reduction in susceptibility) and reduces the risk of severity of disease (hospitalization) if infected by the disease the vaccine is intended.

The flow diagram of the model formulation is shown in Figure 1. Note that in each class other than susceptibles (S_u, V_c, V_f, and V_{cf}) and completely recovered R classes, there are further four sub-classes that are tracked. For instance, the I_c class has I_{cvc}, I_{cvf}, I_{cu}, and I_{cvcf} sub-classes. In these sub-classes, the first letter in sub-index represents the disease class "f" for flu and "c" for COVID-19. Other indices represent the status of the vaccine. For example,

I_{cvc}: First letter "c" in sub-index means infectious with COVID-19, next "vc" represents vaccinated against COVID-19.

I_{cvf}: First letter "c" in sub-index means infectious with COVID-19, next "vf" represents vaccinated against flu.

I_{cu}: First letter "c" in sub-index means infectious with COVID-19, next "u" represents unvaccinated.

I_{cvcf}: First letter "c" in sub-index means infectious with COVID-19, next "vcf" represents vaccinated against both flu and COVID.

Due to the co-circulation of both influenza and SARS-CoV-2 in a single season and due to their common symptoms, COVID-19 testing can be slowed down with a large number of individuals infected with either flu and COVID-19. Therefore, the diagnostic rate to confirm the disease type is a decreasing function of the total number of infections (with either flu or COVID-19) modeled by the available testing capacity. This can be modeled using the Holling type II functions:

$$F_c(I_c, I_f, I_f^c, I_c^f, Q) = \frac{\delta_{Ic} I_c}{1 + w(I_c + I_f + I_f^c + I_c^f + Q)}$$

$$F_f(I_c, I_f, I_f^c, I_c^f, Q) = \frac{\delta_{If} I_f}{1 + w(I_c + I_f + I_f^c + I_c^f + Q)}$$

$$F_f^c(I_c, I_f, I_f^c, I_c^f, Q) = \frac{\delta_{If} I_f^c}{1 + w(I_c + I_f + I_f^c + I_c^f + Q)}$$

$$F_c^f(I_c, I_f, I_f^c, I_c^f, Q) = \frac{\delta_{I_c} I_c^f}{1 + w(I_c + I_f + I_f^c + I_c^f + Q)}$$

$$F_Q(I_c, I_f, I_f^c, I_c^f, Q) = \frac{\delta_Q Q}{1 + w(I_c + I_f + I_f^c + I_c^f + Q)},$$

where w is the constant such that $1/w$ is the maximum number of people who can be tested per day (testing capacity per day). These functions characterize the saturation phenomenon of limited testing resources.

2.1.1. Forces of infections: COVID-19 and influenza

We use standard incidence to describe the forces of infections:

$$\lambda_c = \frac{\beta_c C(I_{cu} + I_{cvf} + I_{cvc} + I_{cvcf} + I_{cu}^f + I_{cvf}^f + I_{cvc}^f + I_{cvcf}^f)}{N}$$

$$\lambda_f = \frac{\beta_f C(I_{fu} + I_{fvf} + I_{fvc} + I_{fvcf} + I_{fu}^c + I_{fvf}^c + I_{fvc}^c + I_{fvcf}^c)}{N}.$$

In summary, the mathematical model for the co-circulation of COVID-19 and flu is given below, with two diseases coupled through their impact on testing speed (and thus isolation duration of the patients with flu):

$$S_u' = -\lambda_c S_u - \lambda_f S_u \tag{1}$$

$$V_c' = -\lambda_c(1 - \phi_c)V_c - \lambda_f V_c \tag{2}$$

$$V_f' = -\lambda_f(1 - \phi_f)V_f - \lambda_c(1 - \eta)V_f \tag{3}$$

$$V_{cf}' = -\lambda_f(1 - \phi_f)V_{cf} - \lambda_c(1 - \phi_c)(1 - \eta)V_{cf} \tag{4}$$

Note that there are 16 subclasses in Q compartment, that is,

$$Q' = Q_{cvc}' + Q_{cvf}' + Q_{cu}' + Q_{cvcf}' + Q_{fvc}' + Q_{fvf}' + Q_{fu}' + Q_{fvcf}' +$$
$$Q_{fvc}'^c + Q_{fu}'^c + Q_{fvf}'^c + Q_{fvcf}'^c + Q_{cvc}'^f + Q_{cu}'^f + Q_{cvf}'^f + Q_{cvcf}'^f$$

where,

$$\begin{cases} Q_{cvc}' = q\lambda_c(1 - \phi_c)V_c - F_{Qcvc} \\ Q_{cvf}' = q\lambda_c(1 - \eta)V_f - F_{Qcvf} \\ Q_{cu}' = q\lambda_c S_u - F_{Qcu} \\ Q_{cvcf}' = q\lambda_c(1 - \phi_c)(1 - \eta)V_{cf} - F_{Qcvcf} \\ Q_{fvc}' = q\lambda_f V_f - F_{Qfvc} \\ Q_{fvf}' = q\lambda_f(1 - \phi_f)V_f - F_{Qfvf} \\ Q_{fu}' = q\lambda_f S_f - F_{Qfu} \\ Q_{fvcf}' = q\lambda_f(1 - \phi_f)V_{cf} - F_{Qfvcf} \\ Q_{fvc}'^c = q\lambda_f R_{vc}^c - F_{Qfvc}^c \\ Q_{fvf}'^c = q\lambda_f(1 - \phi_f)R_{vf}^c - F_{Qfvf}^c \\ Q_{fu}'^c = q\lambda_f R_u^c - F_{Qfu}^c \\ Q_{fvcf}'^c = q\lambda_f(1 - \phi_f)R_{vcf}^c - F_{fvcf}^c \\ Q_{cvc}'^f = q\lambda_c(1 - \phi_c)R_{vc}^f - F_{Qcvc}^f \\ Q_{cvf}'^f = q\lambda_c(1 - \eta)R_{vf}^f - F_{Qcvf}^f \\ Q_{cu}'^f = q\lambda_c R_u^f - F_{Qcu}^f \\ Q_{cvcf}'^f = q\lambda_c(1 - \phi_c)(1 - \eta)R_{vcf}^f - F_{Qcvcf}^f \end{cases} \tag{5}$$

Furthermore, we consider subclasses of I_c as follows:

$$\boxed{I_c' = I_{cu}' + I_{cvf}' + I_{cvc}' + I_{cvcf}'}$$

$$\begin{cases} I_{cvc}' = (1 - q)\lambda_c(1 - \phi_c)V_c - \gamma_c I_{cvc} - F_{cvc}(I_c, I_f, I_f^c, I_c^f, Q) \\ \qquad - (1 - \kappa_c)\tau_c I_{cvc} \\ I_{cvf}' = (1 - q)\lambda_c(1 - \eta)V_f - \gamma_c I_{cvf} - F_{cvf}(I_c, I_f, I_f^c, I_c^f, Q) \\ \qquad - \tau_c I_{cvf} \\ I_{cu}' = (1 - q)\lambda_c S_u - \gamma_c I_{cu} - F_{cu}(I_c, I_f, I_f^c, I_c^f, Q) - \tau_c I_{cu} \\ I_{cvcf}' = (1 - q)\lambda_c(1 - \phi_c)(1 - \eta)V_{cf} - \gamma_c I_{cvcf} \\ \qquad - F_{cvcf}(I_c, I_f, I_f^c, I_c^f, Q) - (1 - \kappa_c)\tau_c I_{cvcf} \end{cases} \tag{6}$$

$$\boxed{D_c' = D_{cu}' + D_{cvf}' + D_{cvc}' + D_{cvcf}'}$$

TABLE 1 Parameters definitions and values with references.

Parameter	Definition	Value	Source
	Parameter list		
β_c	Probability of transmission of COVID	0.1351	Estimated for BA.5
β_f	Probability of transmission of Influenza	0.02–0.035	(32)
C	Contact rate	11.58	(33)
η	Protective effect against infection by the coronavirus if vaccinated with influenza	0.297	(18)
ϕ_c	COVID-19 vaccine effectiveness against infection	Section 2.1.2	(34)
ϕ_f	Influenza Vaccine effectiveness against infection (reduction in susceptibility)	0.4–0.6	(35, 36)
γ_c	Recovery rate of COVID-19 infectious individuals	1/7	(28)
γ_{cD}	Recovery rate of COVID-19 diagnosed isolated individuals	1/5	Section 2.1.3
γ_{cH}	Recovery rate of COVID-19 hospitalized individuals	1/12	(31)
γ_f	Recovery rate of influenza infectious individuals	1/5	(16)
q	Fraction of individuals isolated on symptoms before testing	0.5	Assumed
γ_{fH}	Recovery rate of influenza hospitalized individuals	1/10	Assumed
δ_c	Diagnose rate of symptomatic infected with COVID	1/2	(37)
δ_f	Diagnose rate of symptomatic infected with influenza	1/2	Assumed
$1/w$	Maximum testing capacity per day	5,000	(38)
θ_c	Rate at which COVID-19 confirmed cases hospitalized	0.0305	Calculated from (39)
τ_c	Rate at which COVID-19 infectious cases hospitalized	0.0305	Calculated from (39)
τ_f	Rate at which influenza infectious cases hospitalized	0.0305	Assumed
κ_c	Vaccine effectiveness against hospitalization with COVID	0.8–0.95	(34)
κ_f	Vaccine effectiveness against hospitalization with influenza	0.35–0.5748	(26, 40)

$$
\begin{cases}
D'_{cvc} = F_{cvc}(I_c, I_f, I_f^c, I_c^f, Q) + F_{Qcvc}(I_c, I_f, I_f^c, I_c^f, Q) \\
\quad - \gamma_{cD}D_{cvc} - (1 - \kappa_c)\theta_c D_{cvc} \\
D'_{cvf} = F_{cvf}(I_c, I_f, I_f^c, I_c^f, Q) + F_{Qcvf}(I_c, I_f, I_f^c, I_c^f, Q) \\
\quad - \gamma_{cD}D_{cvf} - \theta_c D_{cvf} \\
D'_{cu} = F_{cu}(I_c, I_f, I_f^c, I_c^f, Q) + F_{Qcu}(I_c, I_f, I_f^c, I_c^f, Q) \\
\quad - \gamma_{cD}D_{cu} - \theta_c D_{cu} \\
D'_{cvcf} = F_{cvcf}(I_c, I_f, I_f^c, I_c^f, Q) + F_{Qcvcf}(I_c, I_f, I_f^c, I_c^f, Q) \\
\quad - \gamma_{cD}D_{cvcf} - (1 - \kappa_c)\theta_c D_{cvcf}
\end{cases}
\tag{7}
$$

$$
\begin{cases}
H'_{cvc} = (1 - \kappa_c)\theta_c D_{cvc} + (1 - \kappa_c)\tau_c I_{cvc} - \gamma_{cH}H_{cvc} \\
H'_{cvf} = \theta_c D_{cvf} + \tau_c I_{cvf} - \gamma_{cH}H_{cvf} \\
H'_{cu} = \theta_c D_{cu} + \tau_c I_{cu} - \gamma_{cH}H_{cu} \\
H'_{cvcf} = (1 - \kappa_c)\theta_c D_{cvcf} + (1 - \kappa_c)\tau_c I_{cvcf} - \gamma_{cH}H_{cvcf}.
\end{cases}
\tag{8}
$$

Here, confirmed/diagnosed cases go to hospitals at rate θ_c and infected but not diagnosed at rate τ_c, and κ_c is the effectiveness of the COVID vaccine to the severity of infection to breakthrough infection.

$$
H'_c = H'_{cu} + H'_{cvf} + H'_{cvc} + H'_{cvcf}
$$

$$
R'^c = R'^c_{vc} + R'^c_{u} + R'^c_{vf} + R'^c_{vcf}
$$

FIGURE 2
The impact of varying the maximum testing capacity per day on COVID-19 and influenza disease prevalence. Left plot represents the total number of infections with COVID-19 ($I_c(t) + I_c^f(t)$) and right plot shows the total number of infections with influenza ($I_f(t) + I_f^c(t)$).
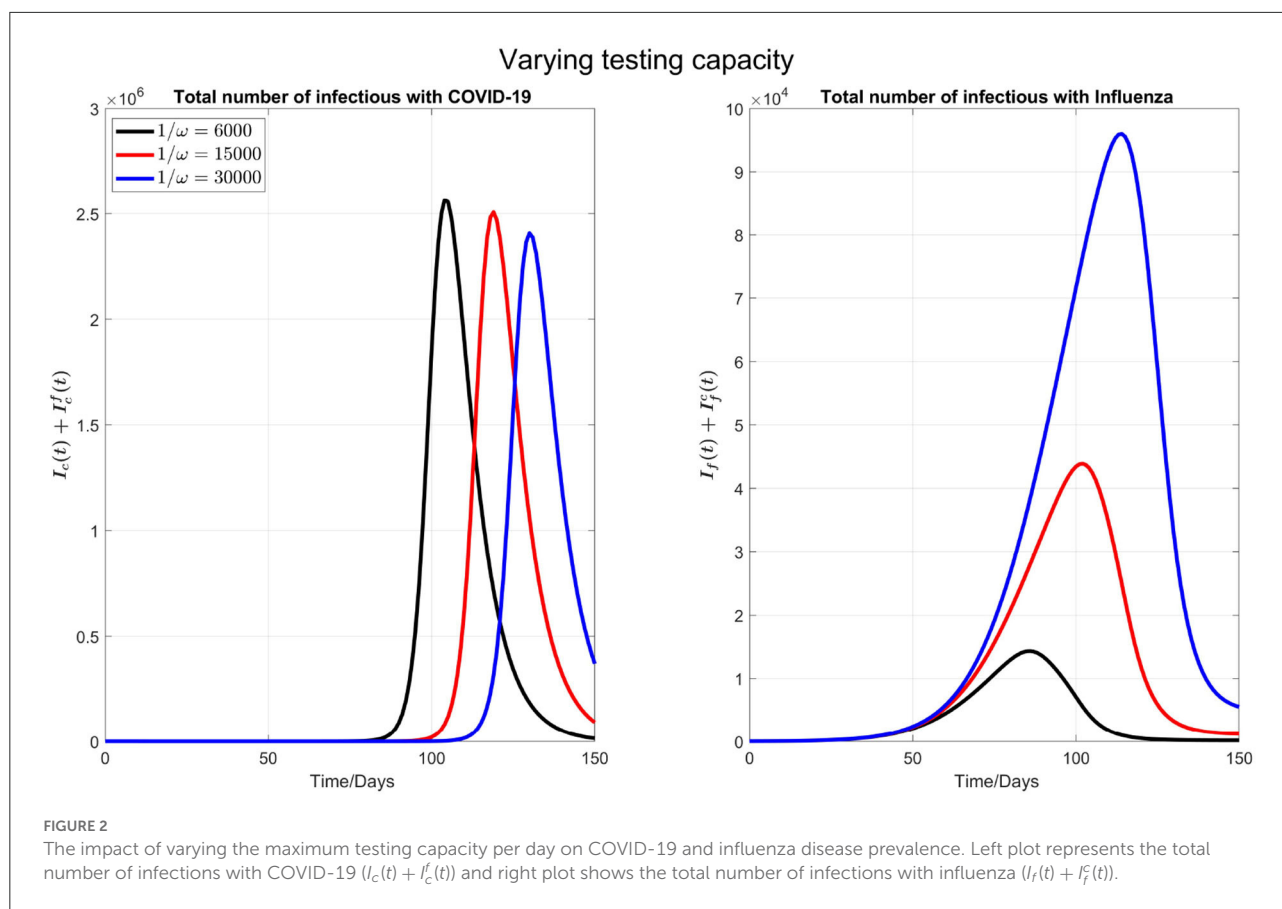
$$
\begin{cases}
R_{vc}^{\prime c} = -\lambda_f R_{vc}^c + \gamma_{cH} H_{cvc} + \gamma_{cD} D_{cvc} + \gamma_c I_{cvc} \\
R_{vf}^{\prime c} = -\lambda_f (1-\phi_f) R_{vf}^c + \gamma_{cH} H_{cvf} + \gamma_{cD} D_{cvf} + \gamma_c I_{cvf} \\
R_u^{\prime c} = -\lambda_f R_u^c + \gamma_{cH} H_{cu} + \gamma_{cD} D_{cu} + \gamma_c I_{cu} \\
R_{vcf}^{\prime c} = -\lambda_f (1-\phi_f) R_{vcf}^c + \gamma_{cH} H_{cvcf} + \gamma_{cD} D_{cvf} + \gamma_c I_{cvcf}
\end{cases}
\tag{9}
$$

$$
\boxed{I_f^{\prime c} = I_{fvc}^{\prime c} + I_{fu}^{\prime c} + I_{fvf}^{\prime c} + I_{fvcf}^{\prime c}}
$$

$$
\begin{cases}
I_{fvc}^{\prime c} = (1-q)\lambda_f R_{vc}^c - \gamma_f I_{fvc}^c - \tau_f I_{fvc}^c + F_{Qfvc}^c(I_c, I_f, I_f^c, I_c^f, Q) \\
I_{fvf}^{\prime c} = (1-q)\lambda_f (1-\phi_f) R_{vf}^c - \gamma_f I_{fvf}^c - \tau_f(1-\kappa_f) I_{fvf}^c \\
\qquad + F_{Qfvf}^c(I_c, I_f, I_f^c, I_c^f, Q) \\
I_{fu}^{\prime c} = (1-q)\lambda_f R_u^c - \gamma_f I_{fu}^c - \tau_f I_{fu}^c + F_{Qfu}^c(I_c, I_f, I_f^c, I_c^f, Q) \\
I_{fvcf}^{\prime c} = (1-q)\lambda_f (1-\phi_f) R_{vcf}^c - \gamma_f I_{fvcf}^c - \tau_f(1-\kappa_f) I_{fvcf}^c \\
\qquad + F_{fvcf}^c(I_c, I_f, I_f^c, I_c^f, Q)
\end{cases}
\tag{10}
$$

$$
\boxed{H_f^{\prime c} = H_{fvc}^{\prime c} + H_{fu}^{\prime c} + H_{fvf}^{\prime c} + H_{fvcf}^{\prime c}}
$$

$$
\begin{cases}
H_{fvc}^{\prime c} = \tau_f I_{fvc}^c - \gamma_{fH} H_{fvc}^c \\
H_{fvf}^{\prime c} = \tau_f (1-\kappa_f) I_{fvf}^c - \gamma_{fH} H_{fvf}^c \\
H_{fu}^{\prime c} = \tau_f I_{fu}^c - \gamma_{fH} H_{fu}^c \\
H_{fvcf}^{\prime c} = \tau_f (1-\kappa_f) I_{fvcf}^c - \gamma_{fH} H_{fvcf}^c
\end{cases}
\tag{11}
$$

$$
\boxed{I_f^{\prime} = I_{fvc}^{\prime} + I_{fu}^{\prime} + I_{fvf}^{\prime} + I_{fvcf}^{\prime}}
$$

$$
\begin{cases}
I_{fvc}^{\prime} = (1-q)\lambda_f V_c - \gamma_f I_{fvc} + F_{Qfvc}(I_c, I_f, I_f^c, I_c^f) - \tau_f I_{fvc} \\
I_{fvf}^{\prime} = (1-q)\lambda_f (1-\phi_f) V_f - \gamma_f I_{fvf} + F_{Qfvf}(I_c, I_f, I_f^c, I_c^f) \\
\qquad - \tau_f(1-\kappa_f) I_{fvf} \\
I_{fu}^{\prime} = (1-q)\lambda_f S_u - \gamma_f I_{fu} + F_{Qfu}(I_c, I_f, I_f^c, I_c^f) - \tau_f I_{fu} \\
I_{fvcf}^{\prime} = (1-q)\lambda_f (1-\phi_f) V_{cf} - \gamma_f I_{cvcf} + F_{Qfvcf}(I_c, I_f, I_f^c, I_c^f) \\
\qquad - \tau_f(1-\kappa_f) I_{cvcf}
\end{cases}
\tag{12}
$$

$$
\boxed{H_f^{\prime} = H_{fvc}^{\prime} + H_{fu}^{\prime} + H_{fvf}^{\prime} + H_{fvcf}^{\prime}}
$$

FIGURE 3
The impact of varying the transmission probability β_f on COVID-19 and influenza disease prevalence. Left plot represents the total number of infections with COVID-19 ($I_c(t) + I_c^f(t)$) and right plot shows the total number of infections with influenza ($I_f(t) + I_f^c(t)$).
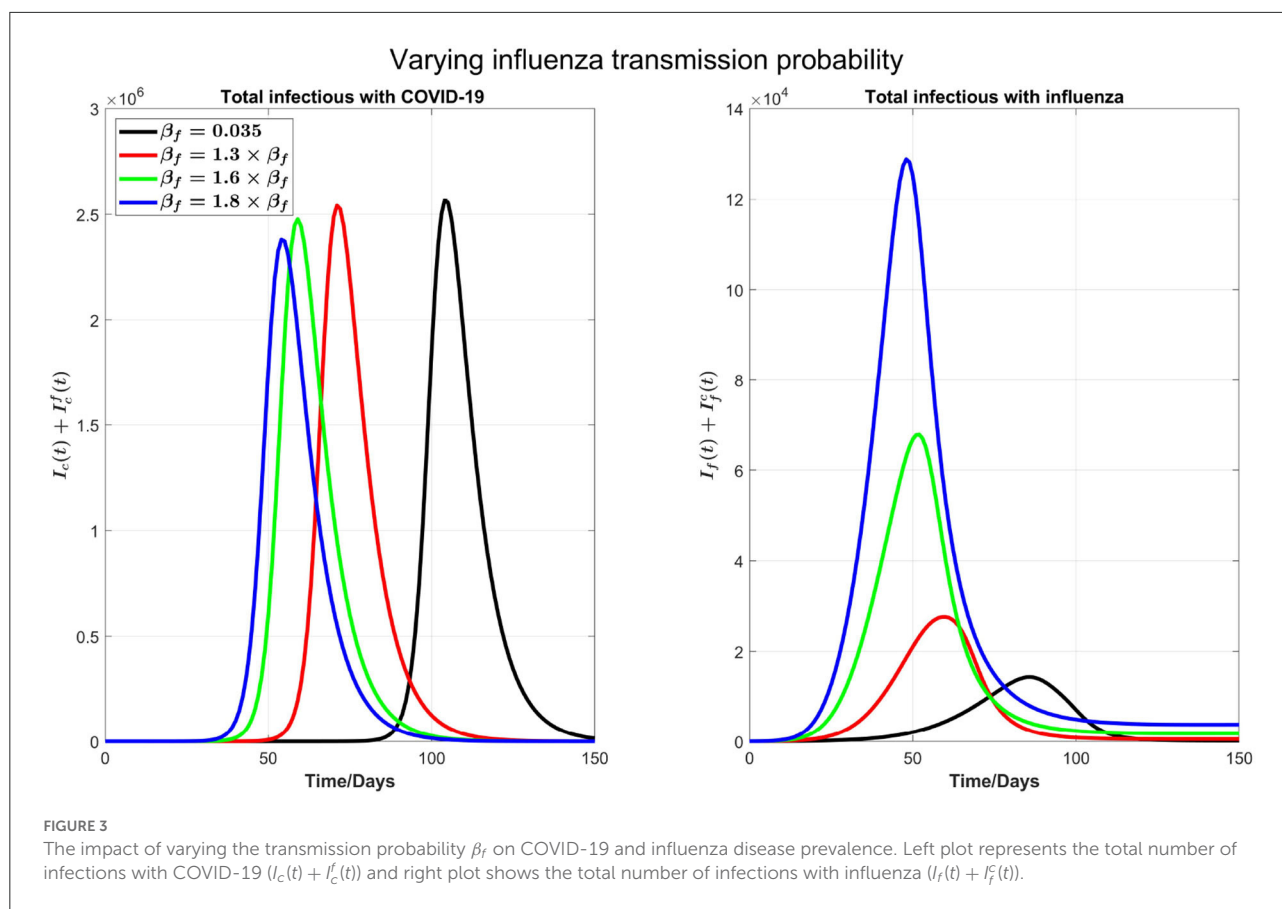
$$\begin{cases} H'_{fvc} = \tau_f I_{fvc} - \gamma_{fH} H_{fvc} \\ H'_{fvf} = \tau_f(1 - \kappa_f)I_{fvf} - \gamma_{fH} H_{fvf} \\ H'_{fu} = \tau_f I_{fu} - \gamma_{fH} H_{fu} \\ H'_{fvcf} = \tau_f(1 - \kappa_f)I_{fvcf} - \gamma_{fH} H_{fvcf} \end{cases} \quad (13)$$

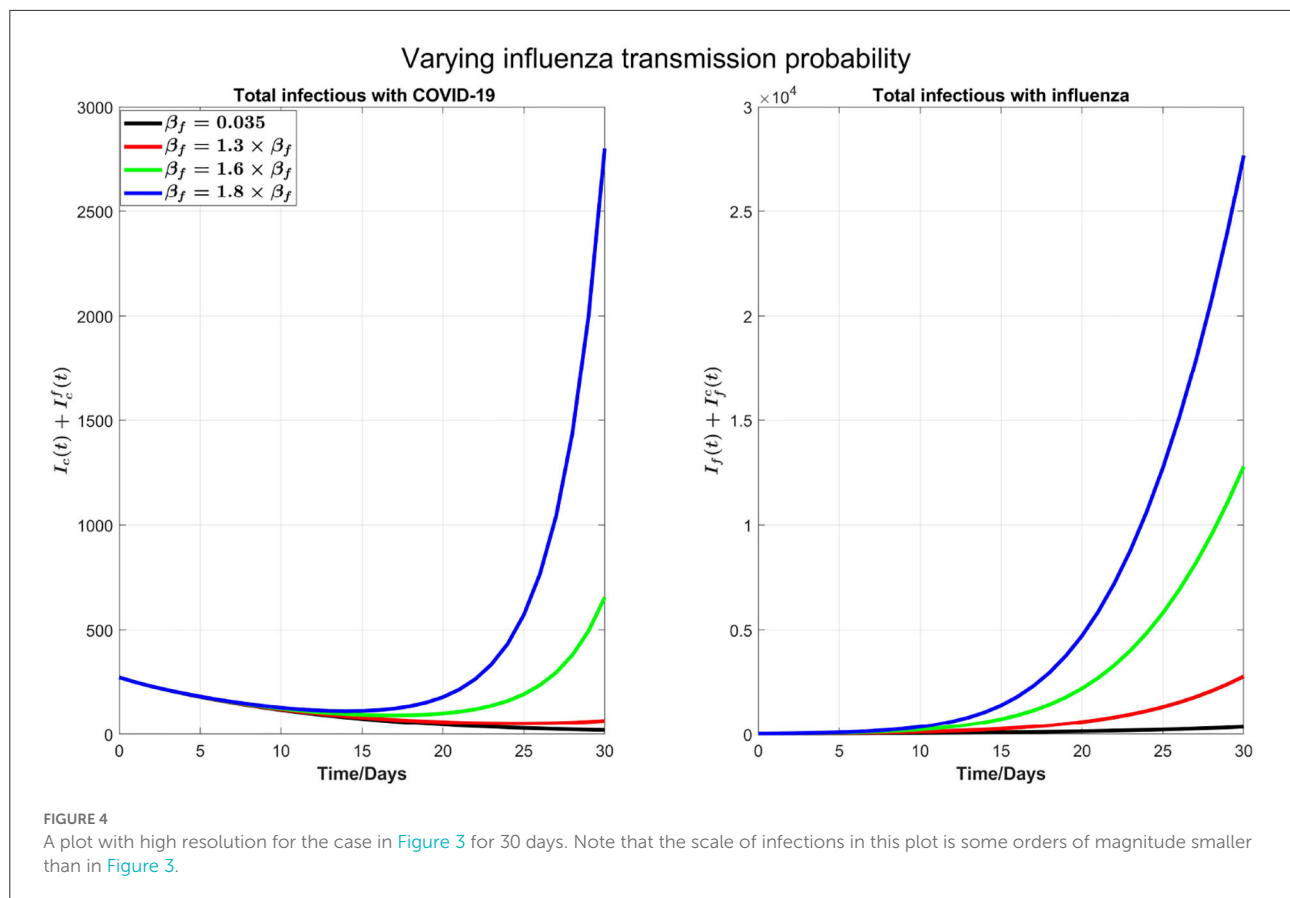$$\boxed{R'^f = R'^f_{vc} + R'^f_u + R'^f_{vf} + R'^f_{vcf}}$$

$$\begin{cases} R^f_{vc} = -\lambda_c(1 - \phi_c)R^f_{vc} + \gamma_{fH} H_{fvc} + \gamma_f I_{fvc} \\ R^f_{vf} = -\lambda_c(1 - \eta)R^f_{vf} + \gamma_{fH} H_{fvf} + \gamma_f I_{fvf} \\ R^f_u = -\lambda_c R^f_{fu} + \gamma_{fH} H_{fu} + \gamma_f I_{fu} \\ R^f_{vcf} = -\lambda_c(1 - \phi_c)R^f_{vcf} + \gamma_{fD} D_{fvcf} + \gamma_f I_{fvcf} \end{cases} \quad (14)$$

$$\boxed{I'^f_c = I'^f_{cvc} + I^f_{cu} + I^f_{cvf} + I'^f_{cvcf}}$$

$$\begin{cases} I'^f_{cvc} = (1 - q)\lambda_c(1 - \phi_c)R^f_{vc} - \gamma_c I^f_{cvc} - \tau_c(1 - \kappa_c)I^f_{cvc} \\ \qquad - F^f_{cvc}(I_c, I_f, I^c_f, I^f_c) \\ I'^f_{cvf} = (1 - q)\lambda_c(1 - \eta)R^f_{vf} - \gamma_c I^f_{cvf} - \tau_c I^f_{cvf} \\ \qquad - F^f_{cvf}(I_c, I_f, I^c_f, I^f_c) \\ I'^f_{cu} = (1 - q)\lambda_c R^f_u - \gamma_c I^f_{cu} - \tau_c I^f_{cu} - F^f_{cu}(I_c, I_f, I^c_f, I^f_c) \\ I'^f_{cvcf} = (1 - q)\lambda_c(1 - \phi_c)(1 - \eta)R^f_{vcf} - \gamma_c I^f_{cvcf} - \tau_c(1 - \kappa_c) \\ \qquad I^f_{cvcf} - F^f_{cvcf}(I_c, I_f, I^c_f, I^f_c) \end{cases}$$
$$(15)$$

$$\boxed{D'^f_c = D'^f_{cvc} + D'^f_{cu} + D'^f_{cvf} + D'^f_{cvcf}}$$

$$\begin{cases} D'^f_{cvc} = F^f_{cvc}(I_c, I_f, I^c_f, I^f_c) + F^f_{Qcvc} - \gamma_{cD} D^f_{cvc} - \theta_c(1 - \kappa_c)D^f_{cvc} \\ D'^f_{cvf} = F^f_{cvf}(I_c, I_f, I^c_f, I^f_c) + F^f_{Qcvf} - \gamma_{cD} D^f_{cvf} - \theta_c D^f_{cvf} \\ D'^f_{cu} = F^f_{cu}(I_c, I_f, I^c_f, I^f_c) + F^f_{Qu} - \gamma_{cD} D^c_{fu} - \theta_c D^f_{cu} \\ D'^f_{cvcf} = F^f_{cvcf}(I_c, I_f, I^c_f, I^f_c) + F^f_{Qcvcf} - \gamma_{cD} D^c_{fvcf} - \theta_c(1 - \kappa_c)D^f_{fvcf} \end{cases}$$
$$(16)$$

FIGURE 4
A plot with high resolution for the case in Figure 3 for 30 days. Note that the scale of infections in this plot is some orders of magnitude smaller than in Figure 3.

$$\boxed{H_c'^f = H_{cvc}'^f + H_{cu}'^f + H_{cvf}'^f + H_{cvcf}'^f}$$

$$\begin{cases} H_{cvc}'^f = \theta_c(1-\kappa_c)D_{cvc}^f + \tau_c(1-\kappa_c)I_{cvc}^f - \gamma_{cH}H_{cvc}'^f \\ H_{cvf}'^f = \theta_c D_{cvf}^f + \tau_c I_{cvf}^f - \gamma_{cH}H_{cvf}'^f \\ H_{cu}'^f = \theta_c D_{cu}^{fc} + \tau_c I_{cu}^f - \gamma_{cH}H_{cu}'^f \\ H_{cvcf}'^f = \theta_c(1-\kappa_c)D_{cvcf}^f + \tau_c(1-\kappa_c)I_{cvcf}^f - \gamma_{cH}H_{fvcf}'^f \end{cases} \quad (17)$$

Taking

$$I_c^f = I_{cvc}^f + I_{cu}^f + I_{cvf}^f + I_{cvcf}^f$$

$$D_c^f = D_{cvc}^f + D_{cu}^f + D_{cvf}^f + D_{cvcf}^f$$

$$H_c^f = H_{cvc}^f + H_{cu}^f + H_{cvf}^f + H_{cvcf}^f$$

$$I_f^c = I_{fvc}^c + I_{fu}^c + I_{fvf}^c + I_{fvcf}^c$$

$$H_f^c = H_{fvc}^c + H_{fu}^c + H_{fvf}^c + H_{fvcf}^c$$

We get the equation for fully recovered class from both diseases.

$$R' = \gamma_c I_c^f + \gamma_{cD}D_c^f + \gamma_{cH}H_c^f + \gamma_f I_f^c + \gamma_{fH}H_f^c \quad (18)$$

2.1.2. Dose-specific (COVID-19) vaccine effectiveness

Most existing vaccines are found to be effective against COVID-19 disease; however, emergence and persistent spread of new SARS-CoV-2 variants render these vaccines less effective against the circulating strain (19, 20). The effectiveness of COVID-19 vaccines against infection also varies by the number of doses. Our study considers the case where the circulating strain is Omicron (B.1.1.529) variant. The first dose of COVID-19 has been deemed ineffective against Omicron infection, but the vaccine effectiveness against second, third, and fourth doses increases and will be assumed to be 0.06, 0.39, and 0.49, respectively (21, 22). To be more generic, the COVID-19 vaccine effectiveness in our model is defined as follows:

$$\phi_c = \rho_1 * ef_1 + \rho_2 * ef_2 + \rho_3 * ef_3 + \rho_4 * ef_4,$$
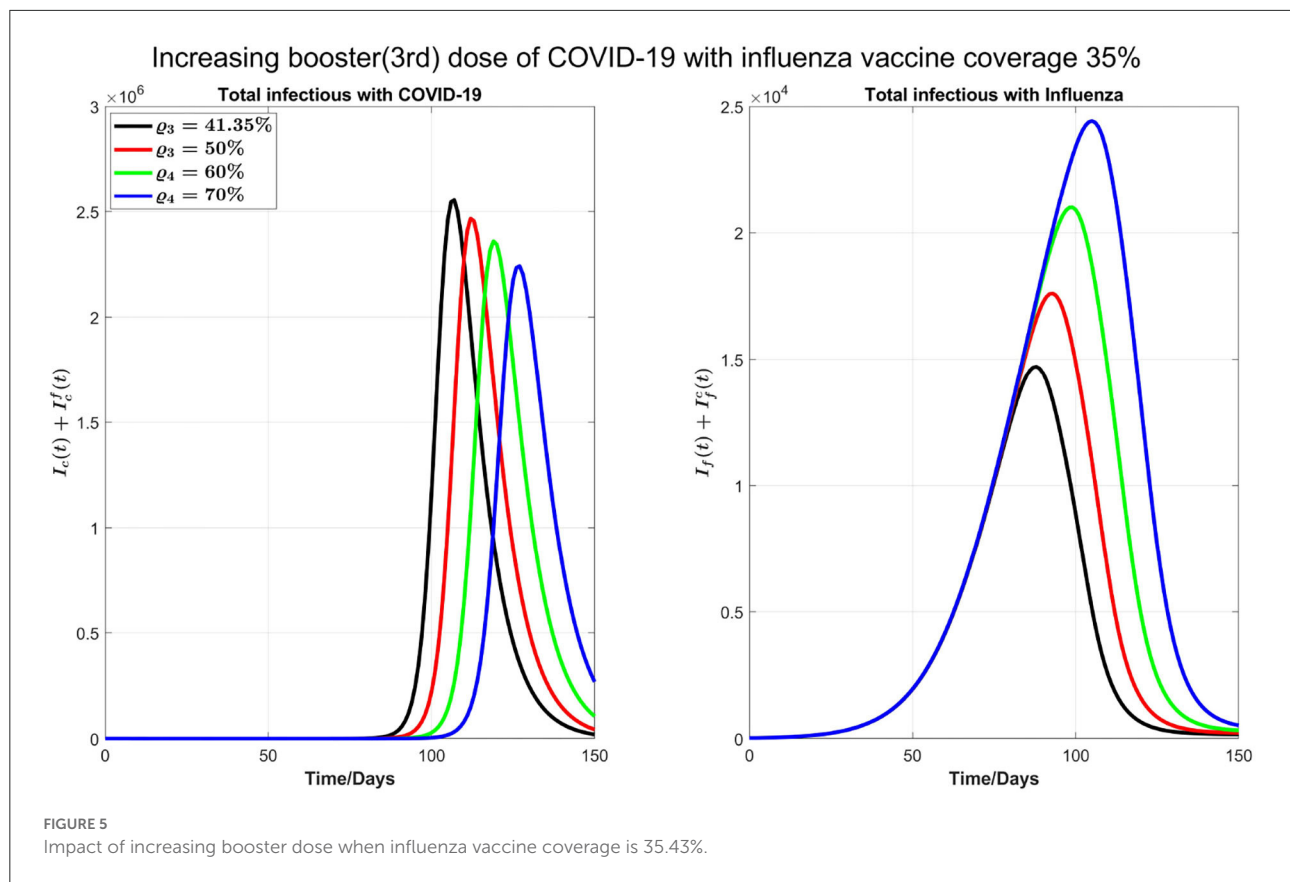
where ef_1, ef_2, ef_3, and ef_4 represent the effectiveness of doses 1, 2, 3, and 4, respectively, against infection and ρ_1, ρ_2, ρ_3, and ρ_4 represent vaccine coverage with doses 1, 2, 3, and 4, respectively.

2.1.3. Recovery rate of diagnosed/isolated individuals

As documented in (23), for COVID-19 laboratory-based testing, patient samples must be first transported to the laboratory, and therefore, it takes 1–3 days to receive test results. So, in what follows, we take diagnose rate $\delta_c = 1/2$ day^{-1} as the baseline assumption. Since we are assuming only flu and COVID-19 are co-circulating, a COVID-19 test negative result implies the flu test positive; hence, we assume the same diagnostic rate for influenza. Furthermore, to calculate the average infectious period $1/\gamma_{cD}$ of the COVID-19 test positive/isolated cases, we use

$$1/\gamma_c = \text{average infectious period for COVID-19}$$

and

$$1/\delta_c = \text{average time to test positive of a COVID-19 infection,}$$

Therefore,

$$1/\gamma_{cD} = 1/\gamma_c - 1/\delta_c, \quad \gamma_{cD} = \frac{\gamma_c \delta_c}{\delta_c - \gamma_c}.$$

As individuals are tested before they recover, $1/\gamma_c > 1/\delta_c$.

2.2. Parameters and initial conditions

We have used some of the data from the province of Ontario, Canada, as a case study. We take the following initial values for vaccinated and unvaccinated state variables. We assume N_0 is the total population of Ontario. From (24), 84.81% of the total population of Ontario will be vaccinated against COVID-19 with dose one, two, three, or four by 19 July 2022. So, initially, a proportion of the population vaccinated against COVID-19 is $V_c = 0.8481 \times N_0$. As a baseline, we assume that 40% (25) of the population who is vaccinated against COVID-19 also gets the influenza vaccine. Thus, this fraction will move to the V_{cf0} initially vaccinated against both COVID-19 and influenza classes. This gives:

$$V_{c0} = 0.8481 \times N_0 - V_{cf0}, V_{cf0} = 0.4 \times 0.8481 \times N_0.$$

Furthermore, because a large proportion of the population has received COVID-19 vaccine and only a small proportion

FIGURE 6

Increasing booster dose when influenza vaccine coverage also increases to 55%.
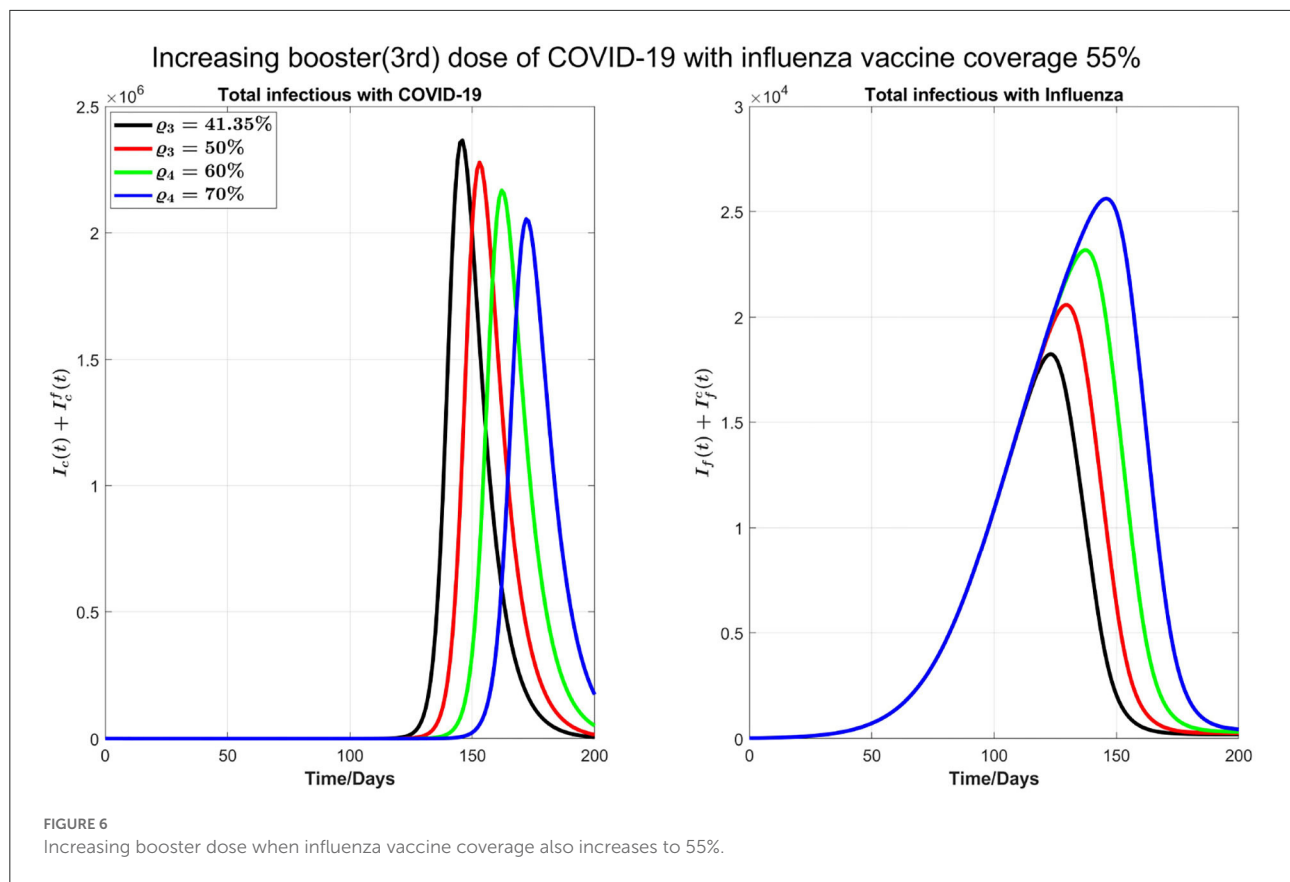
(15.19%) is left unvaccinated, we assume that 10% of the remaining (COVID-19) unvaccinated population is vaccinated against influenza only. Hence, initially vaccinated against influenza is as follows:

$$V_{f0} = 0.1 \times (N_0 - 0.8481 \times N_0).$$

Thus, unvaccinated class has the initial value:

$$S_{u0} = N_0 - V_{c0} - V_{cf0} - V_{f0}.$$

The effectiveness of influenza vaccine against hospitalization range about 35–60% by age (26, 27). In our study, we assume a homogeneous population, so in our simulations, we use 50% as a baseline.
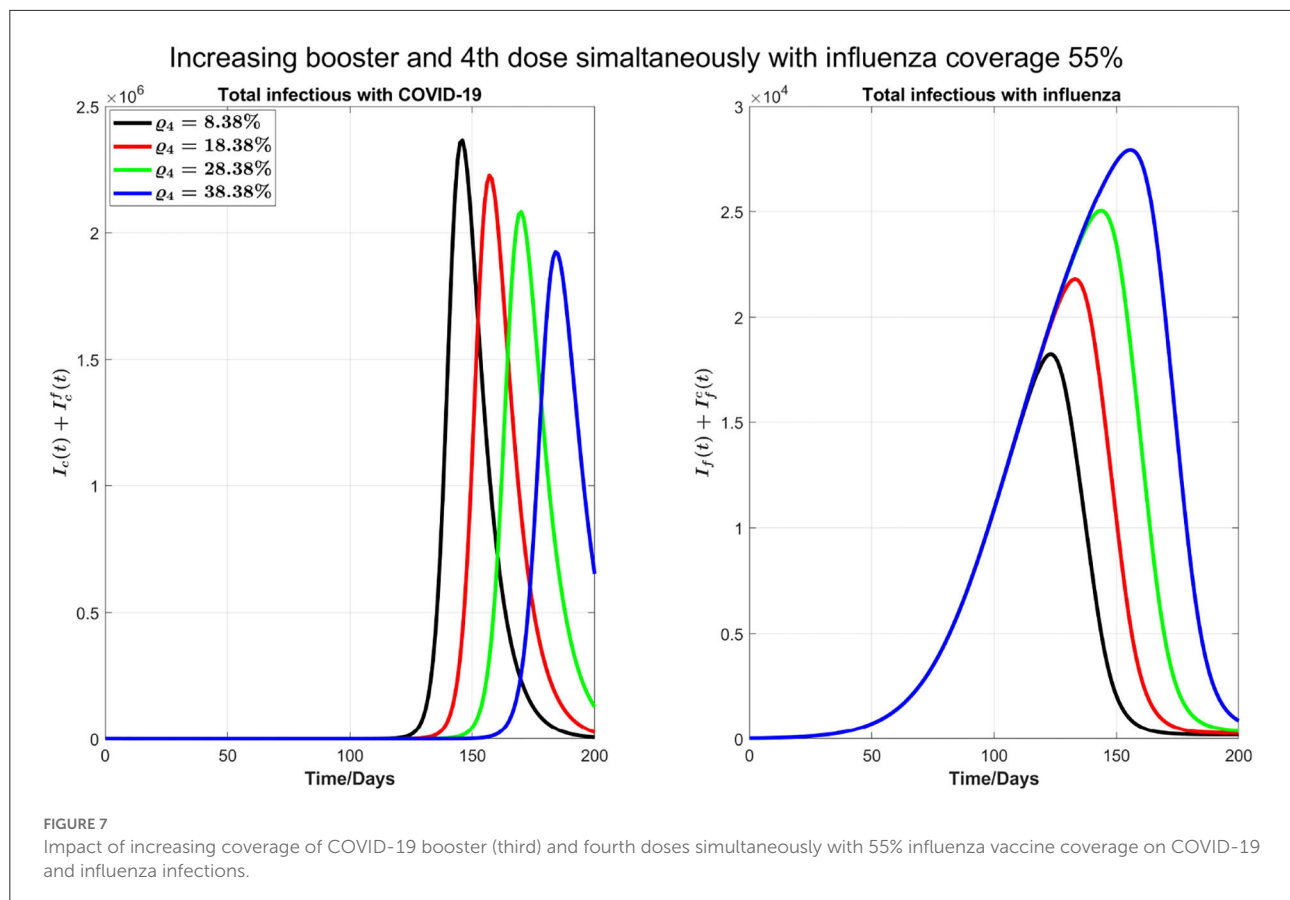
Most of the other parameters are taken from the literature. In particular, we numerically estimated the baseline transmission probability β_c for BA.5 by inverting the formula of reproduction number Ro and following the approach in (28) and using the change in the transmission probability from the ancestral strain to BA.5 in (29, 30). Some of the parameters relevant to influenza, including τ_f, the rate at which influenza infectious cases hospitalized, and δ_f, diagnosis time of symptomatic infected with influenza, are assumed to be the same as COVID-19 for simplifications. On average, the time a person with

COVID-19 stays in hospital is 12 days (31), and 10 days for influenza. Furthermore, a fraction q of individuals who are isolated immediately on symptoms before testing is chosen at 50% as a baseline. The parameters of the co-circulation model is given in Table 1.

3. Simulation results

Increasing COVID-19 test capacity: For simulations below, first, we vary the testing capacity from 6,000 to 30,000 in Figure 2.

We observe that by increasing the testing capacity, the peak time of COVID-19 is postponed and the peak value is also reduced (left plot in the Figure 2). On the other side (right plot of the Figure 2), the peak time of influenza cases is delayed but the peak number of influenza infections increases when the COVID-19 test capacity is increased. This is because, when more tests are done per day, more individuals are isolated because the common flu and COVID-19 symptoms are diagnosed earlier, and those who tested negative for COVID-19 will terminate their isolation. So, an early conclusion of COVID-19 negative will increase the force of infection for flu.

FIGURE 7
Impact of increasing coverage of COVID-19 booster (third) and fourth doses simultaneously with 55% influenza vaccine coverage on COVID-19 and influenza infections.

Varying flu transmission rate: Next, as shown in Figure 3, we consider varying the transmission probability of influenza β_f. This consideration is motivated by the significantly low flu cases during the COVID-19 pandemic due to social distancing. Consequently, there is a possibility that in the coming flu season, the population has larger than normal susceptibility to the flu. In our simulations, we kept testing capacity 6,000 as a baseline.
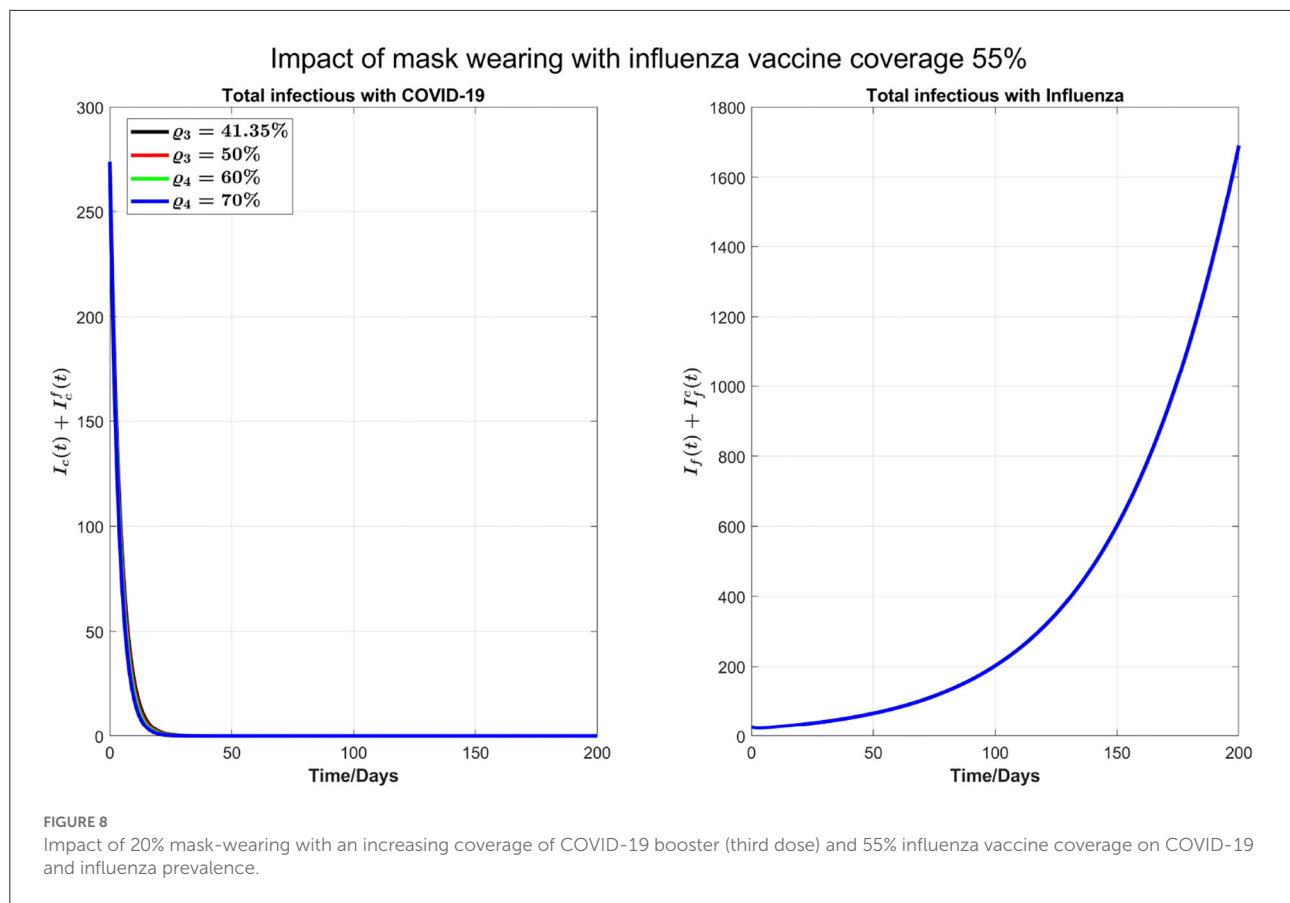
We can see from the right plot of Figure 4 that influenza cases increases and peak earlier by increasing β_f, which is obvious. However, by increasing transmission probability of influenza, COVID-19 cases (left plot) peak earlier with visible higher peak value. For example, if β_f is increased by 30%, from 0.035 to 0.0455, then COVID-19 infections peak about 30 days earlier. Again this is attributed by the testing delay due to the high volume of influenza infections: The higher the transmission probability for the flu, the earlier the peak time for flu outbreak and the higher the number of infections with flu-like symptoms to slow down the COVID-19 testing, leading to a delay in full isolation of the patients with COVID-19. Figure 4 shows the impact of increasing β_f on influenza and COVID-19 infections with high resolution in first 30 days.

3.1. The influenza vaccine and COVID-19 booster coverage

We now consider the issue of minimal and optimal coverage of the booster (third) and fourth doses of COVID-19 and influenza vaccines to control both diseases or to reduce the burden of both diseases together in a single season.

In Figure 5, we start by considering 40% of individuals who are vaccinated against COVID-19 with first, second, or third dose also get the influenza vaccine. As of 19 July 2022, 84.81% of Ontario residents are fully vaccinated against COVID-19. As such, we consider 33.92% of Ontario residents are vaccinated against influenza from this fraction of population, and 10% of the remaining 15.19 % who are not vaccinated against COVID-19 receive influenza vaccine. Thus, the overall population initially vaccinated against influenza is 33.92% + 1.519% = 35.43%. Finally, the proportion of the Ontario population receiving booster dose is 41.35%. In our simulations, we will consider this coverage to be increased to 50, 60, and 70%, respectively, to see the impact of increasing booster coverage on the prevalence of both COVID-19 and influenza.

In the left plot of Figure 6, we observe delaying the COVID-19 peak and reducing the peak size by increasing COVID-19

FIGURE 8
Impact of 20% mask-wearing with an increasing coverage of COVID-19 booster (third dose) and 55% influenza vaccine coverage on COVID-19 and influenza prevalence.

booster coverage. On the other hand, in the right plot, we observe an increase of influenza peak value with increasing COVID-19 booster coverage; increasing the coverage from 41.35 to 70% will see flu peak value changes from close to 15,000 to 25,000. This is because when we increase the COVID-19 booster coverage, there will be less COVID-19 infections and hence less isolation, so the increase in the population susceptibility to flu.

Next, as shown in Figure 7, we investigate the impact of increasing the influenza vaccine coverage to 55% from 35%, along with the increase of booster dose of COVID-19. We observe that, though there is no significant impact of increasing influenza vaccine coverage on peak values of both diseases and the peak size of influenza infections increases by increasing influenza vaccine coverage, peak times for both diseases are remarkably delayed. Similar results in Figure 7 are observed by increasing 10% coverage of booster (third dose) and fourth doses of COVID-19 and influenza coverage by 55%.

We also carry out some simulations to assess the impact of personal protections such as mask-wearing along with increasing the booster dose of COVID-19 and influenza vaccine coverage by 55%. We note that mask-wearing is no longer compulsory in Ontario; however, mask-wearing is strongly recommended. Reducing β_f and β_c by 20%, as shown in

Figure 8, there will be no COVID-19 outbreak under a variety of booster coverage, and we also see a noticeable reduction in influenza cases. On the other hand, we observed through simulations result (not shown here) that if influenza coverage is kept at 35%, then there is a high increase in influenza infections. So, we conclude that to control outbreaks of both diseases, an increase in influenza vaccine coverage and continuing the mask-wearing to reduce the COVID-19 and flu transmission by 20% can play a considerable role even without increasing the COVID-19 booster dose or fourth dose.

We plot some color-coded simulations to visualize the variation of total infections of COVID-19 and influenza at the peak sizes and overall total infections in Figure 9. We vary the proportion of individuals vaccinated with the third dose of COVID-19 vaccine (horizontal axis) with the proportion vaccinated against influenza (vertical axis) and examine the impact on individuals infected with COVID-19 ($I_c + I_c^f$), the upper left panel in Figure 9, and influenza ($I_f + I_f^c$), right panel in Figure 9. The peak size of individuals infected with COVID-19 decreases with an increase in both the proportion of individuals vaccinated with COVID-19 and influenza vaccines. Based on our model results, it is possible to lower the overall peak size and that of individuals infected with COVID-19 when we vaccinate about

70% of the population with the third dose of the COVID-19 vaccine, together with vaccinating about 65% of the population with influenza vaccine (see Figure 9 upper left panel and bottom panel). Alternatively, increasing the proportion of individuals vaccinated with COVID-19 and/or influenza vaccines increase the peak size of individuals infected with influenza (see Figure 9, top right panel). Furthermore, by varying the same proportion of individuals with the booster dose of COVID-19 and influenza vaccine, we plot Figure 10 [the peak time (the time when there would be a maximum number of total COVID-19 infections) (left panel), and total influenza infections (right panel)].

We also consider the impact of the fraction of isolation "q" of infected individuals with influenza and COVID-19 symptoms before the diagnosis. We vary this fraction from 10 to 60% to produce Figure 11 with influenza vaccine coverage of 35%. In Figure 11, we can observe that the peak time for COVID-19 is postponed; and peak value is reduced with increasing fraction "q." The peak time for influenza is also postponed, however, against the intuition, the peak number of infections with influenza for $q = 20\% - 30\%$ decreases and then increases again by increasing isolation before testing. This is because of the intensive transmission rate of COVID-19 and dominance at the beginning, with much higher number of initial infections as compared with the influenza. So, after getting recovered from COVID-19, there would be more susceptibility to influenza, and influenza infections start to increase and COVID-19 infections start to decrease. With $q = 40\% - 60\%$, there is a substantial delay in the COVID-19 peak. So, more infections with influenza can be seen earlier because of more susceptibility to influenza. Although there is also a delay in influenza peak with increasing "q," similar results (not shown here) are found with influenza vaccine coverage of 55%.

3.2. Sensitivity analysis

We also conduct the sensitivity analysis using Latin Hypercube Sampling (LHS) and partial rank correlation coefficients (PRCC) method on different important epidemic outcomes to public health, like the total number of infections, peak number of infections (peak magnitude), and peak time to identify the key parameters of epidemic with the hope of determining public health measures that can be implemented to control or eliminate the outbreaks of these diseases. PRCC is an efficient sensitivity analysis method based on sampling, which assigns a value between -1 and $+1$ for each parameter. A positive PRCC value indicates a positive correlation of the parameter with disease maintenance, whereas a negative value indicates a negative correlation with the disease. The parameters studied are as follows: β_f, β_c, ϕ_f, ϕ_c, q, and C. Figures 12, 13 show the PRCC indices of these selected parameters on total number of infections (Figure 12A), peak time (Figure 12B),
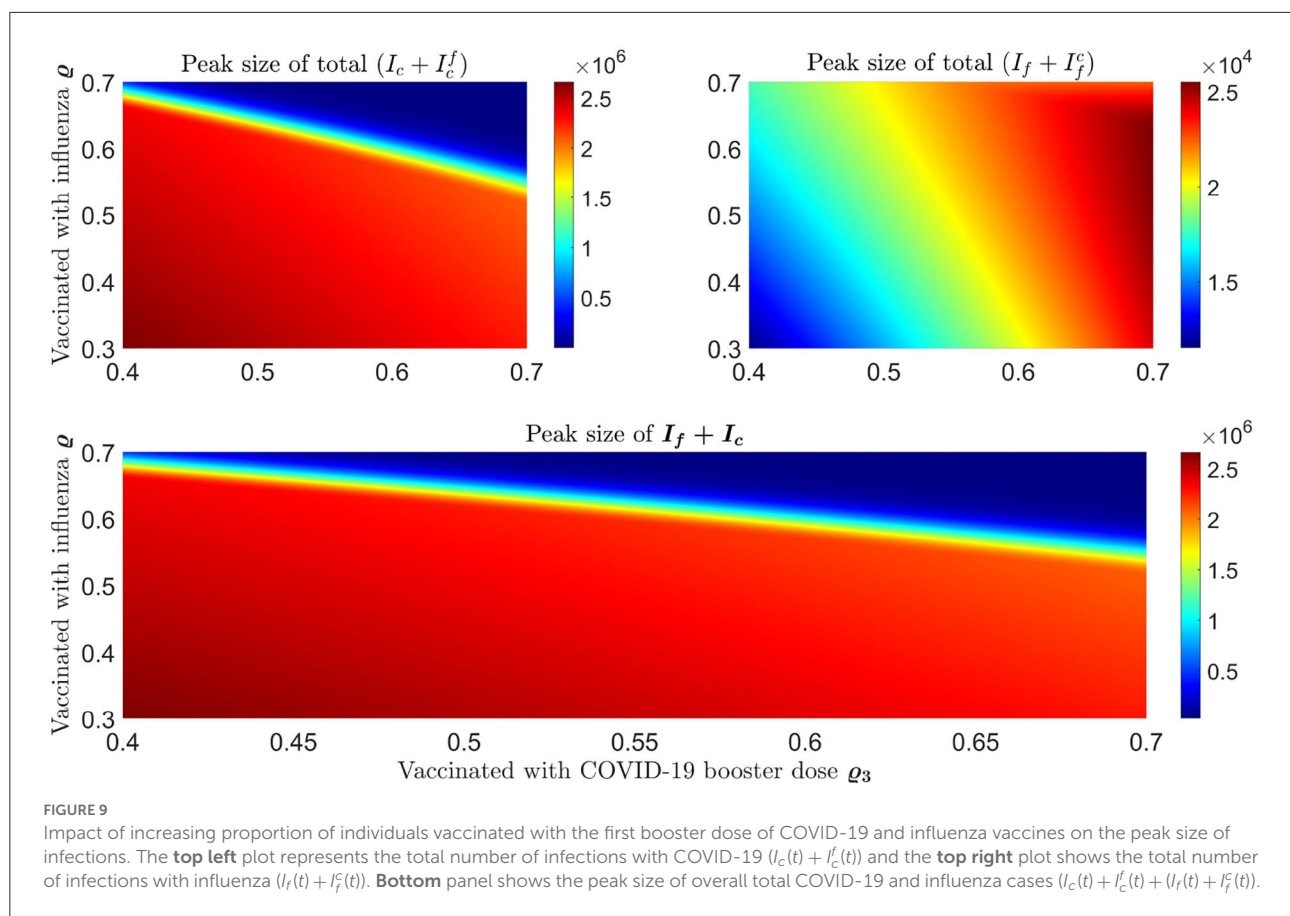
and peak number of infections (Figure 12C) of COVID-19 and influenza, respectively.

Figure 13A determines that the disease transmission probabilities, β_f and β_c, and contact rate C have a positive impacts, while vaccine effectiveness ϕ_f, ϕ_c, and q have negative impacts. A significant positive correlation among COVID-19 infections and contact rate C and β_f, and a negative correlation with COVID-19 vaccine effectiveness ϕ_c are evident. A positive correlation with β_f means more susceptibility to influenza in the population leading to more infections with influenza would also lead to more infections with COVID-19 because of common symptoms and delay in diagnosis. On the other hand, a strong negative correlation of q means as many individuals on symptoms immediately isolate themselves before diagnosis, there would be fewer infections with COVID-19. We can also see that COVID-19 infections are less sensitive to influenza vaccine but increase in influenza coverage would reduce COVID-19 infections. This is because influenza vaccines are assumed to provide some protection against COVID-19 infection too. Figure 13B shows the effects of the PRCC indices on the peak time. The effects of these parameters are reversed, but their relative levels of influence show a similar pattern. Finally, Figure 13C shows the effects of the PRCC indices on the epidemic peak size. The impacts of these parameters closely resemble to those observed in Figure 13C. Next, Figure 13A shows the sensitivity of the chosen parameters to total influenza infections. Here, a strong positive correlation of β_f and contact rate C and a significant negative correlation of ϕ_f and β_c with total influenza infections are understandable. But a positive impact of ϕ_c on influenza infections means that if there are more COVID-19 booster or fourth doses administrated (considering booster and fourth doses because first dose has no and second dose has very low, 0.06%, effectiveness against COVID-19 infection), then it will result in more infections with influenza. Furthermore, here q has a positive correlation with influenza infections. Similarly, sensitivity analysis results for influenza epidemic peak magnitude (Figure 13C) and a reverse pattern for influenza peak time (Figure 13B) are obtained.

4. Discussion

In the present study, we developed a deterministic mathematical model to describe the dynamics of COVID-19 and influenza transmission when both are present in the same season, and a certain percentage of the population has received the influenza, COVID-19, or both vaccines.

Worldwide, the influenza virus, through seasonal waves of infection, generates a significant toll of cases and deaths (41). According to the World Health Organization, before the COVID-19 pandemic, there were an estimated 1 billion cases yearly, of which 3–5 million resulted in 290,000–650,000 influenza-related respiratory deaths (42). In the Northern and

FIGURE 9
Impact of increasing proportion of individuals vaccinated with the first booster dose of COVID-19 and influenza vaccines on the peak size of infections. The **top left** plot represents the total number of infections with COVID-19 ($I_c(t) + I_c^f(t)$) and the **top right** plot shows the total number of infections with influenza ($I_f(t) + I_f^c(t)$). **Bottom** panel shows the peak size of overall total COVID-19 and influenza cases ($I_c(t) + I_c^f(t)) + (I_f(t) + I_f^c(t))$).
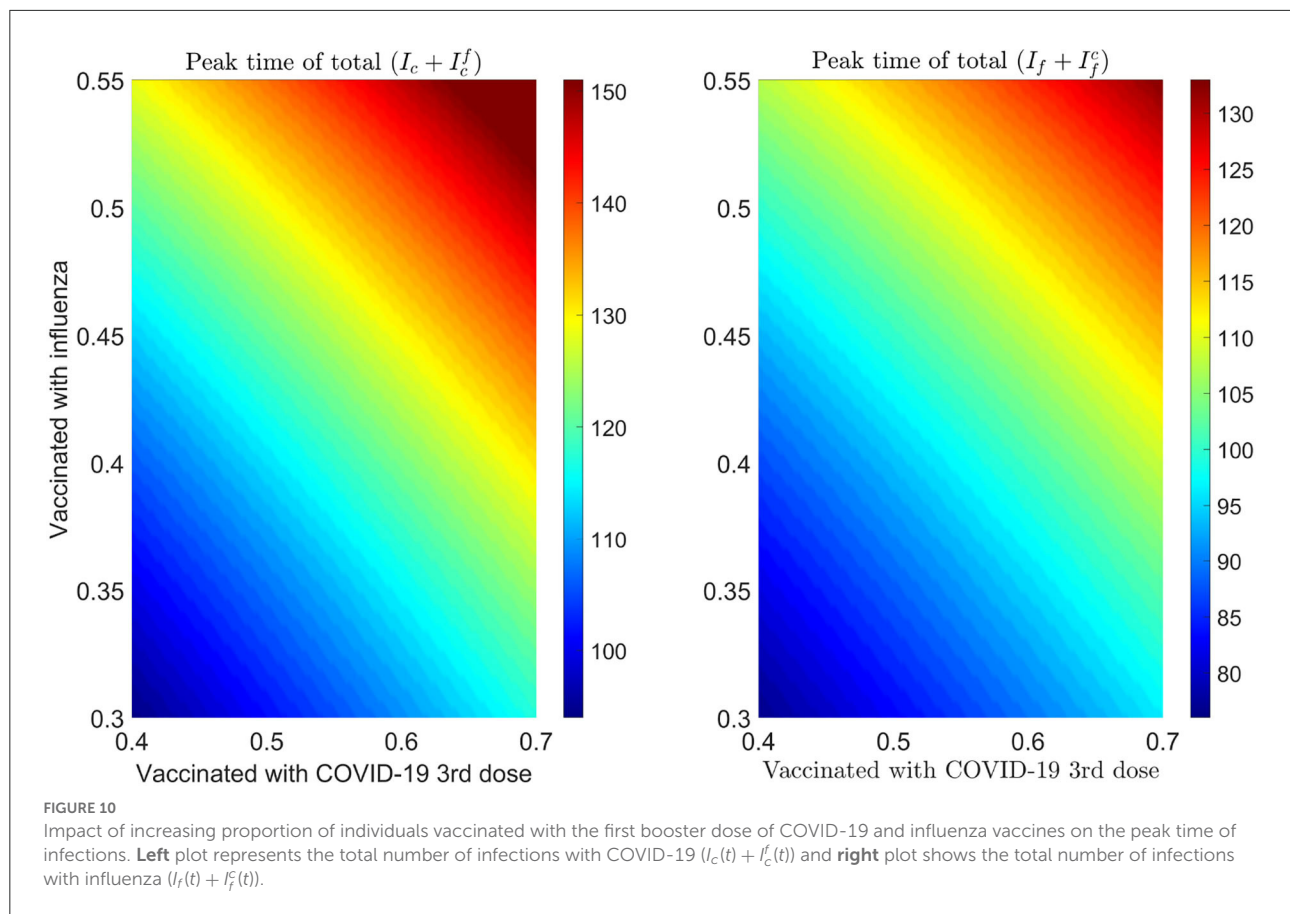
Southern Hemispheres, the flu season usually runs from October to April and April to September, respectively.

COVID-19 has perturbed seasonal influenza activity (43): in 2020 and 2021, this remained low historically at international levels (44–49). These changes were attributed to the widespread implementation of NPIs such as physical distancing, masking requirements, and lockdowns to mitigate the transmission of SARS-CoV-2. More recently, however, and coinciding with relaxation in public health mitigation, a substantial simultaneous burden of influenza and SARS-CoV-2 has been observed in the Southern Hemisphere. The June 2022 Australian Influenza Surveillance Report stated that from mid-April 2022, the weekly number of laboratory-confirmed influenza cases in Australia had occurred earlier than usual and exceeded the 5-year average (50). In the Northern Hemisphere, in Canada, after lifting public health measures at the beginning of March 2022, influenza virus circulation increased, reaching the seasonal epidemic threshold in April with an unusual peak in May 2022 (51).

Although influenza and COVID-19 viruses belong to different families and have some differences, with SARS-CoV-2 having a much higher basic reproduction number, longer incubation period, and shorter interval between symptoms

onset and infectivity (52), COVID-19 shares many significant clinical and epidemiological features with influenza, such as transmission routes and symptoms. Clinical differentiation of these two respiratory diseases can be a challenge, particularly in the early stage, based on a common diagnosis in the absence of laboratory evidence and isolation of a specific pathogen. Therefore, a laboratory diagnostic test may be required to rule out the suspicion and establish a definitive diagnosis. Some sophisticated mathematical/statistical techniques, such as machine learning-based decision modeling approaches, have been able to distinguish between influenza and COVID-19 cases (53, 54).

In terms of clinical public and global health, even though available vaccines do not provide complete protection against infection, they can reduce the severity of the disease significantly. As influenza virus also mutates frequently, an updated flu vaccine is required every year, considering which strains are anticipated to circulate in the community. Thus, the effectiveness of the flu vaccine varies depending on how well the vaccine matches with circulating strains or who is being vaccinated (age or health characteristics of the vaccinated individual) (55). According to health professionals, receiving influenza and COVID-19 vaccines is a cornerstone to protect
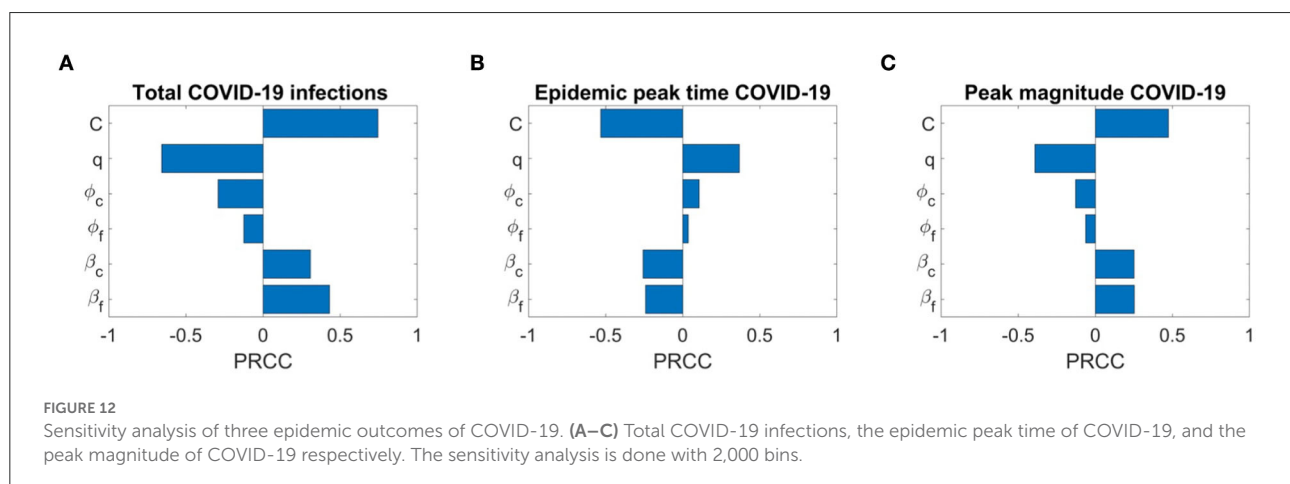
FIGURE 10
Impact of increasing proportion of individuals vaccinated with the first booster dose of COVID-19 and influenza vaccines on the peak time of infections. **Left** plot represents the total number of infections with COVID-19 ($I_c(t) + I_c^f(t)$) and **right** plot shows the total number of infections with influenza ($I_f(t) + I_f^c(t)$).
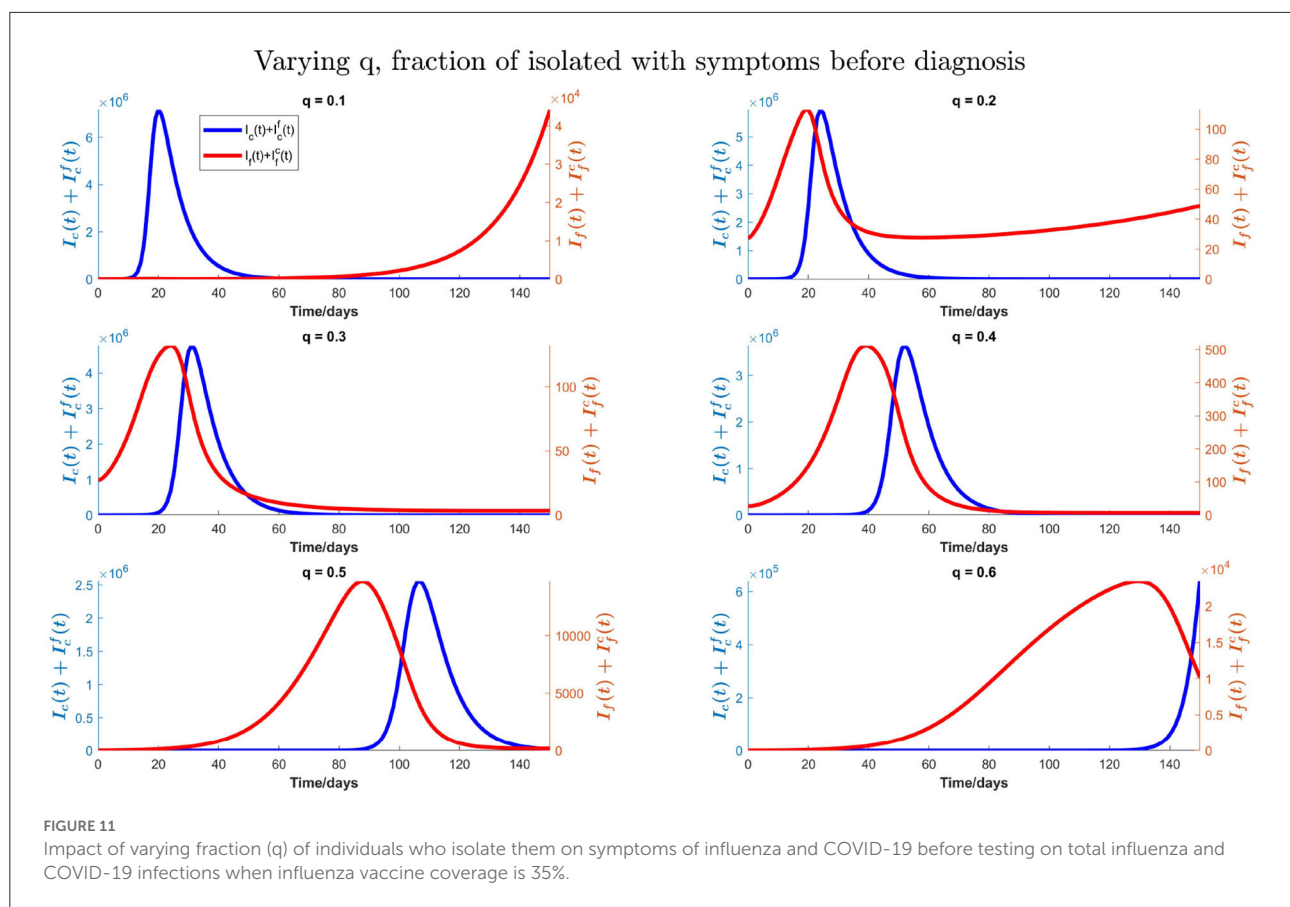
against illness of both diseases and potentially their severe consequences, including hospitalization and fatality (56, 57).

Therefore, reducing the burden of COVID-19 and influenza will depend primarily on increasing vaccination coverage. A mathematical model (58) has modeled the interaction between SARS-CoV-2 and influenza during the early phase of the COVID-19 pandemic. Using a population-based modeling approach, the authors found a 2- to 2.5-fold population-level increase in COVID-19 transmission associated with influenza co-circulation, warranting the importance of being immunized against influenza. Influenza vaccination can represent, indeed, an important public health intervention to, at least partially, relieve and mitigate against the burden generated by SARS-CoV-2 and influenza co-circulation (14). As found by a systematic review of the literature, influenza immunization can also, indeed, confer protection against COVID-19 (59).

In our model, we took into account the multi-factorial interactions between influenza and COVID-19 in terms of additional epidemiological, clinical, and organizational burdens due to their comparable early symptoms, which can present a clinical challenge in determining the patient's disease type and overwhelm testing capacity, lowering the diagnosis rate. In this scenario, we observe that if testing capacity is increased

per day, it will delay the peak of COVID-19 in addition to reducing it. On the other hand, because of early detection and lack of isolation, influenza cases can grow. Agreeing with (60), the low levels of influenza activity in the previous 2 years of the COVID-19 pandemic may result in an increased proportion of susceptible individuals. Our simulations with increasing influenza transmissibility show that this will bring the peaks of influenza and COVID-19 earlier, slightly decrease the COVID-19 peak size, but significantly increase the flu peak. Here, we can see the real influence of co-circulation and interaction of COVID-19 and influenza on each other. It is because of the testing delay brought on by the high number of influenza cases. The earlier the flu season's peak and greater the number of infections with flu-like symptoms, the greater the risk of flu transmission, which slows down COVID-19 testing, resulting in the delay of complete isolation of patients with COVID-19, who have not been isolated before the clinical presentation of symptoms and have been continuing their normal daily activities. Furthermore, our simulations stress the importance of vaccine uptake for preventing infection, severe illness, and hospitalization at the individual level and for disease outbreak control at the population level to avoid putting strain on already weak and overwhelmed healthcare systems. As such, ensuring

FIGURE 11
Impact of varying fraction (q) of individuals who isolate them on symptoms of influenza and COVID-19 before testing on total influenza and COVID-19 infections when influenza vaccine coverage is 35%.

FIGURE 12
Sensitivity analysis of three epidemic outcomes of COVID-19. (A–C) Total COVID-19 infections, the epidemic peak time of COVID-19, and the peak magnitude of COVID-19 respectively. The sensitivity analysis is done with 2,000 bins.

optimal vaccine coverage for COVID-19 and influenza to reduce the burden of these infections is paramount. We showed that by keeping the influenza vaccine coverage about 35% and increasing the coverage of booster or fourth dose of COVID-19 not only reduces the infections with COVID-19 but also can delay its peak time. If the influenza vaccine coverage is increased to 55%, unexpectedly, it increases the peak size of influenza

infections slightly, while it reduces the peak size of COVID-19 as well as significantly delays the peaks of both of these diseases. Also, we have shown that personal protection decisions like mask-wearing can mitigate the COVID-19 outbreak and can avert an outbreak of seasonal influenza significantly. In conclusion, an increase in vaccine uptake of both diseases, particularly influenza, can significantly delay the peak time of
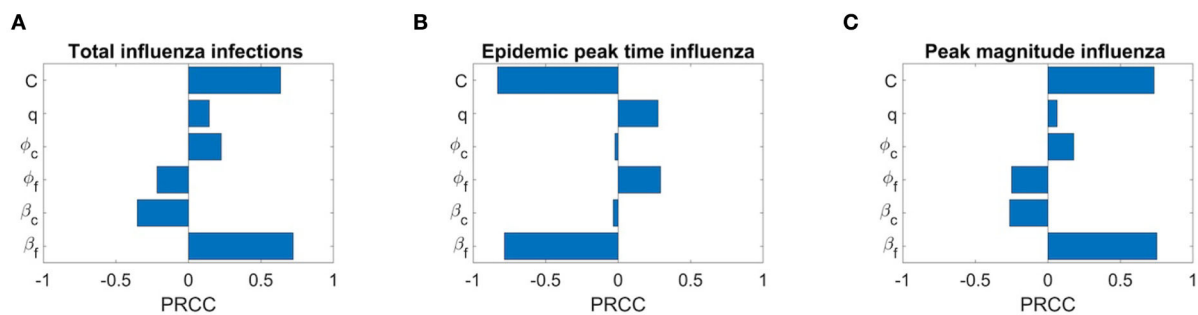
FIGURE 13
Sensitivity analysis of three epidemic outcomes of influenza. **(A–C)** Total number of influenza infections, epidemic peak time of influenza, and peak magnitude of influenza, respectively.

both COVID-19 and influenza. Mask-wearing coupled with a moderate increase in the vaccine uptake may mitigate COVID-19 and prevent an influenza outbreak.

However, there are some limitations in the present study. To keep this model parsimonious, we did not include latent or asymptomatic stages of infection. We believe including these stages would not overly affect our predictions due to the majority of COVID-19 and flu cases being symptomatic, and therefore requiring diagnosis (61). Future studies may be extended to add these states if needed. We did not consider outflow from the isolation class (before testing) to the recovered class because of the long delay in diagnosis. Although we did some simulations by incorporating it into our model (not included in the article), there is not much difference between these and our key findings of the given model in this study, except we see more delay in peak times of both diseases. And by increasing influenza vaccine coverage from 35 to 55%, there is a delay in influenza cases instead of a slight increase in influenza peak as in our original model. Also, with 55% influenza vaccine and 20% masking, there would be no outbreak of COVID-19 and influenza even if booster and fourth doses coverage are kept moderate.

Our study has important practical implications for public health policy: It shows that effectively managing and controlling both influenza and COVID-19 outbreaks during the same season depend on ensuring optimal strategies in terms of vaccine coverage.

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

BM, WW, and JW originated the research idea. The model formulation was carried out by BM. Numerical simulations were performed by BM with the help of JD and WW. The writing of the manuscript was completed by BM, reviewed and edited by JW, NB, and WW. The literature review was done by BM and NB. The project was supervised by JW, and the project was led by WW. All authors provided critical feedback and helped shape the research and analysis. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

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. Iezadi S, Gholipour K, Azami-Aghdash S, Ghiasi A, Rezapour A, Pourasghari H, et al. Effectiveness of non-pharmaceutical public health interventions against COVID-19: a systematic review and meta-analysis. *PLoS ONE*. (2021) 16:e0260371. doi: 10.1371/journal.pone.0260371

2. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys Rep*. (2021) 913:1–52. doi: 10.1016/j.physrep.2021.02.001

3. Feikin DR, Higdon MM, Abu-Raddad LJ, Andrews N, Araos R, Goldberg Y, et al. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression. *Lancet*. (2022) 399, 924–44. doi: 10.2139/ssrn.3961378

4. Scudellari M. How the pandemic might play out in 2021 and beyond. *Nature*. (2020) 584, 22–25. doi: 10.1038/d41586-020-02278-5

5. Araf Y, Akter F, Tang Yd, Fatemi R, Parvez MSA, Zheng C, et al. Omicron variant of SARS-CoV-2: genomics, transmissibility, and responses to current COVID-19 vaccines. *J Med Virol*. (2022) 94:1825–32. doi: 10.1002/jmv.27588

6. Zou Y, Huang D, Jiang Q, Guo Y, Chen C. The vaccine efficacy against the SARS-CoV-2 omicron: a systemic review and meta-analysis. *Front Public Health*. (2022) 10:940956. doi: 10.3389/fpubh.2022.940956

7. Buchan SA, Chung H, Brown KA, Austin PC, Fell DB, Gubbay JB, et al. Estimated effectiveness of COVID-19 vaccines against omicron or delta symptomatic infection and severe outcomes. *JAMA Netw Open*. (2022) 5:e2232760. doi: 10.1001/jamanetworkopen.2022.32760

8. Saha S, Tanmoy AM, Tanni AA, Goswami S, Al Sium SM, Saha S, et al. New waves, new variants, old inequity: a continuing COVID-19 crisis. *BMJ Global Health*. (2021) 6:e007031. doi: 10.1136/bmjgh-2021-007031

9. Antia R, Halloran ME. Transition to endemicity: understanding COVID-19. *Immunity*. (2021) 54:2172–6. doi: 10.1016/j.immuni.2021.09.019

10. Cong B, Deng S, Wang X, Li Y. The role of respiratory co-infection with influenza or respiratory syncytial virus in the clinical severity of COVID-19 patients: a systematic review and meta-analysis. *J Glob Health*. (2022) 12:05040. doi: 10.7189/jogh.12.05040

11. Ozaras R, Cirpin R, Duran A, Duman H, Arslan O, Bakcan Y, et al. Influenza and COVID-19 coinfection: report of six cases and review of the literature. *J Med Virol*. (2020) 92:2657–65. doi: 10.1002/jmv.26125

12. Huang QS, Wood T, Jelley L, Jennings T, Jefferies S, Daniells K, et al. Impact of the COVID-19 nonpharmaceutical interventions on influenza and other respiratory viral infections in New Zealand. *Nat Commun*. (2021) 12:1–7. doi: 10.1038/s41467-021-21157-9

13. Dhanasekaran V, Sullivan S, Edwards KM, Xie R, Khvorov A, Valkenburg SA, et al. Human seasonal influenza under COVID-19 and the potential consequences of influenza lineage elimination. *Nat Commun*. (2022) 13:1–11. doi: 10.1038/s41467-022-29402-5

14. Ehrlich H, Boneva D, Elkbuli A. The intersection of viral illnesses: a seasonal influenza epidemic amidst the COVID-19 pandemic. *Ann Med Surg*. (2020) 60:41–3. doi: 10.1016/j.amsu.2020.10.021

15. Opatowski L, Baguelin M, Eggo RM. Influenza interaction with cocirculating pathogens and its impact on surveillance, pathogenesis, and epidemic profile: a key role for mathematical modelling. *PLoS Pathog*. (2018) 14:e1006770. doi: 10.1371/journal.ppat.1006770

16. Zegarra MAA, Lopez MN, Cibrian MS, Garcia AC, Velasco-Hernandez JX. Co-circulation of SARS-CoV-2 and influenza under vaccination scenarios. *medRxiv*. (2021). doi: 10.1101/2020.12.29.20248953

17. Domnich A, Orsi A, Sticchi L, Panatto D, Dini G, Ferrari A, et al. Effect of the 2020/21 season influenza vaccine on SARS-CoV-2 infection in a cohort of Italian healthcare workers. *Vaccine*. (2022) 40:1755–60. doi: 10.1016/j.vaccine.2022.02.013

18. Tayar E, Abdeen S, Alah MA, Chemaitelly H, Bougmiza I, Ayoub H, et al. Effectiveness of influenza vaccination against SARS-CoV-2 infection among healthcare workers in Qatar. *medRxiv*. (2022). doi: 10.1101/2022.05.09.22274802

19. Hayawi K, Shahriar S, Serhani MA, Alashwal H, Masud MM. Vaccine versus Variants (3Vs): are the COVID-19 vaccines effective against the variants? A systematic review. *Vaccines*. (2021) 9:1305. doi: 10.3390/vaccines9111305

20. Chenchula S, Karunakaran P, Sharma S, Chavan M. Current evidence on efficacy of COVID-19 booster dose vaccination against the Omicron variant: a systematic review. *J Med Virol*. (2022) 94:2969–76. doi: 10.1002/jmv.27697

21. Grewal R, Kitchen SA, Nguyen L, Buchan SA, Wilson SE, Costa AP, et al. Effectiveness of a fourth dose of covid-19 mRNA vaccine against the omicron variant among long term care residents in Ontario, Canada: test negative design study. *BMJ*. (2022) 378. doi: 10.1101/2022.04.15.22273846

22. Yao L, Zhu KL, Jiang XL, Wang XJ, Zhan BD, Gao HX, et al. Omicron subvariants escape antibodies elicited by vaccination and BA. 2.2 infection. *Lancet Infect Dis*. (2022) 22:1116–7. doi: 10.1016/S1473-3099(22)00410-8

23. Government of Canada. *Testing for COVID-19: When to Get Tested and Testing Results*. (2022). Available online at: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/symptoms/testing/diagnosing.html#a4

24. Government of Canada. *COVID-19 Vaccination in Canada*. (2022). Available online at: https://health-infobase.canada.ca/covid-19/vaccination-coverage/

25. Public Health Canada. *Seasonal Influenza (Flu) Vaccination Coverage Survey Results, 2019-2020*. (2022). Available online at: https://www.canada.ca/en/public-health/services/publications/vaccines-immunization/2019-2020-seasonal-influenza-flu-vaccine-coverage-survey-results.html (accessed August 26, 2022).

26. Andrew MK, Shinde V, Hatchette T, Ambrose A, Boivin G, Bowie W, et al. Influenza vaccine effectiveness against influenza-related hospitalization during a season with mixed outbreaks of four influenza viruses: a test-negative case-control study in adults in Canada. *BMC Infect Dis*. (2017) 17:805. doi: 10.1186/s12879-017-2905-8

27. Nichols MK, Andrew MK, Hatchette TF, Ambrose A, Boivin G, Bowie W, et al. Influenza vaccine effectiveness to prevent influenza-related hospitalizations and serious outcomes in Canadian adults over the 2011/12 through 2013/14 influenza seasons: a pooled analysis from the Canadian Immunization Research Network (CIRN) Serious Outcomes Surveillance (SOS Network). *Vaccine*. (2018) 36:2166–75. doi: 10.1016/j.vaccine.2018.02.093

28. David J, Bragazzi NL, Scarabel F, McCarthy Z, Wu J. Non-pharmaceutical intervention levels to reduce the COVID-19 attack ratio among children. *R Soc Open Sci*. (2022) 9:211863. doi: 10.1098/rsos.211863

29. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science*. (2021) 372:eabg3055. doi: 10.1126/science.abg3055

30. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England. *Nature*. (2021) 593:266–9. doi: 10.1038/s41586-021-03470-x

31. Canadian Institute for Health Information. *COVID-19 Hospitalization and Emergency Department Statistics, 2021-2022*. (2022). Available online at: https://www.cihi.ca/en/covid-19-hospitalization-and-emergency-department-statistics (accessed: July 25, 2022).

32. McCarthy Z, Athar S, Alavinejad M, Chow C, Moyles I, Nah K, et al. Quantifying the annual incidence and underestimation of seasonal influenza: a modelling approach. *Theoret. Biol. Med. Modell*. (2020) 17:1–16. doi: 10.1186/s12976-020-00129-4

33. Tang B, Scarabel F, Bragazzi NL, McCarthy Z, Glazer M, Xiao Y, et al. De-escalation by reversing the escalation with a stronger synergistic package of contact tracing, quarantine, isolation and personal protection: feasibility of preventing a COVID-19 rebound in Ontario, Canada, as a case study. *Biology*. (2020) 9:100. doi: 10.3390/biology9050100

34. Buchan SA, Chung H, Brown KA, Austin PC, Fell DB, Gubbay J, et al. Effectiveness of COVID-19 vaccines against Omicron or Delta infection. *medRxiv*. (2022) 2021–12. doi: 10.1101/2021.12.30.21268565

35. Nah K, Alavinejad M, Rahman A, Heffernan JM, Wu J. Impact of influenza vaccine-modified infectivity on attack rate, case fatality ratio and mortality. *J Theoret Biol.* (2020) 492:110190. doi: 10.1016/j.jtbi.2020.110190

36. Center for Disease Control and Prevention (CDC). *Vaccine Effectiveness: How Well Do Flu Vaccines Work*? (2022). Available online at: https://www.cdc.gov/flu/vaccines-work/vaccineeffect.htm (accessed August 23, 2022).

37. Govermment of Canada. *Testing for COVID-19.* (2022). Available online at: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/symptoms/testing/diagnosing.html (accessed August 5, 2022).

38. Public Health Ontario. *Ontario COVID-19 Data Tool.* (2022). Available online at: https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool?tab=overview (accessed August 17, 2022).

39. Public Health Ontario. *COVID-19 Vaccine Data in Ontario-Hospitalizations by Vaccination Status.* (2022). Available online at: https://data2.ontario.ca/en/dataset/covid-19-vaccine-data-in-ontario/resource/274b819c-5d69-4539-a4db-f2950794138c (accessed July 7, 2022).

40. Kalligeros M, Shehadeh F, Mylona EK, Dapaah-Afriyie C, van Aalst R, Chit A, et al. Influenza vaccine effectiveness against influenza-associated hospitalization in children: a systematic review and meta-analysis. *Vaccine.* (2020) 38:2893–903. doi: 10.1016/j.vaccine.2020.02.049

41. Paget J, Spreeuwenberg P, Charu V, Taylor RJ, Iuliano AD, Bresee J, et al. Global mortality associated with seasonal influenza epidemics: new burden estimates and predictors from the GLaMOR Project. *J Glob Health.* (2019) 9. doi: 10.7189/jogh.09.020421

42. World Health Organization. *Influenza (Seasonal).* World Health Organization (2022). Available online at: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) (accessed August 15, 2022).

43. Lee SS, Viboud C, Petersen E. Understanding the rebound of influenza in the post COVID-19 pandemic period holds important clues for epidemiology and control. *Int J Infect Dis.* (2022) 122:1002–4. doi: 10.1016/j.ijid.2022.08.002

44. Lei H, Xu M, Wang X, Xie Y, Du X, Chen T, et al. Nonpharmaceutical interventions used to control COVID-19 reduced seasonal influenza transmission in China. *J Infect Dis.* (2020) 222:1780–3. doi: 10.1093/infdis/jiaa570

45. Groves HE, Piché-Renaud PP, Peci A, Farrar DS, Buckrell S, Bancej C, et al. The impact of the COVID-19 pandemic on influenza, respiratory syncytial virus, and other seasonal respiratory virus circulation in Canada: a population-based study. *Lancet Region Health Am.* (2021) 1:100015. doi: 10.1016/j.lana.2021.100015

46. Nwosu A, Lee L, Schmidt K, Buckrell S, Sevenhuysen C, Bancej C. Influenza vaccine: National Influenza Annual Report, Canada, 2020-2021, in the global context. *Canada Commun Dis Rep.* (2021) 47:405. doi: 10.14745/ccdr.v47i10a02

47. Olsen SJ, Azziz-Baumgartner E, Budd AP, Brammer L, Sullivan S, Pineda RF, et al. Decreased influenza activity during the COVID-19 pandemic–United States, Australia, Chile, and South Africa, 2020. *Am J Transpl.* (2020) 20:3681–5. doi: 10.1111/ajt.16381

48. Itaya T, Furuse Y, Jindai K. Does COVID-19 infection impact on the trend of seasonal influenza infection? 11 countries and regions, from 2014 to 2020. *Int J Infect Dis.* (2020) 97:78–80. doi: 10.1016/j.ijid.2020.05.088

49. Kim J, Gómez REG, Hong K, Yum S, Jang J, Chun BC. Changing influenza activity in the Southern hemisphere countries during the COVID-19 pandemic. *Int J Infect Dis.* (2021) 108:109–11. doi: 10.1016/j.ijid.2021.05.039

50. Australian Department of Health and Aged Care. *Australian Influenza Surveillance Report and Activity Updates.* (2022). Available online at: https://www1.health.gov.au/internet/main/publishing.nsf/Content/ozflu-surveil-no06-22.html (accessed July 5, 2022).

51. Public Health Agency of Canada (PHAC). *Influenza Weekly Reports 2021-22 Season.* Ottawa, ON: PHAC (2022). Available online at: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/fluwatch/2021-2022/weeks-16-18-april-17-may-7-2022.html (accessed July 5, 2022).

52. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis.* (2020) 20:e238–44. doi: 10.1016/S1473-3099(20)30484-9

53. Zhou X, Wang Z, Li S, Liu T, Wang X, Xia J, et al. Machine learning-based decision model to distinguish between COVID-19 and influenza: a retrospective, two-centered, diagnostic study. *Risk Manage Healthcare Policy.* (2021) 14:595. doi: 10.2147/RMHP.S291498

54. Farhang-Sardroodi S, Ghaemi M, Craig M, Ooi HK, Heffernan JM. A machine learning approach to differentiate between COVID-19 and influenza infection using synthetic infection and immune response data. *medRxiv.* (2022). doi: 10.1101/2022.01.27.22269978

55. CDC. *Seasonal Flu Vaccine Effectiveness Studies.* Center for Disease Control and Prevention (2022). Available online at: https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm (accessed August 15, 2022).

56. Segaloff HE, Leventer-Roberts M, Riesel D, Malosh RE, Feldman BS, Shemer-Avni Y, et al. Influenza vaccine effectiveness against hospitalization in fully and partially vaccinated children in Israel: 2015-2016, 2016-2017, and 2017-2018. *Clin Infect Dis.* (2019) 69:2153–61. doi: 10.1093/cid/ciz125

57. Arriola CS, El Omeiri N, Azziz-Baumgartner E, Thompson MG, Sotomayor-Proschle V, Fasce RA, et al. Influenza vaccine effectiveness against hospitalizations in children and older adults–data from South America, 2013-2017. A test negative design. *Vaccine X.* (2019) 3:100047. doi: 10.1016/j.jvacx.2019.100047

58. de Cellés MD, Casalegno JS, Lina B, Opatowski L. Influenza may facilitate the spread of SARS-CoV-2. *medRxiv.* (2020). doi: 10.1101/2020.09.07.20189779

59. Del Riccio M, Lorini C, Bonaccorsi G, Paget J, Caini S. The association between influenza vaccination and the risk of SARS-CoV-2 infection, severe illness, and death: a systematic review of the literature. *Int J Environ Res Public Health.* (2020) 17:7870. doi: 10.3390/ijerph17217870

60. Larrauri A, Trilar KP. Preparing for an influenza season 2021/22 with a likely co-circulation of influenza virus and SARS-CoV-2. *Eurosurveillance.* (2021) 26:2100975. doi: 10.2807/1560-7917.ES.2021.26.41.2100975

61. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. *J Med Virol.* (2021) 93:820–30. doi: 10.1002/jmv.26326

Check for updates

Investigating the spatio-temporal variation of hepatitis A in Korea using a Bayesian model

Jaehong Jeong[1,2], Mijeong Kim[3]* and Jungsoon Choi[1,2]*

[1]Department of Mathematics, Hanyang University, Seoul, Republic of Korea, [2]Research Institute for Natural Sciences, Hanyang University, Seoul, Republic of Korea, [3]Department of Statistics, Ewha Womans University, Seoul, Republic of Korea

Hepatitis A is a water-borne infectious disease that frequently occurs in unsanitary environments. However, paradoxically, those who have spent their infancy in a sanitary environment are more susceptible to hepatitis A because they do not have the opportunity to acquire natural immunity. In Korea, hepatitis A is prevalent because of the distribution of uncooked seafood, especially during hot and humid summers. In general, the transmission of hepatitis A is known to be dynamically affected by socioeconomic, environmental, and weather-related factors and is heterogeneous in time and space. In this study, we aimed to investigate the spatio-temporal variation of hepatitis A and the effects of socioeconomic and weather-related factors in Korea using a flexible spatio-temporal model. We propose a Bayesian Poisson regression model coupled with spatio-temporal variability to estimate the effects of risk factors. We used weekly hepatitis A incidence data across 250 districts in Korea from 2016 to 2019. We found spatial and temporal autocorrelations of hepatitis A indicating that the spatial distribution of hepatitis A varied dynamically over time. From the estimation results, we noticed that the districts with large proportions of males and foreigners correspond to higher incidences. The average temperature was positively correlated with the incidence, which is in agreement with other studies showing that the incidences in Korea are noticeable in spring and summer due to the increased outdoor activity and intake of stale seafood. To the best of our knowledge, this study is the first to suggest a spatio-temporal model for hepatitis A across the entirety of Korean. The proposed model could be useful for predicting, preventing, and controlling the spread of hepatitis A.

KEYWORDS

hepatitis A virus, spatio-temporal analysis, spatio-temporal models, zero-inflated Poisson, Bayesian hierarchical modeling, Korea

1. Introduction

Unlike Hepatitis B or C, Hepatitis A virus (HAV) is not transmitted through blood, but by consuming food or water contaminated with HAV or by contact with an infected person (1). World Health Organization (1) reported that the number of deaths from HAV in 2015 was estimated to be about 11,000 worldwide, accounting for 0.8% of deaths from viral infections. A case-control study of the HAV outbreak in Shanghai in 1988 supported that clams were a carrier of the virus (2). In Korea, HAV is often reported to spread through shellfish consumption (3). In India, a case-control study showed a high association between pipe water contamination and HAV infection (4). It particularly occurs in underdeveloped areas where personal hygiene management is poor, and HAV infection cases are decreasing in countries with an improvement in socioeconomic level, clean water management system, and HAV vaccination (5). However, the incidence rate has recently increased rapidly, in young adults who grew up in a hygienic environment in Korea (6).

Several studies have investigated the effects of socioeconomic and epidemiological factors such as age, medical level, and hygiene level on HAV in various countries. For example, it was noted that a significantly lower rate of HAV infection in people coupled with moderate to high socioeconomic conditions in Brazil, Argentina, and Mexico in Tapia-Conyer et al. (7). Mantovani et al. (8) discovered that a region with a high incidence of HAV had a weak socioeconomic condition in Brazil, thus emphasizing the need for hygiene improvement and better water treatment in the western Brazilian Amazon to reduce infectious disease outbreaks. Dogru et al. (9) classified children under the age of 15 in Turkey into three categories and analyzed the spatial patterns of HAV occurrence. In Turkey, the incidence of HAV was reported to be high in areas where water and sewage facilities are not well equipped. Copado-Villagrana et al. (10) pointed out that HAV infections were mainly found in the metropolitan areas of southern and western Mexico, noting that it may be associated with poor medical services in the most marginalized areas. Zheng et al. (11) characterized changes in the incidence and mortality of HAV in various age groups and regions in China from 1990 to 2018 and evaluated the effectiveness of the nationwide expanded program on immunization. The spread of the disease was decreased by expanding vaccinations and improving hygiene facilities. Shanmugam et al. (12) showed that primary infection of HAV among the older age group in India has recently decreased with improved living conditions.

Weather-related variables are associated with the incidence of HAV. According to Cann et al. (13), in extreme weather conditions such as hurricanes, cross-contamination of water supply and sewage may affect the transmission of waterborne diseases. In Brazil, cases of HAV infection increased during the rainy season (14). In state of Pará, Brazil, monthly accumulated precipitation was positively correlated with the incidence of HAV (15). Tosepu (16) found a strong relationship between HAV and weather change, particularly rainfall and floods, in several areas, such as Spain, India, China, and Egypt. Baek et al. (17) showed that weekly precipitation and maximum temperature tended to decrease the incidence rate ratio of HAV in Seoul, Korea from the analysis of time series with past 1–6 week lags. In Seoul, the capital of South Korea, nearly 100% of households receive sterilized tap water and most citizens drink purified or clean spring water. In addition, since Seoul is geographically less affected by typhoons and floods, HAV is not likely to be transmitted due to cross-contamination of water supply and sewerage due to heavy seasonal rains as in other regions (17). In this respect, HAV incidence seems to be related to weather conditions, but the pattern may differ across regions. Fares (18) pointed out that some specific months are associated with a higher incidence of HAV in most countries around the world, but the exact reason for the seasonality of HAV has yet to be known. Several researchers have suggested climatic and behavioral factors such as swimming and traveling may play an important role in seasonal disease incidence (18). Moon et al. (19), based on the dataset in Korea from 2011 to 2013, reported that most HAV cases occurred from March to June. People have more outdoor activities as the weather becomes warmer during this period; thus, they are at risk of being exposed to tainted drinking water and uncooked seafood. Both are well-known risk factors for HAV infections (20). Thus, weather-related variables such as temperature and precipitation should be considered when studying HAV occurrences because weather conditions can affect people's behavior.

Various spatio-temporal analyses have been conducted to explore and understand the risk of HAV in terms of spatial and temporal structures. Gomez-Barroso et al. (21) analyzed the space-time risk of HAV using standardized incidence ratios (SIR; the ratio between actual and expected cases) and the posterior probability of the smoothed relative risk (RR; the ratio of the outcome probability for the exposed group to the probability for the unexposed group) in Spain at the municipal level from 1997 to 2007. Stoitsova et al. (22) applied the Global Moran's I index for spatial autocorrelation across Bulgaria concerning the risk of HAV infection and SIR across the nation for the whole period from 2003 to 2013 and two divided periods (2003–2008 and 2008–2013). Leal et al. (23) explored the spatio-temporal patterns of HAV outbreaks before (2008–2013) and after (2014–2017) the implementation of the national public immunization program in Pará State, a region of Brazil with severe endemic disease. Space-time scan statistics were applied to detect spatio-temporal clusters. Moreover, Leal et al. (15) investigated the association between environmental and socio-demographic data in HAV transmission in Pará State, Brazil, using various models, including generalized linear

models, multilayer perceptron (MPL) deep-learning algorithm, gradient boost, decision tree, and histogram gradient boost (HGB). To reflect the spatial variation, the longitude and latitude of each municipality were used as covariates in the model.

As discussed above, HAV is related to many factors, such as socioeconomic and epidemiological factors, weather-related factors, and spatio-temporal variations; therefore, referring to the status of diseases in other countries is not enough. For a better understanding of the HAV of the country, we should consider not only its local and national characteristics but also its social and hygienic situation.

With rapid urbanization, Korea has become cleaner. Since the first sewage treatment plant was established in Korea in 1976 (24), the number of sewage treatment plants has gradually increased, and accordingly, the number of HAV infections caused by contaminated water has decreased rapidly. As of 2020, the water supply rate is 97% nationwide, 100% in Seoul and 45.7% in Cheongyang-gun, Chungcheongnam-do. In the same year, the sewage supply rate was 94.5% nationwide, 100% in Seoul, and 5.4% in Ulleung-gun, Gyeongsangbuk-do. Although Korea's large cities have well-equipped water and sewage facilities, some areas of rural and fishing villages are in poor condition. Children with acute HAV are asymptomatic or mildly symptomatic, and antibodies (IgG anti-HAV) develop, resulting in lifelong immunity (6). According to Yoon et al. (6), IgG anti-HAV seropositivity in the Korean young adult population was low and a clean environment may lead to a decrease in the natural immune system population. Since the HAV vaccination began in 1997, the vaccination rate for 3-year-old infants in Korea exceeded 95% in 2019, according to the Korea Disease Control and Prevention Agency. People without antibodies are at an increased risk of exposure to HAV during active adolescence because they do not contract HAV as a child. When infected as an adult, the immune response is severe and symptoms such as jaundice appear; in severe cases, acute liver failure can lead to death (6). People born after 1980, when the environment was cleaner and the HAV vaccine was yet developed, have been reported to be more susceptible to HAV (25, 26). Other similar studies on the seroprevalence of HAV antibodies have been reported steadily over time (27–30).

Some efforts have been made to investigate and understand the status of HAV in Korea using statistical approaches. Research on the frequency analysis of the number of HAV infections by year, region, and age have been steadily reported (19, 31–33). Moon et al. (19) studied the epidemiological status of HAV cases in Korea between 2011 and 2013. They described significant differences in the incidence of HAV between months, regions, sexes, and age groups. In particular, they classified regions into five clusters according to the RR. Using Moran's I and scan statistics, they found clear and existing regional differences in the incidence of HAV; however, their approach

was limited to exploratory data analysis. In addition, RR does not always show correct risks (34). Seo et al. (25) studied the effect of socioeconomic status and environmental hygiene by region on the incidence of HAV based on the registered national population of Korea and national health insurance data from 2004 to 2008 using a Poisson regression model. Choi (35) conducted spatial hotspot detection of monthly incidences of HAV using spatial scan statistics and investigated the effects of socioeconomic factors using the Bayesian spatial Poisson regression model. Even though the HAV data were monthly, the proposed spatial model considered yearly incidence data and was independently applied. Thus, this study did not consider temporal and spatial variations for consecutive periods. Choi (36) analyzed HAV incidence data from 2007 to 2012 in Korea using a Bayesian spatio-temporal Poisson regression model. In Korea, HAV occurs more in summer and less in winter; therefore, seasonal factors are reflected in the model using a sine/cosine function. However, this study did not consider socioeconomic or weather factors. Baek et al. (17) conducted a time series analysis to explain the influence of factors, such as temperature and precipitation, on the incidence of HAV in Seoul, Korea. By minimizing the influence of other factors and limiting the study region to a place with a similar lifestyle, they could explain the association between weather-related factors and HAV incidence. However, it is questionable whether the same result can hold for other regions rather than Seoul in Korea. Also, further investigation of the association between HAV cases and variables other than weather is required.

In this study, we aimed to investigate the spatio-temporal variation of HAV in Korea and effects of socioeconomic and weather-related factors using a flexible spatio-temporal model. We proposed a Bayesian spatio-temporal zero-inflated Poisson regression model of weekly HAV incidence in Korea to estimate the effects of risk factors. This study is the first to develop a spatio-temporal model of HAV incidence across the entirety of Korean, with various socioeconomic factors. The advantage of this study is that the proposed model could be useful in predicting, preventing, and controlling the spread of hepatitis A.

2. Materials and methods

2.1. Description of data

We considered HAV cases as a response variable and socioeconomic, environmental, and weather-related factors as explanatory variables. The dataset for 250 nationwide districts from 2016 to 2019 was obtained from the Korea Disease Control and Prevention Agency. The weekly HAV cases at the district level (called si/gun/gu) had many zero counts (73.5%), suggesting zero-inflated statistical modeling.

We considered income, education level, and fertility rate as socioeconomic factors, because they affect the quality of life. The income variable is defined as the average monthly income per person at the district level. We calculated the high education rate as the proportion of educational attainment of a university degree or higher among the population in their 20s or an older age. The fertility rate was obtained from the actual fertility rate of women 15–49 years in Statistics Korea. We also considered the male proportion because people with active social activities are more likely to be exposed to HAV, and previous studies, including Moon et al. (19), found that the infection occurred more often among men than that among women. Since people born around the 1980s in Korea have a weak tendency toward HAV immunity (26), we considered the age group of 30–49 years. Jacobsen (20) mentioned that diverse epidemiological profiles should be treated as risk factors for HAV, and the number of registered foreigners was used for analyzes. Based on the result of Choi (35), we included the number of medical doctors per 1,000 people as a factor.

For environmental factors, we considered the water supply and sewage treatment facility rates obtained annually from Statistics Korea. Water supply plays an important role not only in terms of health and sanitation but also in industry and firefighting. Waterworks are essential for daily life, but the water system capacity varies depending on several conditions, such as the residential environment (including the housing structure) and the financial status of the local government. Therefore, the water supply rate can be used to evaluate the quality of the local living environment. According to statistics published by the Ministry of Environment of Korea, the water supply rate in Korea increased from 80.1% in 1991 to 97.3% in 2019 due to the government's continuous infrastructure expansion. However, the water supply rate varied between regions. For example, Seoul and Daegu reached 100%, but Jeju Island did not even reach 90%. The gap between the city and rural (and small town) areas is sufficiently large to be ignored. The public sewage treatment facility rate for the population is the ratio of the population beneficiaries of public sewerage services, and the closer it is to 100%, the higher the ratio of the population beneficiaries of public sewage services.

The three weather variables of interest, average temperature, total precipitation, and average humidity, were obtained from the Korea Meteorological Administration. The weather datasets were measured using two systems, an automatic weather system (AWS) and automated synoptic observing system (ASOS), at distinct weather stations (up to 510 and 103 stations, respectively). These measured values should be located in the same spatial domain as other factors; thus, we need to fit a surface to irregularly spaced weather values. Here, we combined the datasets from the two systems and subsequently predicted the weather values at each time and location of interest. To achieve this goal, we used the Kriging model based on a Gaussian

TABLE 1 Description of data set (outcome, socioeconomic factors, environmental factors, and weather-related factors).

Variable type	Variable	Time unit
Outcome	HAV cases	Weekly
Socioeconomic factor	Total income per person (1 million won)	Yearly
	High education rate (%)	2015
	Total fertility rate per woman aged 15–49 years	Yearly
	Proportion of males (%)	Yearly
	Proportion of people aged 30–49 years (%)	Yearly
	Log(number of foreigners)	Yearly
	Number of doctors per thousand people	Yearly
Environmental factor	Water supply rate (%)	Yearly
	Sewage treatment facility rate (%)	Yearly
Weather-related factor	Average temperature (°C)	Weekly
	Total precipitation (mm)	Weekly
	Average humidity (%)	Weekly

spatial random process with a Matérn covariance function at a given time (37).

We report a list of factors for this study in Table 1 and their data sources in Supplementary Table S1. Here, HAV cases and weather-related factors were measured weekly, but socioeconomic and environmental factors were collected yearly. For the socioeconomic and environmental factors, we obtained datasets from different statistics from various institutions. However, we can conveniently access all statistics through the Korean Statistical Information Service (KOSIS).

2.2. Spatial and temporal association measures

First, weekly HAV cases in a given district are now becoming time series data. Thus, we considered the autocorrelation function (ACF) and partial autocorrelation function (PACF) to examine the temporal association. The ACF corresponds to the correlation between the time series with a lagged version of itself, whereas the PACF measures the additional correlation explained by each successive lagged term. Although these two functions are slightly different, they are both measures of the association between current and past series values. For more details on the ACF and PACF, we refer to Brockwell and Davis (38).

In contrast, HAV cases in districts given at a time are areal data. In this case, we used Moran's I index measuring the strength of spatial associations among districts [39] to examine the spatial association. It is a spatial analog of the measure of association in a time series. In addition, Moran's I index explores a specific type of spatial clustering [40]. The proximity matrix consists of weights that spatially connect two districts in a certain manner. Here, two districts closer to one another have more weight than those farther away. Moran's I index coupled with the proximity matrix can be interpreted as follows: a negative value corresponds to some clustering of dissimilar values, a zero value corresponds to perfect randomness, and a positive value indicates some clustering of similar values.

All the analyzes were performed using R software (version 4.1.0; https://www.r-project.org). We used "ape" [41] and "fields" [42] packages to compute the distances among districts and the global Moran's I index.

2.3. Statistical model

A Bayesian space-time regression model was developed to investigate the association between socioeconomic, environmental, and weather-related factors and HAV cases and to account for the space-time-dependent structures in the data. As there were many zero values in the weekly district-level HAV cases data, we used a zero-inflated Poisson (ZIP) distribution. Moreover, we considered the two-stage framework proposed by Lawson et al. [43] to overcome the spatial confounding bias problem.

In the first stage, the number of cases for district area s ($=$ 1, 2, $\cdots$, S) and weekly time index t ($=$ 1, 2, $\cdots$, T), $y_{s,t}$, follows a zero-inflated Poisson distribution as follows:

$$y_{s,t} \sim \text{ZIP}(p_{s,t}, \lambda_{s,t}),$$

where $p_{s,t}$ is the probability of structural zeros and $\lambda_{s,t}$ is the mean term of the Poisson distribution without structural zeros. The hierarchical structure of the ZIP model can be expressed as

$$y_{s,t}|z_{s,t} \sim \text{Poisson}(\mu_{s,t} = \lambda_{s,t}(1 - z_{s,t})),$$
$$z_{s,t} \sim \text{Bernoulli}(p_{s,t}),$$
$$\text{logit}(p_{s,t}) = \gamma_0 + \mathbf{X}_{s,j(t)}^T \boldsymbol{\gamma} + \mathbf{W}_{s,t}^T \boldsymbol{\delta},$$

where $j(t)$ is the yearly time index for socioeconomic and environmental factors and $z_{s,t}$ is a binary variable with probability $p_{s,t}$, representing whether it is a structural zero or not. The $\text{logit}(p_{s,t})$ is the linear combination of the intercept γ_0, and the fixed socioeconomic and environmental factors $\mathbf{X}_{s,j(t)}$ and weather-related factors $\mathbf{W}_{s,t}$ with the corresponding coefficient vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, respectively. The log RR, $\log(\lambda_{s,t})$, is

modeled using fixed factors with the corresponding coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

$$\log(\lambda_{s,t}) = \beta_0 + \mathbf{X}_{s,j(t)}^T \boldsymbol{\beta} + \mathbf{W}_{s,t}^T \boldsymbol{\alpha} + \log(N_{s,j(t)}), \quad (1)$$

where β_0 is the intercept and $N_{s,j(t)}$ is the population density as the off-set. The model of the first-stage only considers fixed factors without spatio-temporal variations.

After fitting the first-stage model using a Bayesian approach, the estimates $\hat{\mu}_{s,t}$ were computed using the posterior means. Continuous-type residuals were calculated as follows:

$$\hat{r}_{s,t} = \log(y_{s,t} + 0.1) - \log(\hat{\mu}_{s,t}),$$

where an extra value of 0.1 is added in the residual calculation because of the zero values of $y_{s,t}$.

In the second-stage, the residuals are modeled to explain the spatio-temporal variations over the first-stage covariates-only model.

$$\hat{r}_{s,t} \sim \text{N}(ST_{s,t}, \sigma_r^2),$$
$$ST_{s,t} = r_0 + u_s + v_s + \eta_t + \tau_t, \quad (2)$$

where $r_0 \sim \text{N}(0, 100)$ is the intercept. The spatial random component $u_s \sim \text{N}(0, \sigma_u^2)$ explains the spatially uncorrelated structures, and v_s explains the spatially-correlated structures with conditional intrinsic auto-regressive (CIAR) distribution from Besag et al. [44], $v_s \sim \text{CIAR}(\sigma_v^2)$. The random components $\eta_t \sim \text{N}(0, \sigma_\eta^2)$ and $\tau_t \sim \text{N}(\tau_{t-1}, \sigma_\tau^2)$ explain the temporal-uncorrelated and temporal-correlated structures, respectively. After fitting the residual model with a Bayesian approach, the estimated means of the spatio-temporal structures, $\widehat{ST}_{s,t}$, were obtained. We incorporated the estimated spatio-temporal variations into the fixed covariates. The final model is expressed as follows:

$$\log(\lambda_{s,t}) = \beta_0 + \mathbf{X}_{s,j(t)}^T \boldsymbol{\beta} + \mathbf{W}_{s,t}^T \boldsymbol{\alpha} + \log(N_{s,j(t)}) + \widehat{ST}_{s,t} + \epsilon_{s,t}, \quad (3)$$

where $\epsilon_{s,t} \sim \text{N}(0, \sigma_\epsilon^2)$ is the uncorrelated space-time random component that is not explained by the estimated space-time structure. Finally, the restricted ZIP regression model was fitted using a Bayesian approach to obtain the final estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

For the parameter estimation, we use non-informative priors, Normal(0, 100), for the coefficient parameters β_0, γ_0, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$. The standard deviations σ_u, σ_v, σ_η, and σ_τ are assigned to a uniform distribution, Uniform (0, 10). The NIMBLE package developed by de Valpine et al. [45] in the statistical software R was used to produce posterior samples. After discarding samples as a burn-in, 5,000 posterior samples with thin 50 were collected. Codes for the models can be found at https://github.com/JungsoonChoi/STmodeling_HepA.git.

TABLE 2 Descriptive statistical analysis.

Variable	Mean	SD	Q1	Q3	IQR
Total income per person	33.68	6.25	29.45	36.40	6.95
High education rate	31.12	11.19	21.70	38.12	16.42
Total fertility rate	1.14	0.27	0.95	1.30	0.35
Proportion of males	50.09	1.29	49.26	50.75	1.49
Proportion of people aged 30–49 years	27.66	5.22	23.14	31.77	8.63
Log(number of foreigners)	7.83	1.22	6.88	8.68	1.80
Number of doctors per thousand people	2.69	2.25	1.70	2.80	1.10
Water supply rate	81.05	30.39	78.97	99.90	20.93
Sewage treatment facility rate	84.69	17.25	74.18	99.50	25.32
Average temperature	12.82	9.69	4.32	21.05	16.73
Total precipitation	21.45	38.02	0.39	24.77	24.38
Average humidity	68.46	10.63	61.01	76.66	15.65

3. Results

Table 2 presents a summary of the statistical analysis of all the variables. The proportion of men had a 1.49% interquartile range (IQR = Q3-Q1), and the proportion of people aged 30–49 years had an IQR of 8.63%. Among the socioeconomic factors, these variables showed relatively smaller variations over districts and years. The average temperature and total precipitation had larger standard deviation (SD) values than that of the mean values. Water supply and sewage treatment facility rates have high mean values of 81.05 and 84.69, respectively. Supplementary Figure S1 shows the spatial variation of the average socioeconomic and environmental factors for 2016–2019. Supplementary Table S2 presents the number of HAV cases per 1,000 people divided into five groups for each socioeconomic and environmental factor. For the level of high education, the average number of HAV cases per 1,000 people 0.470 in the G1 districts and 0.615 in the G5 districts, respectively.

3.1. Spatial and temporal distributions of HAV

Figure 1A represents the weekly number of HAV cases from 2016 to 2019. Time series plots of weekly HAV cases in selected districts, Seoul-si Jongno-gu, Busan-si Sasang-gu, Daejeon-si Seo-gu, and Gyeonggi-do Bucheon-si, are shown in Figure 1B. We found that the temporal distribution of weekly HAV cases varied across districts. Figure 1C compares the

temporal variation in the cases from 2016 to 2019. In 2016, the number of HAV cases was high in the spring, but in 2019, it was high in the summer. There was little change in the number of HAV cases in 2018 compared with that of the other years.

Figures 2A–D illustrate the number of cases per 1,000 people in 2016, 2017, 2018, and 2019, respectively. Overall, the central region of Korea, the Seoul metropolitan area, had a higher number of cases than that in the other regions. The Chungcheong Province, which is close to the Seoul metropolitan region, had a greater number of cases per 1,000 people than that in the other regions.

Figure 2 shows the existing connections of observations between the different districts. The more correlated the cases, the closer the districts were to the map. Table 3 presents the global Moran's I values for each year. Here, we calculated the distance-based global Moran's I value and found a positive correlation in space. Figure 3 presents the time series plot of weekly Moran's I and its P-values. We also found spatial dependence of HAV cases at most time points.

Figure 4 shows the ACF and PACF for all districts and each selected district. They are important tools in the exploratory data analysis of time series and, in particular, help to understand the correlation between observations at different time points. It is evident that the time series values at Daejeon-si Seo-gu and Gyeonggi-do Bucheon-si are related to their past values. Thus, considering temporal dependencies when modeling HAV cases is required. Figures 3, 4 show that we must not ignore not only temporal dependence but also spatial dependence. Thus, we included both spatial- and temporal-dependent structures in the model for a better fit.

3.2. Bayesian spatio-temporal model

We evaluated the performance of the proposed model, with several competing models as follows: Poisson and ZIP distributions were considered. Type 1 included only fixed factors without a random component. Type 2 included both fixed factors and spatio-temporal random components $ST_{s,t}$ on the right side of (2). Type 3 also includes fixed factors and spatio-temporal random components within the two-stage framework. The proposed model was a ZIP model with type 3.

Table 4 presents the comparison results of the models, with the mean absolute error (MAE), mean squared prediction Error (MSPE), and deviance information criterion (DIC=Dbar+pD) (46). In general, smaller values of the model fit criterion indicate a better model than that of the competitors. The proposed model (ZIP with type 3) had a smaller MAE and MSPE than that of the other models. Overall, the ZIP models provide slightly better performance than that of the Poisson models in terms of MAE, MSPE, and DIC. The ZIP model with type 2 has a slightly smaller DIC than that of the proposed model. However, the parameter estimation result of a ZIP
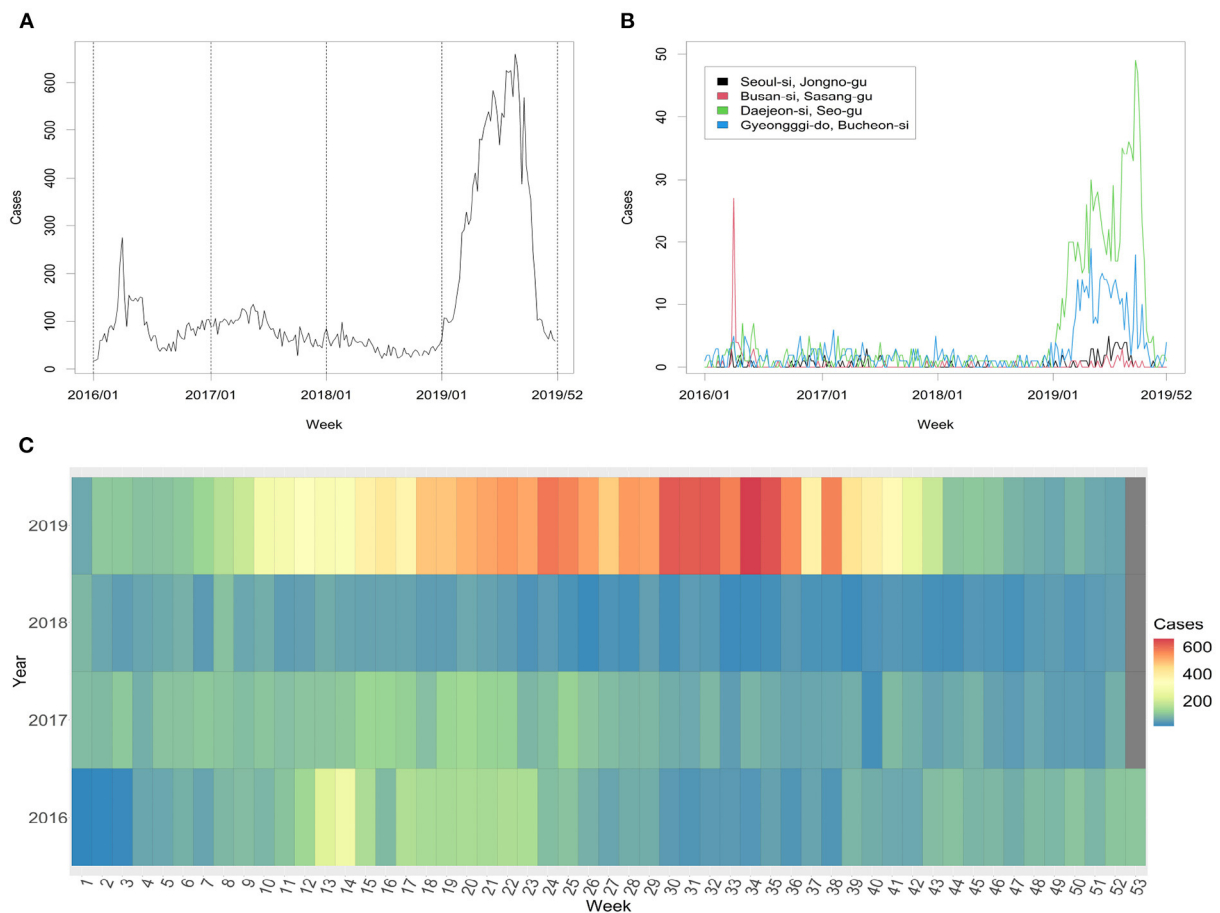
FIGURE 1
(A) Number of HAV cases from 2016 to 2019. **(B)** Weekly HAV cases at selected districts. **(C)** Temporal distribution of weekly total cases from 2016 to 2019.

with the ST model suffered from the spatial confounding bias issue, providing many insignificant coefficients due to the spatio-temporal random components. Thus, we preferred the proposed model in terms of model fit and better interpretation.

All the variables considered in our proposed model were statistically significant at a significance level of 0.05. The parameter estimates are presented in Table 5. Total income per person, high education rate, total fertility rate, the proportion of males, and the number of foreigners were positively associated with the number of HAV cases. However, the proportion of people aged 30–49 years and the number of doctors per 1,000 people were negatively associated with the number of cases. In addition, we found a negative association between environmental factors, including water supply rate, sewage treatment facility rate, and cases. This indicates that the higher the water quality of the environment, the lower the HAV incidence rate. For weather-related factors, the coefficient of average temperature had a positive value, and the coefficient of

precipitation and humidity had a negative value with a small absolute value.

4. Discussion

We investigated the spatio-temporal distribution of the HAV incidence data in Korea from 2016 to 2019 with visualization methods and various statistical methods such as ACF, PACF, and Moran's I. The results showed that the spatial distribution of HAV incidence varied dynamically over the temporal period of interest and that the temporal distribution varied across districts. We also found that the yearly temporal distribution of HAV cases in Korea is quite different. We found a high peak and significant temporal variation in 2019. Son et al. (47) reported that the ingestion of salted clams significantly increased the risk of HAV in Korea in 2019.

Several HAV studies have been limited to frequency analysis, spatial correlation exploration using Moran's I, and comparisons

FIGURE 2
Number of HAV cases per 1,000 people in **(A)** 2016, **(B)** 2017, **(C)** 2018, and **(D)** 2019.

TABLE 3 Distance-based global Moran's I from 2016 to 2019.

	2016	2017	2018	2019
Moran's I	0.138	0.304	0.246	0.186

of SIR and RR. Frequency analysis is useful for finding the frequency of a variable in the entire data, but it is difficult to find a pattern when multiple variables are given conditionally. For example, our frequency analysis in Supplementary Table S2 showed that HAV cases increases up to the 80% quantile of

water supply rates and sewage treatment facility rates, and then decreases in the quantile beyond that. However, such results were inconsistent with previous studies (4, 9), and it was known that these factors were related to other socioeconomic factors and had spatial variations. Therefore, frequency analysis alone has a limitation in investigating the association between environmental factors and HAV. Moran's I index is useful for investigating spatial correlation at a fixed time point, but it has a limitation in that it cannot simultaneously determine spatio-temporal correlation. SIR and RR are mainly used to identify patterns of disease occurrence. By representing the SIR

FIGURE 3
Time series plot of weekly Moran's I and its P-value from 2016 to 2019.
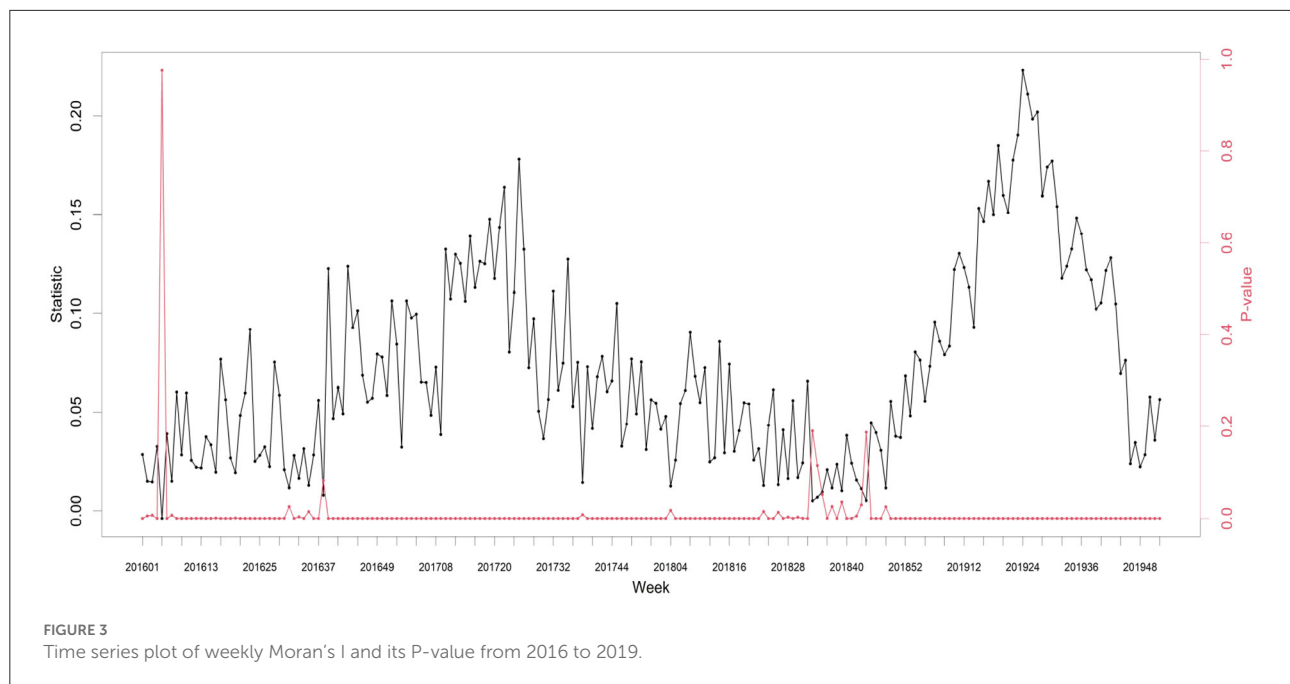
and RR values on a map, it is easy to identify the regions at high risk for disease. However, it is difficult to reflect the spatio-temporal correlation using SIR and RR simultaneously. For example, Moon et al. (19) investigated the incidence rates in Korea from 2011 to 2013 by year and age groups and represented RR in specific regions by year. They focused more on frequency analysis and could not consider spatio-temporal dependent structures simultaneously. Examining each variable separately, without considering multiple variables simultaneously, may result in a biased conclusion. In this respect, a regression model with multiple variables is better than that of frequency analysis. Furthermore, it is important to consider spatio-temporal association for epidemic data simultaneously.

Our study proposed spatio-temporal modeling of weekly incidence data in Korea using a Bayesian approach to better explain the complicated spatio-temporal dependence structures of HAV cases. The model assessed the effects of socioeconomic, environmental, and weather factors on weekly HAV cases by adjusting the spatio-temporal dynamics.

The contribution of this study is to examine the spatio-temporal distribution of HAV cases using various exploratory data analysis methods and to develop a Bayesian spatio-temporal model for considering simultaneous spatio-temporal dependent structures of the data. In terms of modeling, our contributions are as follows. Because the onset of infectious diseases has spatial and temporal correlations, a regression model that does not reflect these variations may result in a poor model fit. Considering this point, we applied a regression model with

spatio-temporal variations to HAV data in Korea. We attempted to find a model that best reflects spatio-temporal variation. We applied the two-stage framework following (43) to avoid spatial confounding bias issues; thus, we obtained a better model fit than that of the other models. Moreover, the HAV cases are counted and contain many zero values, and we consider the ZIP regression as the base model. Using the ZIP regression model coupled with two-stage and spatio-temporal structures, we demonstrated that spatio-temporal variation could not be neglected in analyzing an epidemic disease.

In the proposed spatio-temporal model, various socioeconomic, environmental, and weather-related factors were statistically significant for HAV occurrence in Korea from 2016 to 2019. The results showed that the higher the level of income and education, the more social activities, and the more frequent contact with people, the higher the possibility of exposure to HAV in Korea. It also showed that the higher the male ratio and number of foreigners residing, the higher the HAV incidence rate. Our findings agree with the studies mentioned earlier in Korea (19, 25, 35). Moreover, the number of medical doctors was negatively associated with the HAV incidence rate, as mentioned by Choi (35). We found that the proportion of people aged 30–49 years and incidence had a negative association after adjusting various socioeconomic and environmental factors. This result is somewhat inconsistent with the previous study of Yoon et al. (6). There might be confounding factors that were not considered in our study. Explanations for the present findings warrant further study on the association of the proportion of people with specific age groups and HAV.
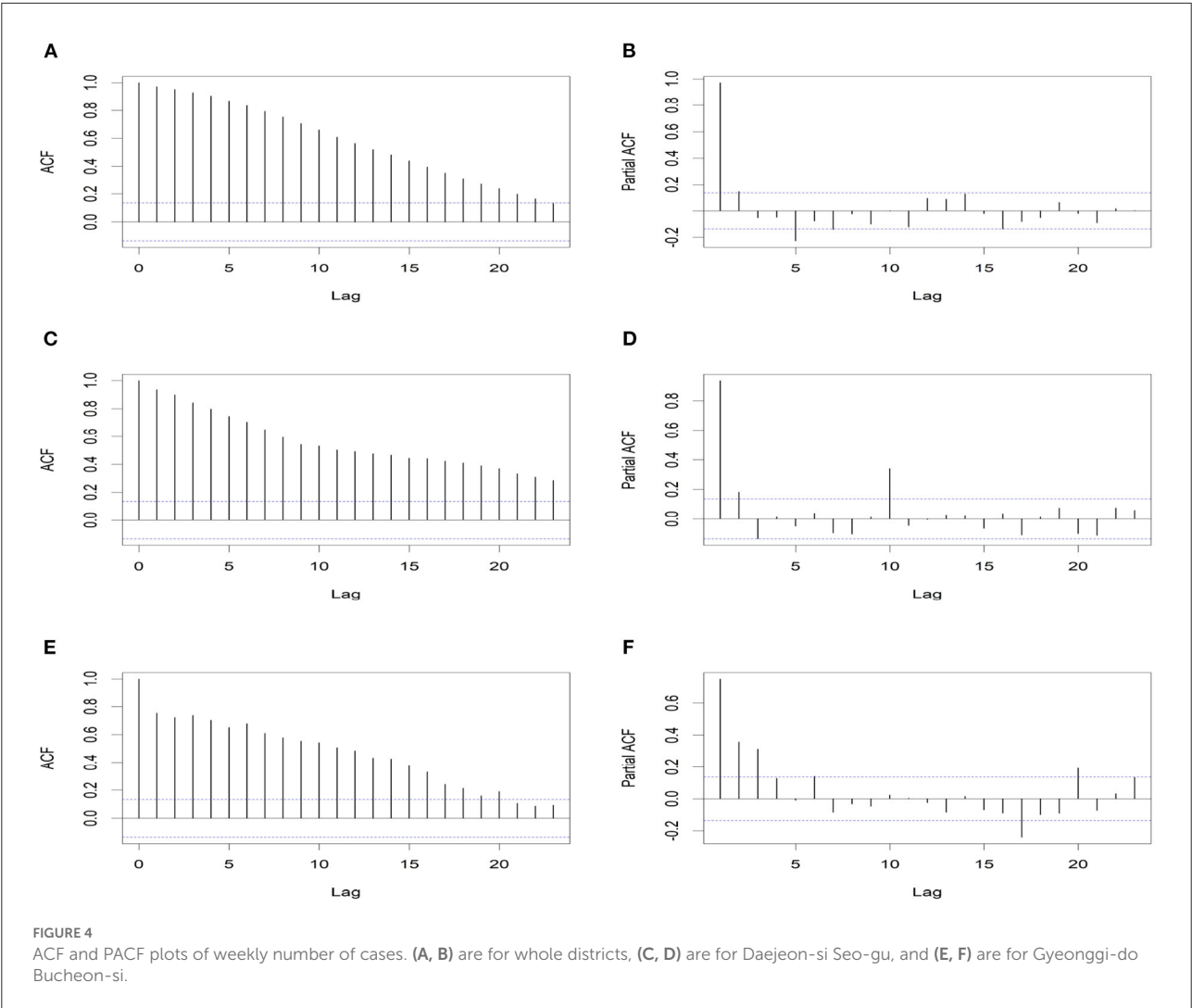
FIGURE 4
ACF and PACF plots of weekly number of cases. **(A, B)** are for whole districts, **(C, D)** are for Daejeon-si Seo-gu, and **(E, F)** are for Gyeonggi-do Bucheon-si.

TABLE 4 Model comparison results.

Distribution	Type	MAE	MSPE	Dbar	DIC	pD
Poisson	1	0.745	2.595	128,717	128,727	11
	2	0.479	0.917	76,911	78,239	1,328
	3	0.366	0.329	66,743	73,175	6,432
ZIP	1	0.742	2.583	128,017	128,028	11
	2	0.465	0.858	72,699	**73,049**	350
	3	**0.365**	**0.328**	66,657	73,119	6,462

The bold values indicate the smallest value of MAE, MSPE, and DIC.

We found that the coefficients of the water supply and sewage treatment facility rates were negative, indicating that the higher the water quality and hygiene conditions, the lower the incidence rate of HAV. These results are in line with previous studies (8, 9), even though our exploratory data analysis in Supplementary Table S2 looked like a positive association. Thus,

we again confirmed the importance of spatio-temporal multiple regression modeling to examine the association between factors and HAV simultaneously.

For weather-related factors, the coefficient of average temperature had a positive value, and the coefficient of precipitation and humidity had a negative value with a

TABLE 5 Parameter estimates from the proposed model.

Variable type	Variable	Posterior mean	2.5%	97.5%
Socioeconomic factors	Total income per person	0.042	0.039	0.047
	High education rate	0.025	0.019	0.030
	Total fertility rate	0.437	0.332	0.769
	Proportion of males	0.367	0.286	0.408
	Proportion of people with 30–49 years old	−0.138	−0.150	−0.127
	Log(number of foreigners)	0.517	0.493	0.577
	Number of doctors per thousand people	−0.171	−0.183	−0.161
Environmental factors	Water supply rate	−0.028	−0.034	−0.024
	Sewage treatment facility rate	−0.031	−0.034	−0.019
Weather-related factors	Average temperature	0.038	0.036	0.040
	Total precipitation	−0.001	−0.002	−0.001
	Average humidity	−0.006	−0.008	−0.003

small absolute value. Supplementary Figure S3 showed the positive association between average temperature and HAV cases, although the associations between other weather factors and HAV were not clearly shown. Moreover, Supplementary Table S3 indicated that the number of HAV cases was relatively large in the spring and summer seasons. A clear seasonal variation was observed in 2019. In Korea, the incidence of HAV is relatively high during spring and summer because of increased outdoor activity and ingestion of not clean food handling (19). Thus, we conclude that temperature is more associated with HAV outbreaks than precipitation or humidity in Korea.

As in the existing studies on HAV in Korea (19, 25, 35), we confirmed that there are risk factors for HAV occurrence. The distribution status of HAV varies by region and time. Additionally, differences in socioeconomic variables, such as education level, sex, number of medical doctors, and water quality, affect the number of HAV cases. Environmental and weather-related factors are also important; however, we found that the contribution of socioeconomic factors is more crucial for HAV occurrence. Therefore, we should recognize the different factors in different regions and prepare region-specific control and prevention strategies for HAV infection. Furthermore, Kang et al. (26) mentioned that a particular age group has a low antibody cultivation rate and is more vulnerable to infection. Therefore, we must consider an age-specific strategic vaccine plan.

Association between socioeconomic factors and HAV prevalence may vary from region to region because the different areas have different characteristics. For example, Jacobsen and Koopman (48) described that a higher level of education leads to a sustained decrease in the incidence of HAV, whereas there

was no statistically significant difference in education when examining HAV antibodies between sewage workers in France and the control group in Cadilhac and Roudot-Thoraval (49). While Rachiotis et al. (50) showed that people with higher education levels had higher rates of anti-HAV in stratified analysis among municipal waste collectors, Arvanitidou et al. (51) showed that the prevalence of anti-HAV was significantly higher in less educated persons. Our exploratory data analysis (Supplementary Table S2) and modeling results provided the positive association between higher education rate and HAV in Korea during 2016–2019. In Korea, people with higher education background tend to have more active social lives and more frequent contact with people so they may have highly exposure to HAV. Seo et al. (25) reported a similar result in Korea. Thus, it is important to consider regional characteristics along with weather-related factors to better understand HAV across Korea.

There was a limitation concerning the data in this study. We focused on regional aggregated data, which could lead to biased results. Thus, ascertaining the direct relationship between factors and outcomes can be limited. If we obtain individual-level HAV case data with individual-level risk factors and conduct spatio-temporal data analyzes, we can find more features that influence the HAV cases and draw clearer pictures of the infection spread problem. Thus, this is one of the future research directions the authors intend to pursue.

Data availability statement

The number of weekly HAV cases can be found in the Korea Disease Control and Prevention Agency database (https://

Author contributions

JJ and JC performed the statistical analyzes. All authors contributed to the conception and design of the study, organized the database, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

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.2022.1085077/full#supplementary-material

References

1. World Health Organization. *Global Hepatitis Report*. (2017). Geneva: World Health Organization (2017).

2. Halliday ML, Kang LY, Zhou TK, Hu MD, Pan QC, Fu TY, et al. An epidemic of hepatitis a attributable to the ingestion of raw clams in Shanghai, China. *J. Infect. Dis*. (1991) 164:852–9. doi: 10.1093/infdis/164.5.852

3. Ki M, Son H, Choi BY. Causes and countermeasures for repeated outbreaks of hepatitis A among adults in Korea. *Epidemiol Health*. (2019) 41:e2019038. doi: 10.4178/epih.e2019038

4. Rakesh P, Sherin D, Sankar H, Shaji M, Subhagan S, Salila S. Investigating a community-wide outbreak of hepatitis a in India. *J Glob Infect Dis*. (2014) 6:59. doi: 10.4103/0974-777X.132040

5. Franco E, Meleleo C, Serino L, Sorbara D, Zaratti L. Hepatitis A: Epidemiology and prevention in developing countries. *World J Hepatol*. (2012) 4:68. doi: 10.4254/wjh.v4.i3.68

6. Yoon JG, Choi MJ, Yoon JW, Noh JY, Song JY, Cheong HJ, et al. Seroprevalence and disease burden of acute hepatitis A in adult population in South Korea. *PLoS ONE*. (2017) 12:e0186257. doi: 10.1371/journal.pone.0186257

7. Tapia-Conyer R, Santos JI, Cavalcanti AM, Urdaneta E, Rivera L, Manterola A, et al. Hepatitis A in Latin America: a changing epidemiologic pattern. *Am J Trop Med Hyg*. (1999) 61:825–829. doi: 10.4269/ajtmh.1999.61.825

8. Mantovani SA, Delfino BM, Martins AC, Oliart-Guzmán H, Pereira TM, Branco FL, et al. Socioeconomic inequities and hepatitis A virus infection in Western Brazilian Amazonian children: spatial distribution and associated factors. *BMC Infect Dis*. (2015) 15:1–12. doi: 10.1186/s12879-015-1164-9

9. Dogru AO, David RM, Ulugtekin N, Goksel C, Seker DZ, Sözen S. GIS based spatial pattern analysis: Children with Hepatitis A in Turkey. *Environ Res*. (2017) 156:349–57. doi: 10.1016/j.envres.2017.04.001

10. Copado-Villagrana ED, Anaya-Covarrubias JY, Viera-Segura O, Trujillo-Ochoa JL, Panduro A, José-Abrego A, et al. Spatial and temporal distribution of hepatitis A virus and hepatitis E virus among children with acute hepatitis in mexico. *Viral Immunol*. (2021) 34:653–7. doi: 10.1089/vim.2021.0045

11. Zheng B, Wen Z, Pan J, Wang W. Epidemiologic trends of hepatitis A in different age groups and regions of China from 1990 to 2018: observational population-based study. *Epidemiol Infect*. (2021) 149:1552. doi: 10.1017/S0950268821001552

12. Shanmugam N, Sathyasekaran M, Rela M. Pediatric liver disease in India. *Clin Liver Dis*. (2021) 18:155. doi: 10.1002/cld.1138

13. Cann K, Thomas DR, Salmon R, Wyn-Jones A, Kay D. Extreme water-related weather events and waterborne disease. *Epidemiol Infect*. (2013) 141:671–86. doi: 10.1017/S0950268812001653

14. Villar LM, De Paula VS, Gaspar AMC. Seasonal variation of hepatitis A virus infection in the city of Rio de Janeiro, Brazil. *Revista do Instituto de Medicina Tropical de S ao Paulo*. (2002) 44:289–92. doi: 10.1590/S0036-46652002000500011

15. Leal PR, de Paula RJ, Guimarães S, Kampel M. Associations between environmental and sociodemographic data and Hepatitis-A transmission in Pará State (Brazil). *GeoHealth*. (2021) 5:e2020GH000327. doi: 10.1029/2020GH000327

16. Adibin, Aisnah, Indrawati, Sumriati, Ramadhan T. Increased risk of hepatitis a due to weather changes: a review. In: *IOP Conference Series: Earth and Environmental Science. Vol. 755*. IOP Publishing (2021). p. 012085. doi: 10.1088/1755-1315/755/1/012085

17. Baek K, Choi J, Park JT, Kwak K. Influence of temperature and precipitation on the incidence of hepatitis A in Seoul, Republic of Korea: a time series analysis using distributed lag linear and non-linear model. *Int J Biometeorol*. (2022) 66:1725–36. doi: 10.1007/s00484-022-02313-2

18. Fares A. Seasonality of hepatitis: a review update. *J Family Med Primary Care*. (2015) 4:96. doi: 10.4103/2249-4863.152263

19. Moon S, Han JH, Bae GR, Cho E, Kim B. Hepatitis A in Korea from 2011 to 2013: current epidemiologic status and regional distribution. *J Korean Med Sci*. (2016) 31:67–72. doi: 10.3346/jkms.2016.31.1.67

20. Jacobsen KH. Globalization and the changing epidemiology of hepatitis A virus. *Cold Spring Harb Perspect Med.* (2018) 8:a031716. doi: 10.1101/cshperspect.a031716

21. Gomez-Barroso D, Varela C, Ramis R, Del Barrio J, Simon F. Space-time pattern of hepatitis A in Spain, 1997-2007. *Epidemiol Infect.* (2012) 140:407–16. doi: 10.1017/S0950268811000811

22. Stoitsova S, Gomez-Barroso D, Vallejo F, Ramis R, Kojouharova M, Kurchatova A. Spatial analysis of hepatitis a infection and risk factors, associated with higher hepatitis a incidence in Bulgaria: 2003-2013. *Compt Rend Acad Bulg Sci.* (2015) 68:1071–8.

23. Leal PR, de Paula RJ, Guimarães S, Kampel M. Sociodemographic and spatiotemporal profiles of hepatitis-A in the state of Pará Brazil, based on reported notified cases. *Geospat Health.* (2021) 16:981. doi: 10.4081/gh.2021.981

24. Kim I, Ryu J, Lee J. Status of construction and operation of large wastewater treatment plants in South Korea. *Water Sci Technol.* (1996) 33:11–18. doi: 10.2166/wst.1996.0292

25. Seo JY, Seo JH, Kim MH, Ki M, Park HS, Choi BY. Pattern of hepatitis a incidence according to area characteristics using national health insurance data. *J Prevent Med Public Health.* (2012) 45:164. doi: 10.3961/jpmph.2012.45.3.164

26. Kang SH, Kim MY, Baik SK. Perspectives on acute hepatitis a control in Korea. *J Korean Med Sci.* (2019) 34:e230. doi: 10.3346/jkms.2019.34.e230

27. Moon H, Noh J, Hur M, Yun Y, Lee CH, Kwon SY. High prevalence of autoantibodies in hepatitis A infection: the impact on laboratory profiles. *J Clin Pathol.* (2009) 62:786–8. doi: 10.1136/jcp.2009.064410

28. Moon HW, Cho JH, Hur M, Yun YM, Choe WH, Kwon SY, et al. Laboratory characteristics of recent hepatitis A in Korea: ongoing epidemiological shift. *World J Gastroenterol.* (2010) 16:1115. doi: 10.3748/wjg.v16.i9.1115

29. Cho HC, Paik SW, Kim YJ, Choi MS, Lee JH, Koh KC, et al. Seroprevalence of anti-HAV among patients with chronic viral liver disease. *World J Gastroenterol.* (2011) 17:236. doi: 10.3748/wjg.v17.i2.236

30. Kim H, Ryu J, Lee YK, Choi MJ, Cho A, Koo JR, et al. Seropositive rate of the anti-hepatitis A immunoglobulin G antibody in maintenance hemodialysis subjects from two hospitals in Korea. *Korean J Intern Med.* (2019) 34:1297. doi: 10.3904/kjim.2017.293

31. Kim YJ, Lee HS. Increasing incidence of hepatitis in Korean adults. *Intervirology.* (2010) 53:10–14. doi: 10.1159/000252778

32. Lee H, Cho HK, Kim JH, Kim KH. Seroepidemiology of hepatitis A in Korea: changes over the past 30 years. *J Korean Med Sci.* (2011) 26:791–6. doi: 10.3346/jkms.2011.26.6.791

33. Yoon EL, Sinn DH, Lee HW, Kim JH. Current status and strategies for the control of viral hepatitis A in Korea. *Clin Mol Hepatol.* (2017) 23:196. doi: 10.3350/cmh.2017.0034

34. Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. *BMJ.* (2010) 341:c4830. doi: 10.1136/bmj.c4830

35. Choi SY. *Space and Time pattern of Hepatitis A in South Korea, 2004-2012.* [dissertation]. Seoul Korea: Hanyang University (2015).

36. Choi J. Bayesian spatiotemporal modeling in epidemiology: Hepatitis A incidence data in Korea. *Korean J Appl Stat.* (2014) 27:933–45. doi: 10.5351/KJAS.2014.27.6.933

37. Stein ML. *Interpolation of Spatial Data: Some Theory for Kriging.* New York, NY: Springer-Verlag (1999).

38. Brockwell PJ, Davis RA. *Time Series: Theory and Methods.* New York, NY: Springer-Verlag (2009).

39. Moran PA. Notes on continuous stochastic phenomena. *Biometrika.* (1950) 37:17–23. doi: 10.1093/biomet/37.1-2.17

40. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data.* 2nd ed. Florida: Chapman and Hall/CRC (2014).

41. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* (2019) 35:526–8. doi: 10.1093/bioinformatics/bty633

42. Nychka D, Furrer R, Paige J, Sain S. fields: Tools for spatial data. *R package version.* (2017).

43. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu J. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat.* (2012) 17:417–41. doi: 10.1007/s13253-012-0100-3

44. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math.* (1991) 43:1–20. doi: 10.1007/BF00 116466

45. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. Programming with models: writing statistical algorithms for general model structures with NIMBLE. *J Comput Graph Stat.* (2017) 26:403–13. doi: 10.1080/10618600.2016.1172487

46. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc B.* (2002) 64:583–639. doi: 10.1111/1467-9868.00353

47. Son H, Lee M, Eun Y, Park W, Park K, Kwon S, et al. An outbreak of hepatitis A associated with salted clams in Busan, Korea. *Epidemiol Health.* (2022) 44:e2022003. doi: 10.4178/epih.e2022003

48. Jacobsen KH, Koopman JS. Declining hepatitis A seroprevalence: a global review and analysis. *Epidemiol Infect.* (2004) 132:1005–22. doi: 10.1017/S0950268804002857

49. Cadilhac P, Roudot-Thoraval F. Seroprevalence of hepatitis A virus infection among sewage workers in the Parisian area, France. *Eur J Epidemiol.* (1996) 12:237–40. doi: 10.1007/BF00145411

50. Rachiotis G, Papagiannis D, Thanasias E, Dounias G, Hadjichristodoulou C. Hepatitis A virus infection and the waste handling industry: a seroprevalence study. *Int J Environ Res Public Health.* (2012) 9:4498–503. doi: 10.3390/ijerph9124498

51. Arvanitidou M, Mamassi P, Vayona A. Epidemiological evidence for vaccinating wastewater treatment plant workers against hepatitis A and hepatitis B virus. *Eur J Epidemiol.* (2004) 19:259–62. doi: 10.1023/B:EJEP.0000020444.64546.3b

Optimizing non-pharmaceutical intervention strategies against COVID-19 using artificial intelligence

Vito Janko[1]*, Nina Reščič[1,2], Aljoša Vodopija[1,2], David Susič[1,2], Carlo De Masi[1], Tea Tušar[1], Anton Gradišek[1], Sophie Vandepitte[3], Delphine De Smedt[3], Jana Javornik[4], Matjaž Gams[1,2] and Mitja Luštrek[1,2]*

[1]Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia, [2]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, [3]Department of Public Health and Primary Care, Ghent University, Ghent, Belgium, [4]Leeds University Business School, University of Leeds, Leeds, United Kingdom

One key task in the early fight against the COVID-19 pandemic was to plan non-pharmaceutical interventions to reduce the spread of the infection while limiting the burden on the society and economy. With more data on the pandemic being generated, it became possible to model both the infection trends and intervention costs, transforming the creation of an intervention plan into a computational optimization problem. This paper proposes a framework developed to help policy-makers plan the best combination of non-pharmaceutical interventions and to change them over time. We developed a hybrid machine-learning epidemiological model to forecast the infection trends, aggregated the socio-economic costs from the literature and expert knowledge, and used a multi-objective optimization algorithm to find and evaluate various intervention plans. The framework is modular and easily adjustable to a real-world situation, it is trained and tested on data collected from almost all countries in the world, and its proposed intervention plans generally outperform those used in real life in terms of both the number of infections and intervention costs.

KEYWORDS

COVID-19, multi-objective optimization, epidemiological modeling, machine learning, intervention plans

1. Introduction

The first line of defense against the spread of the SARS-CoV-2 virus was the introduction of Non-Pharmaceutical Interventions (NPIs) by national governments. With the virus being aerosol-borne, some of the key measures included the use of face masks and restrictions on gatherings, which have often resulted in partial or full lockdowns. While effective at reducing the number of infections (1, 2), restrictive NPIs also presented immense Socio-Economic Costs (SECs) to the population (3). Policy-makers were faced with an almost impossible task of carefully balancing NPI costs against the predicted NPI benefits, largely without having appropriate tools and data for evidence-based decisions.

To add complexity to the problem, in a typical intervention plan adopted by policy-makers, a combination of NPIs would be used, each of them taking place for different periods of time. These plans were usually prepared by expert panels who had the challenge of selecting intervention plans without assurance that they would really flatten the infection curve enough to be lifted within the expected period (4, 5).

While many models for the prediction of daily infections and the impact of NPIs on the spread of the pandemic have been proposed ([1](#), [2](#)), little work has been done regarding the prescription of intervention plans—especially taking into account the NPI costs and how to best combine NPIs. Yousefpour et al. ([6](#)), for example, proposed a framework based on SEIRD models and multi-objective optimization to prescribe NPIs. However, the optimization did not operate on real-life NPIs, and as such, this approach cannot be directly used by policy-makers. Chen et al. ([7](#)) created a linear programming tool to explore the trade-off between the expected mortality rate of COVID-19 and return to normal activities, while Yaesoubi et al. ([8](#)) developed a decision tool to determine when to trigger, continue, or stop physical distancing intervention in order to minimize both the deaths from COVID-19 and intervention duration. Both studies combined the objectives into a single function and the final result was a single intervention plan. Such approaches require a strong predefined preference on how to balance the objectives, which is often difficult to define in practice. In addition, none of the three approaches was extensively tested on various epidemiological scenarios. For this reason, their generalization to real-world situations is not known.

A more structured attempt to research the possibility of using artificial intelligence (AI) to automatically prescribe intervention plans was made by the $500K Pandemic Response Challenge ([9](#)), organized by XPRIZE and sponsored by Cognizant. The participants were tasked with finding good trade-offs between the costs of NPIs and their benefits—and assemble three-month intervention plans for each territory (all countries and some sub-country regions). An approach proposed by the sponsor [Miikkulainen et al. ([10](#))] involved the use of evolutionary algorithms to evolve neural networks that prescribe intervention plans. This approach was intended to point the way for the competitors, who would go on to develop better-performing approaches. The competition ended with two "Grand Prize Winners." One of them ([11](#)) combined two prescriptors: the first selected the most cost-effective intervention plans from a subset of possible plans with precomputed effectiveness, and the second greedily composed intervention plans from most cost-effective individual NPIs. The other winning submission—submitted by some of this paper's authors—was the starting point for the approach described here.

In this paper, we describe a framework to help policy-makers design reasonable intervention strategies by dynamically adjusting NPIs. The framework is comprised of three components: a predictor based on the SEIRD epidemiological model that predicts infection trends, a compilation of SECs of NPIs, as found in the literature, and a prescriptor that finds diverse optimized intervention plans. The main methodological novelty of the predictor is that the key parameters of the SEIRD model can be dynamically adapted to any set of given NPIs using a machine-learning model. Intuitively, the machine-learning model decreases the disease transmission rate in the SEIRD model when strict NPIs are in place, and vice versa. In contrast to most related work, our prescriptor uses multi-objective optimization and does not combine the objectives into a single function. As such, it can find near-optimal trade-offs between the costs (SEC) and benefits (reduced number of infections) of NPIs, and presents the results in the form of a Pareto-front approximation. Ideally, the obtained Pareto-front approximation ranges from costly intervention plans, which significantly decrease infections, to cheap but not as effective

TABLE 1　Social and economic costs for OxNPIs.

OxNPI	Economic	Social	Combined
C1: School closing	3.9	11	0.55
C2: Workplace closing	22.0	11	0.96
C3: Cancel public events	1.4	7	0.32
C4: Restrictions on gatherings	1.4	10	0.45
C5: Close public transport	0.3	2	0.09
C6: Stay at home requirements	5.2	12	0.62
C7: Restrictions on internal movement	7.8	10	0.59
C8: International travel controls	6.6	2	0.20
H1: Public information campaigns	0.0026	1	0.04
H2: Testing policy	0.6	1	0.05
H3: Contact tracing	0.1	1	0.04
H6: Facial coverings	0.03	5	0.21

Economic costs are shown as % of GDP loss in the period the NPI was implemented. The social costs are based on domain knowledge and expressed on a 1–12 scale. The combined column is the average of the two costs, when both are normalized to the [0, 1] range.

ones—presenting a set of plans for the policy-maker to choose from. Our methodology was extensively tested: the predictor was tested on data from 194 territories and the prescriptor on data from 50 territories. It yields semantically sensible results, achieves similar or better prediction accuracy than previously proposed models, and furthermore, proposes better plans—at least based on our simulations—than those actually implemented by policy-makers in the studied period (March 2020 to April 2021).

2. Methods

We defined an intervention plan as a prescription of which NPIs, and with what strictness, are to be used on each day in a time period. For this study we considered 12 NPIs listed in [Table 1](#), and we denoted this set as *OxNPIs* as it is derived from Oxford's OxCGRT dataset ([12](#)) introduced in the following subsection. The task of finding good intervention plans could then be framed as a multi-objective optimization problem—trying to minimize both the number of infections and the SEC that would result from a given plan.

Given this formulation, we had to solve the following three problems: 1) how to estimate the number of infections in a specific territory, given an intervention plan; 2) how to estimate the SEC of an intervention plan; and 3) how to use both these estimators and a multi-objective optimization algorithm to find different intervention plans. We start by describing the dataset used and then our solution to each of the listed problems in the following subsections.

2.1. Dataset

The NPIs used in this study were derived from the "COVID-19 Government response tracker" database, collected by Blavatnik School of Government at Oxford University (12). This database defines the periods in which different NPIs (e.g., "C1: School closing" and "C8: International travel controls") were implemented in each territory (entities such as countries, US states, or counties of the UK). It also defines their "strictness" in the form of numbers usually ranging from 0 to 3 or 4, which can represent, for example, if all or only some schools were closed. From the NPI list available in the Oxford database, we selected 12 for analysis in this study: H1, H2, H4, H6, C1–C8 (OxNPIs). Their description and the reasoning for their selection can be found in the Supplementary material—Non-Pharmaceutical Interventions.

The number of infections and deaths (note that we are working with "reported cases" which is only an approximation for the actual number of infections) was queried from the same database for the period between March 1, 2020, and April 14, 2021. This database contained 235 territories, of which different subsets were used in different stages of our methodology. For fitting the epidemiological model, all 235 territories were used. Then, some territories were excluded as their data could not be accurately fit with an epidemiological model (e.g., if the number of reported infections were too low or data was missing). This resulted in 194 territories on which we evaluated the predictive model. For each of them, we chose fifty 70-day time intervals, thus generating 9,700 test cases for the task.

In addition to the already described OxNPIs and infection numbers, the following attributes were used to train the machine-learning models: vaccination (13) (one shot, two shots), strains (14, 15) of concern and interest as defined by the World Health Organization (16) testing rate (17), number of hospitalized patients (18), number of patients in intensive care (18), mask use (19), mobility (20, 21), weather (22), holidays (19), and 93 static features characterizing countries and regions (e.g., development, culture, and health) from our previous study (23). "Duration" features were also constructed to capture how long each NPI had been active to date and how much time had elapsed since the first recorded infection case.

Finally, for the prescriptor evaluation, we chose a representative sample of 50 different 60-day intervals. This sample was selected by first defining the "category" for each time interval: the categories were created based on the size of the territory (small/large) and the derivative of the number of infections (slope). The slopes were either constant, moderately steep (falling/raising), or very steep (falling/raising). Altogether, we had 10 categories, and we randomly selected five time-intervals from each. An additional condition for an interval to be selected was to have at least 0.5 average number of daily new infections per 100k of population.

2.2. Hybrid machine-learning epidemiological model

To predict the future number of infections we used an epidemiological model that can model the course of the disease given some parameters (infection rate, incubation period, mortality) in combination with a machine-learning model that can estimate these parameters from the active NPIs.

2.2.1. Epidemiological model

We used the SEIRD (24) model, which originates from the SIR family of standard epidemiological models used to study the dynamics of infectious diseases. Even if the SEIRD model is more complex than the basic SIR or SIRD models, it has proven to be more numerically stable than the other two for our purpose, and in addition, the numbers for all five categories were available. The model consists of a set of differential equations (Equation 1). Letters represent the size of a given compartment (*S*usceptible, *E*xposed, *I*nfected, *R*ecovered, and *D*eceased), N is the sum of all compartments, β is the infection rate, σ is the incubation period (1/days), γ is the recovery rate, μ is the mortality rate, and t is time. The reproduction number can be estimated as $\frac{\beta}{\sigma}$.

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta \frac{SI}{N} & \frac{dR}{dt} &= \gamma I \\
\frac{dE}{dt} &= \beta \frac{SI}{N} - \sigma E & \frac{dD}{dt} &= \mu I \\
\frac{dI}{dt} &= \sigma E - (\gamma + \mu)I
\end{aligned}
\tag{1}
$$

In a standard SEIRD model, the parameters β, μ, and σ are constant. In reality—especially in the case of COVID-19—they are highly dependent on various factors, including the NPIs. In related work, there were several attempts at modeling β as a function of interventions. In the DELPHI model developed by COVID Analytics (25), the effect of interventions was modeled using an arctan function (26). Zou et al. (27) used machine learning to learn the epidemiological model parameter values from the number of infected and removed (deceased and recovered) cases at time t. In our model, we used machine-learning models that used several different features to achieve this task—allowing us a greater flexibility in dynamically changing the parameters, as opposed to what could be achieved with other methods from related work.

2.2.2. Predicting the model parameters with machine learning

The first step of the process was to fit the β, μ, and σ parameters to different territory/time intervals. This was done by finding parameter values that minimize the least squares error in predicting the reported number of infections and deaths. The time series of data for each of the 235 territories were split into intervals based on two criteria: *NPI change* (at least two NPIs change on the same day) and *infection trend* (a 7-day moving-average number of infections that was previously raising, starts falling—or vice versa), and each was fitted separately.

These fitted values were then used as prediction targets for three machine-learning regression models (one model per parameter). When trained, these models would be used to predict the parameters when evaluating different NPIs by the prescriptor.

For the prediction of each parameter, we used the features described in the Dataset section, and some of their subsets. We performed an initial feature selection on the available dataset by employing Recursive Feature Elimination (RFE) with a 10-fold cross-validation. We evaluated both 1) straightforward feature selection

(i.e., running the algorithm on all available features), and 2) including the OxNPIs in the selected features and running the RFE only on the remaining features. However, the results showed no significant improvement after the RFE algorithm. For the sake of model interpretability, we selected the features presenting the strongest correlation with the reported number of infections, and ended up with OxNPIs, duration features, historical infections, COVID-19 strains, and vaccination features.

We tested linear regression (28), ridge regression (28), decision tree (28), LGBM (29), XGB (30), CatBoost (31), Elastic Net (28), Bayesian ridge (28), SVR (28), and Random Forest models (28). The models were compared with 10-fold cross-validation where the train/test splits were performed territory-wise, meaning that all instances of a territory were in either the test or the train set. Keeping all instances of one territory in the same set was important since consecutive instances were typically similar.

In the cases of linear and ridge regression, the regression coefficients for the final model were calculated as the mean values of the coefficients generated in the 10-fold cross-validation. The "H1: Public information campaigns" regression coefficient initially had an excessive value because the corresponding NPI was essentially always present (and was thus used by the model almost as the intercept). We, therefore, manually adjusted it based on Haug et al.'s (32) study. Specifically, we used the four NPIs for which there was a good match between our categorization and the one presented by Haug et al.: "C1: School closing," "C7: Restrictions on internal movement," "C3: Cancel public events," and "C5: Close public transport." We computed the ratio between the decrease in reproduction rate (β/γ) for these four NPIs (32), and the decrease for "H1: Public information campaigns." We then multiplied our coefficients for the same NPIs with these ratios, which yielded four possible values for the H1 coefficient. We used the average of these. We then re-ran the regression with fixed relations between the NPI coefficients, so that the relation between them and other coefficients could be readjusted.

Since the parameter β (infection rate) was most strongly affected by NPIs, and since we are aware of no strong reason why the other two should be, we also considered predicting β only. And since the parameters of the model are not independent, we considered using some as features for the prediction of others. However, both of these approaches gave worse results.

2.2.3. Prediction pipeline

The goal of the prediction pipeline is to predict the number of infections given an intervention plan (which OxNPIs are used on a given day). To do so, we create a feature vector by joining the OxNPI data with the remaining features. Then, for each day, a prediction of all three parameters is made with the three respective machine-learning models.

Next, for the time interval leading to (but not including) the prediction interval, the fitted parameters are queried. We assume that the parameters at the beginning of the prediction interval should be the same as the fitted parameters at the end of the last one directly preceding it. Thus, the machine-learning predictions are normalized as $\beta_i = \beta_{last}/\beta_0$, where β_i is the value of the predicted parameter β on the i-th day, and β_{last} is the last known fitted value of β preceding the prediction interval. Parameters σ and μ are normalized similarly.

If the parameters for any day are such that the reproduction rate exceeds five, then the value of β is reduced until the reproduction rate falls to this threshold value. This is done because such high reproduction rates do not appear in real-life data, but they might be predicted due to some edge case in machine learning. All parameters are smoothed using weighted decay ($\alpha = 0.2$), as we assume that all parameters are changing smoothly.

When the parameters are estimated for each day, they are inserted into the SEIRD model, which can then produce the number of infections for each day. The starting value of the "Exposed" compartment is set in a way such that the predicted and reported numbers of infections match on day zero.

2.3. Socio-economic costs of different NPIs

The collection of socio-economic costs (SEC) of individual OxNPIs was not the primary focus of our work, but nonetheless we compiled a sensible set in order to properly test our methodology. The collected SECs were derived from a set of costs from related work and from the opinion of a domain expert. Due to the available literature, the costs are likely to contain a bias toward Western countries, and most data is based on reports and gray literature.

In the study, we used the values listed in Table 1, but the methodology is rather general and a policy-maker can easily adapt it to produce a set of SECs for a specific territory—possibly also implicitly expressing their preferences on what NPIs to avoid (by assigning them higher costs). The combined SEC is made simply by normalizing both costs to the [0, 1] range and then averaging both. While this number does not have a good interpretation, it does rank the OxNPIs according to their SECs. The costs are given for the case in which the NPI is implemented with its maximum strictness. For other strictness levels, the costs were linearly scaled down (in rare cases, a custom social cost was defined and used instead of the linearly scaled value). In addition, the "C6: Stay at home requirements" NPI requires the implementation of the C1, C2, C3, C4, C5, and C8 NPIs. Thus, even if it did not have the highest cost, the overall cost implicitly includes the costs of all other listed NPIs.

2.3.1. GDP loss

Because the available findings on economic cost of NPIs differ in terms of the setting and time, they were normalized to represent the % of GDP loss caused by the NPI while it was in effect. Country-specific GDP values (US $) were used (33). For example, if the "C3: Cancel public events" NPI is active for 1 month and it has the cost of 1.4, then our method assumes that the GDP in this month is 1.4% lower than usual—note that this is not the annual GDP loss but that for the predicted period. The complete overview of the cost data used can be found in the Supplementary material—GDP cost. While there is some overlap between the NPIs, we have explicitly modeled this only in case of C6 as previously described.

2.3.2. Social impact

While economic costs were available for most OxNPIs, the literature on social costs was far more scarce. We thus placed the ranking of OxNPIs by social costs on a theoretical foundation, but we could not justify the numerical costs as solidly. In addition, according to the literature, these costs may vary across countries (e.g., collectivistic vs. individualistic countries); however, we applied

standard levels for all WEIRD countries (i.e., for Western, Educated, Industrialized, Rich, and Democratic, a common grouping in psychological studies). To estimate the social costs, we ranked the OxNPIs from the highest to the lowest based on the perceived strain, dread and loss, perceptions of restricted freedoms, and constraining behaviors (i.e., on the negative impact of each measure on behavior, attitudes, and one's well-being). Using the rational choice theory, we assumed that the higher the perception of strain, dread and loss, the more negative is the impact and the higher are the social costs. Understanding human behavior and risk perception is central to effective pandemic management, and thus we applied insights from social and behavioral sciences to inform our assumptions on social impact. For determining the cost of individual policies see Supplementary material—Social cost discussion.
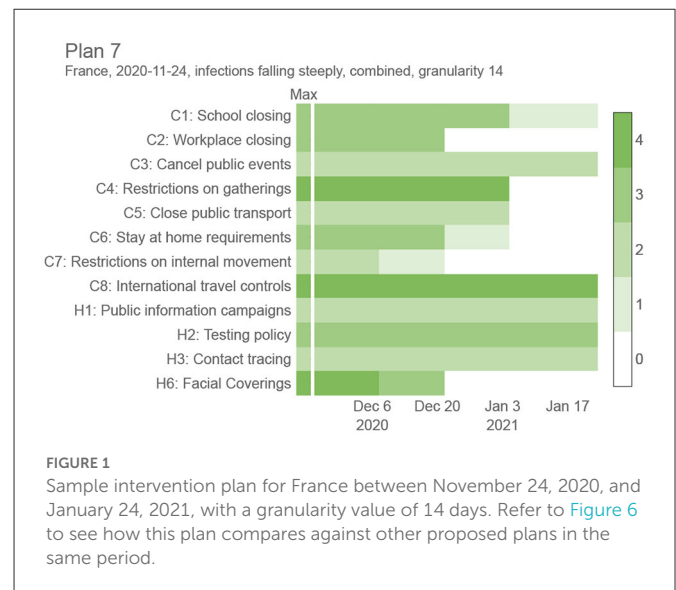
2.4. Proposing interventions

The task of proposing intervention plans can be mathematically formulated as a multi-objective optimization problem with two objectives that need to be minimized: the total number of infections (f_1) and the SECs of the proposed plan (f_2). The two objectives are conflicting since an effective way to slow down the spread of infection requires a stringent intervention plan with expensive NPIs. The first objective is expressed as the total number of infections predicted from the HMLE model, while the second objective is the cost of NPIs averaged over the plan's duration. The problem is constrained by limiting the number of new daily infections to 150 per 100k residents. This is done as the plans with more infections were not considered useful to policy-makers and almost never appear ($< 1\%$) in real-life data in the studied period.

The proposed intervention plans are composed of OxNPIs that can vary over time, but are restricted to last at least g days in a row, where g is a predetermined parameter we refer to as *granularity*. An NPI, for example, "C2: Workplace closing," can be applied with different levels of strictness (0—no policy, 1— closure recommended, 2—closure for specific sectors, 3—closure for all-but-essential workplaces). With this in mind we can formally define the intervention plan—a solution to the proposed optimization problem—as a $12 \times n$ integer-valued matrix, P, where its 12 rows correspond to the 12 OxNPIs and n is the number of time slots determined by the given granularity and the whole period (e.g., Figure 1 contains $n = 4$ time slots resulting from a granularity value of 14 days and an interval length of 60 days). In detail, P_{ij} indicates the strictness of the i-th NPI in the j-th time slot. In particular, we tested five values for granularity: 1, 3, 7, 14, and 30.

Based on the multi-objective formulation of the proposed optimization problem, the experimental evaluation aimed at finding sets of trade-off intervention plans representing approximations for Pareto fronts. For this purpose, we used the Nondominated Sorting Genetic Algorithm II (NSGA-II) (34) equipped with a Constrained Dominance Principle (CDP) (34) to handle the constraint. NSGA-II belongs to the group of evolutionary algorithms, and as such, it imitates the biological evolution to search the space of possible intervention plans and find plans with good trade-offs between the two objectives.

The optimization problem was solved using two NSGA-II internal solution representations: the full representation defined by
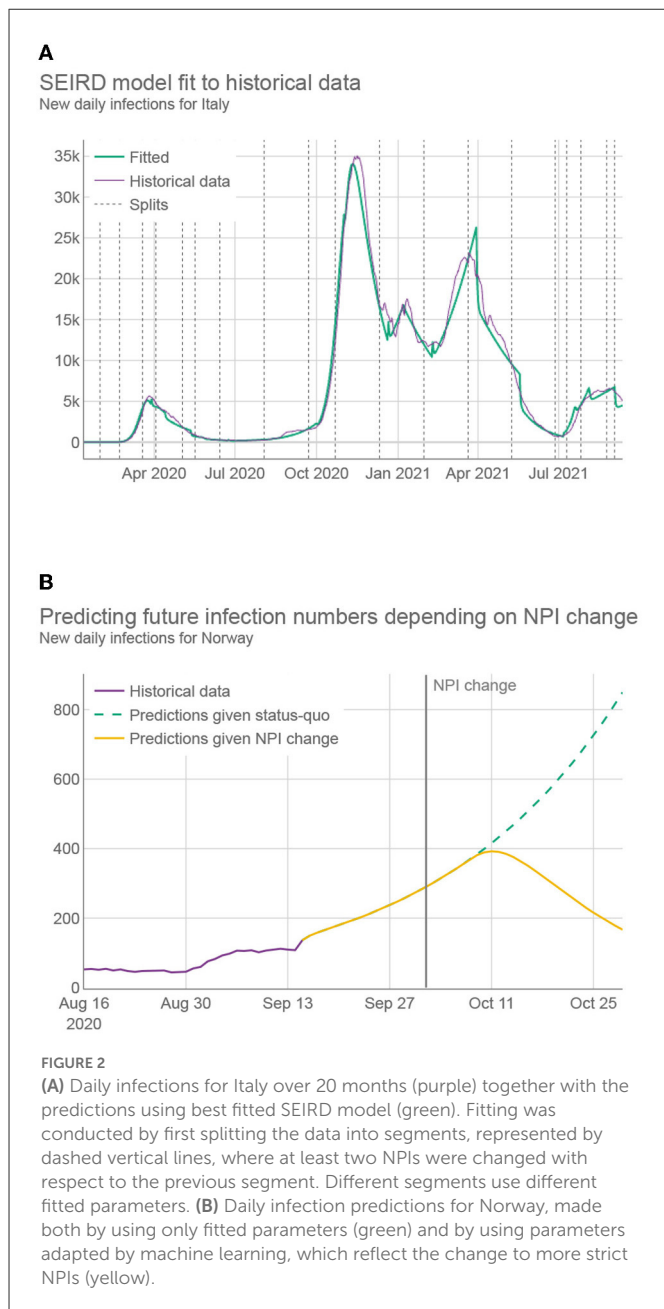
Sample intervention plan for France between November 24, 2020, and January 24, 2021, with a granularity value of 14 days. Refer to Figure 6 to see how this plan compares against other proposed plans in the same period.

the matrix P and the condensed representation defined by a vector of length n where the j-th variable corresponds to the maximum SEC allowed at the j-th time slot. The second representation was considered due to the significant reduction in the search space dimensionality (from $12n$ to n), allowing for much faster convergence than the high-dimensional search space for the full representation. While the full representation can be used without modifications, the condensed representation needs to be decoded to the intervention plan before evaluation. This is achieved by replacing the SECs with OxNPI values. The OxNPI combination to replace each SEC is selected as the one with the lowest projected infections out of those within the allowed SECs. This mapping is computed in advance, by having all OxNPIs combinations sorted based on their effectiveness (by using linear model's coefficients for each NPI), so that the most effective combination that does not exceed the cost threshold can easily be selected.

The one-point crossover was used as the crossover operator and the random resetting as the mutation operator. Additionally, the crossover probability was set to 0.9 and the mutation probability to $1/D$, where D equals $12n$ for the full representation and n for the condensed representation.
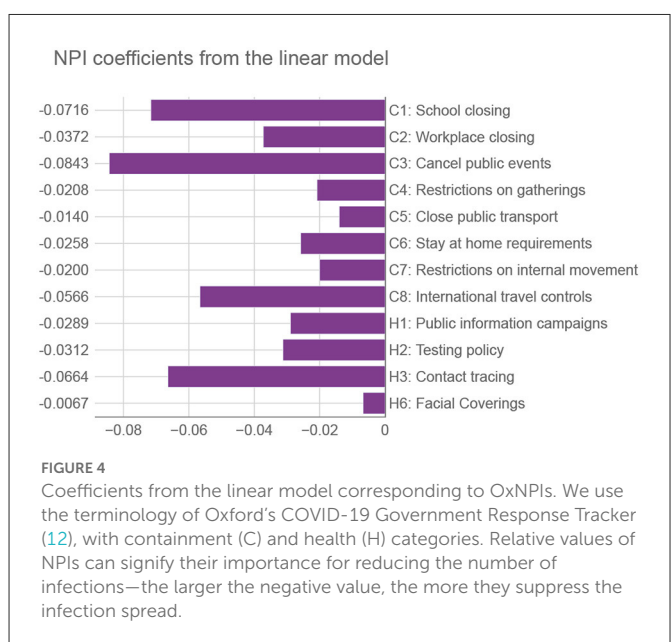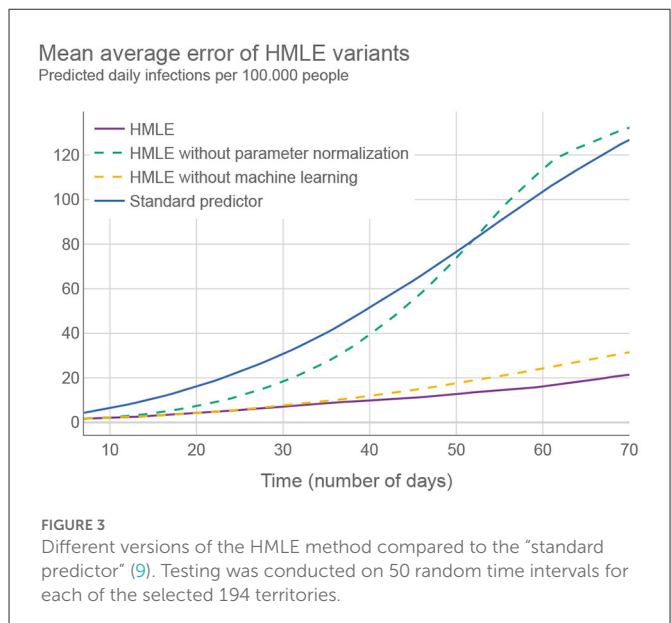
3. Results

3.1. Predicting infections

While the SEIRD model on its own is accurate in predicting the future in cases where NPIs are not changing and historically-fitted parameters can be used (see Figure 2 and Supplementary material— Estimating the prediction error), it does not correctly predict the infection trends following a change of the NPIs— which is essential if the framework is to propose which NPIs to use in the future. Ideally, as the NPIs change, the parameters of the SEIRD system would be adjusted accordingly, taking into account their changed impact on the disease transmission rate. An example of such behavior can be seen in Figure 2, as generated by our HMLE method.
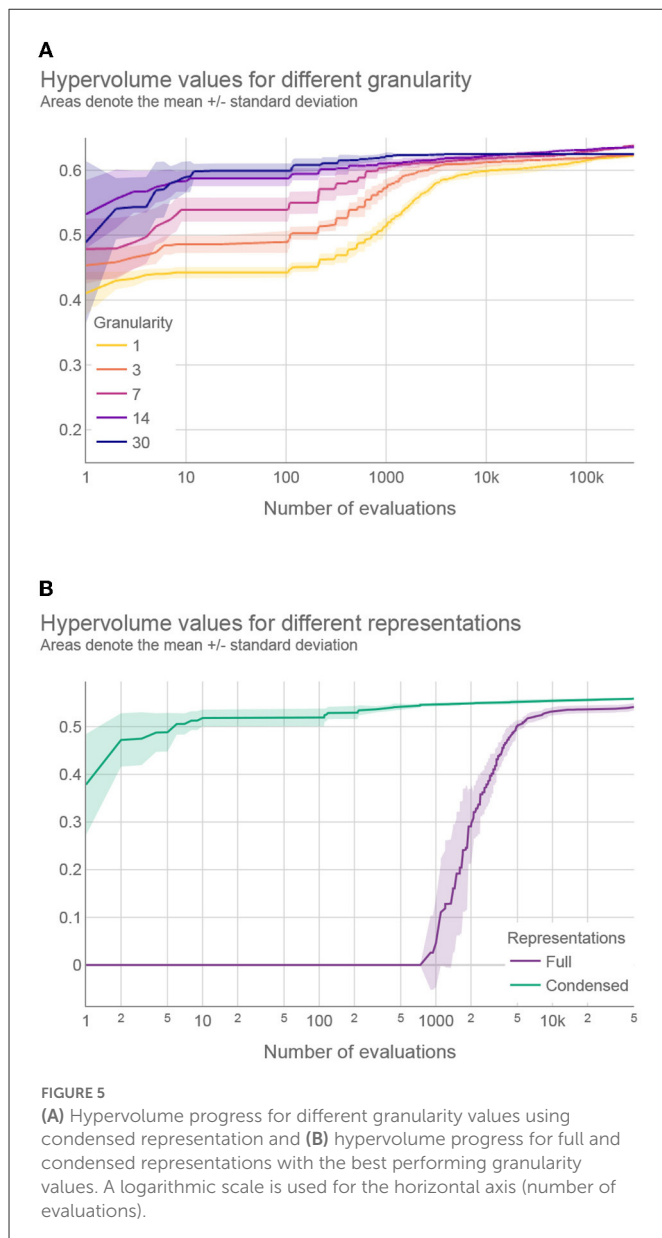
FIGURE 2

(A) Daily infections for Italy over 20 months (purple) together with the predictions using best fitted SEIRD model (green). Fitting was conducted by first splitting the data into segments, represented by dashed vertical lines, where at least two NPIs were changed with respect to the previous segment. Different segments use different fitted parameters. **(B)** Daily infection predictions for Norway, made both by using only fitted parameters (green) and by using parameters adapted by machine learning, which reflect the change to more strict NPIs (yellow).

FIGURE 3

Different versions of the HMLE method compared to the "standard predictor" (9). Testing was conducted on 50 random time intervals for each of the selected 194 territories.

FIGURE 4

Coefficients from the linear model corresponding to OxNPIs. We use the terminology of Oxford's COVID-19 Government Response Tracker (12), with containment (C) and health (H) categories. Relative values of NPIs can signify their importance for reducing the number of infections—the larger the negative value, the more they suppress the infection spread.

To assess the performance of the HMLE method, we show in Figure 3 that our predictor significantly outperforms the "standard predictor" provided by Cognizant in the second phase of the XPRIZE competition (9) (for details of this test, see Supplementary material— Estimating the prediction error). This is a predictor published prior to the competition (10), which represented the state of the art for NPI-dependent prediction at the time. The mean average error (MAE) of our predictor is 5.9 times lower on day 70. To explore what contributes to the increased performance, we compared the full implementation to two additional versions of our method: 1) one that relies only on machine learning to set the parameter values of the SEIRD model without normalizing them using the last known fitted parameter values, and 2) one that retains the last known fitted parameter values throughout the forecast period, without using

machine learning to account for NPI changes. The experimental results showed that the parameters predicted by the machine-learning model are less appropriate on average, than the last known fitted parameters; when normalized, however, they outperform the last known fitted parameters. The benefit of machine learning does not appear to be huge, but it is significant in case of important NPI changes, as demonstrated in Figure 2.

Of all machine-learning algorithms tested (see Supplementary material—Estimating the prediction error), the Ridge classifier (a type of linear model) had the highest accuracy. Aside from the prediction accuracy, the model has an additional advantage—it is easily interpretable. Figure 4 lists the coefficients corresponding to the normalized OxNPI strictness values. Given

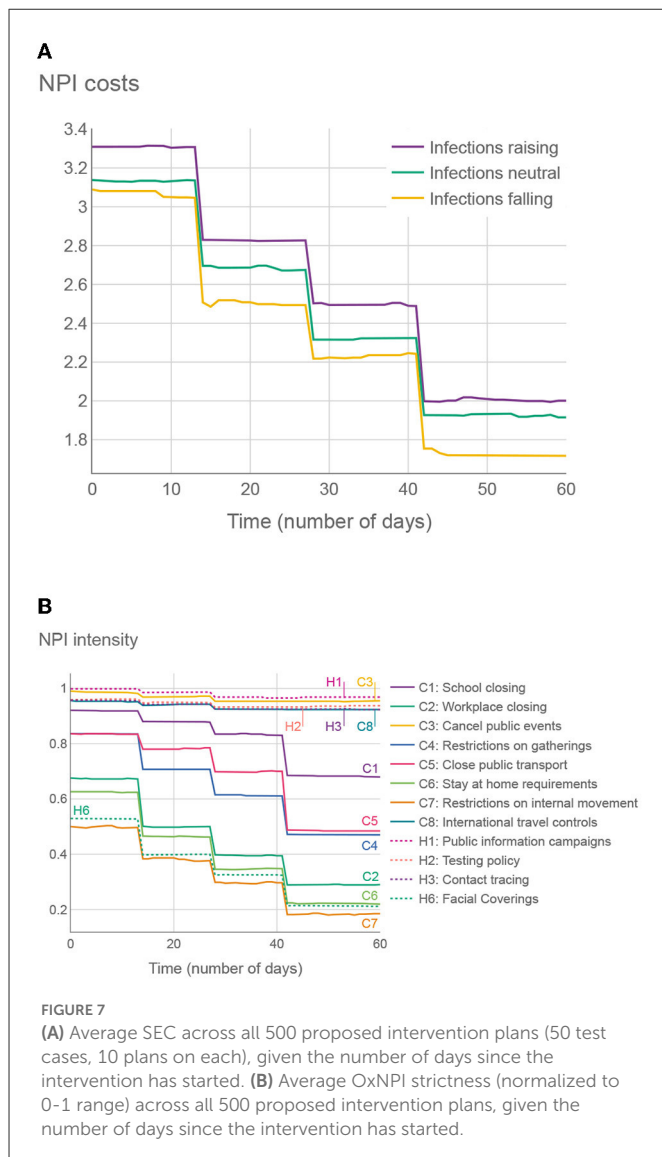**FIGURE 5**

(A) Hypervolume progress for different granularity values using condensed representation and **(B)** hypervolume progress for full and condensed representations with the best performing granularity values. A logarithmic scale is used for the horizontal axis (number of evaluations).

FIGURE 6

Comparing different intervention plans for France. **(A)** Shows the SEC (GDP loss + social cost) over time. **(B)** Shows the predicted number of infections, **(C)** Shows the trade-offs between the two criteria (SEC and the number of infections) for different plans.

this normalization, the model's coefficient magnitude can indicate relative NPI importance. Our model's most important intervention is the cancellation of public events, which is consistent with the related work that typically ranks it among the top NPIs (32). Next is school closure, which additionally results in some parents staying at home, so its importance is not surprising. These two are followed by contact tracing—which is difficult to execute well, and other sources do not rate this NPI as high. In the fourth place are international travel controls, which played a big role in some countries, particularly in the early stages of the pandemic. The importance of this NPI was corroborated by Haug et al. (32). Other NPIs have notably lower coefficient values. This may come as a surprise for "C2: Workplace closing," "C4: Restrictions on gatherings," and "C6: Stay at home requirements," but it should be noted that 1) these three NPIs have a large overlap with each other and with other NPIs, and 2) they were usually instituted when the epidemiological situation was grave, with many NPIs in force simultaneously, thus making it very difficult to
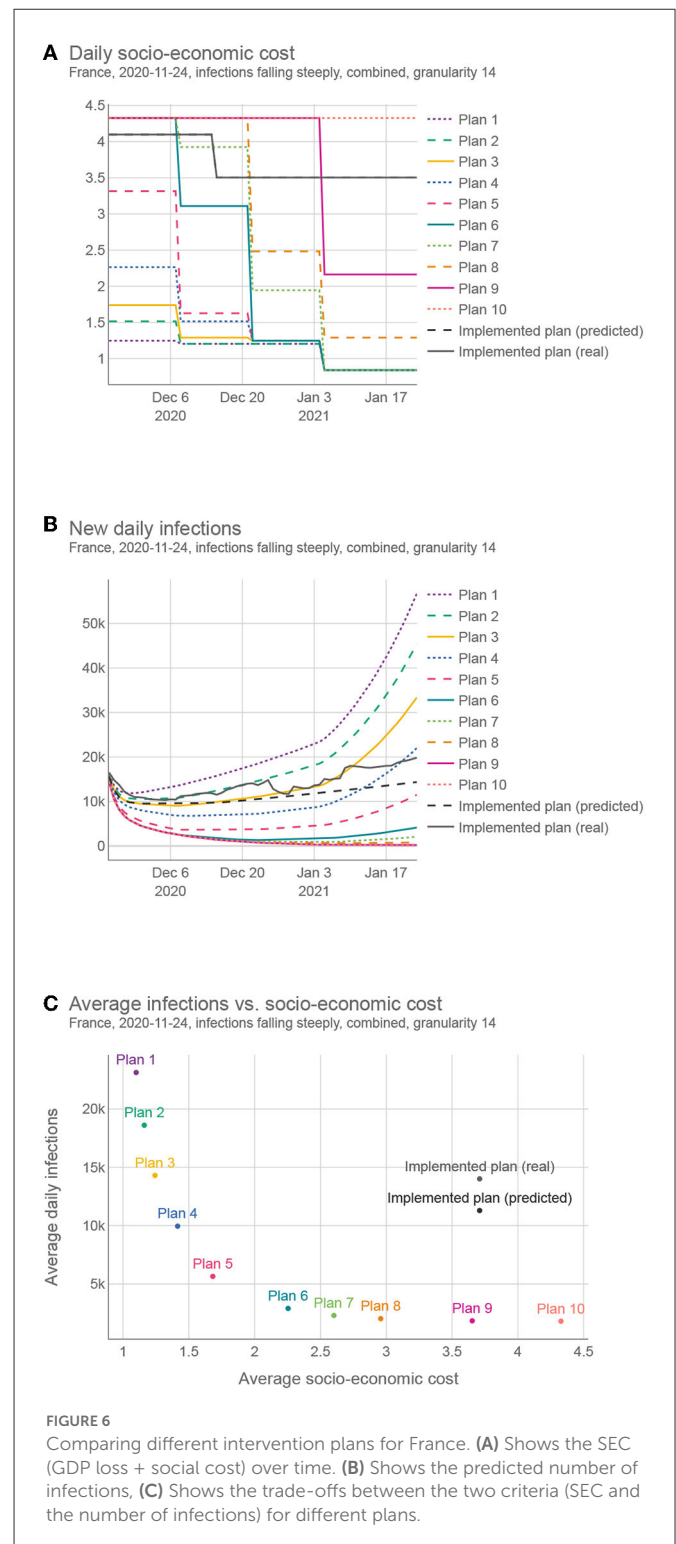
properly isolate the importance of each of them. This is why in these cases the assigned regression coefficient do not necessarily correctly reflect their relative importance. Nonetheless, their sum is close to the largest single coefficient. Of note, the NPI features were not the only ones included in the model, but the coefficient values of the others were an order of magnitude lower than those listed here.

Finally, for a direct comparison with related work, the HMLE model described here is an improved version of the one used in the

FIGURE 7
(A) Average SEC across all 500 proposed intervention plans (50 test cases, 10 plans on each), given the number of days since the intervention has started. **(B)** Average OxNPI strictness (normalized to 0-1 range) across all 500 proposed intervention plans, given the number of days since the intervention has started.

this reason, 50k evaluations represented a good trade-off between the framework's effectiveness and efficiency.

We tried to identify the best value for granularity and we compared five values: 1, 3, 7, 14, and 30. Theoretically, with a finer granularity, we can achieve at least as good intervention plans as with a coarser granularity. However, with finer granularity, aside from being impractical in real-life use, the search space of the optimization problem increases significantly, and the optimization cannot always find the best solutions. Then, we compared the two ways of representing intervention plans during optimization: full vs. condensed.

In all experiments in this section, the optimization was tested on 50 representative territory/time interval examples (see Section 2.1.). Due to the stochastic nature of the employed optimization approach, the presented results were obtained after running the optimization 31 times on each example, as this is enough to obtain statistically relevant results. To measure the effectiveness of the multi-objective optimization, we used the well-known hypervolume indicator (36)— the volume of the area bounded by the Pareto front approximation and a user-defined reference point. The medians of the obtained hypervolumes were used for testing the statistical significance of one granularity/representation being better than the other.

We first compared different granularity values when using the condensed representation. According to the Friedman test, we observed statistically significant differences between granularity values: $\chi^2(3) \approx 150.678$ and $p < 0.01$ for social weights, $\chi^2(3) \approx 119.309$ and $p < 0.01$ for GDP weights, and $\chi^2(3) \approx 106.139$ and $p < 0.01$ for combined weights. *Post-hoc* analysis with Wilcoxon signed-rank test and Holm's correction to adjust the p-values indicated that the granularity of 14 days was the most effective among the tested values (see Supplementary material—Details about the multi-objective optimization results).

Our results confirm that the optimization algorithm struggles to find near-optimal interventions plans with fine granularity values, due to the increase in search space dimensionality. For example, Figure 5 shows the hypervolume progress—the improvement of the results during the optimization—averaged over 31 optimization runs where the number of intervention plan evaluations was experimentally increased from the default 50 to 300 k. This was done to estimate the optimization behavior and convergence when using a large number of evaluations. As we can see, although the results obtained with a granularity value of 7 days eventually surpassed those results obtained with a granularity value of 14 days (at around 230 k evaluations), the computational time required to obtain better results using finer granularity values was almost five times longer, and the gain in the solutions' quality was negligible compared to the additional computational resources spent (Figure 5). In addition, the extremely small differences between the granularity value of 7 or 14 days are practically irrelevant since, in a real-world scenario, the objectives cannot be measured and predicted with such accuracy. Moreover, it is easier to implement intervention plans that change with coarse granularity values (37); therefore, a granularity value of 14 days seems to be a reasonable choice.

A similar investigation was devoted to finding the best granularity value for the full representation. The results of the statistical analysis revealed significant differences in hypervolume values and showed that the granularity of 30 days is the best performing value for this representation. The complete results can be found in the Supplementary material.

XPRIZE challenge, which was ranking between the 1st and 4th place during the 2 month prediction period on real data for 235 territories (35).

3.2. Proposing interventions

Figure 1 shows a sample trade-off intervention plan consisting of NPIs changing in time ($g = 14$), to provide a better intuition for the end goal of this work. It lists all 12 OxNPIs, their maximum value, and some sample values. For example, the intervention plan depicted in Figure 1 suggests to close all-but-essential workplaces from November 24, 2020, to December 20, 2020, but relaxes most countermeasures after that.

The experimental setup was established based on some initial experiments. NSGA-II was run with a population of 100 solutions for 500 generations (50k plan evaluations in total). This number of evaluations proved to be sufficient for convergence using coarser granularity values. Moreover, increasing function evaluations did not significantly improve the results, even for finer granularity values. For

Finally, we compared the full and condensed representations with the best performing granularity values. According to the Wilcoxon signed-rank test, the condensed representation outperformed the full representation for all types of weights ($p < 0.01$). Moreover, Figure 5 compares the hypervolume progress between the two representations on a typical problem instance, where a much faster convergence can be observed with the condensed representation. This was not unexpected since the applied optimization approach performs significantly faster for low-dimensional search spaces. The results provided in the following sections were obtained using the condensed representation with a granularity of 14 days since this was the best performing setting.

3.3. Intervention plan interpretation

To better understand how different intervention plans compare, we generated 10 different intervention plans for the same territory/time interval as that shown in Figure 1 (among all intervention plans obtained by the optimization, we selected the 10 that are the furthest from each other in the objective space). Figure 6 shows for each plan 1) the strictness of the interventions over time, 2) the resulting infection curve, and 3) the comparison of the 10 plans in terms of the number of infections and strictness. This example was done with the granularity of 14 days using the "combined" cost for the interventions. However, we generated plans using all different intervention costs and both 7 and 14 granularities for the same 50 test cases that were used for testing multi-objective optimization. This complete set of results can be found on the results webpage (38). For a subset of these results, see Supplementary material—Sample intervention plans.

The proposed plans present a wide range of trade-offs between the two objectives, and policy-makers can choose the one most suited to their needs. In addition, they can change a portion of the plan if deemed necessary and evaluate it again. This whole framework is available as a web tool (39), currently implemented for Slovenia.

The proposed solutions were compared with the real-life solution implemented in the same territory/time. This real-life solution was estimated in two ways, *(real)* using the actual reported number of infections and *(predicted)* using the predicted number of infections given the implemented NPIs. As the real SEC was, in most cases, unknown, we used the same estimation function for the *real* case as for the proposed plans. In all 50 test cases, the proposed solutions compared favorably against the *predicted* case, and in 47 test cases, the proposed solutions compared favorably against the *real* case. On average, we could find a solution with the same number of infections but with 47.1% lower SEC, or a solution with the same SEC but 68.8% lower number of infections (for details, see Supplementary material— Comparison of the proposed and implemented solutions).

To explore the trends in the structure of the intervention plans, we considered two experiments. First, we averaged the OxNPIs costs across all plans in all test examples, aggregated on a daily basis. The results in Figure 7 show that, on average, the intervention plans are the strictest at the beginning and then gradually become more relaxed. It also shows that in test intervals where the infections were falling, the overall strictness is lower than in cases where infections were raising. The difference might not be as big as expected,

again due to the optimizer providing a wide range of intervention plans.

In the second experiment in Figure 7, we show the average strictness of individual OxNPIs, again averaged across all intervention plans in all test cases. The NPIs with high average intensities can be considered to provide good trade-offs between their cost and effect.

The structure of the proposed plans was generally quite consistent from one territory to another. One can reason that—since the NPIs tend to have similar cost and benefit (at least in relative terms) regardless of the current epidemiological picture, and the prescriptor is designed to create solutions with a wide range of costs—the resulting plans will, in most cases, share a common structure that will be somewhat adjusted for different territories/time intervals. Another way of looking at it is to consider that reducing the number of infections when there are, for example, 1,000 daily infections has the same importance to the algorithm as reducing the number when there are 3,000 daily infections. It is up to the policy-maker to consider when the situation merits selecting a different proposed intervention plan with a lower/higher SEC.

4. Conclusion

The presented framework can generate efficient intervention plans to fight a pandemic, and can evaluate their effect and costs. This can greatly help policy-makers to pursue sensible intervention strategies and reason about their strengths and weaknesses. We showed that intervention plans it generates—at least when evaluated with our methodology—are better than past interventions generated by policy-makers. Since very few NPIs are still used against COVID-19, the main value of our framework is in pandemic preparedness: both as a tool to fight future pandemics (for which it would probably not require many modifications), and as a demonstration of the value of artificial intelligence in this area in general. All data used to generate the figures is available in our repository (38). The same repository also contains all final results. All code used in the production of the results is available in our code repository (40).

4.1. Intervention plan insights

In general, the most effective NPIs were school closing, canceling of public events, workplace closing, contact tracing, and international travel controls. This list is not surprising as it is similar to the findings in the literature (1, 2). When accounting for cost (which is usually not done), the most efficient NPIs were information campaigns, canceling of public events, and international travel controls, followed by school closing. The least efficient were the restrictions on internal movement, facial coverings, stay-at-home requirements and workplace closing. The latter two are on the list due to their high cost; in particular, the former can usually be substituted with a combination of other more socially acceptable NPIs. The low placement of facial coverings was surprising. Perhaps this is due to masks being somehow inconsistently applied, which may result in bad training data—or alternatively due to "facial coverings" NPI being almost always active, which made it difficult to isolate its effect. Finally, it could be the case that its social cost was overestimated in this study and it should be reduced in potential future analysis.

An additional benefit of the framework, aside from calculating the cost benefit of individual NPIs, is that it can present a timeline of NPI changes that adapts to the current epidemiological situation. In most cases, the approach "start with a strict policy and reduce it over time" seems to be the most effective. We have also shown that adapting the NPI policy every 14 days is enough to get almost ideal cost/benefit as with finer granularities (e.g., adapting every 3 days provides negligible benefits). Intervention plans made and changed on a monthly basis were found still acceptable; however, using a granularity value of 14 proved to be generally more robust. This could be a valuable finding as frequent changes in NPI policy make adherence difficult and can probably increase socio-economic costs (although we did not model this explicitly). For comparison, we analyzed how often were NPIs changing in real-life situations. For 80% of countries, the median time before changing at least one NPI was somewhere between 14 and 30 days and approximately 90% of countries changed their NPIs at least once, under 14 days of the last change.

4.2. Technical advantages

The following are the key innovations introduced: 1) combining machine-learning and SEIRD models in a way that allows the SEIRD parameters to be adapted to different NPIs and thus simulate their effect on infections; 2) using historically fitted parameters to normalize the values output by machine learning in order to adapt predictions for each territory; 3) using multi-objective optimization for finding the best intervention plans in combination with a "condensed" solution representation—facilitating a highly efficient search.

We argue our predictor to be state-of-the-art. However, it was designed and trained for the whole world, and it is almost certain that for many specific territories, a better predictor could be/was developed.

Similarly, while the proposed OxNPI costs are carefully considered, they can certainly be improved upon, especially for specific territories. In future work, the whole SEC model can even be made more complex, i.e., non-linearly accounting for the NPI duration. To take all of this in consideration, we made our whole methodology highly modular, so that each part can be substituted by a similar one if necessary—or one can simply adjust the parameter values of the current components.

4.3. Limitations

A drawback of the proposed framework is the negligible effect of vaccinations in the models. While we used some vaccination data, the vaccinations were not widespread at the time of data collection. This can be remedied in future work by using more recent data and probably adding another compartment that models vaccinations to the epidemiological model.

Second, the infection predictor can sometimes become unreliable when predicting for two or more months in advance. We thus recommend that it should be mostly used for shorter periods (30–45 days in advance) and then the predictions should be updated in real time as new data become available. The predictor also becomes unreliable when the number of infections is growing very quickly.

Due to the nature of exponential growth, even a small misprediction of a parameter of the SEIRD model can quickly lead the model astray. The problem is compounded by people spontaneously behaving more cautiously during severe disease breakouts, which affects the infections but is not recorded in NPI data. This effect is difficult to avoid, so it should be taken into consideration when analyzing the proposed plans. It should also be noted that infection prediction is used as the basis for NPI prescription (it is used to simulate the effects of different intervention plans), and thus any error in the former affects the latter. This effect is also difficult to avoid or even evaluate, as only one intervention plan can be executed at the same time in practice.

Models were made based on the COVID strains active in the studied period and would have to be slightly adjusted in order to be used for the currently emerging or future COVID strains.

Last, we used the reported number of the infections as one of the objectives—and one can argue that some other metric, such as the number of hospitalizations or deaths might be more appropriate. The hospitalizations were rejected in this study as the data needed was available for only 33% of the studied territories, while infections were preferred over deaths to match the Pandemic Response Challenge competition. Nonetheless, effectively the same methodology (with some tweaks to the epidemiological model) could be used to study the other mentioned criteria.

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

VJ, ML, and MG conceived the experiments. NR and DS collected and prepared the dataset. NR, AV, DS, and CD conducted the experiments. SV, DD, and JJ provided analysis of NPI costs. TT provided visualizations. VJ, AG, and ML analyzed the results. All authors reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

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.2023.1073581/full#supplementary-material

References

1. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. (2020) 584:257–61. doi: 10.1038/s41586-020-2405-7

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. Osterrieder A, Cuman G, Pan-Ngum W, Cheah PK, Cheah PK, Peerawaranun P, et al. Economic and social impacts of COVID-19 and public health measures: results from an anonymous online survey in Thailand, Malaysia, the UK, Italy and Slovenia. *BMJ Open*. (2021) 11:863. doi: 10.1136/bmjopen-2020-046863

4. Berger L, Berger N, Bosetti V, Gilboa I, Hansen LP, Jarvis C, et al. Rational policymaking during a pandemic. *Proc Natl Acad Sci USA*. (2021) 118:e2012704118. doi: 10.1073/pnas.2012704118

5. Lazzerini M, Putoto G. COVID-19 in Italy: momentous decisions and many uncertainties. *Lancet Glob Health*. (2020) 8:e641–2. doi: 10.1016/S2214-109X(20)30110-8

6. Yousefpour A, Hadi J, Stelios B. Optimal policies for control of the novel coronavirus disease (COVID-19) outbreak. *Chaos Solitons Fractals*. (2020) 136:109883. doi: 10.1016/j.chaos.2020.109883

7. Chen VCP, Zhou Y, Fallahi A, Viswanatha A, Yang J, Ghasemi Y, et al. An optimization framework to study the balance between expected fatalities due to COVID-19 and the reopening of U.S. Communities. *medRxiv*. (2020) doi: 10.1101/2020.07.16.20152033

8. Yaesoubi R, Havumaki J, Chitwood MH, Menzies NA, Gonsalves G, Salomon JA, et al. Adaptive policies to balance health benefits and economic costs of physical distancing interventions during the COVID-19 pandemic. *Med Decis Making*. (2021) 41:386–92. doi: 10.1177/0272989X21990371

9. XPRIZE.org. *Pandemic Response Challenge*. XPRIZE (2021). Available online at: ww. xprize.org/challenge/pandemicresponse (accessed February 1, 2023).

10. Miikkulainen R, Francon O, Meyerson E, Qiu X, Sargent D, Canzani E, et al. From prediction to prescription: evolutionary optimization of non-pharmaceutical interventions in the COVID-19 pandemic. *IEEE Trans Evolut Comput*. (2021) 25:386–401. doi: 10.1109/TEVC.2021.3063217

11. Lozano MA, Orts ÓGI, Piñol E, Rebollo M, Polotskaya K, Garcia-March MA, et al. Open Data science to fight COVID-19: winning the 500k XPRIZE pandemic response challenge. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Bilbao: Springer International Publishing (2021). p. 384–99.

12. Hale T, Webster S, Petherick A, Phillips T, Kira B. *Oxford COVID-19 Government Response Tracker (OxCGRT)*. (2020). Available online at: https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker

13. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. *Nat Hum Behav*. (2021) 5:947–53. doi: 10.1038/s41562-021-01122-8

14. Mullen JL, Tsueng G, Latif AA, Alkuzweny M, Cano M, Haag E, et al. *Outbreak.Info*. (2021). Available online at: https://outbreak.info/ (accessed on April 14, 2021).

15. Yuelong S, John M. GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill*. (2017) 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494

16. World Health Organization. *Tracking SARS-CoV-2 Variants*. (2021). Available online at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed on July 27, 2021).

17. Hasell J, Mathieu E, Beltekian D, Macdonald B, Giattino C, Ortiz-Ospina E, et al. A cross-country database of COVID-19 testing. *Sci Data*. (2020) 7:345. doi: 10.1038/s41597-020-00688-8

18. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B, et al. *Coronavirus Pandemic (COVID-19)*. Our World in Data (2020). Available online at: https://ourworldindata.org/coronavirus

19. Fan J, Li Y, Stewart K, Kommareddy AR, Garcia A, O'Brien J, et al. *The University of Maryland Social Data Science Center Global COVID-19 Trends and Impact Survey,*

in partnership with Facebook. (2021). Available online at: https://covidmap.umd.edu/api.html (accessed April 14, 2021).

20. Google LLC. *Google COVID-19 Community Mobility Reports*. (2020). Available online at: https://www.google.com/covid19/mobility/ (accessed April 14, 2021).

21. Apple. *COVID-19 Mobility Trends Reports*. (2020). Available online at: https://covid19.apple.com/mobility (accessed April 14, 2021).

22. Visual Crossing Corporation. *Visual Crossing Weather; 2017-2019*. Available online at: https://www.visualcrossing.com/weather-data (accessed April 14, 2021).

23. Janko V, Slapničar G, Dovgan E, Reščič N, Kolenik T, Gjoreski M, et al. Machine learning for analyzing non-countermeasure factors affecting early spread of COVID-19. *Int J Environ Res Public Health*. (2021) 18:750. doi: 10.3390/ijerph18136750

24. Martcheva M. An introduction to mathematical epidemiology. In: *Texts in Applied Mathematics*. New York, NY: Springer US (2015).

25. Thornburg H. *Introduction to Bayesian Statistics*. (2001). Available online at: http://ccrma.stanford.edu/jos/bayes/bayes.html

26. Li ML, Bouardi HT, Lami OS, Trikalinos TA, Trichakis NK, Bertsimas D. Forecasting COVID-19 and analyzing the effect of government interventions. *MedRxiv*. (2021). doi: 10.1101/2020.06.23.20138693

27. Zou D, Wang L, Xu P, Chen J, Zhang W, Gu Q. Epidemic model guided machine learning for COVID-19 forecasts in the United States. *MedRxiv*. (2020). doi: 10.1101/2020.05.24.20111989

28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12: 2825–30. doi: 10.5555/1953048.2078195

29. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. (2017) 30:3146–54.

30. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16*. New York, NY: ACM (2016). p. 785–94.

31. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS'18*. Red Hook, NY: Curran Associates Inc. (2018). p. 6639–49.

32. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav*. (2020) 4:1303–12. doi: 10.1038/s41562-020-01009-0

33. Worldbank. *GDP for the US*. (2021). Available online at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD (accessed October 20, 2021).

34. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evolut Comput*. (2002) 6:182–97. doi: 10.1109/4235.996017

35. XPRIZE. *Predictor Model Results as of Monday 22 February 2021*. (2021). Available online at: https://phase1.xprize.evolution.ml/

36. Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans Evol Comput*. (1999) 3:257–71. doi: 10.1109/4235.797969

37. Shen Y, Powell G, Ganser I, Zheng Q, Grundy C, Okhmatovskaia A, et al. Monitoring non-pharmaceutical public health interventions during the COVID-19 pandemic. *Sci Data*. (2021) 8:225. doi: 10.1038/s41597-021-01001-x

38. JSI. *Results Repository*. (2021). Available online at: https://github.com/jsi-dis/ai-covid-interventions

39. JSI. *COVID-19 Intervention Plans-A Web Application for the Ministry of Health*. (2021). Available online at: http://xprize-e9.ijs.si:5555/

40. JSI. *Code Repository*. (2021). Available online at: https://repo.ijs.si/covid/covid-xprize/

Mathematical epidemiological modeling and analysis of monkeypox dynamism with non-pharmaceutical intervention using real data from United Kingdom

Mercy Ngungu[1], Emmanuel Addai[2,3], Adejimi Adeniji[4], Umar Muhammad Adam[5] and Kayode Oshinubi[6]*

[1]Human Sciences Research Council (HSRC), Pretoria, South Africa, [2]Department of Biomedical Engineering, College of Biomedical Engineering, Taiyuan University of Technology, Taiyuan, China, [3]Department of Mathematics, Taiyuan University of Technology, Taiyun, China, [4]Department of Mathematics, Tshwane University of Technology, Pretoria, South Africa, [5]Department of Mathematics, Federal University, Dutse, Nigeria, [6]AGEIS Laboratory, University Grenoble Alpes, Saint Martin d'Hères, France

In this study, a mathematical model for studying the dynamics of monkeypox virus transmission with non-pharmaceutical intervention is created, examined, and simulated using real-time data. Positiveness, invariance, and boundedness of the solutions are thus examined as fundamental features of mathematical models. The equilibrium points and the prerequisites for their stability are achieved. The basic reproduction number and thus the virus transmission coefficient $\mathfrak{R}_0$ were determined and quantitatively used to study the global stability of the model's steady state. Furthermore, this study considered the sensitivity analysis of the parameters according to $\mathfrak{R}_0$. The most sensitive variables that are important for infection control are determined using the normalized forward sensitivity index. Data from the United Kingdom collected between May and August 2022, which also aid in demonstrating the usefulness and practical application of the model to the spread of the disease in the United Kingdom, were used. In addition, using the Caputo–Fabrizio operator, Krasnoselskii's fixed point theorem has been used to analyze the existence and uniqueness of the solutions to the suggested model. The numerical simulations are presented to assess the system dynamic behavior. More vulnerability was observed when monkeypox virus cases first appeared recently as a result of numerical calculations. We advise the policymakers to consider these elements to control monkeypox transmission. Based on these findings, we hypothesized that another control parameter could be the memory index or fractional order.

KEYWORDS

Caputo-Fabrizio fractional derivative, reproduction number, parameter estimation, numerical scheme, data fitting

1. Introduction

The unexpected breakout and global spread of monkeypox have drawn the attention of scientists due to the continuing COVID-19 pandemic. The prevalence of the largest and most pervasive monkeypox pandemic outside of Africa as of 22 June 2022, is 3,340 confirmed cases reported across the world. In addition to mother-to-child vertical transmission, the monkeypox virus can spread from person to person by direct contact with infectious skin or

mucosal skin lesions, respiratory droplets, or indirect contact with contaminated objects or materials. The possibility of community transmission cannot be ruled out, and it may also be sexually transferred by semen or vaginal fluid. The virus that causes monkeypox is called the monkeypox virus, and it is an enveloped, linear, double-stranded DNA virus that belongs to the Chordopoxvirinae subfamily of the Poxviridae family. With symptoms of the disease lasting 2–4 weeks and a death rate that previously ranged from 0 to 11 deaths, monkeypox is often a self-limiting sickness. Intense headaches, fever, lesions, and lymphadenopathy are some of the symptoms of monkeypox. Antiviral medications and smallpox vaccines have been approved for use in various nations in response to the monkeypox outbreak, despite the fact that there is no specific treatment or vaccine for monkeypox virus infection. Before allowing the virus to successfully establish person-to-person transmission, quick action is required to stop the local development of the disease and, consequently, the global monkeypox outbreak (1–11). In Peter et al. (12), modeling and optimal control were used to study monkeypox and the cost-effective strategies were investigated. This study shows that, among all competing measures, combining preventative measures to reduce rodent-to-human disease transmission is the most practical and cost-effective option.

Numerous research articles have been published where both classical and fractional models were constructed, and there is a plethora of literature on modeling infectious diseases. Because fractional-order derivative has unique properties such as heredity and memory that enable it to fully comprehend the dynamics of real phenomena, an analysis based on fractional-order derivative is more advantageous and practical than an analysis based on classical derivative (13, 14). At two separate closed locations, the phenomenon is indistinguishable by the standard derivatives. A generalized derivative known as the fractional order was proposed to address the problems with ordinary derivatives (15). Many researchers used fractional- order derivatives in many fields, as shown in Kumar et al. (16), Higazy et al. (17), Djida and Atangana (18), Baba (19), Owolabi and Atangana (20), Mohammadi et al. (21), Baleanu et al. (22), and Wutiphol and Turab (23). In the realm of mathematical biology, the Mittag–Leffler-type kernel has been used continuously over other derivatives, and numerous epidemiological models, such as for dengue fever, smoking, tuberculosis, measles, Ebola, and other diseases, have been studied using this operator as shown in Asamoah et al. (24), Peter et al. (25, 26), Kumar et al. (27), Morales-Delgadoa et al. (28), Atangana and Baleanu (29), and Atangana et al. (30). Most notably, in Zhang et al. (31), the Mittag–Leffler-type kernel modeling for Ebola–malaria co-infection was investigated by the authors with the best possible control. They strongly recommended the Mittag–Leffler-type kernel. In Kumar et al. (32), investigated the COVID-19 model using singular and non-singular fractional operators and compared the results of these operators. In Aslam et al. (33), the authors examined a recent study on the mathematical modeling of HIV/AIDS using the Mittag–Leffler-type kernel and came to the conclusion that the infection rate decreases with decreasing operator. In Evirgen (34), the authors studied the transmission dynamics of the Nipah virus using the Caputo derivative. One of the interesting segments of their study was to focus on tracing the influence of fractional-order derivatives on the manner in which the model responds. In Ucar (35), the

authors investigated a fractional SAIDR model within the framework of the Mittag–Leffler-type kernel. The effectiveness of the fractional operator is shown through a numerical simulation.

Considering the characteristics of exponential decay, the Caputo–Fabrizio fractional-order operator has been preferred over Atangana–Beleanu beta derivatives and a few other operators in the field of mathematical biology with more information (17–19, 24, 36–39). For instance, in Addai et al. (40), the authors studied a novel model of COVID-19 incorporating Alzheimer's disease using the Caputo–Fabrizio fractional-order operator. The results of the aforementioned study revealed that the two diseases have a link and the authors also concluded that the fractional operator is related to the rate of infection. In Shaikh and Nisar (41), the authors also considered the transmission dynamics of a fractional-order typhoid fever model using the Caputo–Fabrizio operator and the existence theory and achieved numerical solutions. In Shah et al. (42), Shah and his co-authors conducted a semi-analytical study of the Pine Wilt Disease (PWD) model with a convex rate *via* fractional order involving a non-singular kernel. To comprehend the trade-off between the lockdown and the transmission of the virus, Ahmed and his co-authors devised a five-term dynamical system (43). Another use of the Caputo–Fabrizio fractional-order operator was indicated, for instance, in Addai et al. (40), Shaikh and Nisar (41), Shah et al. (42), Ahmed et al. (43), Ullah et al. (44), Abboubakar et al. (45).

Furthermore, in Peter et al. (46), the authors used real data from Nigeria to study the dynamics of the transmission of the monkeypox virus using fractional calculus. The authors presented an argument on the modeling system by studying the infection control policies that will help the public to better understand the significance of control parameters in the eradication of the virus in the studied population. Furthermore, the transmission dynamics of the monkeypox virus was studied using a mathematical modeling approach in Peter et al. (47). In their findings, the authors indicated that the isolation of infected individuals in the human population helps reduce the transmission of the disease, which can serve as a form of intervention to control the spread of the virus.

We observed that none of the studies on the monkeypox virus and its modes of transmission took into account the interaction between the isolated and exposed compartments in the human subpopulation and the results of that contact rate with the rodent population and applied the modeling approach to real data from the United Kingdom. The major goals of this research are to calculate the exponential growth rate of the monkeypox virus, to forecast what might occur in future and how to stop it from spreading, and to understand the effects of non-pharmaceutical intervention on infected individuals, which will be able to guide us on how to deploy intervention resources to contain the spread of the disease. The remaining sections of the article are structured as follows: Section 2 presents some basic definitions and preliminary information, Section 3 presents the model formulation, Sections 4 deals with the dynamism of the model, Section 5 computes the basic reproduction number and some basic mathematical analysis, Section 6 present the endemic equilibrium of the model, Section 7 proves the existence and uniqueness of our model, Section 8 deals with the fitting of the model to real data from the United Kingdom, Section 9 presents numerical schemes and numerical simulations, Section 10 deals with sensitivity analysis, and Section 11 provides some perspectives, discussion, and conclusion.

2. Preliminaries

In this section, we review several key definitions, lemmas, and concepts that are necessary to understand the suggested model.

Definition 2.1 Let $f \in Q^1(p,q), q > p$, and $\alpha_* \in (0,1)$ (17), (40). Then, the Caputo–Fabrizio fractional-order derivative can be defined as

$$_p^{CF}D_t^\alpha f(t) = \frac{G(\alpha)}{1-\alpha} \int_p^t f'(x) exp\left[-\alpha \frac{t-s}{1-\alpha}\right]ds.$$

Here, $G(\alpha)$ is a normalization function, where $G(0) = G(1) = 1$. The fractional integral of the Caputo–Fabrizio fractional order is defined by:

$$I_t^\alpha f(t) = \frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}f(t) + \frac{2\alpha}{(2-\alpha)G(\alpha)} \int_0^t f(s)ds, t \geq 0.$$

Lemma 2.2 Assuming there is a function $u(t) \in W_l[0,\eta]$, then the solution of fractional differential equation

$$\begin{cases} ^{CF}D_t^\alpha f(t) = u(t), t \in [0,\eta], \\ f(0) = f_0, \end{cases}$$

is given by

$$f(t) = f_0 + \frac{2(1-\alpha)}{2(1-\alpha)G(t)}f(t) + \frac{2\alpha}{(2-\alpha)G(\alpha)} \int_0^t f(s)ds, t \geq 0$$

(24),(17), (40).

Lemma 2.3 Suppose $A \subset B$ be a closed convex non-empty subset of A and there exist two operators, T_1 and T_2, then it is Krasnoselskii's fixed point theorem (40) and it follows that:

(i) $T_1u + T_1u \in A, \forall u \in A$;
(ii) T_1 is contraction and T_2 continuous and compact. Then quantify at least one solution $u \in A$ such that

$$T_1u + T_2u = u.$$

3. Model formulation

Using a system of differential equations, we studied both human and rodent populations in a closed homogeneous environment. There are five compartments in a human population of size $N_h(t)$: Susceptible $S_h(t)$; Exposed $E_h(t)$; Infected $I_h(t)$; Isolation/Quarantine $Q_h(t)$; and Recovered $R_h(t)$; where $N_h(t) = S_h(t) + E_h(t) + I_h(t) + Q_h(t) + R_h(t)$. The rodent population $N_r(t)$ is split into $S_r(t)$ Susceptible; $E_r(t)$ Exposed; and $I_r(t)$ Infected. Let $N_r(t) = S_r(t) + E_r(t) + I_r(t)$. From the aforementioned description, using the ideas in Yinka-Ogunleye et al. (5), we extend the studies of Peter et al. (46) and (47), then the ordinary differential equations in system (1) describe the dynamics of monkeypox transmission incorporating

TABLE 1 Interpretation of parameters in the model.

Parameter	Interpretation
Λ_h	Human recruitment rate
Λ_r	Rodent recruitment rate
ξ_h	Immunity loss rate for human
θ_h	Undetected rate of human after diagnosis
μ_h, μ_r	Natural death rate for humans and rodents
ν_h, ν_r	Disease-induced death rate for humans and rodents
ϕ_h, ϕ_r	The rate at which humans and rodents move from exposed to infectious stage
ψ_h	The rate of humans recovery from monkeypox
γ_h	The rate of identifying as suspected case of monkeypox
δ_h	The rate of moving from isolated to recovered class
β_{rh}	The rate of transmission within rodents and humans
β_{hh}	The rate of transmission within humans
β_{rr}	The rate of transmission within rodents

non-pharmaceutical intervention;

$$\begin{cases} \frac{dS_h}{dt} = \Lambda_h + \xi_h R_h + \theta_h Q_h - \lambda_h S_h - \mu_h S_h, \\ \frac{dE_h}{dt} = \lambda_h S_h - \gamma_h E_h - \phi_h E_h - \mu_h E_h, \\ \frac{dI_h}{dt} = \phi_h E_h - (\psi_h + \mu_h + \nu_h)I_h, \\ \frac{dQ_h}{dt} = \gamma_h E_h - (\theta_h + \delta_h + \mu_h + \nu_h)Q_h, \\ \frac{dR_h}{dt} = \psi_h I_h + \delta_h Q_h - \xi_h R_h - \mu_h R_h, \\ \frac{dS_r}{dt} = \Lambda_r - \lambda_r S_r - \mu_r S_r \\ \frac{dE_r}{dt} = \lambda_r S_r - \phi_r E_r - \mu_r E_r, \\ \frac{dI_r}{dt} = \phi_r E_r - (\mu_r + \nu_r)I_r, \end{cases} \quad (1)$$

where $\lambda_h = \frac{\beta_{rh}I_r + \beta_{hh}I_h}{N_h}$, $\lambda_r = \frac{\beta_{rr}I_r}{N_r}$. To capture the memory in the predictions of the monkeypox virus transmission model and also to verify that both sides of the fractional equations have exact dimensions, the time-dependent kernel is defined by the power law correlation function, as in Tilahuna et al. (48); therefore, we propose the following fractional-order model for the monkeypox virus transmission model using the Caputo–Fabrizio fractional-order derivative;

$$\begin{cases} ^{CF}D_t^\alpha S_h(t) = \Lambda_h + \xi_h R_h + \theta_h Q_h - \lambda_h S_h - \mu_h S_h, \\ ^{CF}D_t^\alpha E_h(t) = \lambda_h S_h - \gamma_h E_h - \phi_h E_h - \mu_h E_h, \\ ^{CF}D_t^\alpha I_h(t) = \phi_h E_h - (\psi_h + \mu_h + \nu_h)I_h, \\ ^{CF}D_t^\alpha Q_h(t) = \gamma_h E_h - (\theta_h + \delta_h + \mu_h + \nu_h)Q_h, \\ ^{CF}D_t^\alpha R_h(t) = \psi_h I_h + \delta_h Q_h - \xi_h R_h - \mu_h R_h, \\ ^{CF}D_t^\alpha S_r(t) = \Lambda_r - \lambda_r S_r - \mu_r S_r \\ ^{CF}D_t^\alpha E_r(t) = \lambda_r S_r - \phi_r E_r - \mu_r E_r, \\ ^{CF}D_t^\alpha I_r(t) = \phi_r E_r - (\mu_r + \nu_r)I_r. \end{cases} \quad (2)$$

The flow diagram of the model equation is presented in Figure 1 while the parameters used in the model and their signification is presented in Table 1.

4. Dynamics of the model

In this section, we focus on the dynamics of the solutions for the suggested models (1) and (2) that are positive, bounded,
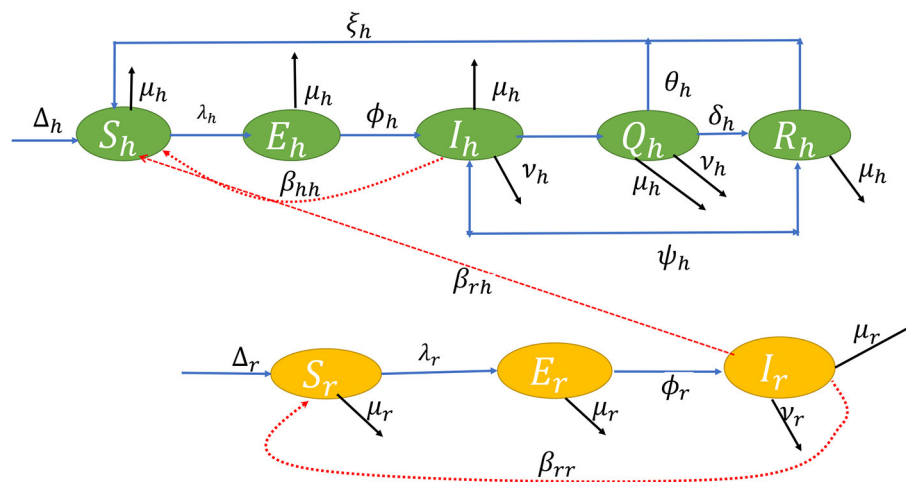
FIGURE 1
Transfer diagram of the dynamic transmission of the monkeypox virus.

and invariant. In an epidemiological model, it is important to evaluate the population survival and the expansion that is naturally constrained by scarce resources. As a result, we demonstrate the following theorem.

Theorem 1. The solution of (1) along with initial conditions is positively invariant and bounded in R_+^8. Therefore,

$$
\begin{cases}
\lim_{t \to \infty} \sup S_h(t) \le S_{h\infty} = \frac{\Lambda_h + \theta_h Q_{h\infty} + \xi_h R_{h\infty}}{\lambda_h + \mu_h}, \\
\lim_{t \to \infty} \sup E_h(t) \le E_{h\infty} = \frac{\lambda_h S_{h\infty}}{(\xi_h + \phi_h + \mu_h)}, \\
\lim_{t \to \infty} \sup I_h(t) \le I_{h\infty} = \frac{\phi_h E_{h\infty}}{(\psi_h + \eta_h + \mu_h)}, \\
\lim_{t \to \infty} \sup Q_h(t) \le Q_{h\infty} = \frac{\gamma_h E_{h\infty}}{(\theta_h + \delta_h + \mu_h + \nu_h)}, \\
\lim_{t \to \infty} \sup R_h(t) \le R_{h\infty} = \frac{\psi_h I_{h\infty} + \delta_h Q_{h\infty}}{(\xi_h + \mu_h)}, \\
\lim_{t \to \infty} \sup S_r(t) \le S_{r\infty} = \frac{\Lambda_r}{\lambda_r + \mu_r}, \\
\lim_{t \to \infty} \sup E_r(t) \le E_{r\infty} = \frac{\lambda_r S_{r\infty}}{(\phi_r + \mu_r)}, \\
\lim_{t \to \infty} \sup I_r(t) \le I_{r\infty} = \frac{\phi_r E_{r\infty}}{(\nu_r + \mu_r)}.
\end{cases}
\tag{3}
$$

Proof. Using the results in Lin (49) and taking into account the initial values given, from model (2), we obtain

$$
\begin{cases}
{}^{CF}D_t^\alpha S_h(t)|_{S_h(0)} = \Lambda_h + \xi_h R_h + \theta_h Q_h \ge 0, \\
{}^{CF}D_t^\alpha E_h(t)|_{E_h(0)} = \lambda_h S_h \ge 0, \\
{}^{CF}D_t^\alpha I_h(t)|_{I_h(0)} = \phi_h E_h \ge 0, \\
{}^{CF}D_t^\alpha I_h(t)|_{Q_h(0)} = \gamma_h E_h \ge 0, \\
{}^{CF}D_t^\alpha R_h(t)|_{R_h(0)} = \psi_h I_h + \delta_h Q_h \ge 0, \\
{}^{CF}D_t^\alpha S_v(t)|_{S_r(0)} = \Lambda_r \ge 0, \\
{}^{CF}D_t^\alpha E_v(t)|_{E_r(0)} = \lambda_r S_r \ge 0, \\
{}^{CF}D_t^\alpha I_v(t)|_{I_r(0)} = \phi_r E_r \ge 0.
\end{cases}
\tag{4}
$$

From Equation (4), we can see that $S_h(0) > 0, E_h(0) > 0, I_h(0) > 0, R_h(0) > 0, S_v(0) > 0, E_v(0) > 0, I_v(0) > 0$, for all $t > 0$. From Equation (2), the first equation gives

$$
{}^{CF}D_t^\alpha S_h(t) \le \Lambda_h + \xi_h R_h + \theta_h Q_h - \lambda_h S_h - \mu_h S_h \ge 0.
$$

Then, by applying the fractional comparison technique, we obtain the first estimate of Equation (4). We continue for the second equation of the system of Equation (2), we obtain

$$
{}^{CF}D_t^\alpha E_h(t) \le \lambda_h S_h - \gamma_h E_h - \phi_h E_h - \mu_h E_h \ge 0.
$$

Therefore, we get the second estimate of Equation (1). We continue again for the third equation of the system of Equation (2), we obtain

$$
{}^{CF}D_t^\alpha I_h(t) \le \phi_h E_h - (\psi_h + \mu_h + \nu_h) I_h \ge 0,
$$

and, consequently, we obtain the third estimate of Equation (4). Similarly, for the fourth to eighth equation, we obtain the estimate of Equation (4). Hence, Theorem 1 is complete.

4.1. Monkeypox equilibrium state

The monkeypox model is studied by obtaining the equilibrium states. To verify the existence of the equilibrium points, the derivatives of the model on the right-hand side are set to zero, which provides the monkeypox disease free equilibrium points.

We assume $E_h, E_r, I_h, I_r, Q_h, R_h, S_h, S_r$ be the solution to the monkeypox model with the initial condition in a feasible region such that

$$
\Gamma_h = E_h, I_h, Q_h, R_h, S_h \in \mathbb{R}^5 : N_h = \frac{\Lambda_h}{\mu_h},
\tag{5}
$$

$$
\Gamma_r = E_r, I_r, S_r \in \mathbb{R}^3 : N_r = \frac{\Lambda_r}{\mu_r},
\tag{6}
$$

where the human population is represented as

$$
N_h = E_h(t) + I_h(t) + Q_h(t) + R_h(t) + S_h(t),
\tag{7}
$$

and the rodent population, respectively,

$$
N_r = E_r(t) + I_r(t) + S_r(t).
\tag{8}
$$

To achieve the disease-free equilibrium state, the derivatives are set to zero as seen in (10) to obtain

$$
E^* = \left(E_h^*, E_r^*, I_h^*, I_r^*, Q_h^*, R_h^*, S_h^*, S_r^* \right).
\tag{9}
$$

By setting the derivatives to zero, we obtain

$$\frac{dE_h}{dt} = \frac{dE_r}{dt} = \frac{dI_h}{dt} = \frac{dI_r}{dt} = \frac{dQ_h}{dt} = \frac{dR_h}{dt} = \frac{dS_h}{dt} = \frac{dS_r}{dt} = 0; \tag{10}$$

hence, Equation (9) is represented as

$$E^* = \left(0, 0, 0, 0, 0, 0, \frac{\Lambda_h}{\mu_h}, \frac{\Lambda_r}{\mu_r}\right). \tag{11}$$

This equation describes a population free of monkeypox infection and is denoted as E^*

5. The basic reproduction number

We derive the basic reproduction number $\mathfrak{R}_0$ by using the next-generation matrix approach (25). Since E_h, I_h, Q_h, and I_r are the disease-infected classes, hence,

$$f = \begin{pmatrix} 0 \\ \lambda_h S_h \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad v = \begin{pmatrix} -\Lambda_h - \xi_h R_h - \theta_h Q_h + \lambda_h S_h + \mu_h S_h \\ \gamma_h E_h + \phi_h E_h + \mu_h E_h \\ -\phi_h E_h + (\psi_h + \mu_h + \nu_h)I_h \\ -\gamma_h E_h + (\theta_h + \delta_h + \mu_h + \nu_h)Q_h \\ -\psi_h I_h - \delta_h Q_h + \xi_h R_h + \mu_h R_h \\ -\Lambda_r + \lambda_r S_r + \mu_r S_r \\ -\lambda_r S_r + \phi_r E_r + \mu_r E_r \\ -\phi_r E_r + (\mu_r + \nu_r)I_r \end{pmatrix}. \tag{12}$$

$$F = \begin{pmatrix} 0 & 0 & \frac{\beta_{hh}\Lambda_h}{\mu_h} & 0 & \frac{\beta_{hh}\Lambda_h}{\mu_h} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$V = \begin{pmatrix} \gamma_h + \phi_h + \mu_h & 0 & 0 & 0 \\ -\phi_h & \psi_h + \mu_h + \nu_h & 0 & 0 \\ -\gamma_h & 0 & \theta_h + \delta_h + \mu_h + \nu_h & 0 \\ 0 & 0 & 0 & \mu_r + \nu_r \end{pmatrix}. \tag{13}$$

$$V^{-1} = \begin{bmatrix} \frac{1}{\gamma_1+\mu_1+\phi_1} & 0 & 0 & 0 \\ \frac{\phi_1}{\gamma_1\mu_1+\gamma_1\nu_1+\gamma_1\psi_1+\mu^2+\nu_1\mu_1+\mu_1\phi_1+\mu_1\psi_1+\nu_1\phi_1+\psi_1\phi_1} & \frac{1}{\nu_1+\mu_1+\psi_1} & 0 & 0 \\ \frac{\gamma_1}{\gamma_1\delta_1+\delta_1\mu_1+\delta_1\phi_1+\gamma_1\mu_1+\nu_1\gamma_1+\gamma_1\theta_1+\mu_1^2+\mu_1\nu_1+\mu_1\phi_1+\mu_1\theta_1+\nu_1\phi_1+\theta_1\phi_1} & 0 & \frac{1}{\nu_1+\mu_1+\theta_1+\delta_1} & 0 \\ 0 & 0 & 0 & \frac{1}{\nu_2+\mu_2} \end{bmatrix}. \tag{14}$$

The next-generation matrix (G) is given by

$$G = F.V^{-1} = \begin{bmatrix} \frac{\beta_1\lambda_1\phi_1}{N_h\mu_1(\gamma_1\mu_1+\gamma_1\nu_1+\gamma_1\psi_1+\mu^2+\nu_1\mu_1+\mu_1\phi_1+\mu_1\psi_1+\nu_1\phi_1+\psi_1\phi_1)} & \frac{\beta_1\lambda_1}{N_h\mu_1(\nu_1+\mu_1+\psi_1)} & 0 & \frac{\beta_1\lambda_1}{N_h\mu_1(\nu_1+\mu_1)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{15}$$

The basic reproduction number $\mathfrak{R}_0$ is the dominant eigenvalue (spectral radius) of the next-generation matrix G, that is, $\mathfrak{R}_0 = \rho(G)$.

$$\mathfrak{R}_0 = \frac{\beta_{hh}\Lambda_h\phi_h}{\mu_h(\gamma_h + \phi_h + \mu_h)(\psi_1 + \mu_h + \nu_h)}$$

5.1. Stability of monkeypox-free equilibrium (MFE)

Investigating the stability of the monkeypox disease-free equilibrium, we compute the Jacobian matrix of the system at the disease-free equilibrium by obtaining the eigenvalues, which will be used to determine the stability of the model.

$$J_{E^*} =$$
$$\begin{pmatrix} -\frac{\beta_{hh}I_h+\beta_{rh}I_r}{N_h}-\mu_h & 0 & 0 & \theta_h & \xi_h & 0 & 0 & -\frac{(\beta_{hh}+\beta_{rh})S_h}{N_h} \\ \frac{\beta_{hh}I_h+\beta_{rh}I_r}{N_h} & \zeta_1 & 0 & 0 & 0 & 0 & 0 & \frac{(\beta_{hh}+\beta_{rh})S_h}{N_h} \\ 0 & \varphi_h & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_h & 0 & \zeta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_h & -\mu_h-\xi_h & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{\beta_{rr}I_r}{N_r}-\mu_r & 0 & -\frac{\beta_{rr}S_r}{N_r} \\ 0 & 0 & 0 & 0 & 0 & \frac{\beta_{rr}I_r}{N_r} & -\varphi_r-\mu_r & \frac{\beta_{rr}S_r}{N_r} \\ 0 & 0 & 0 & 0 & 0 & 0 & \varphi_r & -\mu_r-\nu_r \end{pmatrix}, \tag{16}$$

where ζ_1 and ζ_2 are represented in Equations (17) and (18)

$$\zeta_1 = -\gamma_h - \varphi_h - \mu_h, \tag{17}$$
$$\zeta_2 = -\theta_h - \delta_h - \mu_h - \nu_h. \tag{18}$$

Evaluating J_{E^*} at the monkeypox-free equilibrium (MFE), we obtain

$$J_{MFE^*}$$
$$= \begin{pmatrix} -\mu_h & 0 & 0 & \theta_h & \xi_h & 0 & 0 & -\frac{(\beta_{hh}+\beta_{rh})\Lambda_h}{N_h\mu_h} \\ 0 & -\gamma_h-\varphi_h-\mu_h & 0 & 0 & 0 & 0 & 0 & \frac{(\beta_{hh}+\beta_{rh})\Lambda_h}{N_1\mu_h} \\ 0 & \varphi_h & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_h & 0 & -\theta_h-\delta_h-\mu_h-\nu_h & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_h & -\mu_h-\xi_h & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\mu_r & 0 & -\frac{\beta_{rr}\Lambda_r}{N_r\mu_r} \\ 0 & 0 & 0 & 0 & 0 & 0 & -\varphi_r-\mu_r & \frac{\beta_{rr}\Lambda_r}{N_r\mu_r} \\ 0 & 0 & 0 & 0 & 0 & 0 & \varphi_r & -\mu_r-\nu_r \end{pmatrix}. \tag{19}$$

We compute the eigenvalues from the J_{MFE^*} using the characteristic polynomial of O^8, which will not be represented as a result of its lengthiness. The eigenvalues and characteristic polynomial are calculated by $|J_{MFE^*} - I|$, where I is an 8×8 unit matrix, and the values of λ

are obtained:

$$\lambda = \begin{pmatrix} 0 \\ -\mu_h \\ -\mu_h - \xi_h \\ -\mu_r \\ -\gamma_h - \varphi_h - \mu_h \\ -\theta_h - \delta_h - \mu_h - \nu_h \\ \frac{-2N_r\mu_r^2 + (-\nu_r - \varphi_r)N_r\mu_r + \sqrt{\mu_r(N_r(\nu_r - \varphi_r)^2\mu_r + 4\Lambda_r\beta_{rr}\varphi_r)N_r}}{2N_r\mu_r} \\ \frac{-2N_r\mu_r^2 + (-\nu_r - \varphi_r)N_r\mu_r - \sqrt{\mu_r(N_r(\nu_r - \varphi_r)^2\mu_r + 4\Lambda_r\beta_{rr}\varphi_r)N_r}}{2N_r\mu_r} \end{pmatrix} \quad (20)$$

Let Δ_1 *and* Δ_2 be well represented from Equation (20) in Equations (21) and (22)

$$\Delta_1 = \mu_r(N_r(\nu_r - \varphi_r)^2\mu_r + 4\Lambda_r\beta_{rr}\varphi_r)N_r, \quad (21)$$

$$\Delta_2 = 2N_r\mu_r^2 + (-\nu_r - \varphi_r)N_r\mu_r. \quad (22)$$

Then,

$$\lambda_1 = 0, \quad (23)$$

$$\lambda_2 = -\mu_h, \quad (24)$$

$$\lambda_3 = -(\mu_h + \xi_h), \quad (25)$$

$$\lambda_4 = -\mu_r, \quad (26)$$

$$\lambda_5 = -(\gamma_h + \varphi_h + \mu_h), \quad (27)$$

$$\lambda_6 = -(\theta_h + \delta_h + \mu_h + \nu_h), \quad (28)$$

$$\lambda_7 = \frac{-\Delta_2 + \sqrt{\Delta_1}}{2N_r\mu_r}, \quad (29)$$

$$\lambda_8 = \frac{-\Delta_2 - \sqrt{\Delta_1}}{2N_r\mu_r}. \quad (30)$$

From the calculated eigenvalues, we obtain negative real parts, that is, the monkeypox-free equilibrium is asymptotically stable if

$$\frac{-\Delta_2 - \sqrt{\Delta_1}}{2N_r\mu_r} < 0. \quad (31)$$

Upon simplification, we obtain Equation (31):

$$\frac{N_r\mu_r(2\mu_r - \nu_r - \varphi_r)^2}{N_r\mu_r(\nu_r - \varphi_r)^2 + 4\Lambda_r\beta_{rr}\varphi_r} < 1. \quad (32)$$

Therefore, the monkeypox-free equilibrium state is asymptotically stable.

5.2. Global stability of the equilibrium state

If $\Re_0 < 1$, then the monkeypox-free equilibrium is globally asymptotically stable; otherwise, it is unstable. This is proven by the Lyapunov function such that

$$L(E_h) = E_h \quad (33)$$

Differentiating, we obtain

$$L'(E_h) = E_h' \quad (34)$$

$$= \lambda_h S_h - \gamma_h E_h - \phi_h E_h - \mu_h E_h \quad (35)$$

$$= \lambda_h S_h - (\gamma_h + \phi_h + \mu_h)E_h. \quad (36)$$

At the disease-free equilibrium state as seen in Equation (11), $S_h = \frac{\Lambda_h}{\mu_h}$,

$$L'(E_h) = \lambda_h\left(\frac{\Lambda_h}{\mu_h}\right) - (\gamma_h - \phi_h - \mu_h)E_h \quad (37)$$

$$E_h' = (\gamma_h + \phi_h + \mu_h)\left[\frac{\Lambda_h\lambda_h}{\mu_h(\gamma_h + \phi_h + \mu_h)E_h} - 1\right]E_h \quad (38)$$

$$E_h' = (\gamma_h + \phi_h + \mu_h)(\Re_0 - 1)E_h \leq 0 \text{ if } \Re_0 \leq 0. \quad (39)$$

From the result obtained in Equation (39), we can see that $E_h' \leq 0$ provided $\Re_0 \leq 0$ as well as $E_h' = 0$ provided that $\Re_0 = 0$ or $E_h = 0$. Global stability of the disease-free equilibrium is asymptotically stable, if $\Re_0 \leq 0$; otherwise, it is unstable.

6. Endemic equilibrium state

The endemic equilibrium state occurs when the rate of infection persists in the population and it is represented in Equations (40 - 47) by $E_h^{**}, E_r^{**}, I_h^{**}, I_r^{**}, Q_h^{**}, R_h^{**}, S_h^{**}, S_r^{**}$.

$$E_h^{**} = \frac{\left[\mu_h^3 + k_1\mu_h^2 + (k_2 + k_3)\mu_h + k_4\right]\Lambda_h\lambda_h}{\mu_h^5 + p_1\mu_h^4 + p_2 \cdot \mu_h^3 + p_3 \cdot \mu_h^2 + \mu_h \cdot p_4 + \lambda_h\nu_h\xi_h \cdot p_5} \quad (40)$$

$$E_r^{**} = \frac{\lambda_r\Lambda_r}{(\mu_r + \varphi_r)(\mu_r + \lambda_r)} \quad (41)$$

$$I_h^{**} = \frac{(\mu_h^2 + (\delta_h + \nu_h + \theta_h + \xi_h)\mu_h + \delta_h\xi_h + \nu_h\xi_h + \theta_h\xi_h)\Lambda_h\lambda_h\varphi_h}{\mu_h^5 + p_1\mu_h^4 + p_2 \cdot \mu_h^3 + p_3 \cdot \mu_h^2 + \mu_h \cdot p_4 + \lambda_h\nu_h\xi_h \cdot p_5 + p_6} \quad (42)$$

$$I_r^{**} = \frac{\varphi_r\lambda_r\Lambda_r}{\lambda_r\mu_r^2 + \lambda_r\mu_r\nu_r + \lambda_r\mu_r\varphi_r + \lambda_r\nu_r\varphi_r + \mu_r^3 + \mu_r^2\nu_r + \mu_r^2\varphi_r + \mu_r\nu_r\varphi_r} \quad (43)$$

$$Q_h^{**} = \frac{(\gamma_h\mu_h^2 + \gamma_h(\nu_h + \psi_h + \xi_h)\mu_h + \gamma_h(\nu_h\xi_h + \psi_h\xi_h))\Lambda_h\lambda_h}{\mu_h^5 + p_1\mu_h^4 + p_2 \cdot \mu_h^3 + p_3 \cdot \mu_h^2 + \mu_h \cdot p_4 + \lambda_h\nu_h\xi_h \cdot p_5 + p_6} \quad (44)$$

$$R_h^{**} =$$

$$\frac{(\delta_h\psi_h + \mu_h\psi_h + \nu_h\psi_h + \psi_h\theta_h)\lambda_h\Lambda_h\varphi_h + (\delta_h\gamma_h\mu_h + \delta_h\gamma_h\nu_h + \delta_h\gamma_h\psi_h)\lambda_h\Lambda_h}{\mu_h^5 + p_1\mu_h^4 + p_2 \cdot \mu_h^3 + p_3 \cdot \mu_h^2 + \mu_h \cdot p_4 + \lambda_h\nu_h\xi_h \cdot p_5 + p_6} \quad (45)$$

$$S_h^{**} = \frac{\Lambda_h \mu_h^4 + \Lambda_h \cdot h_1 \cdot \mu_h^3 + \Lambda_h \cdot h_2 \cdot \mu_h^2 + \Lambda_h \cdot h_3 \cdot \mu_h + \Lambda_h \cdot h_4}{\mu_h^5 + p_1 \mu_h^4 + p_2 \cdot \mu_h^3 + p_3 \cdot \mu_h^2 + \mu_h \cdot p_4 + \lambda_h \nu_h \xi_h \cdot p_5 + p_6} \quad (46)$$

$$S_r^{**} = \frac{\Lambda_r}{\lambda_r + \mu_r}, \quad (47)$$

where

$d_1 = (\psi_h + \nu_h),$

$d_2 = (\delta_h + \nu_h + \theta_h + \xi_h),$

$d_3 = (\delta_h + \nu_h + \theta_h),$

$k_1 = (\psi_h + 2\nu_h + \delta_h + \theta_h + \xi_h)$

$k_2 = d_1 + d_2,$

$k_3 = \xi_h \cdot d_3,$

$k_4 = d_1 \cdot \xi_h \cdot d_3,$

$p_1 = \delta_h + \gamma_h + \lambda_h + 2\nu_h + \psi_h + \theta_h + \varphi_h + \xi_h,$

$p_2 = \delta_h\gamma_h + \delta_h\lambda_h + \delta_h\nu_h + \delta_h\psi_h + \delta_h\varphi_h + \delta_h\xi_h + \gamma_h\lambda_h$
$\quad + 2\gamma_h\nu_h + \gamma_h\psi_h + \gamma_h\theta_h + \gamma_h\xi_h + 2\lambda_h\nu_h + \lambda_h\psi_h + \lambda_h\theta_h$
$\quad + \lambda_h\varphi_h + \lambda_h\xi_h + \nu_h^2 + \nu_h\psi_h + \nu_h\theta_h + 2\nu_h\varphi_h + 2\nu_h\xi_h$
$\quad + \psi_h\theta_h + \psi_h\varphi_h + \psi_h\xi_h$
$\quad + \theta_h\varphi_h + \theta_h\xi_h + \varphi_h\xi_h,$

$p_3 = \delta_h\gamma_h\lambda_h + \delta_h\gamma_h\nu_h + \delta_h\gamma_h\psi_h + \delta_h\gamma_h\xi_h + \delta_h\lambda_h\nu_h + \delta_h\lambda_h\psi_h$
$\quad + \delta_h\lambda_h\varphi_h + \delta_h\lambda_h\xi_h + \delta_h\nu_h\varphi_h + \delta_h\nu_h\xi_h + \delta_h\psi_h\varphi_h + \delta_h\psi_h\xi_h$
$\quad + \delta_h\varphi_h\xi_h + 2\gamma_h\lambda_h\nu_h + \gamma_h\lambda_h\psi_h + \gamma_h\lambda_h\xi_h + \gamma_h\nu_h^2 + \gamma_h\nu_h\psi_h$
$\quad + \gamma_h\nu_h\theta_h + 2\gamma_h\nu_h\xi_h + \gamma_h\psi_h\theta_h + \gamma_h\psi_h\xi_h + \gamma_h\theta_h\xi_h + \lambda_h\nu_h^2$
$\quad + \lambda_h\nu_h\psi_h + \lambda_h\nu_h\theta_h + 2\lambda_h\nu_h\varphi_h + 2\lambda_h\nu_h\xi_h + \lambda_h\psi_h\theta_h + \lambda_h\psi_h\varphi_h$
$\quad + \lambda_h\psi_h\xi_h + \lambda_h\theta_h\varphi_h + \lambda_h\theta_h\xi_h + \lambda_h\varphi_h\xi_h + \nu_h^2\varphi_h + \nu_h^2\xi_h + \nu_h\psi_h\varphi_h$
$\quad + \nu_h\psi_h\xi_h + \nu_h\theta_h\varphi_h + \nu_h\theta_h\xi_h + 2\nu_h\varphi_h\xi_h + \psi_h\theta_h\varphi_h + \psi_h\theta_h\xi_h$
$\quad + \psi_h\varphi_h\xi_h + \theta_h\varphi_h\xi_h,$

$p_4 = \delta_h\gamma_h\lambda_h\nu_h + \delta_h\gamma_h\lambda_h\psi_h + \delta_h\gamma_h\nu_h\xi_h + \delta_h\gamma_h\psi_h\xi_h + \delta_h\lambda_h\nu_h\varphi_h$
$\quad + \delta_h\lambda_h\nu_h\xi_h + \delta_h\lambda_h\psi_h\varphi_h + \delta_h\lambda_h\psi_h\xi_h + \delta_h\lambda_h\varphi_h\xi_h + \delta_h\nu_h\varphi_h\xi_h$
$\quad + \delta_h\psi_h\varphi_h\xi_h + \gamma_h\lambda_h\nu_h^2 + \gamma_h\lambda_h\nu_h\psi_h + 2\gamma_h\lambda_h\nu_h\xi_h + \gamma_h\lambda_h\psi_h\xi_h$
$\quad + \gamma_h\nu_h^2\xi_h + \gamma_h\nu_h\psi_h\xi_h + \gamma_h\nu_h\theta_h\xi_h + \gamma_h\psi_h\theta_h\xi_h + \lambda_h\nu_h^2\varphi_h + \lambda_h\nu_h^2\xi_h$
$\quad + \lambda_h\nu_h\psi_h\varphi_h + \lambda_h\nu_h\psi_h\xi_h + \lambda_h\nu_h\theta_h\varphi_h + \lambda_h\nu_h\theta_h\xi_h + 2\lambda_h\nu_h\varphi_h\xi_h$
$\quad + \lambda_h\psi_h\theta_h\varphi_h + \lambda_h\psi_h\theta_h\xi_h + \lambda_h\theta_h\varphi_h\xi_h + \nu_h^2\varphi_h\xi_h$
$\quad + \nu_h\psi_h\varphi_h\xi_h + \nu_h\theta_h\varphi_h\xi_h + \psi_h\theta_h\varphi_h\xi_h,$

$p_5 = \delta_h\varphi_h + \gamma_h\nu_h + \gamma_h\psi_h + \nu_h\varphi_h + \theta_h\varphi_h,$

$p_6 = \delta_h\lambda_h\nu_h\varphi_h\xi_h + \gamma_h\lambda_h\nu_h^2\xi_h + \gamma_h\lambda_h\nu_h\psi_h\xi_h + \lambda_h\nu_h^2\varphi_h\xi_h$
$\quad + \lambda_h\nu_h\theta_h\varphi_h\xi_h,$

$h_1 = \delta_h + \gamma_h + 2\nu_h + \psi_h + \theta_h + \varphi_h + \xi_h,$

$h_2 = \delta_h\gamma_h + \delta_h\nu_h + \delta_h\psi_h + \delta_h\varphi_h + \delta_h\xi_h + 2\gamma_h\nu_h + \gamma_h\psi_h + \gamma_h\theta_h$
$\quad + \gamma_h\xi_h + \nu_h^2 + \nu_h\psi_h + \nu_h\theta_h + 2\nu_h\varphi_h + 2\nu_h\xi_h + \psi_h\theta_h + \psi_h\varphi_h$
$\quad + \psi_h\xi_h + \theta_h\varphi_h + \theta_h\xi_h + \varphi_h\xi_h,$

$h_3 = \delta_h\gamma_h\nu_h + \delta_h\gamma_h\psi_h + \delta_h\gamma_h\xi_h + \delta_h\nu_h\varphi_h + \delta_h\nu_h\xi_h + \delta_h\psi_h\varphi_h$
$\quad + \delta_h\psi_h\xi_h + \delta_h\varphi_h\xi_h + \gamma_h\nu_h^2 + \gamma_h\nu_h\psi_h + \gamma_h\nu_h\theta_h$

$\quad + 2\gamma_h\nu_h\xi_h + \gamma_h\psi_h\theta_h + \gamma_h\psi_h\xi_h + \gamma_h\theta_h\xi_h + \nu_h^2\varphi_h$
$\quad + \nu_h^2\xi_h + \nu_h\psi_h\varphi_h + \nu_h\psi_h\xi_h + \nu_h\theta_h\varphi_h + \nu_h\theta_h\xi_h + 2\nu_h\varphi_h\xi_h$
$\quad + \psi_h\theta_h\varphi_h + \psi_h\theta_h\xi_h + \psi_h\varphi_h\xi_h + \theta_h\varphi_h\xi_h,$

$h_4 = \delta_h\gamma_h\nu_h\xi_h + \delta_h\gamma_h\psi_h\xi_h + \delta_h\nu_h\varphi_h\xi_h + \delta_h\psi_h\varphi_h\xi_h + \gamma_h\nu_h^2\xi_h$
$\quad + \gamma_h\nu_h\psi_h\xi_h + \gamma_h\nu_h\theta_h\xi_h + \gamma_h\psi_h\theta_h\xi_h + \nu_h^2\varphi_h\xi_h + \nu_h\psi_h\varphi_h\xi_h$
$\quad + \nu_h\theta_h\varphi_h\xi_h + \psi_h\theta_h\varphi_h\xi_h,$

$$\quad (48)$$

7. Existence and uniqueness results for the monkeypox transmission model with non-pharmaceutical intervention

We reformulate Equation (2) as follows:

$$
\begin{cases}
\Phi_1(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \Lambda_h + \xi_h R_h \\
\qquad + \theta_h Q_h - \lambda_h S_h - \mu_h S_h, \\
\Phi_2(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \lambda_h S_h \\
\qquad - \gamma_h E_h - \phi_h E_h - \mu_h E_h, \\
\Phi_3(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \phi_h E_h \\
\qquad - (\psi_h + \mu_h + \nu_h) I_h, \\
\Phi_4(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \gamma_h E_h \\
\qquad - (\theta_h + \delta_h + \mu_h + \nu_h) Q_h \\
\Phi_5(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \psi_h I_h \\
\qquad + \delta_h Q_h - \xi_h R_h - \mu_h R_h, \\
\Phi_6(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \Lambda_r - \lambda_r S_r \\
\qquad - \mu_r S_r \\
\Phi_7(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \lambda_r S_r - \phi_r E_r \\
\qquad - \mu_r E_r, \\
\Phi_8(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)) = \phi_r E_r \\
\qquad - (\mu_r + \nu_r) I_r.
\end{cases}
$$

From Equation (10), the developed model of Equation (1) can be written in the form

$$
\begin{cases}
{}^{CF}D_t^\alpha \Phi(t) = \Upsilon(t, \Phi(t)), t \in [0, \eta],\ 0 < \alpha \le 1, \\
\Phi(0) = \Phi_0,
\end{cases}
\quad (49)
$$

$$
\Phi(t) =
\begin{cases}
S_h(t), \\
E_h(t), \\
I_h(t), \\
Q_h(t), \\
R_h(t), \\
S_r(t), \\
E_r(t), \\
I_r(t),
\end{cases}
\Phi_0 =
\begin{cases}
S_h(0), \\
E_h(0), \\
I_h(0), \\
Q_h(0), \\
R_h(0), \\
S_r(0), \\
E_r(0), \\
I_r(0),
\end{cases}
\quad (50)
$$

therefore,

$$
\Upsilon(t, \Phi(t)) = \begin{cases}
\Phi_1(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_2(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_3(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_4(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_5(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_6(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_7(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)), \\
\Phi_8(t, S_h(t), E_h(t), I_h(t), Q_h(t), R_h(t), S_r(t), E_r(t), I_r(t)).
\end{cases}
\tag{51}
$$

With the help of Lemma 2.4, Equation (49) yields

$$
\begin{cases}
\Phi(t) = \Phi_0(t) + \dfrac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\Upsilon(t, \Phi(t)) + \dfrac{2\alpha}{(2-\alpha)G(\alpha)} \\
\quad \times \int_0^t \Upsilon(s, \Phi(s))ds.
\end{cases}
\tag{52}
$$

Furthermore, let ussay $E = C([0, \eta])$ is the Banach space, and supposing that the following assumptions hold;

(H_1), there exists a non-negative constant Q, W, and $k \in [0, 1)$ such that

$$
\Upsilon(t, \Phi(t)) \le Q|\Phi|^k + W.
$$

(H_2) There exists a nonnegative constant $\mathbf{C}_\rho > 0$ for all $\Phi, \tilde{\Phi} \in E$, then

$$
|\Upsilon(t, \Phi(t)) - \Upsilon(t, \tilde{\Phi}(t))| \le \mathbf{C}_\rho[\|\Phi - \tilde{\Phi}\|].
$$

Furthermore, let us define operator $\mathbf{A}_m : E \to E$ such that

$$
\mathbf{A}_m \aleph(t) = M_1 \Phi(t) + M_2 \Phi(t),
$$

therefore, we can see that

$$
\begin{cases}
M_1 \Phi(t) = \Phi_0(t) + \dfrac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\Upsilon(t, \Phi(t)), \\
M_2 \Phi(t) = \dfrac{2\alpha}{(2-\alpha)G(\alpha)} \int_0^t \Upsilon(s, \Phi(s))ds.
\end{cases}
\tag{53}
$$

From this knowledge, Equation (52) can be written as

$$
\begin{cases}
\mathbf{A}_m \Phi(t) = \Phi_0(t) + \dfrac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\Upsilon(t, \Phi(t)) + \dfrac{2\alpha}{(2-\alpha)G(\alpha)} \\
\quad \times \int_0^t \Upsilon(s, \Phi(s))ds.
\end{cases}
\tag{54}
$$

Theorem 2. Suppose that (H_1) and (H_2) hold, such that $\dfrac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\mathbf{C}_\rho < 1$, then, the monkeypox transmission model with non-pharmaceutical intervention has at least one solution.
Proof. For simplicity, we divide the proof into two steps.
Step 1. We prove that operator M_1 is contraction. Then, let $\tilde{\Phi} \in \Omega$, where $\Omega = \{\Phi \in \mathbf{Z} : \|\Phi\| \le \vartheta, \vartheta > 0\}$ is a close convex set, thus

$$
\begin{aligned}
|M_1 \Phi(t) - M_2 \Phi| &= \frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)} \max_{\alpha \in [0, \eta]} \\
&\quad |\Upsilon(t, \Phi(t)) - \Upsilon(t, \tilde{\Phi}(t))|, \\
&\le \frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\mathbf{C}_\rho \|\Phi - \tilde{\Phi}\|.
\end{aligned}
\tag{55}
$$

Thus,

$$
\|M_1 \Phi - M_2 \Phi(t)\| \le \frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\mathbf{C}_\rho \|\Phi - \tilde{\Phi}\|.
$$

Hence, M_1 is contraction since $\dfrac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\mathbf{C}_\rho < 1$.
Step 2. We also prove that M_2 is compact and also continuous; for all $\Phi \in \Omega$, then M_2 will be continuous as Φ is continuous, thus

$$
\begin{aligned}
\|M_2(\Phi)\| &= \max_{t \in [0,\eta]} |\frac{2\alpha}{(2-\alpha)G(\alpha)} \int_0^t \Upsilon(s, \Phi(s))ds|, \\
&\le \frac{2\alpha}{(2-\alpha)G(\alpha)}\eta \int_0^t |\Upsilon(s, \Phi(s))|ds. \\
&\le \frac{2\alpha}{(2-\alpha)G(\alpha)}\eta[Q|\Phi|^k + W].
\end{aligned}
\tag{56}
$$

Hence, M_2 is boundedness. For equicontinuous, let $t_1, t_2 \in [0, \eta]$ such that

$$
\begin{aligned}
|(M_2\Phi)(t_1) - (M_2\Phi)(t_2)| &= \frac{2\alpha}{(2-\alpha)G(\alpha)} \max_{t \in [0,\eta]} \big| \int_0^{t_1} \Upsilon(s, \Phi(s))ds \\
&\quad - \int_0^{t_2} \Phi(s, \aleph(s))ds \big| \\
&\le \frac{2\alpha}{(2-\alpha)G(\alpha)}[Q|\Phi|^k + W]|t_1 - t_2|.
\end{aligned}
\tag{57}
$$

As $t_1 \to t_2$, then $|(M_2\Phi)(t_1) - (M_2\Phi)(t_2)| \to 0$, which makes operator M_2 equicontinuous and compact by the Arzela–Ascoli theorem. Therefore, by Lemma 2.3, the existence for the monkeypox transmission model with non-pharmaceutical intervention has at least one solution. □
Theorem 3. Suppose that $\exists$ is a nonnegative integer Λ_ρ is > 0 such that

$$
\Lambda_\rho = \left[\frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\mathbf{L}_\rho + \frac{2\alpha}{(2-\alpha)G(\alpha)}\eta\mathbf{L}_\rho\right] < 1,
\tag{58}
$$

then operator $\mathbf{A}_m$ has a unique fixed point.
Proof. Let $\Phi, \tilde{\Phi} \in \Omega$, then we say

$$
\begin{aligned}
\|\mathbf{A}_m \Phi - \mathbf{A}_m \tilde{\Phi}\| &\le \|M_1 \Phi - M_1 \tilde{\Phi}\| + \|M_2 \Phi - M_2 \tilde{\Phi}\|, \\
&\le \frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)} \max_{t \in [0,\eta]} \big| \Upsilon(t, \Phi(t)) - \Upsilon(t, \tilde{\Phi}(t)) \big| \\
&\quad + \frac{2\alpha}{(2-\alpha)G(\alpha)} \max_{t \in [0,\eta]} \big| \int_0^t \Upsilon(s, \Phi(s))ds \\
&\quad - \int_0^t \Upsilon(s, \tilde{\Phi}(s))ds \big| \\
&\le \left[\frac{2(1-\alpha)}{2(1-\alpha)G(\alpha)}\mathbf{C}_\rho + \frac{2\alpha}{(2-\alpha)G(\alpha)}\eta\mathbf{C}_\rho\right]\|\Phi - \tilde{\Phi}\|, \\
&= \Lambda_\rho \|\Phi - \tilde{\Phi}\|.
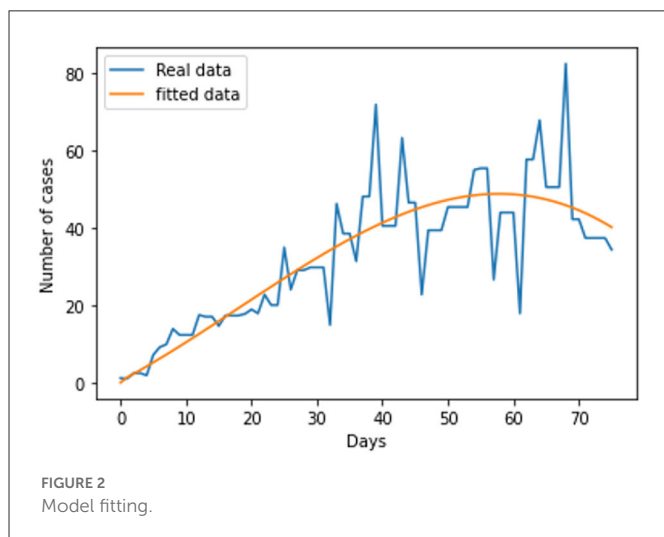\end{aligned}
\tag{59}
$$

Hence, by the Banach contraction principle, $\mathbf{A}_m$ has a unique fixed point. Consequently, the monkeypox transmission model with non-pharmaceutical intervention has a unique solution. □

8. Fitting of model to data

We used the available public database to collect our data while the formulated model of Equation (1) includes 16 parameters. To treat the waggliness of the reported daily new cases, we smoothed the data to remove noise from the data set so as to make it suitable for our analysis. The total population of the United Kingdom is 68,530,739

TABLE 2 Parameter values in the model.

Parameter	Value	Source
Λ_h	8644	Estimated
Λ_r	0.9	Assumed
ξ_h	0.00001	Fitted
θ_h	0.029	Fitted
μ_r	0.00200	(46)
μ_h	0.05	(1)
ν_h, ν_r	0.00008, 0.0001	Fitted
ϕ_h, ϕ_r	0.007	Fitted
ψ_h	0.056	Fitted
γ_h	0.0081	Fitted
δ_h	0.012	Fitted
β_{rh}	0.000009	Fitted
β_{hh}	0.00008	Fitted
β_{rr}	0.0057	Fitted

FIGURE 2
Model fitting.

(1), which was used for calculating the initial number of susceptible humans, while the initial value for the number of infected humans was calculated from the reported daily new cases. Other initial values were assumed.

The link to the data used for this research and the initial values; $S_h(t) = 68530739$; $E_h(t) = 0$; $I_h(t) = 31412$; $Q_h(t) = 0$; $S_r(t) = 1074103$; $E_r(t) = 1074103$; and $I_r(t) = 1074103$; can be found in the Data Availability section. The parameters are fitted based on the smoothed reported daily new cases of infected humans from May to August 2022. This information was taken from the United Kingdom public health database (1). The nonlinear least square technique was used to fit the model using python programming. Table 2 shows all of the parameter values that were fitted, and Figure 2 shows the data fitting of the observed smoothed daily new cases.

9. Numerical scheme

In this section, we present the numerical results for the monkeypox transmission model with non-pharmaceutical intervention based on the Lagrange interpolation. Details about the numerical scheme is presented in Atangana and Owolabi (50). The Cauchy problem of the CF fractional derivative can be given as:

$$^{CF}D_t^\alpha \Phi(t) = \Upsilon(t, \Phi(t)), \tag{60}$$

On the other hand, we can express Equation (60) as

$$\Phi(t) = \Phi_0(t) + \frac{(1-\alpha)}{G(\alpha)}\Upsilon(t, \Phi(t)) + \frac{\alpha}{G(\alpha)} \times \int_0^t \Upsilon(s, \Phi(s))ds. \tag{61}$$

Taking Equation (61) at the point $t_{n+1} = (n+1)h$ and $t_n = nh$, $n = 0, 1, 2, 3, ...,$ with h being the time step, we have

$$\Phi(t_{n+1}) = \Phi(0) + \frac{(1-\alpha)}{G(\alpha)}\Upsilon(t_n, \Phi(t_n)) + \frac{\alpha}{G(\alpha)} \\ \times \int_{t_n}^{t_{n+1}} \Upsilon(s, \Phi(s))ds, \tag{62}$$

$$\Phi(t_n) = \Phi(0) + \frac{(1-\alpha)}{G(\alpha)}\Phi(t_{n-1}, \aleph(t_{n-1})) + \frac{\alpha}{G(\alpha)} \\ \times \int_{t_n}^{t_{n+1}} \Upsilon(s, \Phi(s))ds. \tag{63}$$

Taking the results of Equations (62)-(63) in

$$\Phi(t_{n+1}) - \aleph(t_n) = \frac{(1-\alpha)}{G(\alpha)}\big(\Upsilon(t_n, \Phi(t_n)) - \Upsilon(t_{n-1}, \Phi(t_{n-1}))\big) \\ + \frac{\alpha}{G(\alpha)} \times \int_{t_n}^{t_{n+1}} \Upsilon(s, \Phi(s))ds, \tag{64}$$

Equation (64) in the two-step Lagrange polynomial gives

$$\Phi(t_{n+1}) - \Phi(t_n) = \frac{(1-\alpha)}{G(\alpha)}\big(\Upsilon(t_n, \Phi(t_n)) - \Upsilon(t_{n-1}, \Phi(t_{n-1}))\big) \\ + \frac{\alpha}{G(\alpha)} \times \int_{t_\alpha}^{t_{\alpha+1}} \Big[\frac{\Upsilon(t_n, \Phi(t_n))}{h}(s - t_{n-1}) \\ - \frac{\Upsilon(t_{n-1}, \Phi(t_{n-1}))}{h}(s - t_n)\Big]ds. \tag{65}$$

The aforementioned Equation (65) leads to

$$\Phi(t_{n+1}) - \Phi(t_n) = \frac{(1-\alpha)}{G(\alpha)}\big(\Upsilon(t_n, \Phi(t_n)) - \Upsilon(t_{n-1}, \Phi(t_{n-1}))\big) \\ + \frac{\alpha}{G(\alpha)} \times \Big[\frac{\Upsilon(t_n, \Phi(t_n))}{h}\int_{t_n}^{t_{n+1}}(s - t_{n-1})ds - \frac{\Upsilon(t_{n-1}, \Phi(t_{n-1}))}{h} \\ \int_{t_n}^{t_{n+1}}(s - t_n)ds\Big]. \tag{66}$$
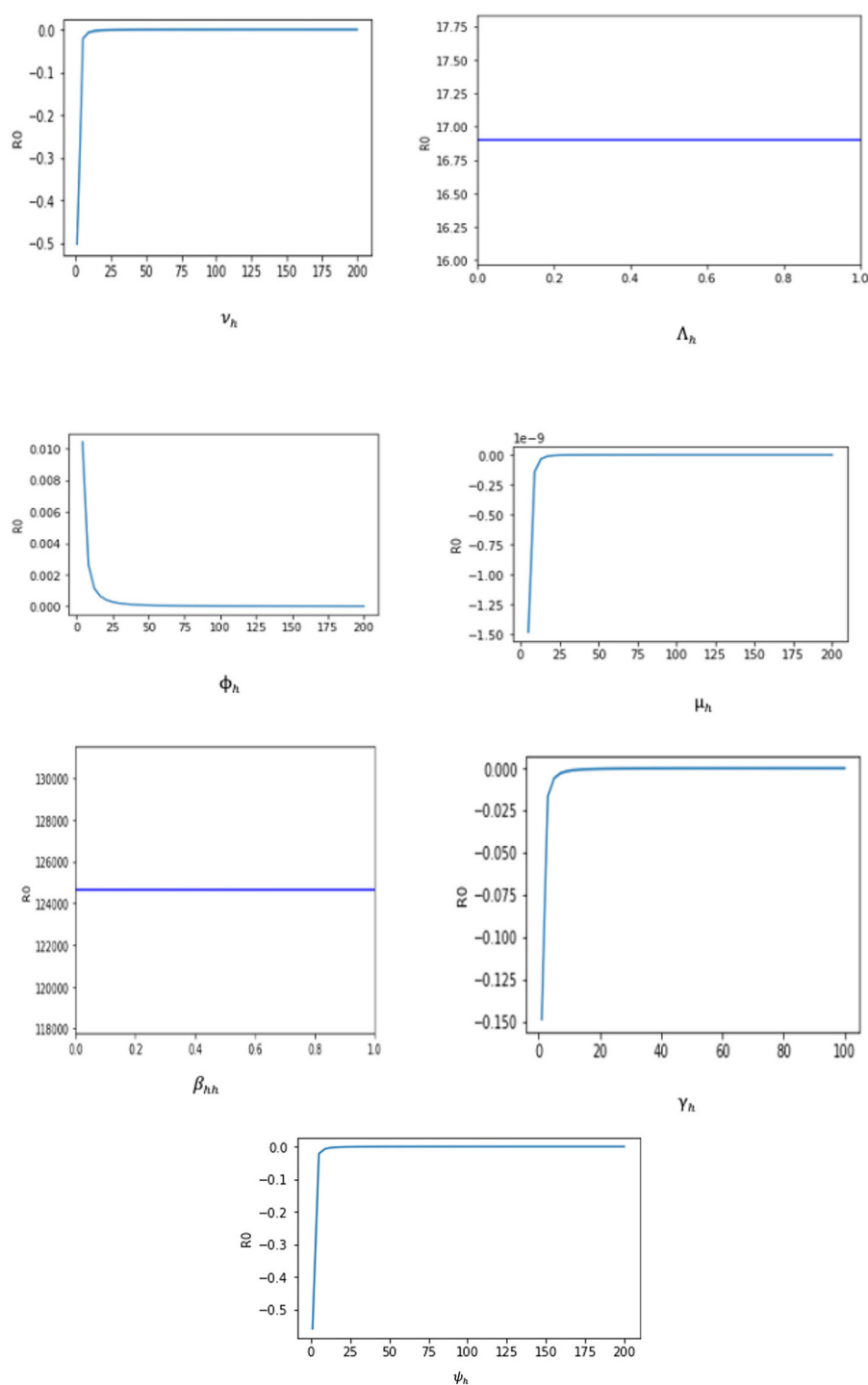
FIGURE 3
The sensitivity analysis of $\mathfrak{R}_0$ with respect to the parameter p of the system (1).

Solving the integrals in Equation (66) yields

Substituting Equation (67) into Equation (66), then generalizing the numerical scheme of CF is as follows:

$$\int_{t_n}^{t_{n+1}} (s - t_{n-1})ds = \frac{3}{2}h^2,$$
$$\int_{t_n}^{t_{n+1}} (s - t_n)ds = \frac{1}{2}h^2. \tag{67}$$

$$\Phi_{n+1} = \aleph_n + \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\right]\Phi(t_n, \Phi_n)$$
$$- \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\right]\Phi(t_{n-1}, \aleph_{n-1}). \tag{68}$$
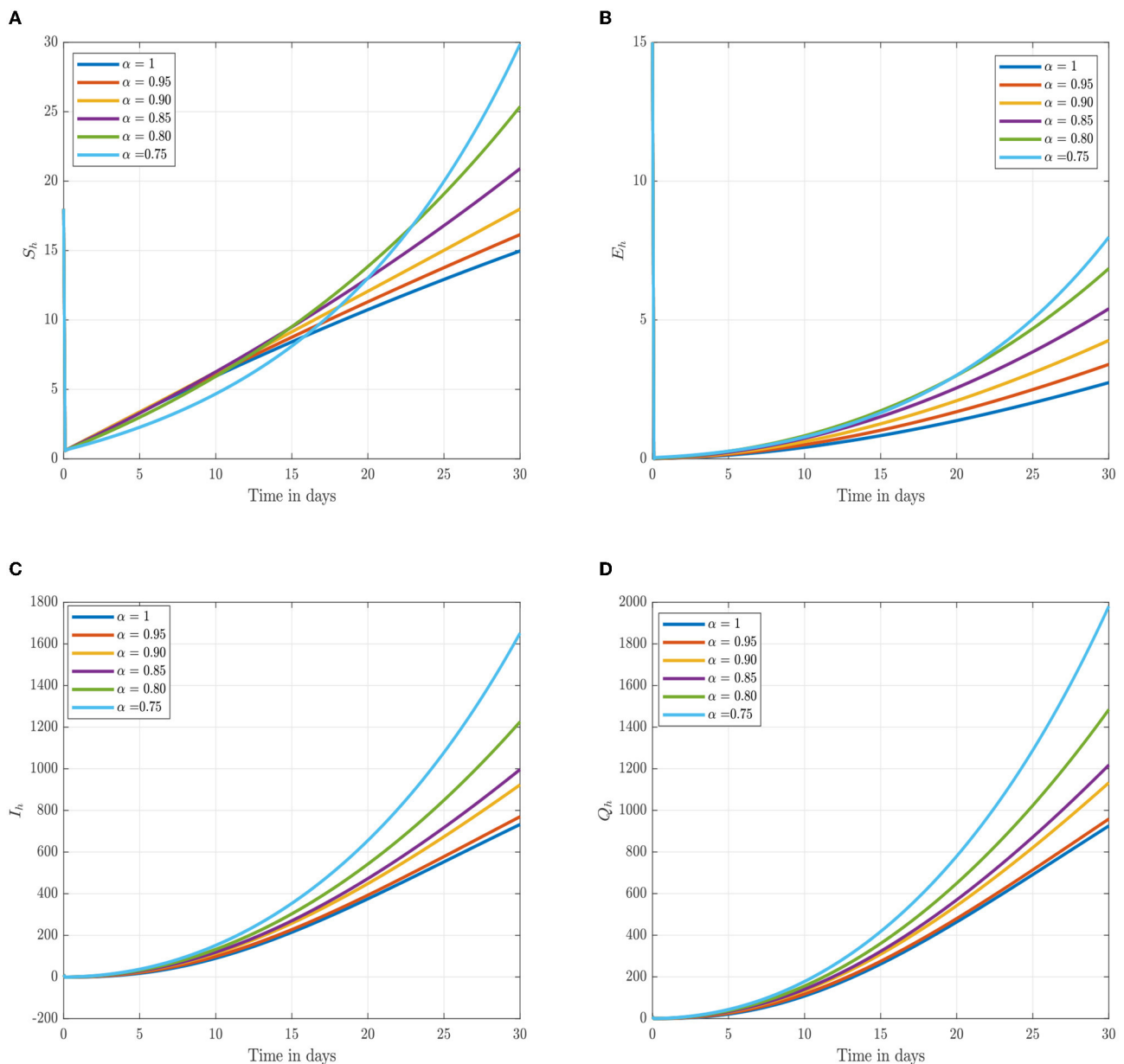
FIGURE 4
Numerical trajectory of the CF-fractional-order derivative, α, of Equation (2). **(A)** Dynamics of susceptible (S_h) humans class. **(B)** Dynamics of expose (E_h) humans class. **(C)** Dynamics of infected (I_h) humans class. **(D)** Dynamics of quarantine (Q_h) class.

Thus, in terms of our CF-fractional monkeypox transmission model with non-pharmaceutical intervention, we obtain;

$$S_{h_{n+1}} = S_{h_n} + \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\Big]\Upsilon(t_n, S_{h_n})$$
$$- \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\Big]\Upsilon(t_{n-1}, S_{h_{n-1}}). \tag{69}$$

$$E_{h_{n+1}} = E_{h_n} + \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\Big]\Upsilon(t_n, E_{h_n})$$
$$- \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\Big]\Upsilon(t_{n-1}, E_{h_{n-1}}). \tag{70}$$

$$I_{h_{n+1}} = I_{h_n} + \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\Big]\Upsilon(t_n, I_{h_n})$$
$$- \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\Big]\Upsilon(t_{n-1}, I_{h_{n-1}}). \tag{71}$$

$$Q_{h_{n+1}} = Q_{h_n} + \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\Big]\Upsilon(t_n, Q_{h_n})$$
$$- \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\Big]\Upsilon(t_{n-1}, Q_{h_{n-1}}). \tag{72}$$

$$R_{h_{n+1}} = S_{h_n} + \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\Big]\Upsilon(t_n, R_{h_n})$$
$$- \Big[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\Big]\Upsilon(t_{n-1}, R_{h_{n-1}}). \tag{73}$$
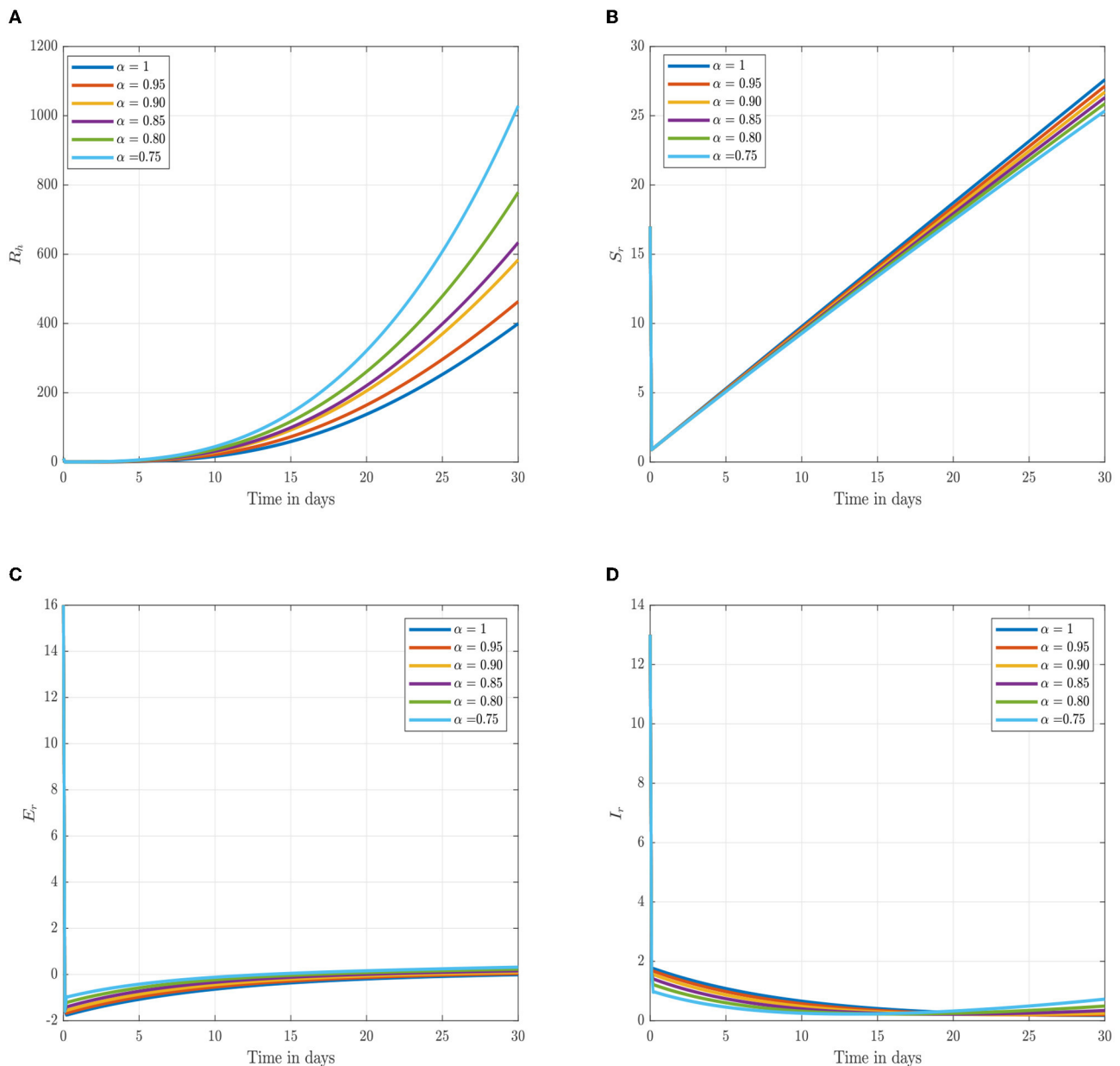
FIGURE 5
Numerical trajectory of the CF-fractional-order derivative, α, of Equation (2). **(A)** Dynamics of recovery (R_h) class. **(B)** Dynamics of susceptible (S_r) rodents class. **(C)** Dynamics of exposed (E_r) rodents class. **(D)** Dynamics of asymptomatic infected (I_r) rodents class.

$$S_{r_{n+1}} = S_{h_n} + \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\right]\Upsilon(t_n, S_{r_n})$$
$$- \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\right]\Upsilon(t_{n-1}, S_{r_{n-1}}). \tag{74}$$

$$E_{r_{n+1}} = S_{h_n} + \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\right]\Upsilon(t_n, E_{r_n})$$
$$- \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\right]\Upsilon(t_{n-1}, E_{r_{n-1}}). \tag{75}$$

$$I_{r_{n+1}} = S_{h_n} + \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{3h\alpha}{2G(\alpha)}\right]\Upsilon(t_n, I_{r_n})$$
$$- \left[\frac{(1-\alpha)}{G(\alpha)} + \frac{h\alpha}{2G(\alpha)}\right]\Upsilon(t_{n-1}, I_{r_{n-1}}). \tag{76}$$

10. Sensitivity analysis

Since an epidemiological system's parameters are either estimated or fitted, there is some degree of uncertainty in the numbers that are utilized to derive conclusions about the underlying epidemic. It is crucial to evaluate the individual effects of each parameter on the dynamics of the epidemic to identify those effects that have the greatest impact on the epidemic's spread or contraction. For biological factors included in the proposed monkeypox model, we perform the sensitivity analysis in this section. This analysis is investigated analytically by computing $\frac{\partial \mathfrak{R}_0}{\partial p}$, where, $p = (\beta_{hh}, \Lambda_h, \phi_h, \mu_h, \gamma_h, \nu_h,$ and $\psi_h)$. The sensitivity of $\mathfrak{R}_0$ to each parameter is

as follows:

$$\frac{\partial \mathfrak{R}_0}{\partial \beta_{hh}} = \frac{\lambda_h \phi_h}{\mu_h(\gamma_h + \mu_h + \phi_h)(\nu_h + \mu_h + \psi_h)} > 0,$$

$$\frac{\partial \mathfrak{R}_0}{\partial \Lambda_h} = \frac{\beta_{hh} \phi_h}{\mu_h(\gamma_h + \mu_h + \phi_h)(\nu_h + \mu_h + \psi_h)} > 0,$$

$$\frac{\partial \mathfrak{R}_0}{\partial \phi_h} = \frac{\beta_{hh} \Lambda_h (\gamma_h + \mu_h)}{\mu_h(\gamma_h + \mu_h + \phi_h)(\nu_h + \mu_h + \psi_h)},$$

$$\frac{\partial \mathfrak{R}_0}{\partial \mu_h} = -\frac{\begin{matrix}\beta_{hh} \Lambda_h(\mu_h(\gamma_h + \mu_h + \phi_h) + \mu_h(\nu_h + \mu_h + \psi_h) \\ +(\gamma_h + \mu_h + \phi_h)(\nu_h + \mu_h + \psi_h))\end{matrix}}{\mu_h^2(\gamma_h + \mu_h + \phi_h)^2(\nu_h + \mu_h + \psi_h)^2} < 0,$$

$$\frac{\partial \mathfrak{R}_0}{\partial \gamma_h} = -\frac{\beta_{hh} \lambda_h \phi_h}{\mu_h(\gamma_h + \mu_h + \phi_h)^2(\nu_h + \mu_h + \psi_h)} < 0,$$

$$\frac{\partial \mathfrak{R}_0}{\partial \nu_h} = -\frac{\beta_{hh} \lambda_h \phi_h}{\mu_h(\gamma_h + \mu_h + \phi_h)^2(\nu_h + \mu_h + \psi_h)^2} < 0,$$

$$\frac{\partial \mathfrak{R}_0}{\partial \psi_h} = -\frac{\beta_{hh} \lambda_h \phi_h}{\mu_h(\gamma_h + \mu_h + \phi_h)^2(\nu_h + \mu_h + \psi_h)^2} < 0,$$

thus,

$$\frac{\partial \mathfrak{R}_0}{\partial \beta_{hh}} = 124645.995943846$$

$$\frac{\partial \mathfrak{R}_0}{\partial \Lambda_h} = 16.9011519923859$$

$$\frac{\partial \mathfrak{R}_0}{\partial \phi_h} = \frac{0.1746837421029776}{(0.008139 + \psi_h)^2}$$

$$\frac{\partial \mathfrak{R}_0}{\partial \mu_h} = -\frac{\begin{matrix}9.912 \times 10^{-7} \mu_h^2 + 4.7035744 \times 10^{-8} \mu_h^2 \\ +2.797853632 \times 10^{-10}\end{matrix}}{\mu_h^2(\mu_h^4 + 0.014236\mu_h^3 + 0.00012055158688\mu_1 \\ +7.17083788864 \times 10^{-7})}, \quad (77)$$

$$\frac{\partial \mathfrak{R}_0}{\partial \gamma_h} = -\frac{0.15096125860751}{(\gamma_h + 0.007039)^2},$$

$$\frac{\partial \mathfrak{R}_0}{\partial \nu_h} = -\frac{0.59600691709814}{(\nu_h + 0.056039)^2},$$

$$\frac{\partial \mathfrak{R}_0}{\partial \psi_h} = -\frac{0.59600691709814}{(\nu_h + 0.056039)^2}.$$

The sensitivity index technique will help measure the most sensitive parameters for the fundamental reproductive number $\mathfrak{R}_0$ (Borgonovo et al. (51) for details about the method). The fundamental reproduction number's normalized sensitivity index is provided by $S_p^{\mathfrak{R}_0} = \frac{\partial \mathfrak{R}_0}{\partial p} \cdot \frac{p}{\mathfrak{R}_0}$, where p is a parameter as defined earlier. We obtain

$$S_{\beta_{hh}}^{\mathfrak{R}_0} = 1,$$

$$S_{\Lambda_h}^{\mathfrak{R}_0} = 1,$$

$$S_{\phi_h}^{\mathfrak{R}_0} = \frac{\gamma_h + \mu_h}{\gamma_h + \mu_h + \phi_h},$$

$$S_{\mu_h}^{\mathfrak{R}_0} = \frac{\begin{matrix}\mu_h(\gamma_h + \mu_h + \phi_h) + \mu_h(\nu_h + \mu_h + \psi_h) \\ +(\gamma_h + \mu_h + \phi_h)(\nu_h + \mu_h + \psi_h)\end{matrix}}{(\gamma_h + \mu_h + \phi_h)(\nu_h + \mu_h + \psi_h)},$$

$$S_{\gamma_h}^{\mathfrak{R}_0} = \frac{\gamma_h}{\gamma_h + \mu_h + \phi_h},$$

$$S_{\nu_h}^{\mathfrak{R}_0} = \frac{\nu_h}{\nu_h + \mu_h + \psi_h},$$

$$S_{\psi_h}^{\mathfrak{R}_0} = \frac{\psi_h}{\nu_h + \mu_h + \psi_h}.$$

TABLE 3 The sensitivity index of $\mathfrak{R}_0$ with respect to parameter p of the system (1).

Parameter	Sensitivity index
ν_h	−0.0014
Λ_h	1
ϕ_h	−1.003
μ_h	−1.003
β_{hh}	1
γ_h	−0.5350
ψ_h	−0.9979

The sensitivity indices using the parameter values given in Table 2 are presented in Table 3. The sensitivity analysis of $\beta_{hh}, \Lambda_h, \phi_h, \psi_h, \nu_h, \gamma_h$, and μ_h with respect to $\mathfrak{R}_0$ and their graphs are presented in Figure 3.

Two of the sensitivity indices are positive while others are negative, as can be seen in Table 3. Additionally, the majority of these indices are functions of the Caputo–Fabrizio fractional monkeypox model parameters. This implies that changing one of the parameters slightly will alter the dynamics of the epidemic. The basic reproductive number $\mathfrak{R}_0$ normalized sensitivity indices to the Caputo–Fabrizio fractional monkeypox model parameters are calculated. We conclude that increasing the rate of recovery and the rate of identifying suspected cases, that is, isolation and quarantining of the monkeypox virus carrier will aid in decreasing the $\mathfrak{R}_0$, which is an affirmation of the effect of non-pharmaceutical intervention to combat the spread of the virus.

11. Discussion and conclusion

Following the estimation of parameter values and data fitting, we simulate the Caputo–Fabrizio fractional monkeypox virus model using the parameter values, as presented in Table 2. The fitted Caputo–Fabrizio curve and $\mathfrak{R}_0$ are given in Figure 2. Figures 4, 5 show dynamic behavior for all the nine compartments involved in the proposed Caputo–Fabrizio fractional monkeypox virus model. We observed a significantly high susceptibility and infection in the solution pathways of individual species. The work of Hammouch et al. (52), Bonyah et al. (53), Peter (54), and Sene (55) have provided a strong basis for the discussion of our results. This indicates that, whenever the memory index increases, the rate at which people get infected with monkeypox virus reduces and vice versa, which then indicates that, using fractional order, we can obtain clear qualitative information on monkeypox virus transmission. In Figure 6, we varied the input parameter γ_h on quarantine and exposed, respectively, to observe variation in the system dynamics. We noticed the contribution of this parameter in the transmission pathways of infected individuals. In a similar way, we varied the input parameters δ_h and ψ_h on individual recovery and noticed the variation in the trajectory of monkeypox recovery. We discovered that the rate at which humans and rodents move from exposed to infectious stage is also important and potentially dangerous in terms of increasing the level of monkeypox infection.

In conclusion, we provided a brief overview of the monkeypox virus and the dynamics of its transmission in this study. We investigated the spread of monkeypox virus and its effect on
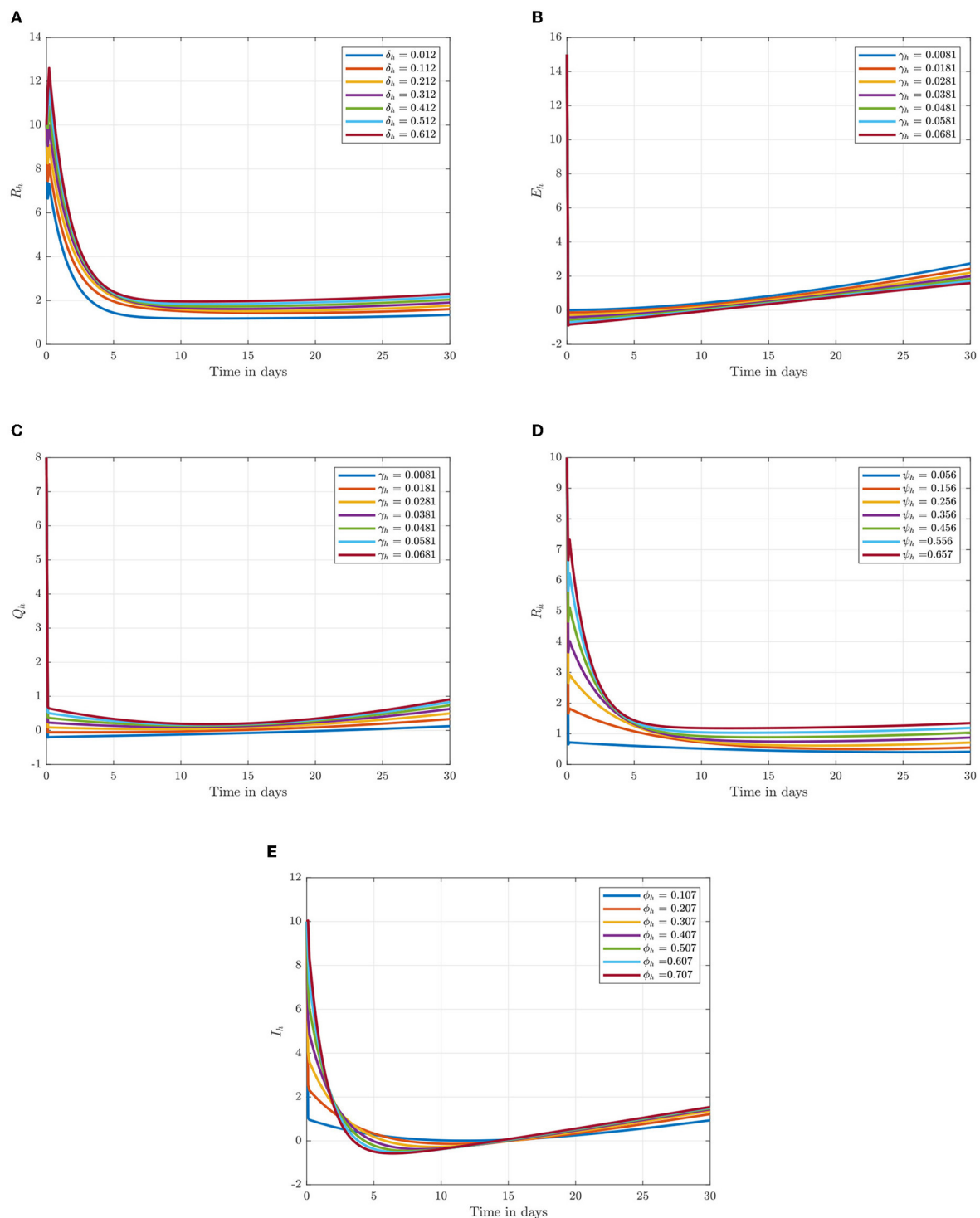
FIGURE 6
Numerical trajectory of varying δ_h, γ_h, ψ_h, and ϕ_h when $\alpha = 0.95$. **(A)** Variation of δ_h on recovery class. **(B)** Variation of γ_h on expose class. **(C)** Variation of γ_h on quarantine class. **(D)** Variation of ψ_h on recovery class. **(E)** Variation of ϕ_h on infected class.

non-pharmaceutical intervention, thus quarantine. Positiveness, invariance, boundedness, and equilibrium points of the solutions are thus examined as fundamental features of mathematical models. We considered real data of the monkeypox virus from the United Kingdom, and the best fit curve has been obtained (see Figure 2). As a result, we created a novel, dimensionally consistent Caputo–Fabrizio fractional-order model. Krasnoselskii's fixed point

theorem has been used to demonstrate that the system has a solution.The Adams–Bashforth method has been used to display numerical simulations of the suggested pandemic model for various fractional orders and parameter values. We looked into the impact of factors on the expansion and contraction of the quarantine compartment, recovery compartment, and infected compartment on the spread and regression of the pandemic with the use of

numerical simulations. As can be inferred from the data, it is clear that the fractional-order equations can help explain this unique effect of the monkeypox. Real-world data can be used to test the accuracy of a mathematical model that has been created. The key challenge, however, is where to find these data and/or how to obtain the right curve for the collected data. The mathematical representation of the monkeypox has been the subject of numerous studies. To the best of our knowledge, there is still no research on fractional modeling that uses actual data on the monkeypox in the United Kingdom. Using actual data on the monkeypox from the United Kingdom, a fractional-order modeling has been shown in this study. The numerical results of this study show that the spread of monkeypox can be stopped if the number of contacts with infected people can be decreased through methods such as effective mass education, improved quarantine facilities, or increased testing of the general population, that is, performing routine tests not only on exposed individuals but also on those who have come into contact with infected patients. As a result, these studies offer other professionals and scientists who focus on infectious diseases insight that may help them in future to control the outbreak of monkeypox and contribute to the development of further treatment options. This study may provide insight into potential future research projects in this regard. Future study of the monkeypox can take into account other fractional operator types, both with and without single kernels. Furthermore, data imputation techniques can be used to fit rodent population parameters from the number of monkeypox disease since the number of rodents cannot be determined.

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

Acknowledgments

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. *Monkeypox Cases Confirmed in England–Latest Updates*. UK Health Security Agency (2022). Available online at: https://www.gov.uk/government/news/monkeypox-cases-confirmed-in-england-latest-updates (accessed August 29, 2022).

2. *Multi-country Monkeypox Outbreak: Situation Update*. World Health Organization (2022). Available online at: https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON392 (accessed August 10, 2022).

3. Mathieu E, Spooner F, Dattani S, Ritchie H, Roser M. Mpox (monkeypox). *Our World Data*. (2022). Available online at: https://ourworldindata.org/monkeypox (accessed August 28, 2022).

4. *Public Health Agencies Issue Monkeypox Guidance to Control Transmission*. UK Health Security Agency (2022). Available online at: https://www.gov.uk/government/news/public-health-agencies-issue-monkeypox-guidance-to-control-transmission (accessed August 13, 2022).

5. Yinka-Ogunleye A, Aruna O, Dalhat M, Ogoina D, McCollum A, Disu Y, et al. Outbreak of human monkeypox in Nigeria in 2017-18: a clinical and epidemiological report. *Lancet Infect Dis* . (2019) 19:872–9. doi: 10.1016/S1473-3099(19)30294-4

6. Hobson G, Adamson J, Adler H, Firth R, Gould S, Houlihan C, et al. Family cluster of three cases of monkeypox imported from Nigeria to the United Kingdom, May 2021. *Euro Surveill*. (2021) 26:2100745. doi: 10.2807/1560-7917.ES.2021.26.32.2100745

7. *Factsheet for Health Professionals on Monkeypox*. UK Health Security Agency (2022). Available online at: https://www.ecdc.europa.eu/en/publications-data/data-mpox-monkeypox-cases-eueea (accessed August 27, 2022).

8. Harris E. What to know about monkeypox. *JAMA*. (2022) 327:2278–9. doi: 10.1001/jama.2022.9499

9. Reynolds MG, Doty JB, McCollum AM, Olson VA, Nakazawa Y. Monkeypox re-emergence in Africa: a call to expand the concept and practice of one health. *Expert Rev Anti Infect Ther*. (2019) 17:129–39. doi: 10.1080/14787210.2019.15 67330

10. Durski KN, McCollum AM, Nakazawa Y, Petersen BW, Reynolds MG, Briand S, et al. Emergence of monkeypox - West and Central Africa, 1970-2017. *MMWR Morb Mortal Wkly Rep*. (2018) 67:306–10. doi: 10.15585/mmwr.mm6710a5

11. Usman S, Adamu II. Modeling the transmission dynamics of the monkeypox virus infection with treatment and vaccination interventions. *J Appl Math Phys*. (2017) 5:2335. doi: 10.4236/jamp.2017.512191

12. Peter OJ, Madubueze CE, Ojo MM, Oguntolu FA, Ayoola TA. Modeling and optimal control of monkeypox with cost-effective strategies. *Model Earth Syst Environ*. (2022) 1–19. doi: 10.1007/s40808-022-01607-z

13. Samko SG, Kilbas AA, Marichev OI. *Fractional Integrals and Derivatives, Theory and Applications*. Yverdon: Gordon and Breach (1993).

14. Katugampola UN. A new approach to generalized fractional derivatives. *Bull Math Anal Appl*. (2014) 6:1–15. doi: 10.48550/arXiv.1106.0965

15. Caputo M, Fabrizio M. A new definition of fractional derivative without singular kernel. *Progr Fract Differ Appl*. (2015) 1:1–13. doi: 10.18576/pfda/020101

16. Kumar S, Ahmadian A, Kumar R, Kumar D, Singh J, Baleanu D, et al. An efficient numerical method for fractional SIR epidemic model of infectious disease by using Bernstein wavelets. *Mathematics.* (2020) 8:558. doi: 10.3390/math8040558

17. Higazy M, Alyami MA. New Caputo-Fabrizio fractional order SEIASqEqHR model for COVID-19 epidemic transmission with genetic algorithm based control strategy. *Alex Eng J.* (2020) 59:4719–36. doi: 10.1016/j.aej.2020.08.034

18. Djida JD, Atangana A. More generalized groundwater model with space-time Caputo Fabrizio fractional differentiation. *Numer Methods Partial Differ Equ.* (2017) 33:1616–27. doi: 10.1002/num.22156

19. Baba IA. Existence and uniqueness of a fractional order tuberculosis model. *Eur Phys J.* (2019) 134:489. doi: 10.1140/epjp/i2019-13009-1

20. Owolabi KM, Atangana A. Mathematical modelling and analysis of fractional epidemic models using derivative with exponential kernel. In: *Fractional Calculus in Medical and Health Science.* London: CRC Press (2020). p. 109–128.

21. Mohammadi H, Kumar S, Rezapour S, Etemad, S. A theoretical study of the Caputo-Fabrizio fractional modeling for hearing loss due to Mumps virus with optimal control. *Chaos Solit Fract.* (2021) 144:110668. doi: 10.1016/j.chaos.2021.110668

22. Baleanu D, Jajarmi A, Mohammad H, Rezapour S. A new study on the mathematical modelling of human liver with Caputo-Fabrizio fractional derivative. *Chaos Solit Fract.* (2020) 134:109705. doi: 10.1016/j.chaos.2020.109705

23. Wutiphol S, Turab A. Mathematical analysis of an extended SEIR model of COVID-19 using the ABC-Fractional operator. *Math Comput Simul.* (2022) 198:65–84. doi: 10.1016/j.matcom.2022.02.009

24. Asamoah JKK, Okyere E, Yankson E, Opoku AA, Adom-Konadu A, Acheampong E, et al. Non-fractional and fractional mathematical analysis and simulations for Q fever. *Chaos Solit Fract.* (2022) 156:111821. doi: 10.1016/j.chaos.2022.111821

25. Peter OJ, Yusuf A, Ojo MM, Kumar S, Kumari N, Oguntolu FA. A mathematical model analysis of meningitis with treatment and vaccination in fractional derivatives. *Int J Appl Comput Math.* (2022) 8:117. doi: 10.1007/s40819-022-01317-1

26. Peter OJ, Shaikh AS, Ibrahim MO, Nisar KS, Baleanu D, Khan I, et al. Analysis and dynamics of fractional order mathematical model of covid-19 in nigeria using atangana-baleanu operator. *Comput Mater Continua.* (2021) 66:1823–48. doi: 10.32604/cmc.2020.012314

27. Kumar S, Ghosh S, Kumar R, Jleli M. A fractional model for population dynamics of two interacting species by using spectral and hermite wavelets methods. *Numer Methods Part Diff Equ.* (2020) 37:1652–72. doi: 10.1002/num.22602

28. Morales-Delgadoa VF, Gómez-Aguilar JF, Taneco-Hernández MA, Escobar Jiménez RF, Olivares Peregrino VH. Mathematical modeling of the smoking dynamics using fractional differential equations with local and nonlocal kernel. *J Nonlinear Sci Appl.* (2018) 11:1004–14. doi: 10.22436/jnsa.011.08.06

29. Atangana A, Baleanu D. New fractional derivatives with nonlocal and non-singular kernel: theory and application to heat transfer model. *Therm Sci.* (2016) 220:763–9. doi: 10.2298/TSCI160111018A

30. Atangana A, Gomez-Aguilar JF. A new derivative with normal distribution kernel: theory, methods and applications. *Physica A Stat Mech Appl.* (2017) 476:1–14. doi: 10.1016/j.physa.2017.02.016

31. Zhang L, Addai E, Ackora-Prah J, Arthur YD, Asamoah JKK. F. Fractional-order ebola-malaria coinfection model with a focus on detection and treatment rate *Comput Math Methods Med.* (2022) 2022:6502598. doi: 10.1155/2022/6502598

32. Kumar S, Chauhan RP, Momani S, Hadid S. Numerical investigations on COVID-19 model through singular and non-singular fractional operators. *Numer Methods Part Diff Equ.* (2020) 2020:1–27. doi: 10.1002/num.22707

33. Aslam M, Murtaza R, Abdeljawad T, Rahman, G, Khan A, et al. A fractional order HIV/AIDS epidemic model with Mittag-Leffler kernel. *Adv Diff Equ.* (2021) 2021:1–5. doi: 10.1186/s13662-021-03340-w

34. Evirgen F. Transmission of Nipah virus dynamics under Caputo fractional derivative. *J Comput Appl Math.* (2023) 419:114654. doi: 10.1016/j.cam.2022.114654

35. Ucar S. Analysis of hepatitis B disease with fractal fractional Caputo derivative using real data from Turkey. *J Comput Appl Math.* (2023) 419:114692. doi: 10.1016/j.cam.2022.114692

36. Khan Y, Khan MA, Fatmawati, Faraz N. A fractional bank competition model in Caputo-Fabrizio derivative through newton polynomial approach. *Alexandria Eng J.* (2021) 60:711–8. doi: 10.1016/j.aej.2020.10.003

37. Fatmawati, Khan MA, Alfiniyah C, Alzahrani E. Analysis of dengue model with fractal-fractional Caputo-Fabrizio operator. *Adv Diff Equ.* (2020) 2020:1–23. doi: 10.1186/s13662-020-02881-w

38. Qureshi S, Yusuf A, Shaikh AA, Inc M, Baleanu D. Mathematical modeling for adsorption process of dye removal nonlinear equation using power law and exponentially decaying kernels. *Chaos.* (2020) 30:043106. doi: 10.1063/1.5121845

39. Sher M, Shah K, Khan ZA, Khan H, Khan A. Computational and theoretical modeling of the transmission dynamics of novel COVID-19 under Mittag-Leffler Power Law. *Alexandria Eng J.* (2020) 59:3133–7. doi: 10.1016/j.aej.2020.07.014

40. Addai E, Zhang LL, Preko AK, Asamoah JKK. Fractional order epidemiological model of SARS-CoV-2 dynamism involving Alzheimer's disease. *Healthc Analyt.* (2022) 2:100114. doi: 10.1016/j.health.2022.100114

41. Shaikh AS, Nisar KS. Transmission dynamics of fractional order Typhoid fever model using Caputo-Fabrizio operator. *Chaos Solit Fract.* (2019) 128:355–65. doi: 10.1016/j.chaos.2019.08.012

42. Shah K, Alqudah MA, Jarad F, Abdeljawad T. Semi-analytical study of Pine Wilt Disease model with convex rate under Caputo-Febrizio fractional order derivative. *Chaos Solit Fract.* (2020) 135:109754. doi: 10.1016/j.chaos.2020.109754

43. Ahmed I, Baba IA, Yusuf A, Kumam P, Kumam W. Analysis of Caputo fractional-order model for COVID-19 with lockdown. *Adv Differ Equ.* (2020) 2020:110007. doi: 10.1186/s13662-020-02853-0

44. Ullah S, Khan MA, Farooq M, Hammouch Z, Baleanu D. A fractional model for the dynamics of tuberculosis infection using Caputo-Fabrizio derivative. *Discrete Cont Dyn Syst.* (2020) 13:975–93. doi: 10.3934/dcdss.2020057

45. Abboubakar H, Kumar P, Rangaig NA, Kumar S. A malaria model with Caputo-Fabrizio and Atangana-Baleanu derivatives. *Int J Mod Simul Sci Comput.* (2021) 12:2150013. doi: 10.1142/S1793962321500136

46. Peter OJ, Oguntolu FA, Ojo MM, Oyeniyi AO, Jan R, Khan I. Fractional order mathematical model of monkeypox transmission dynamics. *Phys Script.* (2022) 97:084005. doi: 10.1088/1402-4896/ac7ebc

47. Peter OJ, Kumar S, Kumari N, Oguntolu FA, Oshinubi K, Musa R. Transmission dynamics of Monkeypox virus: a mathematical modelling approach. *Model Earth Syst Environ.* (2022) 8:3423–34. doi: 10.1007/s40808-021-01313-2

48. Tilahuna GT, Woldegerimab WA, Mohammed N. A fractional order model for transmission dynamics of hepatitis B virus with two-age structure in the presence of vaccination. *Arab J Basic Appl Sci.* (2022) 28:87–106. doi: 10.1080/25765299.2021.1896423

49. Lin W. Global existence theory and chaos control of fractional differential equations. *J Math Anal Appl.* (2007) 332:709726. doi: 10.1016/j.jmaa.2006.10.040

50. Atangana A, Owolabi KM. New numerical approach for fractional differential equations. *Math Model Nat Phenom.* (2018) 13:3. doi: 10.1051/mmnp/2018010

51. Borgonovo E, Plischke E. Sensitivity analysis: a review of recent advances. *Eur J Oper Res.* (2016) 248:869–87. doi: 10.1016/j.ejor.2015.06.032

52. Hammouch Z, Yavuz M, Özdemir N. Numerical solutions and synchronization of avariable-order fractional chaotic system. *Math Model Numer Simul Appl.* (2021) 1:11–23. doi: 10.53391/mmnsa.2021.01.002

53. Bonyah E, Juga M, Fatmawati. Fractional dynamics of coronavirus with comorbidity via Caputo-Fabrizio derivative. *Commun Math Biol Neurosci.* (2022) 2022:12.

54. Peter OJ. Transmission dynamics of fractional order brucellosis model using Caputo-Fabrizio operator. *Int J Differ Equ.* (2020) 2020:2791380. doi: 10.1155/2020/2791380

55. Sene N. Theory and applications of new fractional-order chaotic systemunder Caputo operator. *Int J Opt Cont Theor Appl.* (2022)1:20–38. doi: 10.11121/ijocta.2022.1108

Check for updates

COVID-19 dynamics in an Ohio prison

Wasiur R. KhudaBukhsh[1], Sat Kartar Khalsa[2], Eben Kenah[3], Gregorz A. Rempała[4] and Joseph H. Tien[5]*

[1]School of Mathematical Sciences, The University of Nottingham, Nottingham, United Kingdom, [2]Wexner Medical Center, The Ohio State University, Columbus, OH, United States, [3]Division of Biostatistics, The Ohio State University, Columbus, OH, United States, [4]Division of Biostatistics, Department of Mathematics, The Ohio State University, Columbus, OH, United States, [5]Division of Epidemiology, Department of Mathematics, The Ohio State University, Columbus, OH, United States

Incarcerated individuals are a highly vulnerable population for infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Understanding the transmission of respiratory infections within prisons and between prisons and surrounding communities is a crucial component of pandemic preparedness and response. Here, we use mathematical and statistical models to analyze publicly available data on the spread of SARS-CoV-2 reported by the Ohio Department of Rehabilitation and Corrections (ODRC). Results from mass testing conducted on April 16, 2020 were analyzed together with time of first reported SARS-CoV-2 infection among Marion Correctional Institution (MCI) inmates. Extremely rapid, widespread infection of MCI inmates was reported, with nearly 80% of inmates infected within 3 weeks of the first reported inmate case. The dynamical survival analysis (DSA) framework that we use allows the derivation of explicit likelihoods based on mathematical models of transmission. We find that these data are consistent with three non-exclusive possibilities: (i) a basic reproduction number >14 with a single initially infected inmate, (ii) an initial superspreading event resulting in several hundred initially infected inmates with a reproduction number of approximately three, or (iii) earlier undetected circulation of virus among inmates prior to April. All three scenarios attest to the vulnerabilities of prisoners to COVID-19, and the inability to distinguish among these possibilities highlights the need for improved infection surveillance and reporting in prisons.

KEYWORDS

SARS-CoV-2, correctional facilities, mathematical modeling, mass testing, reproduction number

1. Introduction

The COVID-19 pandemic has demonstrated the tremendous vulnerability of incarcerated individuals to respiratory infections. More than 600,000 COVID-19 cases and close to 3,000 deaths were reported among incarcerated individuals in the United States as of October 2022 (1), and case rates for incarcerated individuals are more than five times higher than for the general population (2). Factors contributing to SARS-CoV-2 transmission in prisons include shared housing, crowding, hygiene challenges, and inability to social distance (3). Outbreak sizes within facilities can be high: infections in more than 80% of prisoners at the Marion Correctional Institution (MCI) in Ohio have been identified (4–7), and similarly high levels of infection have been observed at correctional facilities in other jurisdictions (8, 9). The vulnerability of prisoners and prison staff to COVID-19, the epidemiological connections between prisons and between prisons and surrounding communities, and the potential for prisons to become amplifiers of transmission have been noted by many authors (10–21).

Studies of COVID-19 outbreaks in correctional facilities can help us understand transmission in prisons and jails and identify practices to prevent and control future outbreaks. Research during the COVID-19 pandemic addressed vaccine efficacy and uptake studies among inmates and staff [22, 23], examination of policies by state corrections departments [24], studies of outbreaks started by transfer of infected inmates between prisons [20], quantitative analyses of relationships between correctional facility cases and cases in surrounding communities [18, 19], and analysis of interventions such as decarceration, single-celling, and testing of asymptomatic individuals [25]. Analyses of outbreaks in specific facilities include time series analysis [9], assessment of outbreak response in a California state prison [16], and network analyses based upon inmate housing and staff assignments [26]. Modeling studies include stochastic simulations of transmission among inmates and staff [27] and fitting compartmental models to case time series data [28]. In particular, Puglisi et al. [28] use model fits to estimate the basic reproduction number (R_0) for the ancestral strain of SARS-CoV-2 in a large urban jail. Several of these studies point to the need for improved data collection and reporting [21].

Here, we study the COVID-19 outbreak in MCI in the spring of 2020 using publicly available time series data from the Ohio Department of Rehabilitation and Corrections (ODRC). In particular, our main contribution is a rigorous and statistically principled analysis of the results of mass testing conducted at MCI in April 2020. The analysis is based on a compartmental mathematical model of transmission that is fit to data using a statistical approach called the dynamical survival analysis (DSA) [29, 30], which allows the calculation of explicit likelihoods to summarize uncertainty. Our results highlight the explosive potential for transmission of respiratory infections in prisons as well as the critical need for improved monitoring and reporting of infection in correctional facilities.

2. Data and methods

2.1. Case data

Mass RT-PCR testing of all inmates and partial testing of staff at MCI was conducted on April 16, 2020. The total number of inmates and the number of inmates and staff testing positive for SARS-CoV-2 over time were obtained from public ODRC reports [7]. Results from early SARS-CoV-2 tests were available with a slight time-lag, so we accumulate the cases reported at MCI over April 16–23, 2020 as a single mass testing data point assigned to April 16, which was the date of mass testing. The mass testing event received significant media coverage and was reported widely in numerous news articles [4–6].

2.2. Mathematical model

We use a compartmental susceptible-exposed-infectious-recovered (SEIR) model of SARS-CoV-2 dynamics in MCI. Such compartmental models have been used extensively in the literature because they tend to provide a good approximation to the process of disease spread [31]. Assuming a well-mixed population, under the standard SEIR model, the proportions of individuals in the susceptible (S_t), exposed (E_t), infectious (I_t), and recovered (R_t) compartments as a function of time t satisfy the following system of differential equations:

$$\begin{aligned}
\dot{S}_t &= -\beta S_t I_t, \\
\dot{E}_t &= \beta S_t I_t - \alpha E_t, \\
\dot{I}_t &= \alpha E_t - \gamma I_t, \\
\dot{R}_t &= \gamma I_t,
\end{aligned} \tag{1}$$

where the positive parameters β, α, and γ denote the infection rate, incubation rate, and recovery rate, respectively.

2.3. Statistical analysis

We derive a likelihood function for observing n positives out of N incarcerated individuals on day u as follows: Using the DSA approach of [29, 30, 32, 33], we interpret S_t as an improper survival function. The mathematical justification for such an interpretation is provided by the Sellke construction by which the function S_t can be identified as the limiting probability of an initially susceptible individual not getting infected by time t. Note that the function S_t satisfying (1) is indeed a decreasing function and, when properly scaled, we set $S_0 = 1$. However, unlike proper survival functions that vanish at infinity (i.e., decrease to zero in the limit), the function $S_t \to S_\infty > 0$ as $t \to \infty$ so it is an improper survival function. However, we make it a proper survival function by conditioning on ever being infected. Given observation up to time $T > 0$, the time T_E that an initially susceptible individual becomes infected and enters the E compartment follows the conditional probability density function

$$f_T(t) = -\frac{\dot{S}_t}{\tau_T}, \tag{2}$$

where $\tau_T = 1 - S_T$. The time T_I to becoming infectious has the conditional density
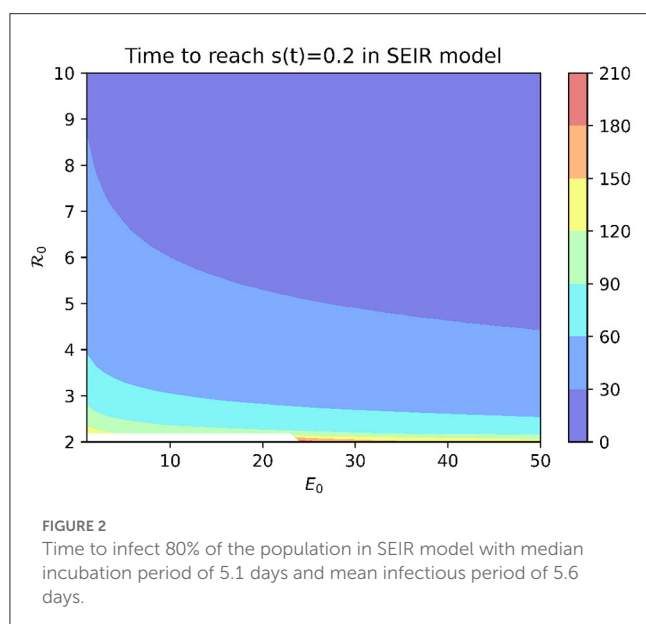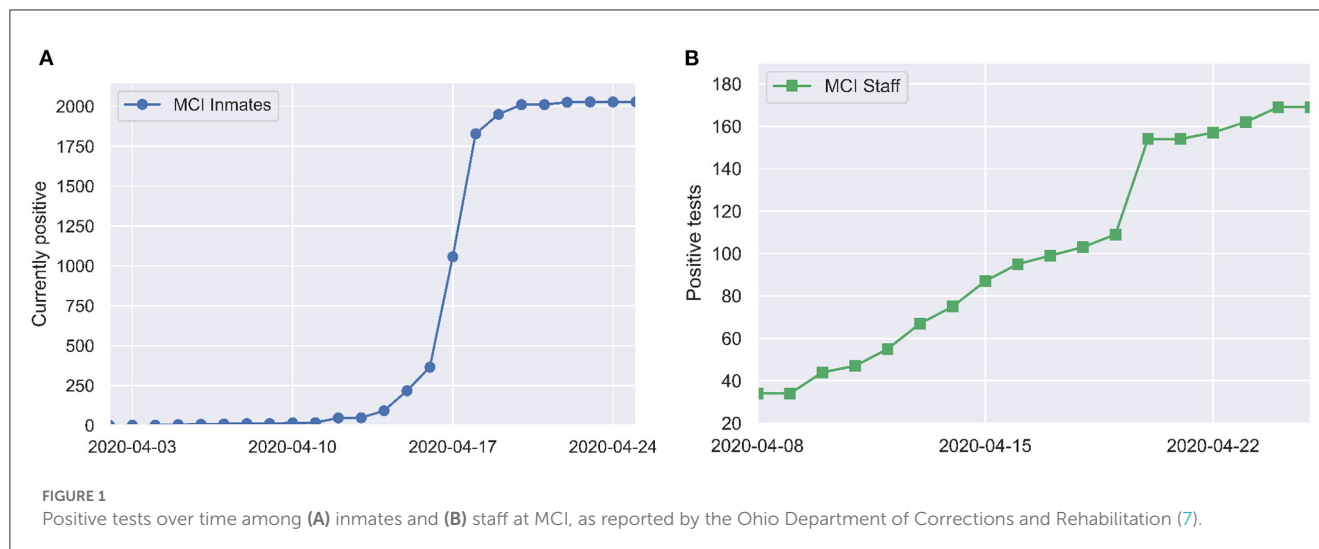
$$g_T(t) = \frac{\alpha E_t}{\tau_t}, \tag{3}$$

and the recovery time T_R has the conditional density

$$h_T(t) = \frac{\gamma(I_t - \rho e^{-\gamma t})}{\tau_T}. \tag{4}$$

Note that the random variables T_E, $T_I - T_E$, and $T_R - T_I$ are mutually independent and that $T_I - T_E$ and $T_R - T_I$ have exponential distributions with rates α and γ, respectively [29]. The parameter ρ is the initial proportion of infectious individuals.

Mass testing yields a number of individuals who test positive and a total number of tests administered on the day of mass testing. To use these data, let T_N denote the time when virus first becomes undetectable in an individual. We then describe the epidemic process by the pair of random variables (T_E, T_N). Let $\varepsilon = T_N - T_E$. Then, the probability of an individual testing positive on the day of the mass testing (at time u) is given by

$$p_u = \Pr(T_E < u < T_N) = \Pr(0 < u - T_E < \varepsilon). \tag{5}$$

FIGURE 1
Positive tests over time among **(A)** inmates and **(B)** staff at MCI, as reported by the Ohio Department of Corrections and Rehabilitation (7).

FIGURE 2
Time to infect 80% of the population in SEIR model with median incubation period of 5.1 days and mean infectious period of 5.6 days.

We fix $\varepsilon = 21$ days, corresponding to detectable virus for 3 weeks following an individual becoming infectious (34, 35). We set $1/\alpha = 5.1/\log(2)$ days [corresponding to a median incubation period of 5.1 days (36)] and assume a mean infectious period $1/\gamma$ of 5.6 days (37).

If n out of N individuals test positive on the day of the mass testing u, the log-likelihood function is given by

$$\ell(\beta \mid n, N) = \log\left(\binom{N}{n} p_u^n (1 - p_u)^{N-n}\right), \qquad (6)$$

with the probability of testing positive p_u as described in Equation (5). Note that the above likelihood function is a consequence of the functional law of the large numbers for Poisson processes and the Sellke construction.

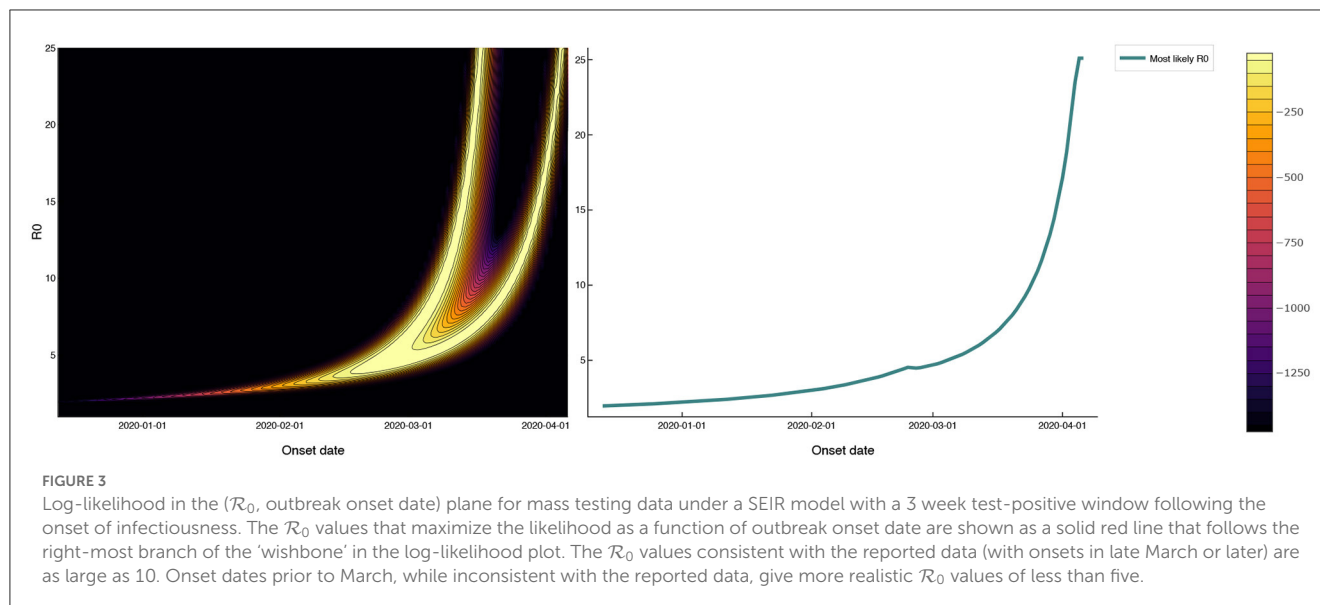The crux of the DSA method is that it allows one to interpret functions that describe the large-population limiting proportions of individuals in different compartments as probabilistic quantities, such as survival functions or probability density functions of transfer times from one compartment to another. This change in perspective has a number of statistical advantages. For instance, it makes available the entire toolkit of survival analysis, by virtue of which it can account for censoring, truncation and aggregation of data in a natural way. Variations of the DSA method have been recently applied to analyze not only COVID-19, but also the 2001 foot-and-mouth disease (FMD) outbreak in the United Kingdom (30) and multiple waves of the 2018–2020 Ebola epidemic in the Democratic Republic of Congo (32). It is important to note that the date considered in this article are from the first phase of the pandemic when vaccines were not yet available. Nevertheless, the DSA method is capable to incorporating vaccination regimes. For instance, the method was applied to assess the potential impact of vaccination in Israel in (38). See also Klaus et al. (39) where the method was applied to COVID-19 data in the state of Ohio, USA.

3. Results

3.1. Reported outbreak time course

According to ODRC reports (7), the first identified COVID-19 case at MCI was an infected staff member on March 29. Following this initial case, precautions such as cohorting and modified movement were enacted in order to restrict mixing and reduce transmission. As stated in the publicly available ODRC report from March 30, 2020:

> Based on a staff member reporting a positive COVID test, MCI is operating under modified movement and the population is being separated by unit along with other precautionary measures. Every inmate at MCI is monitored daily and has their temperature taken along with a check for symptoms. Currently, there are no inmates symptomatic for COVID-19.

FIGURE 3
Log-likelihood in the ($\mathcal{R}_0$, outbreak onset date) plane for mass testing data under a SEIR model with a 3 week test-positive window following the onset of infectiousness. The $\mathcal{R}_0$ values that maximize the likelihood as a function of outbreak onset date are shown as a solid red line that follows the right-most branch of the 'wishbone' in the log-likelihood plot. The $\mathcal{R}_0$ values consistent with the reported data (with onsets in late March or later) are as large as 10. Onset dates prior to March, while inconsistent with the reported data, give more realistic $\mathcal{R}_0$ values of less than five.

The first COVID-19 case among inmates was identified on April 3. Mass RT-PCR testing of all inmates and partial testing of staff was conducted on April 16. By April 20, SARS-CoV-2 infection had been identified in 79% (1,950/2,453) of inmates and 35% (154/446) of staff. These numbers come directly from data on the ODRC website. We take the May 5, 2020 listing of 2,453 inmates at MCI as the denominator. There is a lag of a few days between when mass testing occurred (April 16) and when jumps in case counts are reported in the ODRC data (April 18–19 for inmates and April 20 for staff), which may reflect delay in data entry. Figure 1 shows a time series of reported COVID-19 cases at MCI.

3.2. Basic reproduction number and initial exposure size

The basic reproduction number $\mathcal{R}_0$ is one of the key parameters in models of infectious diseases (31). It is defined as the expected number of secondary cases generated by an infected individual in a population where all individuals are susceptible to infection. When $\mathcal{R}_0 > 1$, disease can spread rapidly and cause a large epidemic with positive probability. When $\mathcal{R}_0 < 1$, the spread of disease dies out stochastically and a large epidemic cannot occur. $\mathcal{R}_0$ can also be used to calculate the so-called "herd immunity threshold" for interventions like vaccination that effectively reduce the susceptible population.

To examine which values of the basic reproduction number $\mathcal{R}_0$ are consistent with the rapid spread of COVID-19 observed at MCI, we use the SEIR model (1). In order for mass screening to identify 80% of the population as positive for COVID-19 on April 16, at least 80% of the population must have been infected by that date. Figure 2 shows the time needed to infect 80% of the population in the SEIR model as a function of $\mathcal{R}_0$ and the initial number of exposed individuals (E_0). While $\mathcal{R}_0$ values of two or larger are able to eventually infect 80% or more of the population, this can take on the order of months for modest values

of $\mathcal{R}_0$. Reproduction numbers >14 are needed before outbreaks originating from a single exposed individual are able to generate a 3-week cumulative incidence consistent with that reported for MCI.

An alternative explanation is that the outbreak involved more than one initially infected prisoner. Figure 2 shows that, for a fixed $\mathcal{R}_0$ value, increasing E_0 decreases the time needed to infect 80% of the population. However, an initial condition of $E_0 > 563$ is needed for an outbreak with $\mathcal{R}_0 = 3$ to infect 80% of the population within 3 weeks.

3.3. Time of initial outbreak circulation

A third possibility is that SARS-CoV-2 was circulating among prisoners prior to April 3. Figure 3 shows the log-likelihood (6) for observing the mass testing results in MCI according to the SEIR model (1) as a function of $\mathcal{R}_0$ and the outbreak onset date, with E_0 fixed at one. The outbreak onset date and $\mathcal{R}_0$ are unidentifiable from the mass testing data alone, with the "wishbone" shape running diagonally across Figure 3 corresponding to pairs of outbreak onset and $\mathcal{R}_0$ that are almost equally likely given the observed data. Outbreak onsets in late March or later correspond to $\mathcal{R}_0 > 10$, while earlier outbreak onsets correspond to smaller $\mathcal{R}_0$ values. Note that onset dates prior to March are required to give $\mathcal{R}_0$ values of less than five.

In general, the larger the value of the parameter $\mathcal{R}_0$, the more difficult it is to control the epidemic. Our analysis is consistent with this. Both the first and the third possibilities explained above suggest that the $\mathcal{R}_0$ values consistent with the reported data must be extremely high, calling attention to the explosive potential for COVID-19 transmission in prisons. Both the second and the third possibilities underscore the implausibility of the reported disease introduction date and/or the initial amount of infection, calling attention to the need for more reliable monitoring and reporting of infection in correctional facilities.

4. Discussion and conclusion

The official reports from ODRC describe widespread infection of MCI inmates with SARS-CoV-2 within the span of 3 weeks. Our primary contribution is a rigorous analysis of the data using an SEIR compartmental model fit to these data using the DSA approach, which allows us to use all of the tools of likelihood-based inference. This analysis indicates three non-exclusive possible explanations for this rapid spread: (i) values for the basic reproduction number that are far higher than the $\mathcal{R}_0$ values between two and three that have been estimated for the ancestral strain of SARS-CoV-2 in non-prison settings in the United States (37), (ii) initial exposure of a large number of infected prisoners as in an extreme superspreading event, or (iii) early undetected circulation of SARS-CoV-2 among prisoners prior to April 3. We note that the $\mathcal{R}_0$ values in (i) are even greater than the already high estimates of the basic reproduction number in a large urban jail (28). All three possibilities speak to the vulnerabilities of prison inmates and staff to COVID-19. Distinguishing between these different scenarios is impossible without improved data collection and reporting. An arguable weakness of our analysis is that it is retrospective in nature. However, we believe studies such as ours will lead to improvements that allow more detailed insight into the transmission of respiratory infections within prisons are critical for protecting the health of prison inmates, staff, and surrounding communities in future pandemics.

Permissive conditions for spread within correctional facilities, challenges for disease surveillance and care in these settings, and the inextricable link between COVID-19 within correctional facilities and disease spread in the surrounding community, have been discussed eloquently by others (10–14). Structural changes such as lower inmate densities (25, 40) and improved ventilation (9, 16) are needed to decrease transmission potential in correctional facilities. Efforts to increase vaccine coverage are also important, particularly among prison staff who may have relatively low vaccine uptake (23). Community case rates are associated with cases in prisons (18), inmate transfers can allow outbreaks to jump from one prison to another (20), and staff can be an epidemiological link between correctional facilities and surrounding communities. Without changes to protect the health of staff and inmates, it is predictable prisons will be vulnerable to extremely rapid spread of future respiratory pathogens.

Improved surveillance and reporting are critical for pandemic preparedness and for preventing or controlling future outbreaks of respiratory diseases in prisons. Testing policies during the COVID-19 pandemic varied widely across state corrections departments (24). Testing protocols changed over time, and state reporting of COVID-19 cases in prisons was often incomplete or absent (41). Swift response is essential for preventing and controlling large outbreaks, and it has been identified as a distinguishing feature for countries with successful COVID-19 pandemic responses (42). This swift response is impossible without pathogen detection and reporting efforts that include correctional facilities. Going forward, we urge health departments and corrections departments to collect accurate data and to make these data available for analysis with appropriate protections for human subjects in this vulnerable population.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/wasiur/PrisonCOVID19Analysis.

Author contributions

SK collected and organized the data. WK, EK, GR, and JT developed the analytical approach. WK carried out the analysis. All authors contributed to the conception and writing of the manuscript and approved the submitted version.

Funding

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. The COVID Prison Project. *National COVID-19 Statistics* (2021) Available online at: https://covidprisonprojectcom/data/national-overview/ (accessed November 01, 2022).

2. Wang EA, Zenilman J, Brinkley-Rubinstein L. Ethical considerations for COVID-19 vaccine trials in correctional facilities. *JAMA*. (2020) 324:1031–2. doi: 10.1001/jama.2020.15589

3. Mhlanga-Gunda R, Rusakaniko S, Chinyuku AN, Pswarayi VF, Robinson CS, Kewley S, et al. "We sleep 10 cm apart so there is no social distancing": COVID-19 preparedness in a Zimbabwean prison complex. *Int J Prison Health*. (2022). doi: 10.1108/IJPH-10-2021-0101/full/html. [Epub ahead of print].

4. Chapell B. *73% of Inmates at an Ohio Prison Test Positive for Coronavirus*. NPR (2020). Available online at: https://www.npr.org/sections/coronavirus-live-updates/2020/04/20/838943211/73-of-inmates-at-an-ohio-test-positive-for-coronavirus (accessed February 12, 2023).

5. Hooks G. *COVID-19 Spread Faster in Counties With Large Prisons–And to Nearby Counties: Marion Correctional Institution (Ohio) is a Disturbing Example* (2020). Available online at: https://www.prisonpolicy.org/reports/covidspread_marion.html (accessed February 12, 2023).

6. Woods J, Cooley P. *Coronavirus in Ohio: More Than 1,800 Inmates at Marion Correctional Test Positive* (2020). Available online at: https://eu.cincinnati.com/story/news/2020/04/19/coronavirus-ohio-more-than-1-800-inmates-marion-correctional-test-positive/5163046002/ (accessed February 12, 2023).

7. Ohio Department of Rehabilitation & Correction. *COVID-19 Inmate Testing* (2020). Available online at: https://coronavirus.ohio.gov/static/reports/DRCCOVID-19Information.pdf (accessed April 30, 2020).

8. Hagan LM, Williams SP, Spaulding AC, Toblin RL, Figlenski J, Ocampo J, et al. Mass testing for SARS-CoV-2 in 16 prisons and jails - six jurisdictions, United States, April-May 2020. *MMWR Morbid Mortal Wkly Rep*. (2020) 69:1139–43. doi: 10.15585/mmwr.mm6933a3

9. Kwan A, Sklar R, Cameron DB, Schell RC, Bertozzi SM, McCoy SI, et al. Respiratory pandemic preparedness learnings from the June 2020 COVID-19 outbreak at San Quentin California State Prison. *Int J Prison Health*. (2022). doi: 10.1108/IJPH-12-2021-0116. [Epub ahead of print].

10. Akiyama MJ, Spaulding AC, Rich JD. Flattening the curve for incarcerated populations – COVID-19 in jails and prisons. *N Engl J Med*. (2020) 382:2075–7. doi: 10.1056/NEJMp2005687

11. Burki T. Prisons are "in no way equipped" to deal with COVID-19. *Lancet*. (2020) 39510234:1411–2. doi: 10.1016/S0140-6736(20)30984-3

12. Oladeru OT, Tran NT, Al-Rousan T, Williams B, Zaller N. A call to protect patients, correctional staff and healthcare professionals in jails and prisons during the COVID-19 pandemic. *Health Just*. (2020) 8:17. doi: 10.1186/s40352-020-00119-1

13. Rubin R. The challenge of preventing COVID-19 spread in correctional facilities. *JAMA*. (2020) 323:1760–1. doi: 10.1001/jama.2020.5427

14. Wurcel AG, Dauria E, Zaller N, Nijhawan A, Beckwith C, Nowotny K, et al. Spotlight on jails: COVID-19 mitigation policies needed now. *Clin Infect Dis*. (2020) 71:891–2. doi: 10.1093/cid/ciaa346

15. Mastrobuoni G, Terlizzese D. Leave the door open? Prison conditions and recidivism. *Am Econ J*. (2022) 14:200–33. doi: 10.1257/app.20190038

16. Duarte C, Cameron DB, Kwan AT, Bertozzi SM, Williams BA, McCoy SI. COVID-19 outbreak in a state prison: a case study on the implementation of key public health recommendations for containment and prevention. *BMC Publ Health*. (2022) 22:977. doi: 10.1186/s12889-022-12997-1

17. Maycock M. 'COVID-19 has caused a dramatic change to prison life'. Analysing the impacts of the COVID-19 pandemic on the pains of imprisonment in the Scottish Prison Estate. *Br J Criminol*. (2022) 62:218–33. doi: 10.1093/bjc/azab031

18. LeMasters K, Ranapurwala S, Maner M, Nowotny KM, Peterson M, Brinkley-Rubinstein L. COVID-19 community spread and consequences for prison case rates. *PLoS ONE*. (2022) 17:e0266772. doi: 10.1371/journal.pone.0266772

19. Reinhart E, Chen DL. Incarceration and its disseminations: COVID-19 pandemic lessons from Chicago's cook county jail. *Health Affairs*. (2020) 39:1412–8. doi: 10.1377/hlthaff.2020.00652

20. Parsons TL, Worden L. Assessing the risk of cascading COVID-19 outbreaks from prison-to-prison transfers. *Epidemics*. (2021) 37:100532. doi: 10.1016/j.epidem.2021.100532

21. Puglisi LB, Brinkley-Rubinstein L, Wang EA. COVID-19 in carceral systems: a review. *Annu Rev Criminol*. (2023) 6:399–422. doi: 10.1146/annurev-criminol-030521-103146

22. Chin ET, Leidner D, Zhang Y, Long E, Prince L, Li Y, et al. Effectiveness of the mRNA-1273 Vaccine during a SARS-CoV-2 Delta Outbreak in a Prison. *N Engl J Med*. (2021) 385:2300–1. doi: 10.1056/NEJMc2114089

23. Prince L, Long E, Studdert DM, Leidner D, Chin ET, Andrews JR, et al. Uptake of COVID-19 vaccination among frontline workers in California state prisons. *JAMA Health Forum*. (2022) 3:e220099. doi: 10.1001/jamahealthforum.2022.0099

24. Zielinski MJ, Cowell M, Bull CE, Veluvolu M, Behne MF, Nowotny K, et al. Policy and public communication methods among U.S. state prisons during the first year of the COVID-19 pandemic. *Health Just*. (2022) 10:27. doi: 10.1186/s40352-022-00187-5

25. Malloy GSP, Puglisi L, Brandeau ML, Harvey TD, Wang EA. Effectiveness of interventions to reduce COVID-19 transmission in a large urban jail: a model-based analysis. *BMJ Open*. (2021) 11:e042898. doi: 10.1136/bmjopen-2020-042898

26. Kirbiyik U, Binder AM, Ghinai I, Zawitz C, Levin R, Samala U, et al. Network characteristics and visualization of COVID-19 outbreak in a large detention facility in the United States – Cook County, Illinois, 2020. *Morbid Mortal Wkly Rep*. (2020) 69:1625–30. doi: 10.15585/mmwr.mm6944a3

27. Ryckman T, Chin ET, Prince L, Leidner D, Long E, Studdert DM, et al. Outbreaks of COVID-19 variants in US prisons: a mathematical modelling analysis of vaccination and reopening policies. *Lancet Publ Health*. (2021) 6:e760–70. doi: 10.1016/S2468-2667(21)00162-6

28. Puglisi LB, Malloy GSP, Harvey TD, Brandeau ML, Wang EA. Estimation of COVID-19 basic reproduction ratio in a large urban jail in the United States. *Ann Epidemiol*. (2021) 53:103–5. doi: 10.1016/j.annepidem.2020.09.002

29. KhudaBukhsh WR, Choi B, Kenah E, Rempala GA. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. *Interface Focus*. (2020) 10:20190048. doi: 10.1098/rsfs.2019.0048

30. Di Lauro F, KhudaBukhsh WR, Kiss IZ, Kenah E, Jensen M, Rempała GA. Dynamic survival analysis for non-Markovian epidemic models. *J R Soc Interface*. (2022) 19:20220124. doi: 10.1098/rsif.2022.0124

31. Brauer F, Driessche P, Wu J. *Mathematical Epidemiology*. Berlin; Heidelberg: Springer (2008).

32. Vossler H, Akilimali P, Pan Y, KhudaBukhsh WR, Kenah E, Rempała GA. Analysis of individual-level data from 2018-2020 Ebola outbreak in Democratic Republic of the Congo. *Sci Rep*. (2022) 12:5534. doi: 10.1038/s41598-022-09564-4

33. KhudaBukhsh WR, Bastian CD, Wascher M, Klaus C, Sahai SY, Weir MH, et al. Projecting COVID-19 cases and hospital burden in Ohio. *J Theor Biol*. (2023) 561:111404. doi: 10.1016/j.jtbi.2022.111404

34. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med*. (2020) 26:672–5. doi: 10.1038/s41591-020-0869-5

35. Xiao AT, Tong YX, Zhang S. Profile of RT-PCR for SARS-CoV-2: a preliminary study from 56 COVID-19 patients. *Clin Infect Dis*. (2020) 71:2249–51. doi: 10.1093/cid/ciaa460

36. 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

37. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc Natl Acad Sci USA*. (2020) 117:9122–6. doi: 10.1073/pnas.2004064117

38. Somekh I, KhudaBukhsh WR, Root ED, Boker LK, Rempala G, Simões EAF, et al. Quantifying the population-level effect of the COVID-19 mass vaccination campaign in Israel: a modeling study. *Open Forum Infect Dis*. (2022) 9:Ofac087. doi: 10.1093/ofid/ofac087

39. Klaus C, Wascher M, KhudaBukhsh WR, Rempała GA. Likelihood-free dynamical survival analysis applied to the COVID-19 epidemic in Ohio. *Math Biosci Eng*. (2023) 20:4103–27. doi: 10.3934/mbe.2023192

40. LeMasters K, Brinkley-Rubinstein L, Maner M, Peterson M, Nowotny K, Bailey Z. Carceral epidemiology: mass incarceration and structural racism during the COVID-19 pandemic. *Lancet Publ Health*. (2022) 7:e287–90. doi: 10.1073/pnas.2026577118

41. Knight K, Bleckner J, Cameron E, Amon JJ. Pandemic treaty should include reporting in prisons. *Health Hum Rights*. (2022) 24:117–9.

42. Sirleaf EJ, Clark H. Report of the independent panel for pandemic preparedness and response: making COVID-19 the last pandemic. *Lancet*. (2021) 398:101–3. doi: 10.1016/S0140-6736(21)01095-3

frontiers | Frontiers in Public Health

A global scale COVID-19 variants time-series analysis across 48 countries

Rachel Yui Ki Chu[1†], Kam Chiu Szeto[2†], Irene Oi Ling Wong[1] and Pui Hong Chung[1]*

[1]School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2]Department of Finance, Business School, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China

Background: The coronavirus disease (COVID-19) pandemic is slowing down, and countries are discussing whether preventive measures have remained effective or not. This study aimed to investigate a particular property of the trend of COVID-19 that existed and if its variants of concern were cointegrated, determining its possible transformation into an endemic.

Methods: Biweekly expected new cases by variants of COVID-19 for 48 countries from 02 May 2020 to 29 August 2022 were acquired from the GISAID database. While the case series was tested for homoscedasticity with the Breusch–Pagan test, seasonal decomposition was used to obtain a trend component of the biweekly global new case series. The percentage change of trend was then tested for zero-mean symmetry with the one-sample Wilcoxon signed rank test and zero-mean stationarity with the augmented Dickey–Fuller test to confirm a random COVID trend globally. Vector error correction models with the same seasonal adjustment were regressed to obtain a variant-cointegrated series for each country. They were tested by the augmented Dickey–Fuller test for stationarity to confirm a constant long-term stochastic intervariant interaction within the country.

Results: The trend series of seasonality-adjusted global COVID-19 new cases was found to be heteroscedastic ($p = 0.002$), while its rate of change was indeterministic ($p = 0.052$) and stationary ($p = 0.024$). Seasonal cointegration relationships between expected new case series by variants were found in 37 out of 48 countries ($p < 0.05$), reflecting a constant long-term stochastic trend in new case numbers contributed from different variants of concern within most countries.

Conclusion: Our results indicated that the new case long-term trends were random on a global scale and stable within most countries; therefore, the virus was unlikely to be eliminated but containable. Policymakers are currently in the process of adapting to the transformation of the pandemic into an endemic.

Introduction

The World Health Organization (WHO) declared the SARS-CoV-2, commonly known for causing COVID-19, a global pandemic crisis on 11 March 2020 (1). According to the WHO, as of 16 September 2022, in total, there were 608,328,548 confirmed cases and 6,501,469 claimed deaths (1). It had been over 2.5 years since the declaration, and this epidemiological crisis has remained a controversial issue worldwide. Numerous challenges

came along with the pandemic. Scientists and researchers across the world were working around the clock to invent vaccines and strategies for curing COVID-19. Despite researchers across the world continuing to investigate the strategies to end the pandemic, including vaccination promotion, vaccine effectiveness toward variants of concern (VOCs), and quarantine effectiveness (2, 3), the pandemic was still ongoing.

The long fight against COVID-19 had led to an economic downturn, as many countries had imposed lockdowns, which had hugely limited global human mobility. In addition to the economic aspects, the healthcare system had been given extra resources and burdens (4). The WHO had provided operational guidance for maintaining essential health services while enhancing medical surveillance to contain the spread of COVID-19 (5). This has resulted in a large number of patients being affected and delayed their medical appointment schedules (6). Furthermore, countries are facing a dilemma in balancing the COVID-19 response and essential healthcare services. Various viewpoints were raised by different countries, with some of them insisting that preventive measures should be upheld while others preferred the world to return to normal without strict preventive measures against the virus. In this study, we aimed to investigate whether there is an uncontrollable, random trend of global new cases and to identify whether the COVID virus behaved like the influenza virus to be long-living and seasonally fluctuating with different VOCs (7) or as a one-off outbreak like severe acute respiratory syndrome (SARS) (8).

Method

Data extraction and processing

Daily reported new case numbers, and the ratio of major concerns of variants per each country from 02 May 2020 to 29 August 2022 were acquired from the GISAID database (https://ourworldindata.org/grapher/covid-variants-bar) (9). There were 10 time series representing the ratio of major concerns of variants per each country, Alpha, Beta, Gamma, Delta, Omicron (BA.1), Omicron (BA.2), Omicron (BA.4), Omicron (BA.5), Omicron (BA.2.12.1), and Omicron (BA.2.75), and 1 extra for other variants. The data were resampled on a biweekly basis due to data sparsity of variant ratio data. There were at most 61 data points per country. The missing data points were filled by the previous data point, which extended for at most 1 month, assuming that the monthly variations were not significant. Only countries with processed data for over two-thirds of the period, i.e., at least 41 data points, were analyzed to avoid misinterpretation of the results. A total of 48 countries met the inclusive criteria after filtering: Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Canada, Chile, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, India, Indonesia, Ireland, Israel, Italy, Japan, Kenya, Latvia, Lithuania, Luxembourg, Malaysia, Mexico, Netherlands, Norway, Peru, Philippines, Poland, Portugal, Romania, Russia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States. Ratios of the submitted sequence of COVID variants of each country were then multiplied with the biweekly reported new case numbers of each country, yielding the expected number of biweekly new cases by a variant type for each country.

Long-term trends of worldwide biweekly new cases

By adding up the biweekly new cases of the included countries, a worldwide biweekly new case series was obtained. It was decomposed for the seasonal component analysis to analyze its trend, seasonality, and noise to remove the seasonality property of COVID (10). An additive model would be used if the series
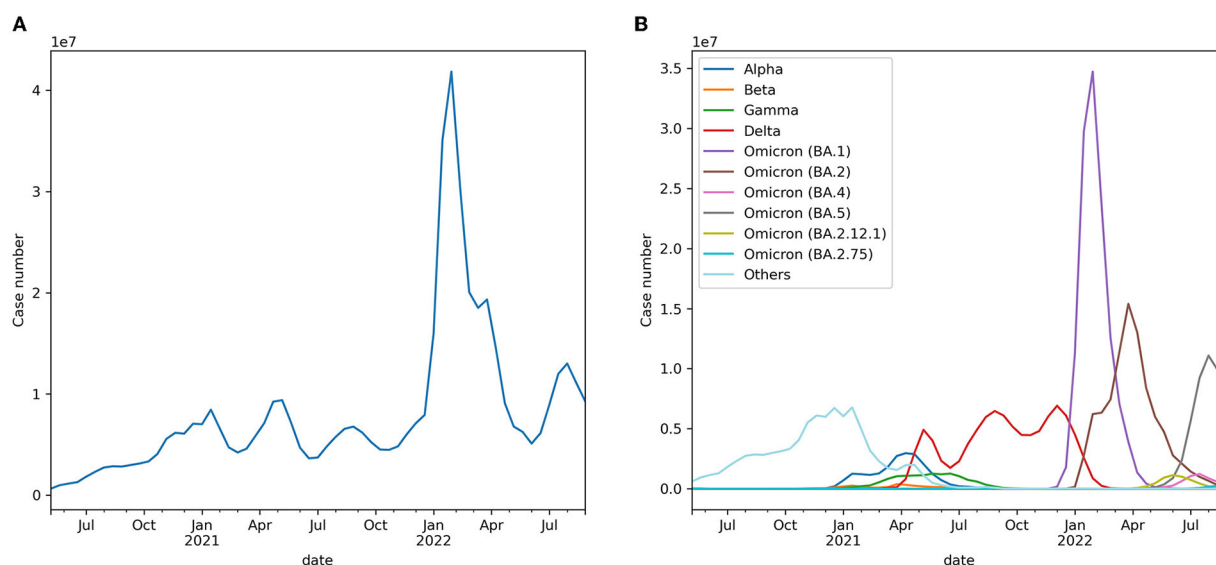

FIGURE 1
(A) Expected worldwide biweekly new cases. **(B)** Expected worldwide biweekly new cases by variants.
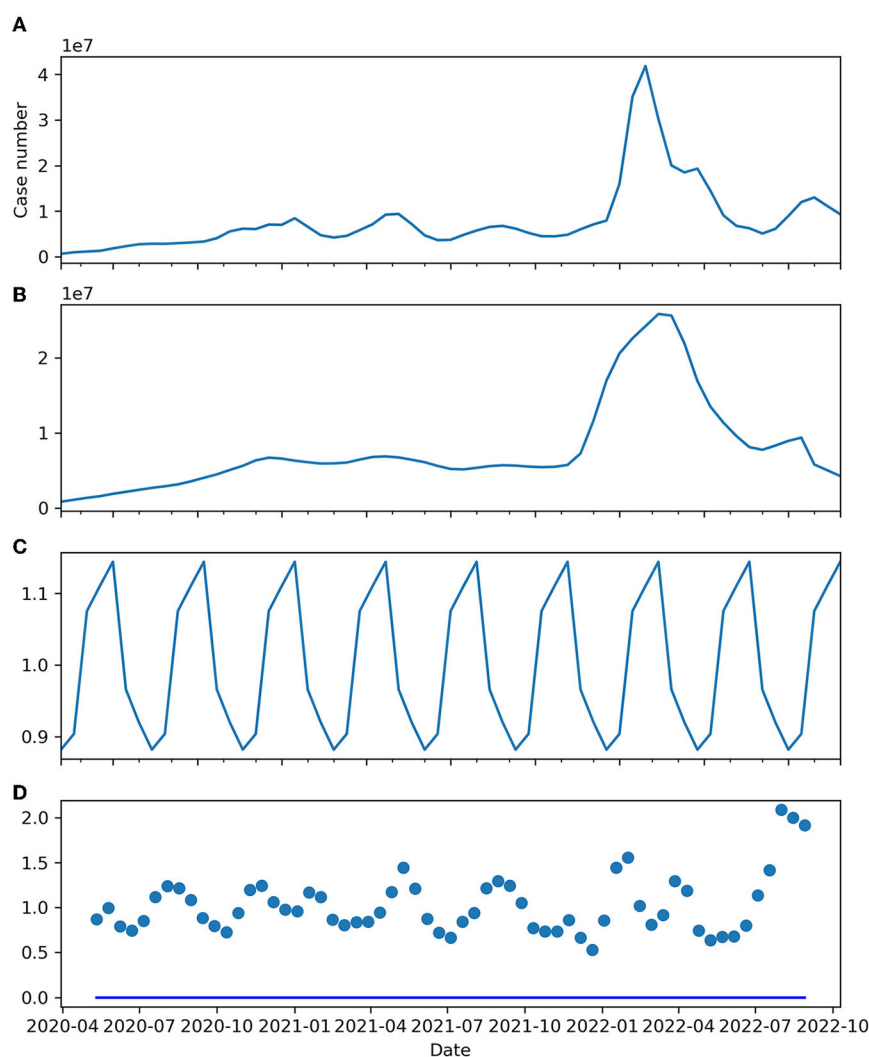
FIGURE 2
(A) Expected worldwide biweekly new cases. **(B)** Trend component. **(C)** Seasonal component. **(D)** Residual.

was homoscedastic (serial independence in regression residual) by time progression, which would be verified by the Breusch–Pagan test (11). Otherwise, a multiplicative model would be used (12). The cycle period was set at seven time steps, assuming a regular quarterly spatiotemporal fluctuation (13).

$$
\begin{cases}
y(t) = T(t) + S(t) + \epsilon & \textit{additive model} \\
y(t) = T(t) \times S(t) \times \epsilon & \textit{multiplicative model}
\end{cases}
$$

where $y(t)$ represented the worldwide biweekly new case series to be decomposed, $T(t)$ was the trend component representing the long-term progression of the series, $S(t)$ was the seasonality component representing the regular seasonal variation of the series, and ϵ was the residual noise.

In this study, the convolution method (linear kernel) was used to filter the trend and seasonality components (14). The trend component was extracted by the following moving average formula:

$$
T(t) = \frac{1}{2 \times 7 + 1} \sum_{i=-7}^{7} x_{t+i} \; \textit{for } t > 7
$$

Then, by removing the trend component from the observed data, the seasonal component was obtained by the mean of every 7th data point from the detrended data starting from the 1st, 2nd, 3rd, 4th, 5th, 6th, and 7th data points:

$$
S(t) = \frac{1}{n(A)} \sum_{i \in A} [x_i - T(i)]
$$

where

$$
A = \{i : i \bmod 7 = t \bmod 7 \textit{ for } 1 \le i \in \mathbb{N} \le n(X)\}
$$
$$
n(A) = \#\textit{of items in } A
$$

The remaining unexplained component by trend and seasonality of the observed data was considered residual noise.

The regressed trend component $T(t)$ was extracted as the smoothened series representing the trend of the worldwide biweekly new cases for studying the long-term trend. The percentage change per time step of this trend was tested with the
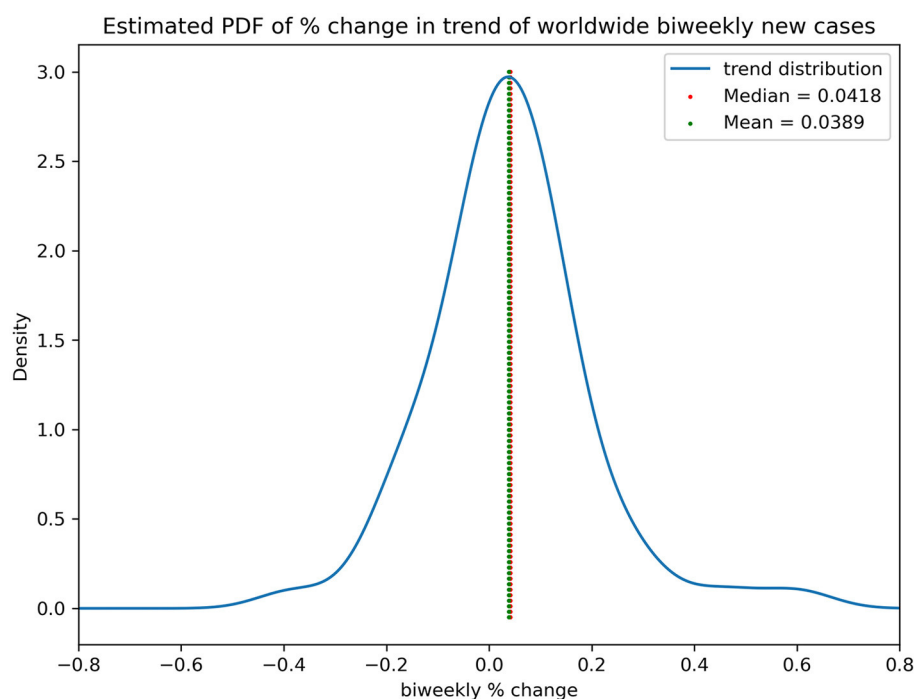
FIGURE 3
Estimated PDF of % change in trend of worldwide biweekly new cases.

one-sample Wilcoxon signed-rank test for zero-mean symmetrical distribution and augmented the Dickey–Fuller test for stationarity, with no lag-level difference allowed and a non-deterministic trend. We could identify the characteristics of the change of trend component to observe if a random or drifted long-term trend of the number of worldwide biweekly new cases existed.

Cointegration between expected new case numbers by variants within countries

Variants of concern (VOCs) were studied by individual countries. To verify if there exist any cointegration relationships between the time series of the estimated number of reported cases of different variants of each shortlisted country, i.e., if the case-by-variant series would have a long-term constant stochastic trend, a Vector Error Correction Model (VECM) with one seasonal lag level (7 data points) was regressed for the case of each country, under the hypothesis that there exist (1) at least one cointegration rank with case numbers by variants and (2) a constant long-term trend of case numbers such that the seasonal difference was a zero-mean normal variable.

$$\Delta y_t = \phi_0 + \Pi_{y_{t-1}} + \Phi \Delta y_{t-1} + \epsilon_t$$
$$\Delta y_t = y_t - y_{t-7}$$

where y_t was the case-by-variant vector at time t, ϕ_0 was the regressed intercept term as the case-by-variant vector at time 0, $\Pi_{y_{t-1}}$ was the error correction term, Φ was the coefficient vector of the auto-correlated case-by-variant term for time $t-1$, and

ϵ_t was white noise at time t. The error correction term $\Pi_{y_{t-1}}$ can be decomposed into $\alpha\beta^T$, where β could be extracted as the cointegrating vector for stationary testing (15).

The normalized dot product of the estimated cases by variants of each country and the corresponding cointegrating vector was obtained as the "cointegrated series" of the country. This series was tested by the augmented Dickey–Fuller test for stationarity, with no lag-level difference allowed and under a non-deterministic trend assumption. A stationary "cointegrated series" confirmed a constant long-term stochastic trend in the number of the combination of cases by variants of that particular country.

All data manipulation, visualization, modeling, and testing were carried out by Python 3.9.7 under Jupyter notebook environment, with the aid of the Pandas, NumPy, Matplotlib, SciPy, and Statsmodel libraries. The alpha values of all statistical tests were set as 0.05.

Results

Long-term trends of worldwide biweekly new cases

The trend of biweekly new cases over the globe and their expected compositions by variants are displayed in Figure 1. It was observed that the biweekly new case series had no deterministic or directional trend, except for the sudden surge in Omicron (BA.1 & BA.2) in early 2022. The Breusch–Pagan test results showed that the series was heteroscedastic ($p = 0.002$).

TABLE 1 Stationarity test on cointegration series by country.

Country	*p*-value	Cointegrated
Argentina	0.000000	✓
Australia	0.000029	✓
Austria	0.161500	
Bangladesh	0.008982	✓
Belgium	0.422995	
Brazil	0.001473	✓
Canada	0.984466	
Chile	0.000000	✓
Croatia	0.000000	✓
Czechia	0.000000	✓
Denmark	0.008688	✓
Estonia	0.000012	✓
Finland	0.000001	✓
France	0.000002	✓
Germany	0.000013	✓
Greece	0.000004	✓
Hong Kong	0.000019	✓
India	0.110762	
Indonesia	0.000001	✓
Ireland	0.000015	✓
Israel	0.683367	
Italy	0.000000	✓
Japan	1.000000	
Kenya	0.000000	✓
Latvia	1.000000	
Lithuania	0.000000	✓
Luxembourg	0.000000	✓
Malaysia	0.002662	✓
Mexico	0.009539	✓
Netherlands	0.735694	
Norway	0.000000	✓
Peru	0.000001	✓
Philippines	0.369270	
Poland	0.002056	✓
Portugal	0.935441	
Romania	0.000000	✓
Russia	0.000068	✓
Singapore	0.011175	✓
Slovakia	0.000000	✓
Slovenia	0.000828	✓
South Africa	0.000000	✓

(Continued)

TABLE 1 (Continued)

Country	*p*-value	Cointegrated
South Korea	0.974641	
Spain	0.000000	✓
Sweden	0.000002	✓
Switzerland	0.000095	✓
Turkey	0.000037	✓
United Kingdom	0.001633	✓
United States	0.044868	✓

As observed in a near-quarterly cyclic fluctuation, the series was decomposed into a trend, seasonal and residual components, as shown in Figure 2, *via* a multiplicative model. The mean and variance of the percentage change in the smoothened trend component were 0.039 (+/− 0.155) but were symmetrically distributed in terms of zero-mean ($p = 0.052$) and non-deterministically stationary ($p = 0.024$). Figure 3 shows the estimated probability density function of the percentage change in the smoothened trend component. A heavy tail on the positive side might have skewed the distribution, explaining why the distribution was only weakly symmetric, given the *p*-value was very close to the threshold of rejecting the null hypothesis.

Cointegration between expected case numbers by variants within countries

Table 1 and Figure 4 show that 37 out of 48 countries had their case-by-variants cointegrated. The cointegrated series of most countries were very stationary until some stirrings were observed in early- and mid-2022, which was the period when worldwide Omicron (BA.1 & BA.2) cases surged, but most of them returned to original levels afterward.

Discussion

Interpretation of research findings

Our results discovered that the percentage change in the underlying trend of the biweekly new case series was a zero-mean symmetrical distribution. The series was heteroscedastic, but meant differently for variance by time. However, the direction of the trend was indeterministic. This increasing randomness over time was very likely coming from the Omicron spike during early 2022. This development suggested that there existed random long-term biweekly COVID new case numbers after seasonal adjustment, which was difficult to model by any distribution. Given the percentage change in the symmetrical zero-mean distribution with non-deterministic variance, one might simulate the trend series by a Heston model (16). It was the collective result of all governments' interventions, people's actions, and environmental factors. As a result, it was likely that the trend in COVID new cases was uncontrollable, random, and unlikely to be diminished by human interference.
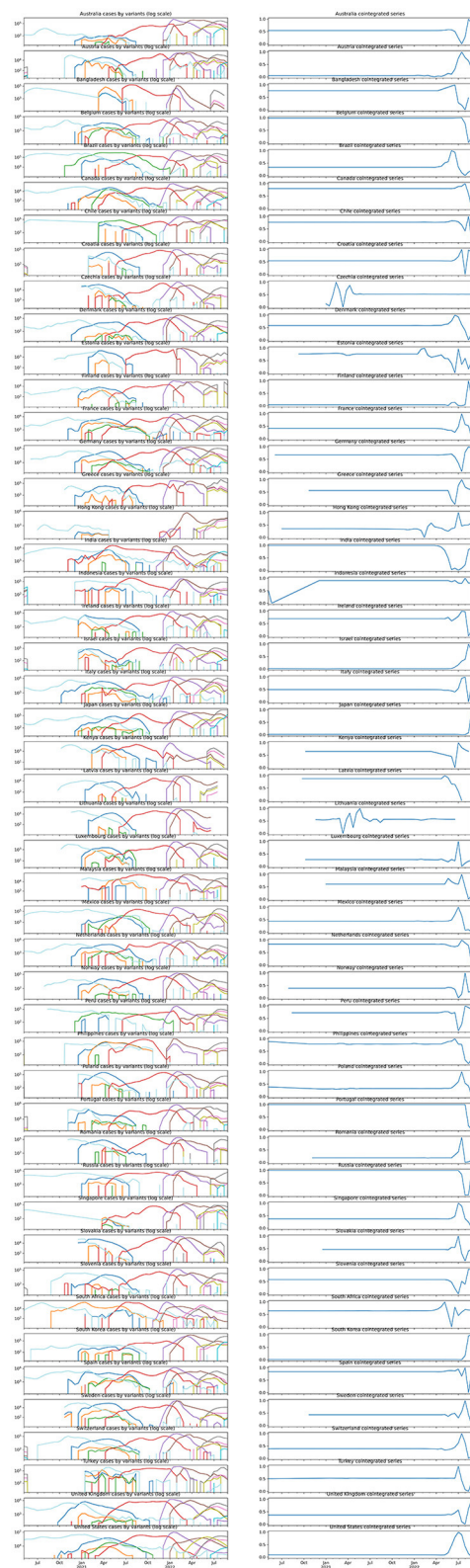
FIGURE 4
Case-by-variant by country and their corresponding normalized cointegration series.

major ethnicity, or population structure and density, exhibited a relatively constant and stable seasonality-adjusted cointegration relationship between different VOCs. A recent study suggested that a variant would dominate a period and subside but then will be replaced by another strand in its ratio of all COVID new cases (17). Our study provides an additional quantitative proof of not only the ratio of the VOCs but also of their newly infected numbers which behaved in that way, resulting in seasonal fluctuation but consistent COVID infection numbers that never ended. The variant-cointegrated countries had a wide range of stringent measures in COVID response policies (18). This could suggest that the strength of COVID control might be able to control the virus spread but not the existence of the virus, as no measure could be taken to prevent the rise of a new variant. The different properties between global and regional scales suggested that the inconsistent policies between countries made the infection uncontrollable, whereas local consistent strategies could contain the spread, regardless of their extent.

It was noticeable that only the Omicron outbreak during early-to-mid 2022 had cause a significant shock in the cointegration series in most countries. That was likely due to the unusually high infectivity and transmitting ability of the VOC (19), causing the infection numbers to ramp up and down sharply. The unsmooth transition of dominating VOC thus disrupted the balance temporarily.

Policies implication

COVID-19 has added an extra burden to the medical system in every country regardless of preventive measures, medical expenses, and research development. Millions of individual lives have been claimed from all walks of life. We are all desperate for a cure to end this pandemic and achieve a healthy community. Owing to the enormous infection numbers and exposure to antigens due to vaccine administration (20), and from our results, the transmission of COVID-19 possibly stayed in a relatively loosely controllable range. Based on our results, over 70% of the country was cointegrated while the VOC continued to surge, and the infection control implemented within a region is sufficient for containment of the disease spread. These measures might be covariates that affect the seasonal property of the disease spread and the infection rate in some regions (10). However, these local interventions remained random in the long-term global biweekly new cases. Thus, extreme preventive measures were unlikely to control the infection number to its aim of total elimination. These results aligned with the results of previous studies (20). Elimination in the community might not be worthwhile given the large amount of medical and social resources allocated. Policymakers should be aware of this issue to balance public health concerns and economical activities.

As stated by the WHO in a media briefing on 14 September 2022, COVID-19 will continue but the wild pandemic situation is coming to an end (21). Our results serve as a quantitative proof of the statement. Our results indicated that the virus appeared to be continuing regardless of the scale and strictness of the implemented infection control policies, but the effectiveness of intracountry

However, there were 77.1% of the involved worldwide nations, regardless of their geographical locations, side of the hemisphere,

containment shall be appreciated. Regional infection control measures and personal hygiene should be sustained to contain the spread. However, further or upgraded anti-virus implementation including lockdowns may not be effective in containing the virus. At the same time, countries might shift their focus from eliminating the virus to avoiding seasonal outbreaks threatening the local healthcare systems. It might also add indications to the direction of preventive measures, especially those measures that are related to vaccine research. Instead of focusing on the current variants of concern, it might be useful to predict and select a few possible virus strands that might be susceptible to a possible outbreak for vaccination, just like the influenza virus.

There are several strengths to this study. This research provided quantitative proof and perspectives on the current trend, seasonal, and cointegration properties of the COVID-19 new case series. Unveiling the underlying structure, it served as a guide to an early adaption of a possible transformation from the COVID pandemic into a regular respiratory endemic. We also pinpointed the need for shifting the policy focus from tackling the current COVID-19 situation to preventing future unknown new variant outbreaks.

The data themselves could also be concerning. Since there were only at most 61 data points per country, the regressed models were sensitive to sharp changes such as the Omicron surge in early-to mid-2022. From Figure 2D, we were able to observe a heteroscedastic, periodic fluctuation in residual noise, which could be due to insufficiency in the decomposition model to capture the full feature of the underlying seasonal signals. Thus, the found trend and seasonal properties are uncertain due to unmodeled factors. Improvement in modeling the decomposition, as well as continued observation, is needed for consolidating the evidence and conclusion. In addition, the data were logged under voluntary input in the database and might not be able to fully reflect the actual ratios of variants. While no data from any African country were available after processing, they aggregated the lack of input data issue. This would affect the representativeness of the data to cover the world's situation.

Conclusion

In this study, a random long-term trend of biweekly global new COVID cases was identified with a seasonal property. There existed cointegration relationships of newly reported cases of different variants of concerns for most countries, regardless of their demographics and responses toward the virus. The results suggested that consistent strategies could contain the spread. In addition, extreme eliminatory measures may not be effective, and a high possibility of the COVID pandemic was transforming into an endemic.

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

RC and KS contributed to the conceptualization, data curation, methodology, writing of the original draft, contributed to the writing and editing, and investigation. PC contributed to editing and supervision. IW contributed to editing and data curation. 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. World Health Organization. *WHO Coronavirus (COVID-19) Dashboard.* Available online at: https://covid19.who.int/ (accessed September 18, 2022).

2. Chu YK, Chung PH, Pang FC. Analysis of the effectiveness of measures on the COVID-19 vaccination rate in Hong Kong. *Vaccines.* (2022) 10:747. doi: 10.3390/vaccines100 50747

3. 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. *New England J Med.* (2021) 385:585–94. doi: 10.1056/NEJMoa2108891

4. Gebru AA, Birhanu T, Wendimu E, Ayalew AF, Mulat S, Abasimel HZ, et al. Global burden of COVID-19: situational analysis and review. *Hum Antibodies.* (2021) 29:139–48. doi: 10.3233/HAB-200420

5. World Health Organization. (2020). *Maintaining Essential Health Services: Operational Guidance for the COVID-19 Context*.

6. Riera R, Bagattini ÂM, Pacheco RL, Pachito DV, Roitberg F, Ilbawi A. Delays and disruptions in cancer health care due to COVID-19 pandemic: systematic review. *JCO Global Oncology*. (2021) 7:311–23. doi: 10.1200/GO.20.00639

7. Lowen AC, Steel J. Roles of humidity and temperature in shaping influenza seasonality. *J Virol*. (2014) 88:7692–5. doi: 10.1128/JVI.03544-13

8. Noy I, Shields S. The 2003 severe acute respiratory syndrome epidemic: A retroactive examination of economic costs. *Asian Develop Bank Econ Working Paper Ser*. (2019) 7:591. doi: 10.22617/WPS190469-2

9. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID's role in pandemic response. *China CDC Weekly*. (2021) 3:1049–51. doi: 10.46234/ccdcw2021.255

10. Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. *Proc Nat Acad Sci*. (2020) 117:27456–64. doi: 10.1073/pnas.2008590117

11. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. *Econ. J. Econ. Soc*. (1979) 1287–94. doi: 10.2307/1911963

12. Dodge Y, Cox D, Commenges D. The Oxford dictionary of statistical terms. Oxford: Oxford University Press on Demand (2006).

13. Townsend JP, Lamb AD, Hassler HB, Sah P, Nishio AA, Nguyen C, et al. *Projecting the Seasonality of Endemic COVID-19*. medRxiv (2022). doi: 10.1101/2022.01.26.22269905

14. Box GE, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis: Forecasting and Control*. New York: John Wiley & Sons (2015).

15. Feng Y, Palomar DP. A signal processing perspective on financial engineering. *Foundations Trends Signal Process*. (2016) 9:1–231. doi: 10.1561/2000000072

16. Heston SL. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev Financial Stud*. (1993) 6:327–43. doi: 10.1093/rfs/6.2.327

17. Barouch DH. Covid-19 vaccines—immunity, variants, boosters. *New England J Med*. (2022). doi: 10.1056/NEJMra2206573

18. Hale T, Angrist N, Kira B, Petherick A, Phillips T, Webster S. *Variation in Government Responses to COVID-19*.

19. Ren SY, Wang WB, Gao RD, Zhou AM. Omicron variant (B.11 529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. *World J Clin Cases*. (2022) 10:1. doi: 10.12998/wjcc.v10.i1.1

20. Murray CJ. COVID-19 will continue but the end of the pandemic is near. *Lancet*. (2022) 399:417–9. doi: 10.1016/S0140-6736(22)00100-3

21. The end of the COVID-19 pandemic is in sight: WHO. UN News. (2022). Available online at: https://news.un.org/en/story/2022/09/1126621 (accessed September 18, 2022).

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

frontiers

Frontiers in
Public Health

frontiers | Research Topics