MATHEMATICAL MODELLING OF THE PANDEMIC OF 2019 NOVEL CORONAVIRUS (COVID-19): PATTERNS, DYNAMICS, PREDICTION, AND CONTROL

EDITED BY: Hui-Jia Li, Lin Wang, Zhen Wang, Zhanwei Du, Bingyi Yang, Chengyi Xia, Aristides (Aris) Moustakas and Sen Pei

PUBLISHED IN: Frontiers in Physics and Frontiers in Applied Mathematics and Statistics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-386-8
DOI 10.3389/978-2-88971-386-8

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 MODELLING OF THE PANDEMIC OF 2019 NOVEL CORONAVIRUS (COVID-19): PATTERNS, DYNAMICS, PREDICTION, AND CONTROL

Topic Editors:
Hui-Jia Li, Beijing University of Posts and Telecommunications (BUPT), China
Lin Wang, University of Cambridge, United Kingdom
Zhen Wang, Hong Kong Baptist University, SAR China
Zhanwei Du, University of Texas at Austin, United States
Bingyi Yang, University of Florida, United States
Chengyi Xia, Tianjin University of Technology, China
Aristides (Aris) Moustakas, University of Crete, Greece
Sen Pei, Columbia University, United States

Citation: Li, H.-J., Wang, L., Wang, Z., Du, Z., Yang, B., Xia, C., Moustakas, A., Pei, S., eds. (2021). Mathematical modelling of the pandemic of 2019 novel coronavirus (COVID-19): Patterns, Dynamics, Prediction, and Control. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-386-8

Table of Contents

frontiers in Physics

EDITORIAL
published: 05 August 2021
doi: 10.3389/fphy.2021.738602

Editorial: Mathematical Modelling of the Pandemic of 2019 Novel Coronavirus (COVID-19): Patterns, Dynamics, Prediction, and Control

Hui-Jia Li[1*†], *Lin Wang*[2†], *Zhen Wang*[3†], *Zhanwei Du*[4†], *Chengyi Xia*[5†], *Aristides Moustakas*[6†] *and Sen Pei*[7†]

[1]School of Science, Beijing University of Posts and Telecommunications, Beijing, China, [2]Department of Genetics, Faculty of Biology, University of Cambridge, Cambridge, United Kingdom, [3]Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China, [4]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, SAR China, [5]Tianjin University of Technology, Tianjin, China, [6]Natural History Museum of Crete, University of Crete, Heraklion, Greece, [7]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York City, NY, United States

Keywords: COVID-19, epidemic spreading, pattern recognition, mathematical modelling, transmission dynamics, disease prediction

Editorial on the Research Topic

Editorial: Mathematical Modelling of the Pandemic of 2019 Novel Coronavirus (COVID-19): Patterns, Dynamics, Prediction, and Control

Appeared as a public health event, the COVID-19 pandemic has triggered massive crises in public health systems and economy. The COVID-19 has fleetly spread to most countries of the world. By the end of May 2021, SARS-CoV-2 has infected 180 million people and caused over three million deaths worldwide. To control the outbreak of COVID-19, travel restrictions, economic lockdowns, and border controls have been taken by many countries. The ongoing COVID-19 pandemic motivates the scientific community to contribute to infectious disease modeling, epidemiological study and outbreak prediction. That is, understanding the mechanism of COVID-19 pandemic spreading, exploiting the infection prevention and mitigation, and evaluating policy implementation, are important research questions for academics and policymakers.

To exploit this urgent question, mathematical models have been widely used to analyze the characteristics, impacts and emergency responsefor COVID-19. Mathematical modeling is a kind of mathematical structure that is made up of various concepts and formulas in mathematics, used for generally or approximately expressing the characteristics or quantitative interdependence of a certain system of things.

This special issue has collected a series of studies on epidemic data analysis and supporting decision-making. Those studies fit mathematical models for COVID-19 to analyze the characters of the COVID-19 pandemic in different regions, for instance, Azanza Ricardo et al. investigate the outbreak of COVID-19 in Mexico with different scenarios. While Miao et al. use mathematical modeling to study options for business reopening during the COVID-19 pandemic. As a commonly used epidemiologic model, the Susceptible, Infectious, and/or Recovered (SIR) model, the reproductive number of the SIR model is discussed to assess COVID-19 spread (Silveira et al.; Espinosa et al.; Schlickeiser et al.). Also, a mathematical model is developed to study what role social heterogeneity plays in the formation of complex infection propagation patterns (Maltsev et al.).

Edited and reviewed by:
Haroldo V. Ribeiro,
State University of Maringá, Brazil

****Correspondence:***
Hui-Jia Li
hjli@bupt.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *09 July 2021*
Accepted: *15 July 2021*
Published: *05 August 2021*

Citation:
Li H-J, Wang L, Wang Z, Du Z, Xia C, Moustakas A and Pei S (2021) Editorial: Mathematical Modelling of the Pandemic of 2019 Novel Coronavirus (COVID-19): Patterns, Dynamics, Prediction, and Control. Front. Phys. 9:738602. doi: 10.3389/fphy.2021.738602

Moreover, statistical analysis has been used to characterize the COVID-19 pandemic and predict the spread trend (Wang et al.; Zhao et al.; Liu et al.). Mathematical models and statistical analysis play important roles in public health emergencies, using these models to analyze and develop strategies for the COVID-19 pandemic is also an essential mission. For example, fast epidemic recognition of the epidemic as soon as it appeared, optimal supply and allocation of medical emergency resources (Wang et al.), and assessing the epidemiological consequences of an emergency evacuation (Butail et al.).

Model estimation also plays an important role in this special issue. Wu et al. propose a generalized-growth model to present the evolution of the number of the total confirmed cases over time. The model provides a simple phenomenological approach, with potential implications for forecasting of the pandemic trend. Zhang et al. quantify the incubation period, transmission rate from close contact to infection, and the properties of multiple-generation transmission from a detailed database in mainland China. Tiwari et al. sought to provide a prediction of the epidemic peak and to evaluate the impact of lockdown on the epidemic peak shift in India. Peker-Dobie et al. argue that the peak of infected individuals coincides with the inflection point of removed individuals. Recent COVID-19 data and the records for Spanish flu and SARS epidemics confirm this observation. Using multistage models the authors provide an explanation for this time shift. Chae et al. estimate the parameters and the initial infections from fitted values, and quantify the infection rate, the basic reproduction number, and the initial number of infected individuals for a number of countries. Kröger and Schlickeiser argue that the Gauss model is the simplest analytically tractable model that allows us to quantitatively forecast the time evolution of infections and fatalities during a pandemic wave and provide relationships between peak time and width, the transient behavior of doubling times, and reproduction factors. Zhuang et al. built a model to estimate the total number of COVID-19 cases in Wuhan, based on the number of cases detected outside Wuhan city in China, with the assumption that cases exported from Wuhan were less likely underreported in other cities/regions. Total cases are determined by the maximum log likelihood estimation and Akaike Information Criterion (AIC) weight.

This special issue aims to understand the impact of COVID-19 spreading on public health, society, and economics, and provides efficient reference values for economic redevelopment and social stability. So far, the authors have fitted mathematical models for COVID-19, studied the connectedness of the global COVID-19 network across countries, captured the virus transmission among cities, analyzed the dynamic characteristics of the COVID-19 pandemic, assessed the impact of COVID-19 spreading, and predicted the spread trends of COVID-19, which alleviates the public panic caused by COVID-19, grasps the trend of COVID-19 transmission, and assists the government or policymakers to make efficient decisions according to the spreading situation of the COVID-19 pandemic.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work is supported by Fundamental Research Funds for the Central Universities of China (No. 2020XD-A01-2), NSFC (No. 71871233).

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

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

frontiers in Physics

ORIGINAL RESEARCH
published: 03 July 2020
doi: 10.3389/fphy.2020.00276

Gaussian Doubling Times and Reproduction Factors of the COVID-19 Pandemic Disease

Martin Kröger[1]* *and Reinhard Schlickeiser*[2,3]*

[1] Polymer Physics, Department of Materials, ETH Zurich, Zurich, Switzerland, [2] Institut für Theoretische Physik, Lehrstuhl IV: Weltraum- und Astrophysik, Ruhr-Universität Bochum, Bochum, Germany, [3] Institut für Theoretische Physik und Astrophysik, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

The Gauss model for the time evolution of the first corona pandemic wave is rendered useful in the estimation of peak times, amount of required equipment, and the forecasting of fade out times. At the same time, it is probably the simplest analytically tractable model that allows us to quantitatively forecast the time evolution of infections and fatalities during a pandemic wave. In light of the various descriptors, such as doubling times and reproduction factors, currently in use to judge the lockdowns and other measures that aim to prevent spreading of the virus, we hereby provide both exact and simple approximate relationships between the two relevant parameters of the Gauss model (peak time and width), the transient behavior of two versions of doubling times, and three variants of reproduction factors, including basic reproduction numbers.

Keywords: coronavirus, statistical analysis, extrapolation, parameter estimation, pandemic spreading

OPEN ACCESS

Edited by:
Aristides Moustakas,
University of Crete, Greece

Reviewed by:
Giorgos P. Tsironis,
University of Crete, Greece
Giovanni Millo,
Generali Investments, Italy

***Correspondence:**
Martin Kröger
mk@mat.ethz.ch
Reinhard Schlickeiser
rsch@tp4.ruhr-uni-bochum.de

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *16 May 2020*
Accepted: *22 June 2020*
Published: *03 July 2020*

Citation:
Kröger M and Schlickeiser R (2020) Gaussian Doubling Times and Reproduction Factors of the COVID-19 Pandemic Disease. Front. Phys. 8:276. doi: 10.3389/fphy.2020.00276

1. INTRODUCTION

Recently [1], we demonstrated that the proposed [2–4] Gaussian time evolution for the daily number of cases (deaths or alternatively infections) at time t

$$c(t) = c_{\max}\, e^{-\left(\frac{t-t_{\max}}{w}\right)^2} \tag{1}$$

provides a quantitatively correct description for the monitored rates in 25 different countries. Here, $c_{\max}$ is the maximum number of daily cases at peak time $t_{\max}$ and w a characteristic duration. The Gauss model (GM) is capable of reproducing reasonably well the monitored time evolution of the Covid-19 disease and, even more importantly, making realistic predictions for the future evolution of the first wave in different countries. An epidemologic foundation of the GM had been suggested by the agent-based model presented in Schüttler et al. [1]. The GM can furthermore be regarded as an approximant of the classical Susceptible-Infected-Recovered/Removed (SIR) model [5, 6] in the limit of large inverse basic reproduction numbers [7]. From a mathematical viewpoint, a sigmoidal time evolution involving a polynomial in the exponent of an exponential can be Taylor expanded about the maximum; the GM corresponds to the lowest non-trivial order of a Taylor expansion.

Values for the parameters of the GM had been extracted by fitting the natural logarithm of the monitored rates with

$$\ln c(t) \;=\; \ln c_{\max} - \left(\frac{t-t_{\max}}{w}\right)^2 = \ln c_{\max} - \frac{t_{\max}^2}{w^2} + \frac{2t_{\max}t}{w^2} - \frac{t^2}{w^2}, \tag{2}$$

which is a polynomial of second order in t, to derive the best fit values and their confidence errors of the three parameters $c_{\max}$, w, and $t_{\max}$. These parameters are country specific and reflect the regional differences in treatment, geographical, political, socioeconomic situations, available equipment, etc. If this fitting and parameter determination is done during the early stage of the pandemic wave, the GM makes predictions for the later time evolution of the wave.

The starting time of the outbreak, t_0 can be defined by the first occurrence of a case, $c(t_0) = 1$, and is thus known from the parameters of the Gaussian. Inverting $c(t_0) = 1$ readily yields $\ln(c_{\max}) = (t_0 - t_{\max})^2/w^2$, or

$$t_0 = t_{\max} - w\sqrt{\ln c_{\max}} \tag{3}$$

To simplify notation, besides absolute time t, we introduce two more times. First, there is the time relative to the peak time, denoted by

$$\Delta = t - t_{\max} \tag{4}$$

so that negative (positive) Δ correspond to times before (after) the peak time. Second, there is the dimensionless time $x = -\Delta/w$. As time unit we choose days throughout, so that $\Delta = +2$ corresponds to 2 days after peak time, and w is also given in units of days; meanwhile, c is a dimensionless number of cases, usually renamed as d or i if we specialize to deaths or infections. The three parameters of the GM are related but not identical for deaths and infections, as discussed earlier [1]. The related—to Equation (3)—starting time

$$\Delta_0 = t_0 - t_{\max} = -wx_0 = -w\sqrt{\ln c_{\max}} \tag{5}$$

is negative, $x_0 = \sqrt{\ln c_{\max}}$ is positive, and $|\Delta_0|$ is the number of days between outbreak and climax of the first pandemic wave. All properties derived for the GM must therefore depend on w, $c_{\max}$, and x or, alternatively, Δ, where $\Delta \in [\Delta_0, \infty]$.

Often monitored data are reported in terms of doubling times and effective reproduction factors. These are also important indicators for the future temporal evolution of the disease, especially if no functional form for the case temporal evolution, such as the GM (2), is adopted. However, there are differently defined doubling times as well as reproduction factors in use. It is the purpose of this manuscript to discuss in detail the properties of differently defined doubling times and the differently defined reproduction factors, their mutual relations to each other, and their temporal behavior for the GM.

2. DAILY INSTANTANEOUS DOUBLING TIME

As before [2], we consider the relative change of the daily number of cases for the GM

$$p(t) = \frac{c'(t)}{c(t)} = [\ln c(t)]' = -\frac{2\Delta}{w^2} = \frac{2x}{w}, \tag{6}$$

where the prime denotes a derivative with respect to time t, $c'(t) = dc(t)/dt$. The monitored data are often given in terms of the instantaneous doubling time d of the corresponding exponential functions at any time for the daily number of cases

$$c_a(t) \propto e^{\frac{t\ln 2}{d}} = 2^{t/d}, \tag{7}$$

with the obvious properties $c_a(t + d) = 2c_a(t)$. With these corresponding exponential functions, we obtain for the relative changes in the daily rate

$$p(t) = [\ln c_a(t)]' = \frac{\ln 2}{d}. \tag{8}$$

Equating the two results, (6) and (8), leads to the time-dependent differential Gaussian doubling time

$$d(t) = -\frac{A}{\Delta}, \qquad A = w^2 \ln\sqrt{2} = 0.35w^2. \tag{9}$$

Apart from the changed notation, these differential Gaussian doubling times agree with the earlier derived Equation (5) in Schlickeiser and Schlickeiser [2]. The differential doubling time is positive for times earlier than the peak time, $\Delta < 0$, and monotonically increases over time until it diverges as it approaches $\Delta = 0$. For later times $\Delta > 0$ the doubling time is formally negatively valued but corresponds to positively valued half-life approaching 0 for $\Delta \to \infty$. Because of the divergence at $\Delta = 0$ and its negative value for $\Delta > 0$, daily doubling times are of limited use only before the peak time of the outburst; instead, in the public discussion, cumulative doubling times are often preferred, which we discuss in the next subsection. Apart from the time Δ relative to the peak time, the daily instantaneous doubling time (9) is determined by the width w of the Gaussian time evolution function (1). **Figure 1** displays the distribution of widths w determined by the best fit of the GM to the death rates in 67 countries, indicating that $w \in [12, 34]$ with a mean value of 18.96.

We emphasize that, at early times t of the time evolution, characterized by $t_0 \le t \ll t_{\max}$ Or, equivalently, $\Delta_0 \le \Delta \ll 0$, the Gaussian time evolution function (1) approaches an exponential distribution because the exponent $-(t - t_{\max})^2 \approx -t_{\max}^2 + 2t_{\max}t$ becomes linear in t and thus also linear in Δ. At such early times, the time relative to the peak time (4) is Δ_0, so that the differential Gaussian doubling time (8) approaches the constant

$$d_0 \simeq d(t_0) = -\frac{A}{\Delta_0} = \frac{0.35w^2}{t_{\max} - t_0}, \tag{10}$$

characterizing the initial, exponential growth.

3. CUMULATIVE DOUBLING TIME

Instead of defining doubling times with daily number of cases one may also define them with the cumulative case rate. From Equations (1) and (4), one has for the corresponding cumulative number of cases at time t

$$C(t) = \int_{-\infty}^{t} ds\, c(s) = \frac{C_{\text{tot}}}{2}\left[1 + \text{erf}\,(\Delta/w)\right], \tag{11}$$

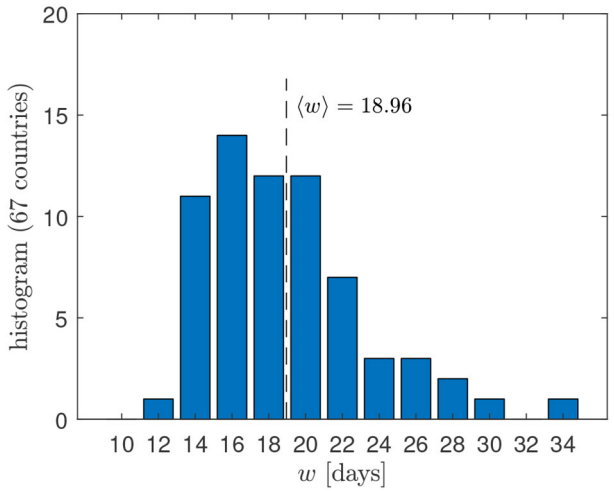


FIGURE 1 | Distribution of widths w obtained by the best fit of the GM to the death rates in 67 different countries [8].

respectively, in terms of the error function, where

$$C_{\rm tot} = \sqrt{\pi}\, c_{\rm max} w \tag{12}$$

denotes the total number of cases. Such values for fatalities, $D_{\rm tot}$, and infections, $I_{\rm tot}$, relevant for the first pandemic wave of the Sars-Cov-2 virus were obtained in Schüttler et al. [1]. With the cumulative numbers (11), we found for the respective relative change

$$P(t) \;=\; [\ln C(t)]' = \frac{C'(t)}{C(t)} = \frac{c(t)}{C(t)}. \tag{13}$$

Equating these results again with Equation (8) for the corresponding exponential function[1] leads to the time-dependent cumulative Gaussian doubling times

$$D(t) \;=\; \frac{C(t)\ln 2}{c(t)} = \chi w e^{(\Delta/w)^2}[1+\mathrm{erf}\,(\Delta/w)], \tag{14}$$

where $\chi = \sqrt{\pi}\ln(\sqrt{2}) \simeq 0.614$ abbreviates the numerical prefactor. Using the identities $1+\mathrm{erf}(x) = 1-\mathrm{erf}(-x) = \mathrm{erfc}(-x)$ in terms of the complimentary error function, we express Equation (14) as

$$D(t) = \chi w F\left(-\frac{\Delta}{w}\right) \tag{15}$$

with the function

$$F(x) = e^{x^2}\,\mathrm{erfc}\,(x) \tag{16}$$

As opposed to doubling times calculated from daily rates, doubling times derived from cumulative numbers are strictly positive, monotonically increase in the course of time, but never diverge, and remain finite at $\Delta = 0$. Because $x = -\Delta/w$, the argument x of $F(x)$ is positive before and negative after the peak time.

In **Appendix 1**, we investigate the properties of the function $F(x)$ and its approximations. It is convenient to consider times t before and after the peak time $t_{\rm max}$, i.e., negative and positive Δ. We consider each in turn.

3.1. Before Peak Time $\Delta < 0$

With the approximation (50) from the Appendix we obtain for the cumulative doubling time (15) at times $t \le t_{\rm max}$

$$D^{\rm before}(t) \simeq \frac{(\chi/3)w^2}{w+0.5|\Delta|}\left[1+\frac{2w^2}{(w+0.5|\Delta|)^2}\right] \tag{17}$$

where $\Delta = t - t_{\rm max}$ is negative, and $\chi/3 \simeq 0.205$. $D^{\rm before}(t)$ continuously increases with time until it reaches $D^{\rm max}$ (21) at peak time. At early times of the time evolution $t \ll t_{\rm max}$, not only d but also the cumulative Gaussian doubling time (15) or (17) approaches the constant

$$D_0 = D^{\rm before}(t_0) \tag{18}$$

reflecting again the result that, at an early time, the Gaussian time distribution function (1) approaches an exponential distribution function with the constant doubling time (10) so that also the cumulative distribution function initially displays an exponential behavior. The ratio of the two, differential (10) and cumulative (18), limits is given by

$$\frac{D_0}{d_0} = \frac{\sqrt{\pi}F(x_0)}{x_0}, \qquad x_0 = \frac{t_{\rm max}-t_0}{w} \tag{19}$$

with F from definition (16).

3.2. After Peak Time, $\Delta > 0$

Here, we use the property (46) and the approximation (50) to obtain the cumulative doubling time (15)

$$\begin{aligned} D^{\rm after}(t) \;&=\; \chi w\left[2e^{(\Delta/w)^2} - F\,(\Delta/w)\right] \\ &\simeq 2\chi w e^{(\Delta/w)^2} - D^{\rm before}(t) \end{aligned} \tag{20}$$

with $2\chi \simeq 1.229$ and where we can make use of $D^{\rm before}(t)$ from Equation (17) because it was written for this purpose in terms of $|\Delta|$. However, this Gaussian cumulative doubling time $D^{\rm after}(t)$ for times $t > t_{\rm max}$ is only a formal indicator for the decreasing slope of the cumulative rate $C(t)$. As the cumulative rate (11) indicates, at the peak times $t_{\rm max}$, it has the value $C(t_{\rm max}) = C_{\rm tot}/2$, so that, for any times greater than $t_{\rm max}$, the cumulative rates can no longer double. This implies that only the maximal cumulative doubling time

$$D^{\rm max} = D(t_{\rm max}) = \chi w \simeq 0.614 w \tag{21}$$

has a real physical meaning. In **Figure 2**, we calculate the Gaussian daily instantaneous and cumulative doubling times as

[1]We note that for a daily exponential rate in time also the cumulative number will be an exponential function in time.

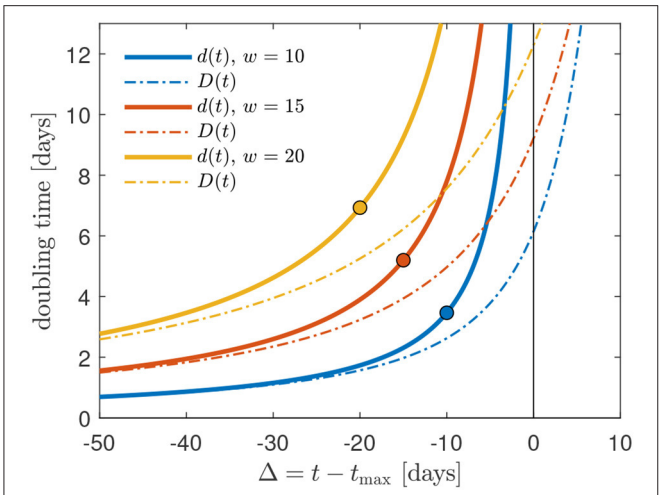


FIGURE 2 | Comparison of the daily instantaneous (d) and cumulative (D) doubling times as a function of the time Δ relative to the peak time for three values of the Gaussian width $w = 10, 15, 20$. The circles mark Δ_0 for $c_{\max} = 1$, cf. Equation (5): the GM should not be used at times smaller than Δ_0.

a function of the time Δ relative to the peak time for three values of the Gaussian width $w = 10, 15, 20$. At times below the peak time $\Delta < 0$, the two doubling times have a similar behavior. As $\Delta \to 0$ the instantaneous doubling times becomes infinitely large, whereas the cumulative doubling times approaches large but finite values. However, as noted before, for times $\Delta > 0$, the Gaussian cumulative doubling time $D^{\rm after}(t)$ is only a formal indicator for the decreasing slope of the cumulative rate $C(t)$ that can no longer double at any times larger than $t_{\max}$.

Nevertheless, cumulative doubling times are issued by health agencies, such as the Robert-Koch-Institut, to the public also at times after the peak time: they can cause much confusion among the public, as they suggest by their name that the cumulative case rate can still double beyond its 50% value, although this is no longer possible. Instead, one should refer to the effective reproduction factor at this stage of the wave time evolution, which we discuss next.

4. BASIC REPRODUCTION NUMBER R_0 AND EFFECTIVE REPRODUCTION FACTOR $R(T)$

In epidemiology, the basic reproduction number R_0 (sometimes called basic reproductive ratio, or incorrectly basic reproductive rate) of an infection can be thought of as the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection [9, 10]. The definition describes the state where no other individuals are infected or immunized (naturally or through vaccination). Some definitions, such as that of the Australian Department of Health, add absence of any deliberate intervention in disease transmission. Within this manuscript, R_0 is thus identical to the $R(t)$ at the starting time of the outbreak, i.e., $R(t_0) = R_0$.

The basic reproduction number R_0 is not to be confused with the effective, time-dependent reproduction factor $R(t)$; this is the number of cases generated in the current state of a population, which does not have to be the uninfected state. By definition, R_0 cannot be modified through vaccination campaigns. Also, it is important to note that R_0 is a dimensionless number and not a rate, which would have units of time like doubling time [9, 10]. Still, the basic reproduction number R_0 will be seen to correspond to $R(t)$ evaluated at time t_0.

The definition of the effective reproduction factor $R(t)$ according to [11, 12] is

$$c(t) = R(t) \sum_{s=-\infty}^{t} W(t-s)c(s) \tag{22}$$

where $c(t)$ is the number of daily cases (deaths or infections, usually the reproduction factor is obtained from the reported number of daily infections) at time t, and $W(s)$ denotes the serial interval distribution [13]. This distribution describes the probability for the time lag between a person's infection and the subsequent transmission of the virus to a second person, and it is known to have an effect on the reproduction factor [14]. The discrete sum in Equation (22) starts from zero rather than unity as in references [11, 12], because $W(0) = 0$ and because we are here interested in the continuous generalization of Equation (22).

Written in terms of integrals, Equation (22) corresponds to [15]

$$R(t) = \frac{c(t)}{\int_0^\infty W(s)c(t-s)\,ds} \tag{23}$$

while the serial interval distribution $W(s)$ has to be properly normalized to unity, i.e.,

$$\int_0^\infty ds W(s) = 1 \tag{24}$$

This normalization is required in order to ensure, according to Equation (23), that a constant stationary $c(t)$ implies $R(t) = 1$. Note that in Scirea et al. [12] they wrote $E[c(t)]$ instead of $c(t)$ on the left hand side of Equation (22), where $E[]$ signals that $c(t)$ is an expectation value involving the serial distribution. Here, we consider $c(t)$ to be defined by Equation (22) so that $c(t)$ can be evaluated for the GM via Equation (1).

In the following we investigate two different choices of the serial interval distribution evaluated for the GM: (i) the gamma distribution, as in previous studies [11, 12], and (ii) the analytically simpler box-shaped serial interval distribution. We consider each in turn.

4.1. Gamma Serial Interval Distribution

Here, the serial interval distribution $W(s)$ is taken to be the gamma distribution [16]

$$W(s) = \frac{\beta^\alpha s^{\alpha-1} e^{-\beta s}}{\Gamma(\alpha)} \tag{25}$$

with the shape parameters $\alpha = 2.785$ and $\beta = \alpha/6.5$ that seem to represent the distribution used in reference [11]. They used

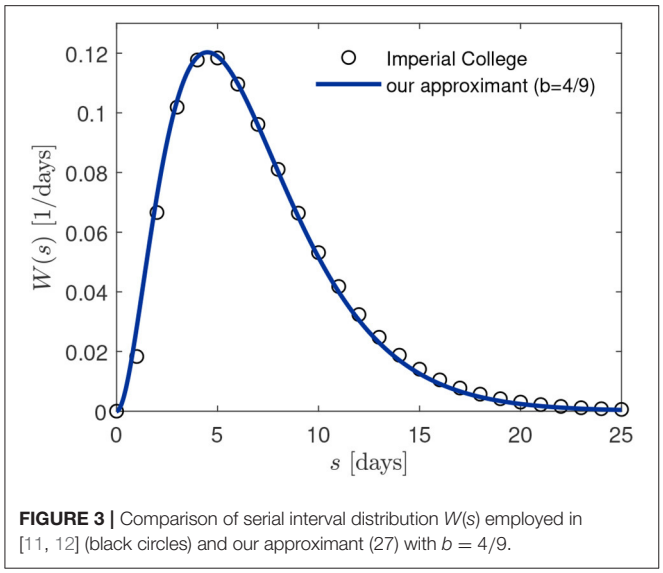


FIGURE 3 | Comparison of serial interval distribution $W(s)$ employed in [11, 12] (black circles) and our approximant (27) with $b = 4/9$.

another convention but mentioned the mean value $\langle s\rangle = 6.5$ and provided an excel file. The mean value of this distribution (25) is

$$\langle s\rangle = \int_1^\infty sW(s)\,ds = \frac{\alpha}{\beta} \tag{26}$$

For the specified parameters, the distribution (25) is very well-approximated (absolute error < 0.006) by the slightly more convenient and, again, properly normalized distribution

$$W(s) = \frac{b^3}{2}s^2e^{-bs}, \tag{27}$$

yielding a mean value $\langle s\rangle = 3/b$. We adopt $b = 4/9 = 0.444$, leading to the mean $\langle s\rangle = 27/4 = 6.75$ days, which is very close to the earlier chosen mean [11] $\langle s\rangle = 6.5$ days. In **Figure 3** we compare the approximation (27) with the discrete distribution used in reference [11] (black circles) during all their calculations.

4.1.1. Reproduction Factor *R(t)*

Here, we use the known $c(t) = c_{\max}\exp[-(\Delta/w)^2]$ for the GM (1). As for the doubling times, $R(t)$ does not depend on the magnitude $c_{\max}$ and absolute time t but can be expressed in terms of the relative time $\Delta = t - t_{\max}$ and w. With $c(t-s) \sim \exp[-(\Delta - s)^2/w^2]$, And, with $W(s)$ from (27), we can thus proceed and calculate $R(t)$ analytically as

$$R(t) = \frac{2}{b^3J(t)} \tag{28}$$

involving the time-dependent integral

$$J(t) = \int_0^\infty ds\,s^2e^{-q_1s-q_2s^2}, \tag{29}$$

where $q_2 = w^{-2}$ and $q_1(t) = b - 2\Delta/w^2$. To this end, it turns out convenient to switch to dimensionless times. We have already introduced x, and we now introduce a characteristic x_c

$$x_c = -\frac{bw}{2}, \tag{30}$$

and the dimensionless, time-dependent distance X between x and x_c via

$$X(t) = x - x_c = \frac{bw}{2} - \frac{\Delta}{w} = \frac{wq_1}{2}. \tag{31}$$

As shown in **Appendix 2**, the integral (29) can be evaluated analytically to yield

$$J(t) = \frac{\sqrt{\pi}w^3}{4}\frac{d}{dX}[XF(X)] = \frac{w^3}{4}\left[(1+2X^2)\sqrt{\pi}F(X) - 2X\right] \tag{32}$$

with the function F given by (16). Consequently, the effective reproduction factor (28) becomes (**Figure 5**)

$$\begin{aligned} R(t) &= \frac{-1}{x_c^3\sqrt{\pi}\frac{d}{dX}[XF(X)]} \\ &= \frac{1}{x_c^3\left\{2(x-x_c) - [1+2(x-x_c)^2]\sqrt{\pi}F(x-x_c)\right\}} \end{aligned} \tag{33}$$

in terms of x and the negatively valued x_c, where we recall that $x = (t_{\max} - t)/w = -\Delta/w$ carries the dependency on time t. The effective reproduction factor approaches the basic reproduction number R_0 at small times and assumes the important value $R(t) = 1$ roughly 4 days after peak time at $\Delta \approx 4$, as **Figure 5** indicates. It is not difficult to show that this Δ asymptotically approaches $2/b = 9/2 = 4.5$ days for large w (**Appendix 4.1**).

4.1.2. Base Reproduction Number R_0

While the basic reproduction number R_0 for the GM can be read off from (33) upon replacing x by $x_0 = \sqrt{\ln c_{\max}}$ (shown in **Figure 4**), it is insightful to make the connection between R_0 and the early doubling time $d_0 = (\ln\sqrt{2})w/x_0$, according to (10). As the Gaussian time distribution is exponential at early times, in the vicinity of $t \simeq t_0$, we can insert the exponential growth (7) into definition (23) with the gamma-shaped serial distribution $W(s)$. This yields a time-independent constant effective reproduction factor

$$\begin{aligned} R_0 = R(t_0) &\approx \frac{2}{b^3\int_0^\infty ds\,s^2\exp[-(b+\frac{\ln 2}{d_0})s]} \\ &= \left(1+\frac{\ln 2}{bd_0}\right)^3 = \left(1-\frac{x_0}{x_c}\right)^3 \end{aligned} \tag{34}$$

where $b = 4/9$ and where we have also mentioned its appearance in terms of dimensionless x_0 and x_c. Since d_0 is positive, the exponential effective reproduction factor at time t_0 (34) is greater than unity and provides an approximation for the exact one (**Figure 4**). It is worthwhile to mention that the same result is obtained without assuming a purely exponential growth but

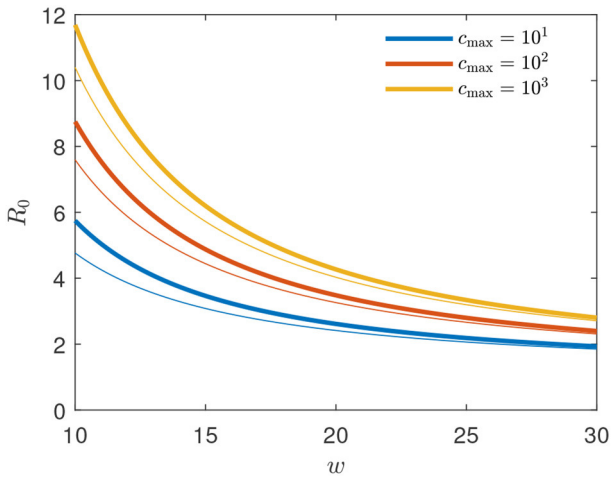


FIGURE 4 | Basic reproduction number $R_0 = R(t_0)$ for the GM as function of peak width w for several peak heights $c_{\max}$ (thick solid lines). For comparison, the thin lines show the approximant (34). The exact R_0 is given by (33) evaluated at $x = x_0 = \sqrt{\ln c_{\max}}$.

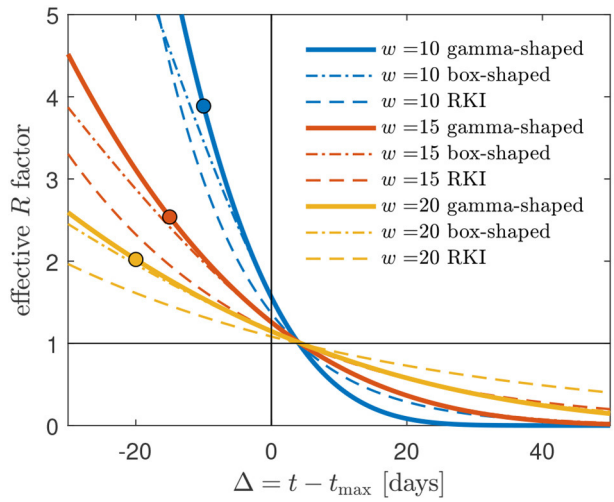


FIGURE 5 | Gamma-shaped $R(t)$ (33) compared with the approximate $R(t)$ (37) (red, dashed), using a box-shaped serial interval distribution $W(s)$, and the RKI formula (41). Cases shown are $w = 10$, $w = 15$, and $w = 20$. Deviations are most pronounced and significant for the smallest $w = 10$. The $R(t)$ curves terminate at $t = t_0$ corresponding to $\Delta = \Delta_0$, see (5). The circles mark Δ_0 for $c_{\max} = 1$.

instead by starting from (33) and assuming $x_0 \gg x_c + 1$ (for a proof see **Appendix 3.1**). For a model with purely exponential growth characterized by a single doubling time d_0, Equation (34) provides the exact relationship between doubling time and basic reproduction number and $R(t) = R_0$.

Adopting $b = 4/9$ and $w = 20$ and thus $x_c = -40/9$ according to (30), provides for the number (34)

$$R_0^{w=20} = \left(1 + 0.225\sqrt{\ln c_{\max}}\right)^3, \tag{35}$$

yielding the estimates 4.77, 4.03, and 3.26 for $c_{\max} = 10^4$, 10^3, and 10^2, respectively, close to the exact values given by $R(t_0)$ from Equation (33). The values of Δ_0 and R_0 for different values of the width w and $c_{\max} = 1$ are marked by circles in **Figure 5**.

4.2. Box-Shaped Interval Distribution

Here, we address the question on how relevant it is to take into account the correct shape of serial interval distribution when calculating $R(t)$ via (23).

If we consider $W(s)$ to be approximated by a constant independent on s on the interval $s \in [0, s_{\max}]$, and consider it to be zero otherwise, the requirement of its proper normalization (24) and unchanged mean value $\langle s \rangle = 6.5$ compared with the unapproximated $W(s)$ from (25) yields

$$W(s) = \frac{\Theta(s;\, 0, s_{\max})}{s_{\max}}, \qquad s_{\max} = 2\langle s \rangle = 13 \tag{36}$$

with the two-sided Heaviside $\Theta(x, A, B) = 1$ for $A \le x \le B$ and $\Theta(x) = 0$ otherwise.

4.2.1. Reproduction Factor $R(t)$

With the Gaussian evolution (1) and the box-shaped serial interval distribution (36) inserted, we obtain with the help of (23)

$$\begin{aligned} R(t) &= \frac{s_{\max}}{\int_0^{s_{\max}} \exp[(2\Delta - s)s/w^2]\, ds} \\ &= \frac{2s_{\max}/w}{\sqrt{\pi} e^{(\frac{\Delta}{w})^2} \left[\mathrm{erf}\,(\Delta/w) - \mathrm{erf}\,([\Delta - s_{\max}]/w)\right]} \\ &= \frac{26/w}{\sqrt{\pi}\left[e^{\frac{26(\Delta-6.5)}{w^2}} F\left(\frac{\Delta-13}{w}\right) - F\left(\frac{\Delta}{w}\right)\right]} \end{aligned} \tag{37}$$

plotted in **Figure 5**. As is visible, the box-shaped $W(s)$ can serve as a good approximation as long as w is sufficiently large, and Δ not too small. It crosses the $R = 1$ line roughly 4 days after peak time and shares this feature with the case of the gamma-shaped serial distribution. This aspect is worked out in detail in **Appendix 4**. Starting from $R(t) = 1$ with $R(t)$ from (37), the exact asymptotic value is $t = t_{\max} + (s_{\max}/3)$ days (proof in **Appendix 4.2**).

4.2.2. Base Reproduction Number R_0

The basic reproduction number is given by $R(t_0)$, which amounts to replacing Δ by Δ_0 in (37). As before, it is useful to consider a regime of exponential growth to come up with a simple approximant for R_0, now using a box-shaped $W(s)$. Inserting the exponential time evolution (7) with constant d_0 and the box-shaped serial interval distribution (36) into Equation (23), we obtain the time-independent constant effective box reproduction

factor that serves an approximant for R_0,

$$R_0 = R(t_0) \approx \frac{s_{\max} \ln 2/d_0}{1 - e^{-s_{\max} \ln 2/d_0}} \tag{38}$$

which is always greater than unity for positive d_0. Since $s_{\max} \ln 2 \approx 9$ days, we thus have

$$R_0 \approx 9/d_0 \tag{39}$$

as long as $d_0 < 9$ days, which is the usual scenario (**Figure 2**). As already mentioned, the box-shaped serial interval distribution is better not used to estimate R_0. It significantly underestimates the R_0 obtained with the gamma serial distribution.

4.3. Robert Koch Institute (RKI)

The RKI estimates an effective reproduction factor from the daily measured number $i(t)$ of people that have been recognized to be infected as follows

$$R(t) = \frac{\int_{t-4}^{t} ds\, i(s)}{\int_{t-8}^{t-4} ds\, i(s)} \tag{40}$$

Here, we again use the continuous version. Because measured data is not available for the future and is not sufficiently reliable if collected within the time frame of a few days, the RKI estimates $R(t)$ for a time t that lies one 8 days the past. A connection between (40) and the true effective reproduction number is based on the assumption that the true number of cases is proportional to the measured ones at any time.

4.3.1. Reproduction Factor *R(t)*

Using the GM instead of measured numbers for $i(t)$, the estimated true number of cases (deaths or infections) in Equation (40) yields

$$R(t) = \frac{\mathrm{erf}[\Delta/w] - \mathrm{erf}[(\Delta-4)/w]}{\mathrm{erf}[(\Delta-4)/w] - \mathrm{erf}[(\Delta-8)/w]} \tag{41}$$

shown in **Figure 5**. With (41) at hand, one can predict the RKI version of $R(t)$ at all times during the first wave of a pandemic. A time of interest is when R drops below unity. Equation (41) readily yields for $\Delta = 4$, with $\mathrm{erf}(0) = 0$,

$$R(t_{\max} + 4) = -\frac{\mathrm{erf}(4/w)}{\mathrm{erf}(-4/w)} = 1, \tag{42}$$

in agreement with **Figure 5**. It is this feature of the RKI, shared with the $R(t)$ for the gamma serial distribution, that may have given rise to the choice of the interval length of 4 days in its definition. **Figure 6** shows, for typical values between $w = 15$ and $w = 20$ days, how the $R(t)$ calculated via the box-shaped $W(s)$, and even more the RKI value, overestimate the $R(t)$ at times beyond peak time.

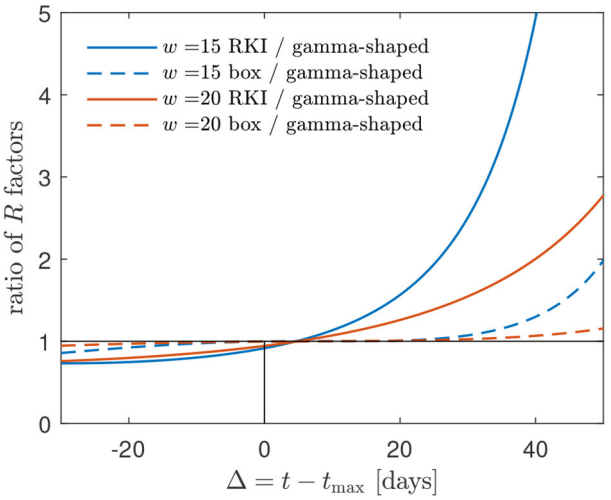


FIGURE 6 | The $R(t)$ factors obtained using (i) the box-shaped $W(s)$ and (ii) the RKI formula greatly overestimate the $R(t)$ using a gamma-shaped serial interval distribution at times beyond the peak time. Alternative representation of the data already shown in **Figure 5**. The mismatch increases with decreasing w. A typical w is in the range between 15 and 20 days for most countries [1] (cf. **Figure 1**).

4.3.2. Base Reproduction Number R_0

As in previous sections, we can read off the basic reproduction number R_0 upon inserting Δ_0 instead of Δ into the expression for $R(t)$ (41), and we can provide an approximate expression for R_0 upon considering purely exponential, initial growth. Following this route, inserting monoexponential $i(t) \propto 2^{t/d_0}$ into (40) yields

$$R_0 = R(t_0) \simeq \frac{2^{t/d_0} - 2^{(t-4)/d_0}}{2^{(t-4)/d_0} - 2^{(t-8)/d_0}} = 2^{4/d_0} \tag{43}$$

While the two approximants (34) and (43) for basic reproduction numbers look different at first glance, they are quantitatively very similar: for $d_0 = 1$, for example, (34) evaluates to 16.77, while (43) equals 16.0. Likewise, for $d_0 = 2$ (34) evaluates to 5.64, while (43) equals 4.0. In the limit of $d_0 \to \infty$, both versions yield $R_0 = 1$. The RKI version generally underestimates R_0, as given by (34), but by no more than about 35%.

5. SUMMARY AND CONCLUSIONS

The Gauss model for the time evolution of the first corona pandemic wave rendered useful in the estimation of peak times, amount of required equipment, and the forecasting of fade out times. At the same time, it is probably the simplest analytically tractable model that allows to quantitatively forecast the time evolution of infections and fatalities during a pandemic wave. For these descriptions and forecasts, various descriptors, such as doubling times and reproduction factors are currently used in order to judge lockdowns and other non-pharmaceutical measures that aim to prevent spreading of the virus. As different definitions of doubling times and reproduction factors and

numbers are used in the literature, we have provided in this manuscript both exact and simple approximate relationships between the two relevant parameters of the Gauss model (peak time $t_{\max}$ and width w) as well as the transient behavior of two versions of doubling times and three variants of reproduction factors $R(t)$, including basic reproduction numbers R_0.

Regarding doubling times, we considered both differential doubling times calculated from the daily number of cases and cumulative doubling times calculated from the cumulative case rates. The former differential doubling time is positive for times earlier than the peak time and monotonically increases in the course of time until it diverges as it approaches the peak time. For later times after the peak time, the differential doubling time is formally negatively valued but corresponds to positively valued half-life. Because of the divergence at the peak time and its negative value beyond, differential doubling times are of limited use only before the peak time of the outburst; instead, in the public discussion, cumulative doubling times are often preferred.

As opposed to doubling times calculated from daily rates, doubling times derived from cumulative numbers of cases are strictly positive, monotonically increase in the course of time, but never diverge, and remain finite at and after the peak time. At times below the peak time, the two doubling times have a similar behavior. However, the Gaussian cumulative doubling time for times after the peak time is only a formal indicator for the decreasing slope of the cumulative rate of cases. The Gaussian cumulative rate at the peak time attains exactly 50% of its maximum value after infinite time so that, for any times larger than the peak time, the cumulative rates can no longer double. This implies that only the maximal cumulative doubling time $0.614w$ has a real physical meaning.

Because of these two drawbacks of differential and cumulative doubling times in characterizing the time evolution of the corona wave after its peak time, health agencies, such as the German Robert-Koch-Institute (RKI) instead refer to the effective reproduction factor of the disease $R(t)$, which is the number of cases infected in the current state of a population by a single individual infected person. As long as this factor remains smaller than unity the number of infections per day decreases with time. The effective reproduction factor is calculated from an integral involving the serial interval distribution $W(s)$ normalized to unity and the differential case time distribution. For the GM, the latter is known analytically, and we investigated three different effective Gaussian reproduction factors: (i) the first is calculated with a gamma-function type serial interval distribution, (ii) the second is calculated with a flat box-shaped serial interval distribution, and (iii) the third, referred to as RKI estimate, involves the ratio of two consecutive 4-days time intervals of the monitored daily cases.

All three discussed effective reproduction factors, calculated with the GM, decrease from the base reproduction number R_0 at the beginning of the pandemic wave to very small values at times much larger than the peak time. They all cross the critical value $R = 1$ about 4 days after the peak times. As the approximated RKI estimate for Germany still, many weeks after the peak times of the infection and death rates, occasionally indicates effective reproduction factors greater than unity, this has to be due to short intraday fluctuations of the rates. Such factors greater unity at late times after the peak time contradict the much smaller (below unity) effective reproduction factors from the GM, as we have demonstrated by **Figure 6**. As the GM provides reasonable descriptions of the overall temporal evolution of the infection and death rates in Germany, we have to conclude that the present RKI estimate of the effective reproduction factor overestimates the influence of short intraday fluctuations in the reported cases.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00276/full#supplementary-material

REFERENCES

1. Schüttler J, Schlickeiser R, Schlickeiser F, Kröger M. Covid-19 predictions using a Gauss model, based on data from April 2. *Physics.* (2020) **2**:197–212. doi: 10.3390/physics2020013
2. Schlickeiser R, Schlickeiser F. A Gaussian model for the time development of the Sars-Cov-2 corona pandemic disease. Predictions for Germany made on March 30, 2020. *Physics.* (2020) **2**:164–70. doi: 10.3390/physics2020010
3. Ciufolini I, Paolozzi A. Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a Gauss error function and Monte Carlo simulations. *Eur Phys J Plus.* (2020) **135**:355. doi: 10.1140/epjp/s13360-020-00383-y
4. Lixiang L, Yang Z, Dang Z, Meng C, Huang J, Meng H, et al. Propagation analysis and prediction of the COVID-19. *Infect Dis Model.* (2020) **5**:282–92. doi: 10.1016/j.idm.2020.03.002
5. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc A.* (1927) **115**:700–21. doi: 10.1098/rspa.1927.0118
6. Kendall DG. Deterministic and stochastic epidemics in closed populations. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 4. Berkeley, CA: University of California Press (1956). p. 149–65.
7. Barmparis GD, Tsironis GP. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos Solit Fract.* (2020) **135**:109842. doi: 10.1016/j.chaos.2020.109842
8. Kröger M. *COVID-19 Real Time Statistics & Extrapolation Using the Gauss Model (GM).* (2020). Available online at: https://www.complexfluids.ethz.ch/corona (accessed May 16, 2020).
9. Milligan GN, Barrett ADT. *An Essential Guide*. Chichester: Wiley Blackwell (2015).

10. Fraser C, Donnelly CA, Cauchemez S. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science.* (2009) **324**:1557–61. doi: 10.1126/science.1176062
11. Flaxman S, Mishra S, Gandy A, Unwin H, Mellan TA, Coupland H, et al. Estimating the effects of nonpharmaceutical interventions on COVID-19 in Europe. *Nature.* (2020). doi: 10.1038/s41586-020-2405-7. [Epub ahead of print].
12. Scirea J, Sarah N, Vaughana TG, Brupbacher G, Fuchs S, Sommer J, et al. Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. *Swiss Med Wkly.* (2020) **150**:w20271. doi: 10.4414/smw.2020.20271
13. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis.* (2020) **93**:284–6. doi: 10.1016/j.ijid.2020.02.060
14. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B.* (2007) **274**:599–604. doi: 10.1098/rspb.2006.3754
15. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE.* (2007) **2**:e758. doi: 10.1371/journal.pone.0000758
16. Abramowitz M, Stegun IA. *Handbook of Mathematical Functions*. Washington, DC: NBS (1972).

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 July 2020
doi: 10.3389/fphy.2020.00289

Accounting for Underreporting in Mathematical Modeling of Transmission and Control of COVID-19 in Iran

Meead Saberi[1]*, *Homayoun Hamedmoghadam*[2], *Kaveh Madani*[3], *Helen M. Dolk*[4], *Andrei S. Morgan*[5,6], *Joan K. Morris*[7], *Kaveh Khoshnood*[8] *and Babak Khoshnood*[5]

[1] Research Centre for Integrated Transport Innovation, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, Australia, [2] Department of Civil Engineering, Monash University, Melbourne, VIC, Australia, [3] Council on Middle East Studies, The Macmillan Center for International and Area Studies, Yale University, New Haven, CT, United States, [4] School of Nursing, Faculty of Life and Health Sciences, University of Ulster, Newtownabbey, United Kingdom, [5] Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (U1153), INSERM, Paris, France, [6] Institute for Womens' Health, University College of London, London, United Kingdom, [7] Population Health Research Institute, St. George's, University of London, London, United Kingdom, [8] Yale School of Public Health, New Haven, CT, United States

Iran has been the country most affected by the outbreak of SARS-CoV-2 in the Middle East. With a relatively high case fatality ratio and limited testing capacity, the number of confirmed cases reported is suspected to suffer from significant underreporting. Therefore, understanding the transmission dynamics of COVID-19 and assessing the effectiveness of the interventions that have taken place in Iran while accounting for the uncertain level of underreporting is of critical importance. In this paper, we developed a compartmental transmission model to estimate the time-dependent effective reproduction number since the beginning of the outbreak in Iran. We associate the variations in the effective reproduction number with a timeline of interventions and national events. The estimation method accounts for the underreporting due to low case ascertainment. Our estimates of the effective reproduction number ranged from 0.66 to 1.73 between February and April 2020, with a median of 1.16. We estimate a reduction in the effective reproduction number during this period, from 1.73 (95% CI 1.60–1.87) on 1 March 2020 to 0.69 (95% CI 0.68–0.70) on 15 April 2020, due to various non-pharmaceutical interventions. The series of non-pharmaceutical interventions and the public compliance that took place in Iran are found to be effective in slowing down the speed of the spread of COVID-19. However, we argue that if the impact of underreporting is overlooked, the estimated transmission and control dynamics could mislead public health decisions, policy makers, and the general public.

Keywords: COVID-19, Iran, mathematical modeling, underreporting, effective reproduction number

OPEN ACCESS

Edited by:
Chengyi Xia,
Tianjin University of Technology, China

Reviewed by:
Ke Hu,
Xiangtan University, China
Zhishuang Wang,
Tianjin University of Technology, China

****Correspondence:***
Meead Saberi
meead.saberi@unsw.edu.au

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *13 May 2020*
Accepted: *26 June 2020*
Published: *10 July 2020*

Citation:
Saberi M, Hamedmoghadam H, Madani K, Dolk HM, Morgan AS, Morris JK, Khoshnood K and Khoshnood B (2020) Accounting for Underreporting in Mathematical Modeling of Transmission and Control of COVID-19 in Iran. Front. Phys. 8:289. doi: 10.3389/fphy.2020.00289

INTRODUCTION

The outbreak of SARS-CoV-2 in Iran was first officially announced in February 2020, 2 months after the initial outbreak in Wuhan, China [1]. Iran's patient zero is believed to have been a merchant from Qom who had traveled to China [2]. Despite the initial signs of a spread in Qom, the government declined to place the city under quarantine to contain the epidemic at an early

stage for various technical, socio-economic, religious, and security reasons [3]. The first local non-pharmaceutical interventions, such as school and university closures, were put in place a few days after the official acknowledgment of the first cases in Qom and Tehran [4]. Since then, various public health control measures at the local and national levels were taken that are believed to have altered the course of the outbreak. See **Figure 1** for a spatial illustration of the spread throughout the country by province in the first week since the official announcement of the first case (**Supplementary Material**).

The relatively high case fatality ratio (CFR), defined as the total number of deaths over the total number of infected cases, in Iran's official reports after the first week since the official declaration of the first case (16.8%) has raised questions on the true number of cases in the country [5, 6]. The testing protocol in Iran at the early stages of the outbreak was limited to hospital admissions of patients with severe symptoms. While Iran later extended the COVID-19 diagnostic testing capacity to patients with milder symptoms, it is believed that the under-ascertainment of cases remains high. While a number of studies have already explored the dynamics of the spread of COVID-19 in different countries with various methodologies [7–10] and its association with travel restrictions and other public health interventions [11, 12], undocumented infections and how they could affect the spread of SARS-CoV-2 in a population has been a less explored area in the rapidly growing field of COVID-19 literature [13].

This study aims to understand the transmission dynamics of COVID-19 in Iran and to assess the effectiveness of the control measures that were put in place over time through estimation of the effective reproduction number $R(t)$, defined as the average number of susceptible persons infected by an infected person during their infectious period at a given time in the course of the epidemic. We assessed $R(t)$ in relation to a timeline of national events and non-pharmaceutical interventions. In the absence of timely and reliable data, modeling can provide helpful answers, including the degree of plausible uncertainty in different estimates and the effectiveness of non-pharmaceutical interventions. By providing explicit and clear information about model assumptions and parameters, modeling can also foster scientific discussion of data gaps and what can be done to improve outbreak-related estimates by borrowing information available elsewhere. Finally, models can be developed and presented using both average estimates and measures of their uncertainty or, alternatively, as scenarios that can illustrate possible developments of the epidemic under various conditions.

METHODS

Data Sources

We use official time-series reports of the number of confirmed cases, recovered cases, and deaths from the World Health Organization (WHO) [1] and Iran's Ministry of Health and Medical Education [14]. The first confirmed case was reported on 19 February 2020, which is assumed as the beginning of the outbreak of COVID-19 in Iran.

Mathematical Model

We describe the dynamics of spread using a variation of the susceptible-exposed-infected-recovered (SEIR) model, distinguishing between fatal and recovered cases combined with an estimate of the percentage of symptomatic cases using delay-adjusted CFR (**Supplementary Material**) (see **Figure 2**). The model accounts for the time between exposure-to-onset of symptoms (or confirmation), also known as an incubation period, assuming a gamma distribution with an average of 5.5 days and a standard deviation of 2.3 days [15]. We also assume the time from symptoms onset-to-death and -to-recovery both follow a gamma distribution with an average of 22.3 days and a standard deviation of 9.4 days, and an average of 22.2 days and a standard deviation of 10 days, respectively [16]. The size of the initial susceptible population is assumed to be 80 million.

Time-Dependent Effective Reproduction Number

The basic reproduction number, R_0, a fundamental measure in infectious disease epidemiology and public health, is defined as the average number of susceptible persons infected by an infected

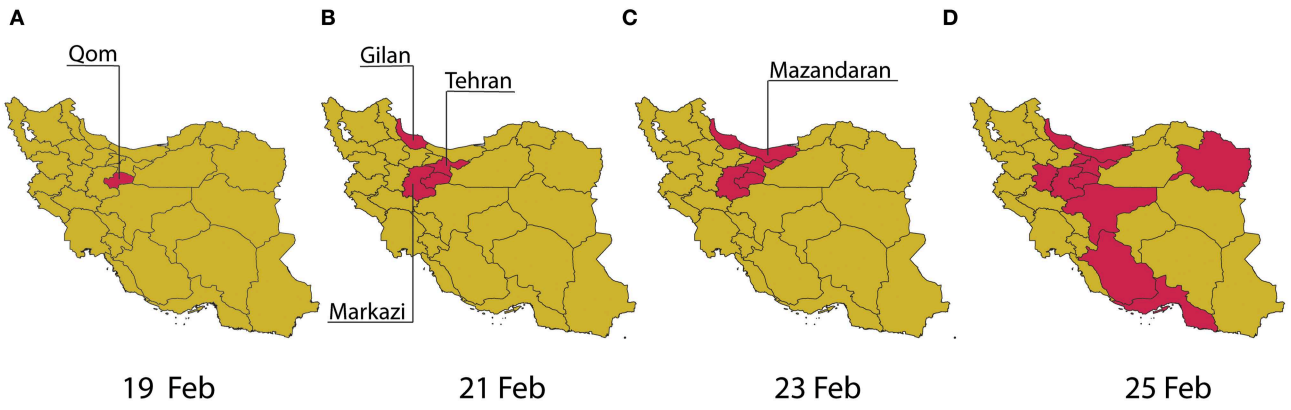


FIGURE 1 | Spatial illustration of the spread of COVID-19 in Iran by province in the first week of the official outbreak announcement from 19 February to 25 February 2020 **(A–D)**.

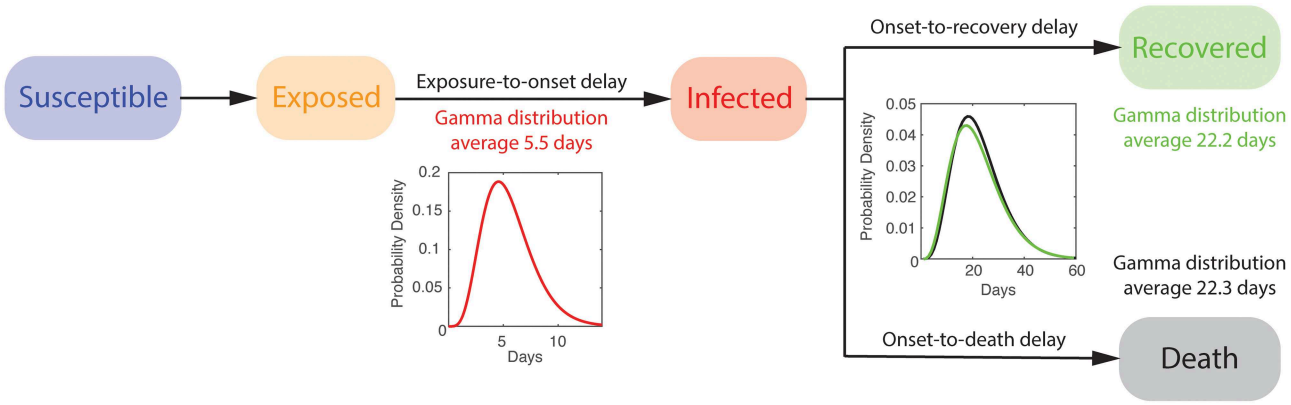


FIGURE 2 | Schematic illustration of the formulated mathematical epidemic model.

person during their infectious period in a fully susceptible population. R_0 is known to fluctuate if the social contact rate in the population changes over time or space. Since R_0 assumes an entirely susceptible population, several studies from the literature [17, 18] suggest using the time-dependent effective reproduction number $R(t)$ when examining the effect of various interventions including vaccination, social distancing, and quarantine. $R(t)$ is defined similarly to R_0 but is not limited to the assumption of a completely susceptible population. Here, we use empirical data from Iran to trace changes in $R(t)$ over a rolling 7-day period since the beginning of the COVID-19 outbreak and describe its association with various interventions (e.g., school closures, social distancing, and bans on public gatherings) that were enacted by the public and government.

Various methods exist to estimate R_0 [and $R(t)$] [19–21]. Here, we use the same framework described in **Figure 2** in which the parameters of the formulated SEIR model are inferred through an optimization problem. To estimate the parameters of the developed SEIR model, we formulate an ordinary least squares (OLS) minimization [22]. We use pattern search as a derivative-free global optimization algorithm [23] to find the model parameters that minimizes the sum of the normalized root mean squared error (RMSE) of the number of infected ε_i, recovered ε_r, and removed cases ε_f. $R(t)$ is then calculated using a rolling time window of 7 days to capture the evolving trend of the spread over time due to various changes in the social network contact rate. The calculated $R(t)$ may be overestimated during the early stage of an outbreak [24] due to different reasons, including the impact of imported cases and heterogeneity in subpopulations (e.g., older than 60 years old) with higher transmission rates.

Accounting for Underreporting

We account for the underreporting of the number of infected cases in the official confirmed data using delay-adjusted case fatality ratio (CFR) approach [25]. This approach assumes that the time from hospitalization-to-death has a known statistical distribution and uses this distribution to estimate when the people who died from COVID-19 would have been reported as being infected (**Supplementary Material**). The case fatality ratio is the ratio of the numbers of deaths over the numbers of reported infections calculated at the time of reporting, not the time of death. This is extremely important for rapidly evolving epidemics. The method, however, does not account for underreporting in fatality cases. The distribution of the time from confirmation-to-death is assumed to follow the same distribution as of the time from hospitalization-to-death, following a lognormal distribution with a mean of 13 days and a standard deviation of 12.7 days [26]. Here, we assume that the estimate for percentage of symptomatic COVID-19 cases reported for Iran follows a lognormal distribution (same distribution type as of the time from hospitalization-to-death) with a mean of 9.9% and a standard deviation of 4% based on the latest estimates in the literature [25]. This is based on the assumption of a baseline CFR of 1.4%. We assume the underreporting level remains constant over time.

Model Limitations

The developed model does not take into account the imported cases. The model also does not capture the impact of quarantine and self-isolation on confirmed cases. While the model produces stochastic outcomes due to various delay distributions used as inputs, the model parameters are deterministic. The model does not take into account the population age distribution. It is also assumed that underreporting remains unchanged over time.

RESULTS

We estimated that $R(t)$ varied between 1.73 and 0.69 during March to April 2020 in Iran, with a median of 1.16. We estimated a reduction in $R(t)$ from early-March to mid-April, from 1.35 (95% CI 1.31–1.39) on 8 March 2020 to 0.69 (95% 0.68–0.70) on 15 April 2020, due to control measures that took place. See **Figure 3A** for a timeline of interventions and events and **Figure 3B** for temporal variations of estimated $R(t)$.

The CFR on 25 February 2020, before adjusting for the time from diagnosis-to-death, was 16.84%. With more data emerging after the second week, the CFR dropped to 4.4% on the 14th day since the declared beginning of the outbreak on February

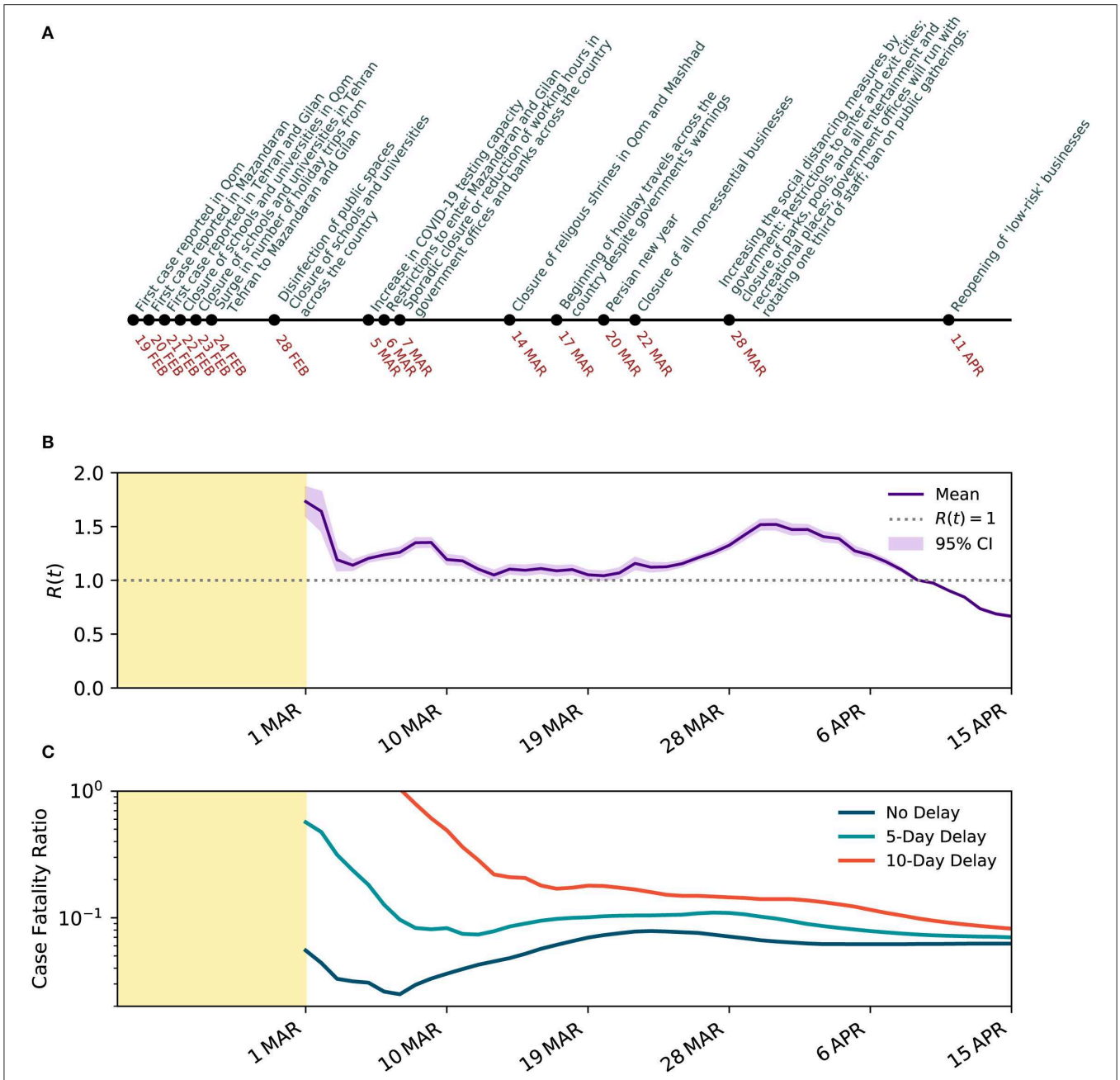


FIGURE 3 | Temporal evolution of SARS-CoV-2 transmission characteristics in Iran. **(A)** Timeline of events and interventions that took place since the first case was reported on 19 February 2020. **(B)** Estimated effective reproduction number $R(t)$ with 95% CI over a 7-day rolling window when underreporting is taken into account. **(C)** Estimated CFR with and without delay adjustment over the same time period in a semi-log scale.

19, 2020. Later, between 15 March 2020 and 12 April 2020, the CFR stabilized between 5.2 and 7.8% with a mean of 6.7% (see **Figure 3C**). With the wider spread of COVID-19 across the country, the CFR increased to 7.86% on 23 March 2020. However, the CFR declined and plateaued around 6.2% between 1 and 15 April 2020. The relatively high CFR could correspond to a significant level of underreporting of the infected cases and an overwhelmed health system. Given the wide distribution of the time from confirmation-to-death of COVID-19, we also explore the delay-adjusted CFR with 5- and 10-day delay periods, as examples. The dynamics of the delay-adjusted CFR with a 10-day delay suggests that the CFR has been gradually reducing in Iran from 17.97% on 16 March 2020 to 8.20% on 12 April 2020.

When underreporting of infected cases is overlooked, the estimated effective reproduction number began from 5.67 (95% CI 5.48–4.86) on 1 March 2020 and reduced to 0.70 (95% CI

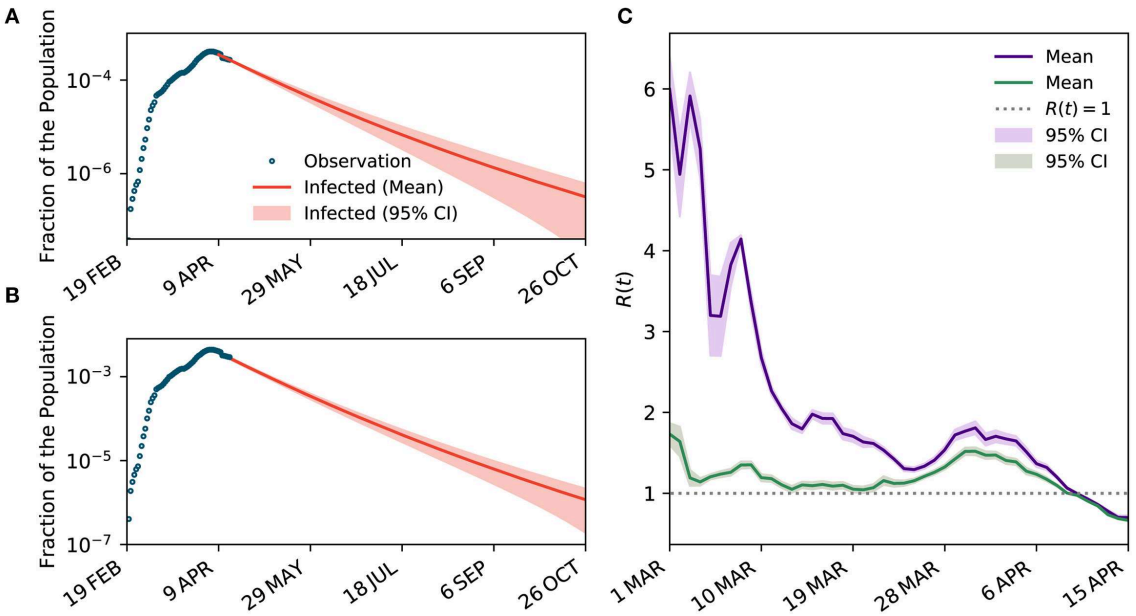


FIGURE 4 | The impact of underreporting on the estimated dynamics of transmission. **(A,B)** The projected curves for the number of infected cases when underreporting is not considered and when it is taken into account, respectively. **(C)** Temporal variations of $R(t)$ when underreporting is overlooked (purple curve) and when it is taken into account (green curve).

0.69–0.71) on 15 April 2020, suggesting the outbreak peak has already occurred on 8 April 2020 when $R(t)$ goes below 1, about 50 days from the confirmation of the first case. The outbreak is also likely to continue until the end of 2020 (see **Figure 4**). The estimates of the effective reproduction number were consistently larger during the early stage of the outbreak when underreporting is overlooked compared to when underreporting is taken into account. However, the estimated effective reproduction numbers converged as the number of infected cases approached the peak. The convergence of the estimates can be partly explained by the fact that the effective reproduction number is more dependent on the rate of change in the infected and recovered cases rather than their absolute numbers. Results also suggest that the impact of control measures on the effective reproduction number is significantly overestimated when underreporting is not taken into account.

With the gradual reduction of the effective reproduction number to below one and the increasing pressure on an already fragile economy because of the implemented control measures, the government is seeking an exit strategy and is considering easing some of the restrictions. Here, we conduct a scenario analysis to understand how three different scenarios could change the projected outlook of the outbreak in Iran: (i) maintaining the same level of control measures as of 12 April 2020; (ii) intensifying the measures to increase physical distancing, represented by a 20% reduction in the reproduction number; and (iii) partially lifting the restrictions to ease physical distancing, represented by a 20% increase in the reproduction number (see **Figure 5**). To estimate the number of ICU beds needed, we assume 5% of confirmed cases require intensive care [27]. We found that in all scenarios the projection of patients requiring ICU admission exceeds the original ICU capacity. Note that no official information is available on the expanded ICU capacity. As of 12 April 2020, both projected curves start above the ICU capacity. Easing the restrictions can quickly push the peak to a level that is five times higher than the scenario where the current level of control measures is maintained and puts additional pressure on the health system. Results clearly suggest that with further restrictions (scenario ii), the projected curve quickly goes back under the ICU capacity, while it takes more than 100 days for the curve to go back under the ICU capacity in scenario i and iii.

CONCLUSIONS AND DISCUSSION

In this study, we used a mathematical epidemic model to provide the first estimates of the changing transmission of SARS-CoV-2 infection in Iran when the underreporting of cases is considered. We used official data and adjusted our estimates for underreporting based on delay-adjusted CFR.

We used a variation of the SEIR model combined with an estimate of the percentage of symptomatic cases using delay-adjusted CFR based on the distribution of the time from confirmation-to-death [19]. In contrast to previous models of the epidemic [28], we did not assume any prior information on the distribution of the effective reproduction number. This, combined with the use of a series of distributions, allowed us to take into account more appropriately the uncertainty and the random variation in both input data and model outcomes.

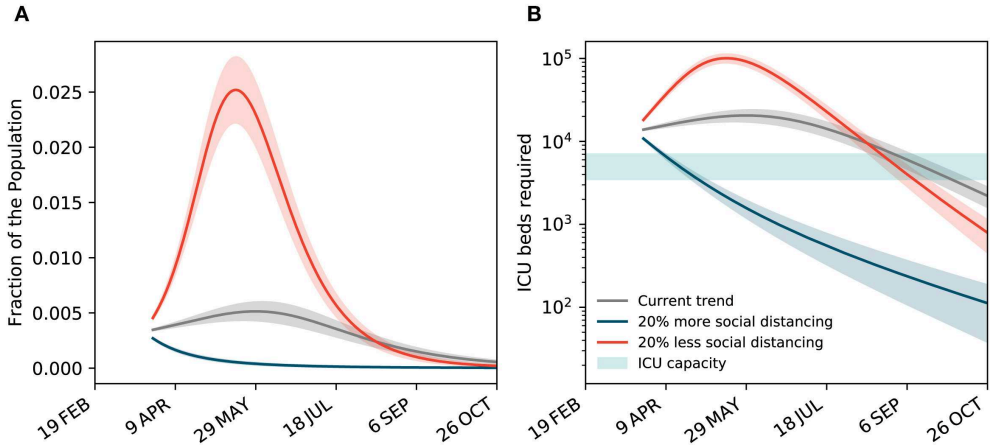


FIGURE 5 | Scenario analysis for maintaining, reinforcing, or easing the restrictions. **(A)** The projected curves for the total number of confirmed and suspected cases under three different scenarios: maintaining the same level of control measures as of 12 April 2020, reinforcing the control measures to increase physical distancing by a 20% increase in the reproduction number, and partial lifting the restrictions to ease physical distancing by a 20% decrease in the reproduction number. **(B)** The projected number of ICU beds needed under the three scenarios. The original ICU capacity is assumed between 3,500 to 7,000 beds [7].

Association of $R(t)$ and Public Health Interventions

The outbreak of COVID-19 in Iran is claimed to have started in Qom province, with the first case officially reported on 19 February 2020. As shown earlier in **Figure 1A**, according to the official data, it only took 2 days for four more provinces, Tehran, Markazi, Mazandaran, and Gilan (Guilan), to report their first cases. The first non-pharmaceutical intervention took place on 22 February 2020, with the closure of schools and universities in Qom, followed by their closure in the capital city of Tehran after a day. With the school closures in the capital combined with low levels of social awareness about the potential risks of COVID-19, a surge in the number of holiday trips from Tehran to northern provinces of Mazandaran and Gilan was observed. Soon after, with no restriction on inter-city travel, the number of identified cases in Gilan and Mazandaran grew rapidly, and the infection spread throughout the rest of the country.

The second major non-pharmaceutical intervention occurred on 28 February 2020 with a wide campaign to disinfect public spaces and for the closure of all schools and universities across all provinces. On 5 March 2020, the government announced an increase in the COVID-19 testing capacity. On 7 March 2020, sporadic closure or reduction of working hours in government offices and banks across the country was reported. On 14 March 2020, two often crowded religious shrines in Qom and Mashhad were closed to visitors. The increased physical distancing as the result of interventions and increase in public awareness of the crisis through major official and unofficial information campaigns, especially on social media, gradually showed its impact by slowing down the speed of the spread, as shown earlier in **Figure 1B**. The estimated effective reproduction number increased from 1.14 on 4 March 2020 to 1.35 on 8 March 2020, perhaps due to the increase in case ascertainment and delay in early interventions to show their impact. The effective reproduction number decreased consistently to 1.04 on 20 March 2020.

On 17 March 2020, only a few days before the Persian New Year (Nowruz, 20 March 2020), in the absence of strict travel restrictions, millions of Iranians began making road trips to various destinations across the country, despite the warnings from the government, and many hotels, restaurants, and the general hospitality industry's refrainment to provide services to any traveler. This is believed to have increased the speed of the spread of COVID-19 in Iran, increasing the effective reproduction number to 1.52 on 30 March 2020.

On 22 March 2020, the government announced more restrictive non-pharmaceutical interventions, leading to the closure of all non-essential businesses for at least 2 weeks, followed by further intensified interventions on March 28, 2020, including restrictions on entry and exit to affected provinces and cities, closure of parks, pools, and all recreational places, and a ban on public gatherings including sport, cultural, and religious events. These restrictions pushed the effective reproduction number further down to 1.0 on 9 April 2020 and 0.69 on 15 April 2020.

Case Fatality Ratio

Iran has had one of the highest CFRs among the affected countries in the world. While the CFR is known to vary significantly between countries due to various reasons, including testing frequency and population age distribution [26], the CFR is still higher in Iran despite having a younger population with a median age of 30.8 years old compared to China with 4% CFR and a median population age of 37.4 and South Korea with 2% CFR and a median population age of 41.8 [26]. Perhaps a comparable country in the Middle East with similar population characteristics (median age of 30.9) is Turkey. The CFR in Turkey

as of 6 April 2020 was 2%, three times lower than that of Iran [26]. While the evidence is indirect, it suggests that the official number of cases reported by Iran may have been significantly underreported, possibly due to relatively low case-ascertainment and underreporting of identified cases [26].

The continuous reduction in the delay-adjusted CFR shows a different trend compared to the CFR with no consideration of the time from confirmation-to-death. The continuous reduction could be explained by the improving case identification practice in Iran over time with various initiatives, including a National Coronavirus Helpline (a.k.a the "4030" service), established in late February 2020 to self-report symptoms and identify suspected cases.

Our results confirmed the significant impact of underreporting in describing the story of the COVID-19 outbreak in Iran, especially in the early stages. We showed how overlooking underreporting can drastically affect the estimation of $R(t)$ and overestimate the impact of control measures.

Final Remarks

Our results showed the reduction in effective reproduction number, a measure of infection transmission, during the study period. This decrease was most likely due to increased physical distancing as the result of multiple non-pharmaceutical public health interventions, including school closures, bans on public gatherings, travel restrictions, and full/partial closure of non-essential businesses, as well as major awareness campaigns over social media. Based on the latest trends, while the first peak of COVID-19 in Iran occurred on 5 April 2020, the post-peak period may continue to the end of 2020. However, these projections assume the continuation of the current level of control measures and the absence of effective therapeutic treatments or vaccination programs. Hence, they can be subject to important shifts depending, in particular, on the public's willingness to continue and the government's success in implementing social distancing measures or easing the restrictions.

Our model provides tangible evidence of the association between the different non-pharmaceutical interventions and their impact on the course of the outbreak. The results showed how the acceptance and, hence, effectiveness of the interventions endorsed by the Iranian government aimed at "flattening the curve" depended in part on the public's level of awareness of the principles behind governmental policies and their trust in the government and its control measures. For example, closure of schools and businesses in the capital city of Tehran near the time of the Persian New Year was followed by a surge in holiday trips to Gilan and Mazandaran and a subsequent increase in the number of cases in these provinces and in other parts of the country. This observation shows how interventions may be associated with unintended and at times counterproductive consequences. These negative consequences can be prevented when there is open and credible communication by competent officials and mutual trust between the public and government. Moreover, intervention measures need to be developed, implemented, and enforced as a whole, for example through strict reinforcement of travel restrictions in conjunction with school and workplace closings.

There is no substitute for high quality data—complete, accurate, and timely—as a basis for public policy. However, in the absence of such data, modeling of the type presented here can help provide reasonable estimates as well as realistic bounds of their uncertainty. The range of uncertainty can be viewed as the margin of error in the model's predictions of the number of cases, ICU admissions, or deaths. Modeling can also illustrate different scenarios – pessimistic vs. optimistic vs. realistic – of how the epidemic may evolve in relation to current and future public health measures and the possible compliance of the public over time.

In conclusion, using a compartmental model of the SARS-CoV-2 epidemic in Iran, we assessed the dynamic of the epidemic in relation to public health measures to increase social distancing. We took into account both the inherent uncertainty in the data and the possible impact of underreporting of true cases due to low case ascertainment and reporting. In the absence of consistently reliable data, the modeling approach as presented here can help to generate reasonable estimates of key public health metrics, such as the number of cases and case fatality ratio. In turn, these metrics and scenarios can help serve the dual purpose of informing public policy and the public, and of fostering discussions and improvements of epidemic modeling.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The data can be found on WHO's website available online at https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports and Iran's Ministry of Health and Medical Education's daily COVID-19 Epidemic Reports available online at http://corona.behdasht.gov.ir/.

AUTHOR CONTRIBUTIONS

MS, HH, KM, and BK designed the study. MS and HH developed the mathematical model and produced the figures. MS, HH, KM, HD, AM, JM, KK, and BK contributed to literature search, data analysis, data interpretation, and writing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at https://www.medrxiv.org/content/10.1101/2020.05.02.20087270v1 [29].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00289/full#supplementary-material

REFERENCES

1. WHO. Coronavirus disease 2019. *(COVID-19). Situation report 31*. Geneva: World Health Organization (2020).
2. World Economic Forum. *Finding 'Patient Zero': The Challenges of Tracing the Origins of Coronavirus*. (2020). Available online at: https://www.weforum.org/agenda/2020/03/coronavirus-covid-19-patient-zero/ (accessed April 13, 2020).
3. The Atlantic. Coronavirus Could Break Iranian Society. (2020). Available online at: https://www.theatlantic.com/ideas/archive/2020/02/iran-cannot-handle-coronavirus/607150/ (accessed April 13, 2020).
4. WSJ. *Iran Takes Emergency Measures After Two Coronavirus Deaths in Qom*. (2020). Available online at: https://www.wsj.com/articles/iran-takes-emergency-measures-after-two-coronavirus-deaths-in-qom-11582202084 (accessed April 13, 2020).
5. Khafaiea MA, Rahim F. Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health Res Perspect*. (2020) **11**:74–80 doi: 10.24171/j.phrp.2020.11.2.03
6. NBC News. *Why is Iran's Reported Mortality Rate for Coronavirus Higher Than in Other Countries?* (2020). Available online at: https://www.nbcnews.com/health/health-news/why-iran-s-reported-mortality-rate-coronavirus-higher-other-countries-n1142646 (accessed April 13, 2020).
7. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science.* (2020) **368**:742–6. doi: 10.1126/science.abb4557
8. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science.* (2020) **368**:493–7. doi: 10.1126/science.abb4218
9. Vespignani A, Tian H, Dye C, Lloyd- Smith JO, Eggo RM, Scarpino SV, et al. Modelling COVID-19. *Nat Rev Phys.* (2020) **2**:279–81. doi: 10.1038/s42254-020-0178-4
10. Xia C, Wang Z, Zheng C, Guo Q, Shi Y, Dehmer M, et al. A new coupled disease-awareness spreading model with mass media on multiplex networks. *Information Sci.* (2019) **471**:185–200. doi: 10.1016/j.ins.2018.08.050
11. Chinazzi M, David JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science.* (2020) **368**:395–400. doi: 10.1126/science.aba9757
12. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis.* (2020) **12**:165–74. doi: 10.21037/jtd.2020.02.64
13. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.* (2020) **368**:489–93. doi: 10.1126/science.abb3221
14. Iran's Ministry of Health and Medical Education. *Daily COVID-19 Epidemic Reports*. Available online at: http://corona.behdasht.gov.ir/ (accessed April 13, 2020).
15. 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
16. Imperial College London COVID-19 Response Team. *Report 4: Severity of 2019-Novel Coronavirus (nCoV)*. (2020). Available online at: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-severity-10-02-2020.pdf (accessed April 13, 2020).
17. Delamater PL, Street EJ, Leslie TF, Yang Y, Jacobsen KH. Complexity of the basic reproduction number (R0). *Emerg Infect Dis.* (2019) **25**:1–4. doi: 10.3201/eid2501.171901
18. Mercer GN, Glass K, Becker NG. Effective reproduction numbers are commonly overestimated early in a disease outbreak. *Stat Med.* (2011) **30**:984–94. doi: 10.1002/sim.4174
19. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In: Chowell G, Hyman JM, Bettencourt LMA, Castillo-Chavez C, editors. *Mathematical and Statistical Estimation Approaches in Epidemiology*. Dordrecht: Springer Netherlands. (2009). p. 103–21. doi: 10.1007/978-90-481-2313-1_5
20. Dietz K. The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res.* (1993) **2**:23–41. doi: 10.1177/096228029300200103
21. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung SM, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med.* (2020) **9**:538. doi: 10.3390/jcm9020538
22. Marinov RR, Mariova RS, Omojola J, Jackson M. Inverse problem for coefficient identification in SIR epidemic models. *Comput Math Appl.* (2014) **67**:2218–27. doi: 10.1016/j.camwa.2014.02.002
23. Audet C, Dennis JE. Analysis of generalized pattern searches. *SIAM J Optim.* (2003) **13**:889–903. doi: 10.1137/S1052623400378742
24. Artalejo JR, Lopez-Herrero MJ. On the exact measure of disease spread in stochastic epidemic models. *Bull Math Biol.* (2013) **75**:1031–50. doi: 10.1007/s11538-013-9836-3
25. CMMID. *Using a Delay-Adjusted Case Fatality Ratio to Estimate Under-Reporting.* (2020). Centre for Mathematical Modelling of Infectious Disease, London School of Hygiene & Tropical Medicine. Available online at: https://cmmid.github.io/topics/covid19/global_cfr_estimates.html (accessed April 28, 2020).
26. CEBM. *Global Covid-19 Case Fatality Rates.* (2020). The Centre for Evidence-Based Medicine, University of Oxford. Available online at: https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/ (accessed April 13, 2020).
27. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease (2019). (COVID-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *JAMA*. (2020) **323**:1239–42. doi: 10.1001/jama.2020.2648
28. CMMID. *Temporal Variation in Transmission During the COVID-19 Outbreak.* (2020). Centre for Mathematical Modelling of Infectious Disease, London School of Hygiene & Tropical Medicine. Available online at: https://epiforecasts.io/covid/ (accessed April 13, 2020).
29. Saberi M, Hamedmoghadam H, Madani K, Dolk HD, Morgan A, Morris JK, et al. Accounting for underreporting in mathematical modelling of transmission and control of COVID-19 in Iran. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.05.02.20087270

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.

frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 11 August 2020
doi: 10.3389/fams.2020.00035

Mathematical Modeling of Business Reopening When Facing SARS-CoV-2 Pandemic: Protection, Cost, and Risk

Hongyu Miao[1], Qianmiao Gao[1], Han Feng[1], Chengxue Zhong[1], Pengwei Zhu[1], Liang Wu[1], Michael D. Swartz[1], Xi Luo[1], Stacia M. DeSantis[1], Dejian Lai[1], Cici Bauer[1], Adriana Pérez[1], Libin Rong[2] and David Lairson[3]*

[1] *Department of Biostatistics and Data Science School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, United States,* [2] *Department of Mathematics University of Florida Gainesville, FL, United States,* [3] *Department of Management, Policy and Community Health School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, United States*

The sudden onset of the coronavirus (SARS-CoV-2) pandemic has resulted in tremendous loss of human life and economy in more than 210 countries and territories around the world. While self-protections such as wearing masks, sheltering in place, and quarantine policies and strategies are necessary for containing virus transmission, tens of millions of people in the U.S. have lost their jobs due to the shutdown of businesses. Therefore, how to reopen the economy safely while the virus is still circulating in population has become a problem of significant concern and importance to elected leaders and business executives. In this study, mathematical modeling is employed to quantify the profit generation and the infection risk simultaneously from a business entity's perspective. Specifically, an ordinary differential equation model was developed to characterize disease transmission and infection risk. An algebraic equation is proposed to determine the net profit that a business entity can generate after reopening and take into account the costs associated of several protection/quarantine guidelines. All model parameters were calibrated based on various data and information sources. Sensitivity analyses and case studies were performed to illustrate the use of the model in practice. The results show that with a combination of necessary infection protection measures implemented, a business entity may stand a good opportunity to generate a positive net profit while successfully controlling the within-business infection prevalence under that in the general population. The use of personal protective equipment (PPE) is also found of significant importance, especially at the early stage of business reopening.

Keywords: SARS-CoV-2 pandemics, mathematical modeling, reopening business, parameter calibration, benefit-cost-risk trade-off

OPEN ACCESS

Edited by:
Sen Pei,
Columbia University, United States

Reviewed by:
Xin Wang,
Beihang University, China
Steve Newbold,
University of Wyoming, United States
Andreas Eilersen,
University of Copenhagen, Denmark

****Correspondence:***
Hongyu Miao
hongyu.miao@uth.tmc.edu

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

Received: *31 May 2020*
Accepted: *13 July 2020*
Published: *11 August 2020*

Citation:
Miao H, Gao Q, Feng H, Zhong C, Zhu P, Wu L, Swartz MD, Luo X, DeSantis SM, Lai D, Bauer C, Pérez A, Rong L and Lairson D (2020) Mathematical Modeling of Business Reopening When Facing SARS-CoV-2 Pandemic: Protection, Cost, and Risk. Front. Appl. Math. Stat. 6:35. doi: 10.3389/fams.2020.00035

INTRODUCTION

Severe Acute respiratory Syndrome Coronavirus 2 (SARS-COV-2) is an ongoing global health threat to every country in the world and has caused significant loss of almost every business entity [1–3]. In the United States alone, as of May 14, 2020, more than 1.3 million confirmed infections and more than 84,000 deaths had been reported since the COVID-19 outbreak, which is associated with a 4.8% drop in gross domestic product (GDP) in the first quarter of 2020.

While the development of effective vaccination, treatment, and prevention strategies is underway, it is still unknown when such efforts will yield clinical and medical practices effective enough to allow a return to usual economic activity [4, 5]. Practicing a variety of stringent quarantine and shutdown policies is an effective way to protect our citizens' health and lives during a pandemic like SARS-CoV-2 [6]. However, there are costs that need to be considered. Besides the extraordinary physical and emotional stress and potentially significant medical expenses that COVID patients and their families must cope with, a large number of people have lost (or may lose) their jobs or businesses. This significant financial loss results in a pressure to reopen the economy prior to the availability of effective vaccination and treatment of COVID-19. Elected leaders and business executives as well as employees must address a critical question: how to reopen the economy in the midst of a SARS-CoV-2 pandemic safely?

In this study, we use mathematical modeling techniques to address particular challenges a business entity may face during reopening. In recognition that mathematical modeling results alone cannot stop pandemics, [7] the following behavioral and social guidelines are still strongly suggested for any business entities in planning to reopen:

1. Social distancing (including dinning in restaurants and manufacturing in factories), e.g., the number of workers (and clients or patrons) that a business environment can support while maintaining a 6ft distance between individuals;
2. Mask, glove, and goggle wearing while not alone;
3. Routine sanitization (e.g., entrance, exit, home, workplace, conference room, bathroom, public transportation, door knobs, shared electronic devices);
4. COVID-19 tests accessible to workers who have symptoms or recent exposure to confirmed infections;
5. Deployment of non-contact sensors (e.g., Kinsa smart thermometer) for real-time fever detection and location;
6. Case reporting and quarantine policy;
7. Determination of a maximum time duration of exposure to working environment;
8. Specific equipment (e.g., stronger ventilation system, UV purification system) for aerosol transmission prevention;
9. Employees in non-contact positions remain working from home;

The implementation of each guideline above can protect employees' health and lives but may add an economical cost. Indeed, the main purpose of reopening businesses is to prevent further deterioration of our economy by generating profits and provide incomes that many citizens desperately need. However, for employees in a contact-based position, the risk of COVID infection and transmission may remarkably increase if they return to work. When the infection rate of COVID-19 within a business entity becomes higher than the prevalence in the general population, it may trigger a domino effect and subsequent infections in a broader community originated from this business entity. Therefore, a delicate balance between profit/income generation and infection/transmission prevention must be the ideal. The focus of this modeling strategy is to provide a quantitative tool for investigating whether a business entity may reopen according to a simple definition of operational reopening feasibility: reopening is considered as feasible if a business entity can generate positive net profit after reopening while keeping the prevalence of COVID-19 infection within this entity less than or equal to the prevalence in the general population. The prevalence in the general population is assumed to remain under a certain threshold after reopening; otherwise, isolation and quarantine orders may be re-issued by government and reopening would halt, disregarding an individual business's performance.

MATERIALS AND METHODS

Mathematical Model

To our knowledge, there exist only a few studies that have modeled the reopening of a country, a state, a city or a local community [8–11], and fewer studies have focused on the reopening of a business entity. This modeling work attempts to address the business-reopening problem by considering a transmission model together with a net profit equation.

Borrowing from the classical susceptible-infected-recovered (SIR) modeling framework as in several previous models [8, 12], the following definitions and notations are introduced. Let N_T denote the total number of employees in a business entity, N_c denote the number of employees in a contact-based position, and N_N denote the number of employees in a non-contact position. Employees in a non-contact position are expected to remain working from home (WFH) according to Guideline #9 and thus are excluded from the transmission model. These WFH employees still actively contribute to profit generation and are counted in the net profit equation. In addition, the total number of employees N_T is assumed variable after the reopening due to various reasons, including infection-related death or other factors. Among the N_c employees in a contact-based position, let S denote the number of susceptible employees; C denote the number of asymptomatic, pre-symptomatic, and very-mildly-symptomatic (VMS) carriers (collectively known as silent spreaders); Q denote the number of infected employees on quarantine, under treatment (i.e., confirmed infection) or awaiting test result (i.e., unconfirmed infection); and D denote the number of deaths or resignations due to infection. Note that: (i) all infected employees confirmed by virus tests should be quarantined immediately, and any company fellow having a recent contact with the infected person is assumed to also be tested for the virus and initially be self-quarantined (see Guideline #4). The asymptomatic, pre-symptomatic, and very-mildly-symptomatic carriers are the three major categories of the so-called "silent spreaders." While it remains unclear how quickly asymptomatic carriers can transmit the infection, some studies in China suggested that ∼25% of those who tested positive without symptoms continued symptomless, and the remaining 75% turned out to be pre-symptomatic [13]. Other ongoing studies suggested that the proportion of asymptomatic infection could be as high as 50% [14]. People with very mild symptoms (e.g., occasional light cough or mild fever) may not

recognize the infection but are also fully capable of spreading the disease like the presymptomatic carriers [15]. Finally, the WFH employees may still be infected and then practice at-home quarantine or receive treatment in hospital; for simplicity, instead of introducing another set of equations for characterizing the WFH transmission, the number of infected WFH employees N_{NQ} is assumed to be collected by the business entity on a self-reported basis and directly accounted for in the net profit calculation equation.

After taking silent spreaders into consideration, [16] the proposed model structure is given below:

$$\frac{dS}{dt} = -\beta_O \cdot \kappa \cdot S - \beta_A \cdot \frac{\alpha C}{N} \cdot S - \beta_P \cdot \frac{(1-\alpha)\, C}{N} \cdot S \quad (1)$$

$$\frac{dC}{dt} = \beta_O \cdot \kappa \cdot S + \beta_A \cdot \frac{\alpha C}{N} \cdot S + \beta_P \cdot \frac{(1-\alpha)\, C}{N} \cdot S - \tau \cdot (1-\alpha)\, C + \omega \cdot Q \quad (2)$$

$$\frac{dQ}{dt} = \tau \cdot (1-\alpha)\, C - (\gamma + \omega) \cdot Q - \delta \cdot Q \quad (3)$$

$$\frac{dR}{dt} = \gamma \cdot Q \quad (4)$$

$$\frac{dD}{dt} = \delta \cdot Q \quad (5)$$

and the net profit is defined as the following:

$$P_{ntot} = \xi_1 \cdot \xi_7 \cdot \rho \cdot \big((N_N - N_{NQ}) + S + C + R\big) - \sum_{i \neq 1 or\ 7} \xi_i \cdot (S + C + R) - w \cdot Q - w \cdot N_{NQ} \quad (6)$$

where $= S + C + R$ denotes the number of employees in contact-based positions who are working on site. The definitions of all model variables and parameters are listed in **Table 1**, and a diagram (**Figure 1**) is also given to illustrate the transmission model mechanisms and assumptions. As suggested in **Figure 1** and Equation (1), susceptible employees can be infected and become virus carriers by contacting people outside of the business at a rate β_O, or interacting with asymptomatic carriers within the business at a rate β_A and with pre-symptomatic and VMS carriers within the business at a rate β_P. Also, α denotes the proportion of asymptomatic carriers among all tested-positive infections, and κ denote the probability of transmission via an average number of contacts per day of one person with other persons. In Equation (2), the infected susceptible become asymptomatic, pre-symptomatic or VMS carriers, and those pre-symptomatic and VMS carriers further progress to symptomatic infections at a rate τ. In Equation (3), symptomatic infections may recover at a rate γ and become immune to virus infection due to, e.g., memory immunity, [17] regress to symptomless carriers (e.g., some patients can test positive and shed viruses after symptoms end [18]) at a rate ω, or die at a rate δ. Also, Equations (4) and (5) characterize the dynamics of recovered and death, respectively. In the net profit Equation (6), ρ denotes the net profit per day generated by healthy WFH, susceptible, and silent carrier employees, ξ_i ($i \neq 1, 7$) is the average cost per person per day in a contact-based position associated with the i-th protection Guideline ($i = 1, \ldots, 9$), and w is the average wage per day per person for employees on quarantine or under treatment. It is assumed that employees on quarantine or under treatment still receive their wages given that the typical quarantine or treatment length is 14–28 days even considering them as sick days.

Data Source and Parameter Values

The datasets generated from recently published studies and surveys are heterogeneous in terms of population demographics, geospatial characteristics, medical resource availability, treatment regimens, prevention polices and implementation, among others. Extraordinary efforts are needed to integrate such heterogeneous datasets and perform a variety of statistical analyses, which is beyond the scope of this study. Instead, the summary statistics or previously-calibrated parameter values from previous studies and surveys are used as the primary information sources for our model parameter calibration. Note that studies reported different accuracy (in terms of decimals), it is thus difficult to standardize the number of decimals without losing accuracy so we keep the original numbers.

To calibrate the transmission model parameters in Equations (1)–(5), we started with the case death rate and the recovery time. In the United States, the case fatality rate is currently 5.65% while Omer et al. [29] previously estimated the infection-related death rate as low as 1.7%. In other countries, the reported case death rates may range from 0.56% (Iceland) to 13.53% (Italy); [29][1,2]. to preserve parameter uncertainty, this wide range of infection fatality rate (0.56∼13.53%) is adopted in our calculations. To calculate the mean recovery time, we consider the following observations:[3] (i) 8 out of 10 infected persons with symptoms have only mild illness; (ii) the average recovery time for mild cases is about 2 weeks; (iii) for severe cases, recovery may take up to 6 weeks; (iv) the overall recovery rate is between 97 and 99.75%. The conservative mean recovery time for all cases is thus $1/[(\frac{0.8}{14} + \frac{0.2}{42}) \times 0.97] = 16.65$ days among people with symptomatic infections, which is longer than the reported median hospitalization period of 12 days [33] among survivors. It was also reported that the recovery time for mild cases can be as short as 7 days, therefore, this study assumes that recovery time ranges from 7 to 42 days[4].

In Equation (1), for the proportion of asymptomatic carriers (denoted by α), multiple studies have been conducted to estimate this parameter among different populations with different methods [14, 19–22]. The range of the estimate of $\hat{\alpha}$ is (13.8%, 75%), and about half of such studies reported a result around 50%. Sun et al. (26) analyzed 391 cases in Zhejiang Province, China from Jan. 20th, 2020 to Feb. 10th, 2020, and found 54 (13.8%) cases were asymptomatic. Nishiura et al. [22] estimated

[1] Available online at: https://ourworldindata.org/coronavirus#what-do-we-know-about-the-risk-of-dying-from-covid-19

[2] Available online at: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html

[3] Available online at: https://www.webmd.com/lung/covid-recovery-overview#1

[4] Available online at: https://www.houstonmethodist.org/blog/articles/2020/apr/recovering-from-coronavirus-what-to-expect-during-and-after-your-recovery/

TABLE 1 | Variable and parameter definitions, values, and sources.

Notation	Definition	Unit	Value	Reference
S	Susceptible	Persons		
C	Carrier	Persons		
Q	Quarantine	Persons		
R	Recovered	Persons		
D	Death	Persons		
N	Sum of S, C and R	Persons		
	Number of employees in contact-based positions	Persons		
	Number of employees in non-contact positions	Persons		
	Number of employees in non-contact positions under quarantine	Persons		
α	Proportion of asymptomatic carriers among all types of carriers	%	50 (13.8, 75)	[14, 19–22][5]
	The product of β_O and κ	day^{-1}	3.10×10^{-7} (2.42×10^{-7}, 3.88×10^{-7})	[23]
κ	Probability of one employee having contacts with infected carriers outside a business entity	%	NA	
β_O	Infection rate associated with activities outside of the business entity	day^{-1}	NA	
β_A	Transmission rate of asymptomatic infection	day^{-1}	0.099 (0.0103, 0.814)	Calibrated from [13, 24, 25]
β_P	Transmission rate of presymptomatic and very-mildly-symptomatic infection	day^{-1}	0.197 (0.0333, 1.18)	Calibrated from[3,4][26]
τ	Rate of progression to symptomatic infection	$day^{-1}person^{-1}$	0.192 (0.143, 0.244)	Calibrated from [13, 24]
ω	Rate of regression to carriage (e.g., even after treatment)	$day^{-1}person^{-1}$	0.00172 (0.000575, 0.00455)	Calibrated from [27, 28]
γ	Clearance rate of symptomatic infection, including the portion of negative test outcomes among exposed employees	day^{-1}	0.0535 (0.0178, 0.141)	[6, 29]
δ	Death rate due to infection	day^{-1}	0.00320 (0.00013, 0.0167)	Calibrated from[4]
ρ	Net profit per capita	$ per person per day	400.73	[6]
w	Average wage of employees	$ per person per da	218.60	[8]
ξ_1	Social distancing	%	67.4	[12]
ξ_2	Personal protective equipment (PPE)	$ per person per day	3.60	[30, 31][10]
ξ_3	Routing sanitization	$ per environmental service staff per day	10.45	[32]
ξ_4	COVID test	$ per person per day	0	Assumed
ξ_5	Non-contact thermometer	$ per person per day	(0.005, 0.01)	[11,12]
ξ_6	Case reporting and quarantine	$ per person per day	0	
ξ_7	Proportion of current hours of exposure to working environment	%	70 (50, 100)	
ξ_8	Specific equipment for aerosol transmission (e.g., UVGI, HEPA filtration)	$ per person per day	0.874	[31]
ξ_9	WFH	$ per person per day	0	Assumed

ξ_i, Protection cost per person associated with the i-th Guideline; UVIG, ultraviolet germicidal irradiation; HEPA, high-efficiency particulate air filtration.

α in the evacuated Japanese citizens to be 30.8% (95% CI: 7.7–53.8%). Mizumoto et al. [14] obtained an estimate 17.9% (95% CI: 15.5–20.2%) using the data from the Diamond Princess cruise. Kimball et al. [21] analyzed the results of symptom assessment and SARS-CoV-2 testing in King County, Washington, and found that 56.5% of the tested positive was asymptomatic. Day [20] suggested that the proportion of asymptomatic infection was between 50 and 75% in northern Italy, and others[5] suggested that α could be between 25 and 50% on Apr. 5th, 2020 [34].

[5]Available online at: https://www.augustahealth.com/health-focused/covid-19-asymptomatic-carriers-and-antibody-tests

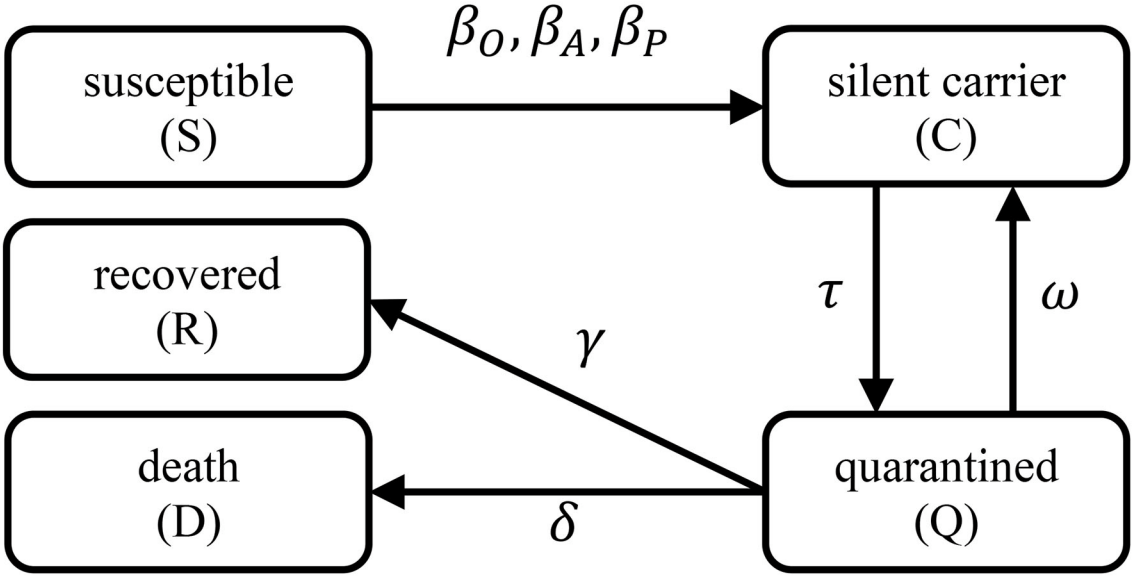


FIGURE 1 | Transmission model diagram. Each rectangle represents a model variable, a directed arrow shows the transition from one to another model variable, and the associated model parameters are labeled next to an arrow.

Furthermore, κ is the probability of one employee having contacts with infected carriers outside a business entity, and β_O denotes the associated transmission rate. A limited number of studies were found helpful for estimating these two parameters, we thus calibrated such parameter values primarily using the simulation results from Tang et al. [23]. In Tang's study, the daily average contact number was estimated to be 14.78 (SE 0.90) contacts per day per person, and the probability of successful transmission per contact was estimated to be 2.10×10^{-8} (SE 1.19×10^{-9}). According to the definitions of κ and β_O, we have $\beta_O\kappa = 14.78 \times 2.10 \times 10^{-8} = 3.10 \times 10^{-7}\ day^{-1}$, and the range of $\beta_O\ \kappa$, $(2.42 \times 10^{-7},\ 3.88 \times 10^{-7})$, can be calculated using the standard errors (i.e., $2.42 \times 10^{-7} = (14.78 - 2 \times 0.90) \times (2.10 \times 10^{-8} - 2 \times 1.19 \times 10^{-9})$, and $3.88 \times 10^{-7} = (14.78 + 2 \times 0.90) \times (2.10 \times 10^{-8} + 2 \times 1.19 \times 10^{-9})$.

The transmission rate (β_P) can be calculated from $\frac{R_0}{recovery\ time}$, where R_0 denotes the reproduction number (i.e., the average number of new infections by one carrier) that has been frequently investigated in many SARS-CoV-2 studies [13, 23, 35–43]. For instance, according to the study of Liu et al. [26], the estimated R_0 of SARS-CoV-2 in China at the early stage of the pandemic ranges from 1.4 to 6.49 with a mean 3.28 and a median 2.79. Note that the population in Liu et al. [26] was mostly not aware of the pandemic and used little protections such as mask wearing, hand washing, and social distancing. Thus, for the unprotected general population, the estimated transmission rate is 3.28/16.65= 0.197 day^{-1}. The corresponding range is between 0.0333 (=1.4/42) and 0.927 (=6.49/7) per day. Note that in the early stage of a pandemic, it can be assumed that most of the cases are pre-symptomatic or mildly-symptomatic such that the parameter value calculated above may be close to the true value of β_P. Another study by Li et al. (43) suggested that the transmission rate β_P could be 1.12 (95% CI: 1.07–1.18), 0.52 (95% CI: 0.42–0.72), or 0.35 (95% CI: 0.28–0.45) per day by fitting data of different pandemic stages in China. For these reasons, we assumed a range for β_P of (0.0333, 1.18) per day.

For the transmission rate of asymptomatic infection (β_A), very little useful information was found in the scientific literature. Here we borrowed the idea of Li 's work, [25] in which the transmission rate of asymptomatic infection was calculated by multiplying a reduction factor μ with the transmission rate of symptomatic infection. The reduction factor μ was estimated as 0.55 (95% CI: 0.46–0.62), 0.50 (95% CI: 0.37–0.69) and 0.43 (95% CI: 0.31, 0.61), corresponding to different stages of SARS-CoV-2 outbreak in China. Using the same reduction factor μ in our model, β_A can be estimated by $\mu\beta_p$, leading to a mean of $0.197 \times 0.5 = 0.099$ per day and a range from 0.0103 (= 0.333×0.31) to 0.814 (= 1.18×0.69) per day.

In Equation (2), τ is the rate of silent carriers progressing to symptomatic infections. The mean incubation period was previously reported to be 5.2 days with a 95% confidence interval (4.1, 7.0) days, [13, 24] or 5.1 days with a 95% confidence interval (2.2, 11.5) days [44]. Based on these results, we estimated τ as the reciprocal of the incubation period with a mean 1/5.2=0.192 per day and a 95% confidence interval (0.143, 0.244) per day.

In Equation (3), for the clearance rate of symptomatic infection γ and the death rate due to COVID infection δ, multiple studies developed various methods to estimate them [6, 23, 45]. In Piguillem's method, (6) the calculations were mainly based on the case mortality rate and the recovery time. After substituting 16.65 days for the mean recovery time and 5.65% for the mean death rate aforementioned, we obtain the estimate of γ as $(1 - 0.0565)^2 / 16.65$ =0.0535 per day, and the death rate δ as $(1 - 0.0565) \times \frac{0.0565}{16.65}$ =0.0032 per day. The range of γ is found

to be (0.0178, 0.141) per day, and the range of δ is (1.3×10^{-4}, 1.67×10^{-2}) per day.

In Equations (2) and (3), we also introduced ω, the rate of regression to carriage, considering the fact that 3.23% of the patients recovered from SARS-CoV-2 infection were tested positive after hospital discharge [27, 28]. However, the potential infectivity of these carriers remains unclear. Here ω is estimated to be 3.23% $\times\hat{\gamma}$ $=0.0323\times0.0535=$ 0.00173 per day, and its range is between 0.000575 ($=0.0178 \times$ 3.23%) and 0.00455 ($=0.141\times3.23\%$) per day.

Now consider the additional parameters in the net-profit Equation (6). The average net profit per capita ρ in the U.S. is found to be \$400.73 per person per day, which is calculated by dividing Jan 2020 U.S. corporate profit \$1908.02 Billion US dollars[6]. by 158,714,000 (the number of employed persons in the United States)[7] and then by 30 days. The average weekly wage of one employee in the U.S. is \$1,093 per person in the third quarter of Year 2019,[8] so the average daily wage of employee w is \$218.60 per day (dividing \$1,109 by the number of weekdays 5). In the previous study of Thumstrom and Newbold (2020), [12] GDP loss was considered as one cost of social distancing; i.e., an immediate GDP decline associated with practicing social distancing alone (i.e., house-hold quarantine) in the United States was predicted to be $13.7 - 6.49 = 7.21$ trillion US dollars. The projected GDP of Fiscal Year 2020 is 22.11 trillion US dollars according to the United States Congressional Budget Office (CBO)[9]. Therefore, for Guideline #1, the cost of social distancing is the loss of productivity by $7.21 \div 22.11 =$ 32.6%, and thus $\xi_1 = 1 - 32.6\% = 67.4\%$. For Guideline #2, the cost of personal protective equipment (PPE), [30] including surgical mask, gloves, goggle wearing, hand sanitization, and soap, is calculated as $\xi_2 =$\$3.60 per person per day [31] under the assumption that each person will consume two surgical masks per day, two pairs of gloves per day, and one goggles. According to the Personal Protective Equipment (PPE) guideline from Perdue University,[10] goggles can last for years if kept clean by using mild soap and water, and if stored in a protected, dry, and temperate storage location. So the use life of goggles is much longer than our typical setting for simulation time window length (i.e., 100 days) and the cost of one goggles per day is $\$5 \div 100 =$ \$0.05 per day for simplicity. The detailed costs of each PPE item as well as hand sanitization and soap can be found from online resources[11,12]. For Guideline #3, the cost of routine sanitization (e.g., cleaning the workplace, bathroom, and shared electronic devices) is calculated as $\xi_3 =$ \$10.45 per environmental service staff per day [32]. Specifically, in the work of White (2019), 11 hospitals with a total of 1,700 environmental service staffs, they found to spend \$11,308 per week to maintain the hygiene by consuming disinfection products. We thus calculate ξ_3 using the following equation:

$$\xi_3 = \frac{\text{cost of disinfection products}}{\left(\frac{\text{Total number of enviromental services staffs}}{\text{Number of hospitals}}\right)} \div 7 \textit{ days}.$$

For Guideline #4, the current policy dictates that the test of COVID is free to the business entity ($\xi_4 =$ \$0 *per person per day*). For Guideline #5, as the current market prices of a non-contact sensor may range from \$50 to \$100, and one non-contact sensor is required for each worksite. With the assumption of having 100 employees per worksite, the cost of deploying non-contact sensor ξ_5 is \$0.005 to \$0.01 per person per day for a 100-day time horizon. We assume that Guideline #6 does not cost any money ($\xi_6 =$ \$0 per person per day). For Guideline #7, the cost is proportional to the current maximum working hours (i.e., $\xi_7 = \frac{\textit{current maximum working hours}}{\textit{regular working hours}}$). In our model, the current maximum working hours is assumed to be 70% of the regular working hours. For Guideline #8, referring to the work of Chen (2013), [31] the cost of deploying specific equipment for reducing aerosol transmission such as ultraviolet germicidal irradiation (UVGI) and high-efficiency particular air (HEPA) filtration are \$182.37 and \$136.78 per person per year, respectively, ξ_8 is thus equal to \$0.874 per person per day via dividing the total cost of the aforementioned equipment by 365 days. Finally, we assume that implementing Guideline #9 does not incur any cost ($\xi_9 =$ \$0 per person per day).

All the parameter definitions, values and ranges are summarized in **Table 1**. However, it should be stressed that the parameter values calibrated above are for heterogeneous populations. More importantly, at the early stage of the pandemic, the estimates of certain transmission model parameters (e.g., the reproduction number and transmission rates) are expected to be larger due to the absence of protection policies and self-protection awareness; and the transmission rates are expected to have a notable drop after the implementation of various protection and quarantine strategies (PQS). Such a hypothesis is supported by several recent studies, which showed that the overall transmission rates may decrease by 2.1–3.2 folds after implementing PQS. Also, the study of Seto et al. [46] quantified the odds ratios of SARS infection as 13 (95% CI: 3, 60), 2 (95% CI: 0.6, 7), or 5 (95% CI: 1, 19), corresponding to the use of masks, gloves or hand-washing, respectively. Thus, the values of certain model parameters (e.g., transmission rates) will be different from (e.g., smaller than) the values calibrated in this section after implementing Guidelines #1-9, which will be elaborated in the result Section.

Implementation and Computing Configuration

All the computing codes were implemented in MATLAB® (MathWorks, Natick, MA), and the ordinary differential equation (ODE) solver ode15s was employed for solving the transmission model numerically. The relative error tolerance was set at 10^{-7}, the absolute error tolerance at 10^{-7}, and the maximum step size at 10^{-2}.

[6]Available online at: https://tradingeconomics.com/united-states/corporate-profits

[7]Available online at: https://dlt.ri.gov/lmi/datacenter/ces.php

[8]Available online at: https://www.bls.gov/regions/southwest/news-release/countyemploymentandwages_texas.htm

[9]Available online at: https://www.cbo.gov/publication/56073#_idTextAnchor148

[10]Available online at: https://www.chem.purdue.edu/chemsafety/Training/PPETrain/ppetonline.htm

[11]Available online at: https://www.shopp.org

[12]Available online at: https://www.amazon.com

RESULTS

Sensitivity Analysis

We evaluated the local sensitivity of the transmission model in Equations (1–5). Specifically, we evaluated the changes in the model outcome variables (i.e. S, C, Q, R, D) corresponding to a 1% increase in one parameter value, with the other seven parameter values fixed at their default values as in **Table 1**. Initial values for the sensitivity analysis were set as 299 susceptible and 1 silent carrier which approximated the proportion of infections in the U.S. general population as reported by May 5, 2020. The initial numbers of quarantined, recovered, and death were all set as 0 for simplicity. The results were visualized in **Figure 2**, showing that the transmission model was most sensitive to the transmission rate of pre-symptomatic and VMS infections (β_P) and the rate of progression to symptomatic infection (τ), and least sensitive to the product $\beta_O \cdot \kappa$. For the most sensitive parameters, a 1% increase in parameters β_P or τ resulted in a change <1.5% in the outcomes throughout a period of 200 days. At the end of the 200 days, all changes in the outcomes reached a plateau or tended to diminish. The results suggested that the transmission model was not locally sensitive to parameter value changes and could make robust predictions. Additional sensitivity analyses were then performed to understand the outcome changes over the entire range of parameter values. As shown in **Figure 2**, *S* was one of the most sensitive outcomes to parameter value changes; therefore, we plotted *S* against time for each of the eight parameter value ranges in **Figure 3**. For each parameter range, the minimum, the 1st quantile, the mean, the 3rd quantile, and the maximum parameter values were used to generate the trajectories of *S*. The results suggested that the dynamic model did not crash into chaos and the outcome *S* varied smoothly and predictably within the entire parameter range.

To evaluate the global sensitivity of this model, the Partial Spearman Rank Correlation Coefficient (PRCC) method [47] was employed. The model parameters were randomly sampled 100 times using the Latin hypercube sampling technique over uniform distribution. The range of the uniform distribution for the global sensitivity for each model parameter was reported in **Table 1**. **Figure 4** showed the plot of the PRCC values (next to the y axis) between model parameters and ODE outputs against time. As suggested in many previous studies, an absolute PRCC value >0.4 was deemed as practically significant [48]. For the susceptible population (*S*), the corresponding subfigure showed that the transmission rates β_A and β_P were strongly and negatively (< -0.7) correlated with *S*. This result was expected as the higher the transmission rates, the smaller the number of the susceptible persons. Please note that the percentage of asymptomatic carriers (α) was positively correlated with *S* initially (> 0.4) and then became negatively correlated with *S* at the end (< -0.6). The initial positive correlation between *S* and α should not be interpreted as that *S* will increase when α increases, but that *S* will decrease less when α increases. This will happen when $\beta_A \cdot \alpha$ is less than $\beta_P \cdot (1 - \alpha)$. Around the end of the time window of the simulation, a larger α corresponded to a smaller $N = S + C + R$ due to the increase in death, and thus *S* will decrease more given a smaller *N* in the denominator at the righthand side of Equation (1). From all the subfigures, the transmission rates β_A and β_P were always found strongly correlated with the outcomes while it was not the case for $\beta_O \cdot \kappa$, the value of which was too small to have a substantial impact on the outcomes. In addition, parameters α, γ and τ were also found strongly correlated with the outcomes, and interventions like more effective drug treatment or vaccination would affect these parameter values.

Case Study

Business executives are strongly recommended to follow the guidelines listed in the Background Section to prevent the reopening from causing any exacerbation of the ongoing pandemic. However, practical difficulties may arise due to, e.g., insufficiency of budget or medical recourses such that only some of the guidelines will be implemented by employers. Through this case study, we illustrated the use of the transmission model and the cost equation in different scenarios to evaluate the feasibility of reopening. Note that our simulation results were obtained under many assumptions and subject to both model structure and parameter value uncertainty; therefore, decisions of business executives should not be made solely based on the results presented in this section.

In this case study, we focused on four scenarios: (i) none of the nine Guidelines was constantly implemented (baseline scenario); (ii) all Guidelines 1–9 were constantly implemented; (iii) PPE and all other indispensable guidelines (1, 2, 3, 4, 6, 9) were constantly implemented; (IV) Indispensable guidelines (1, 3, 4, 6, 9) were constantly implemented but without the use of PPE. Scenarios I was the baseline scenario, corresponding to a complete devaluing of infection risk. Across scenarios II-IV, Guidelines 1, 3, 4, 6, 9 were assumed to be always implemented given the necessity and indispensability of these five Guidelines to business reopening. Scenario III was designed to be less restrictive than Scenario II, considering that some business entities might not have the budget to deploy non-contact sensors for real-time fever detection, reduce the number of working hours, or acquire specific equipment for aerosol transmission prevention. In Scenario IV, the use of PPE was intentionally dropped to account for the possible shortage of such materials on the market and to understand the importance of the use of PPE. The implementation of behavioral and social practice guidelines in Scenarios II-IV led to changes in the values of these three, $\beta_O \cdot \kappa$, β_A, and β_P, parameters in our transmission model. The other model parameters were assumed not directly affected by Guidelines 1-9. For instance, the recovery rate γ primarily depends on subject-specific immunity, the availability, affordability of effective medical intervention and health care, instead of behavioral and social practice patterns. For these reasons, we adjusted the values of $\beta_O \cdot \kappa$, β_A, and β_P for scenarios II–IV, respectively. The guidelines followed in each scenario and the corresponding adjusted parameter values were listed in **Table 2**. For simplicity, we assumed that the effects of different guidelines were independent and remained constant throughout the entire simulation time window. We also assumed that after reopening, every employee

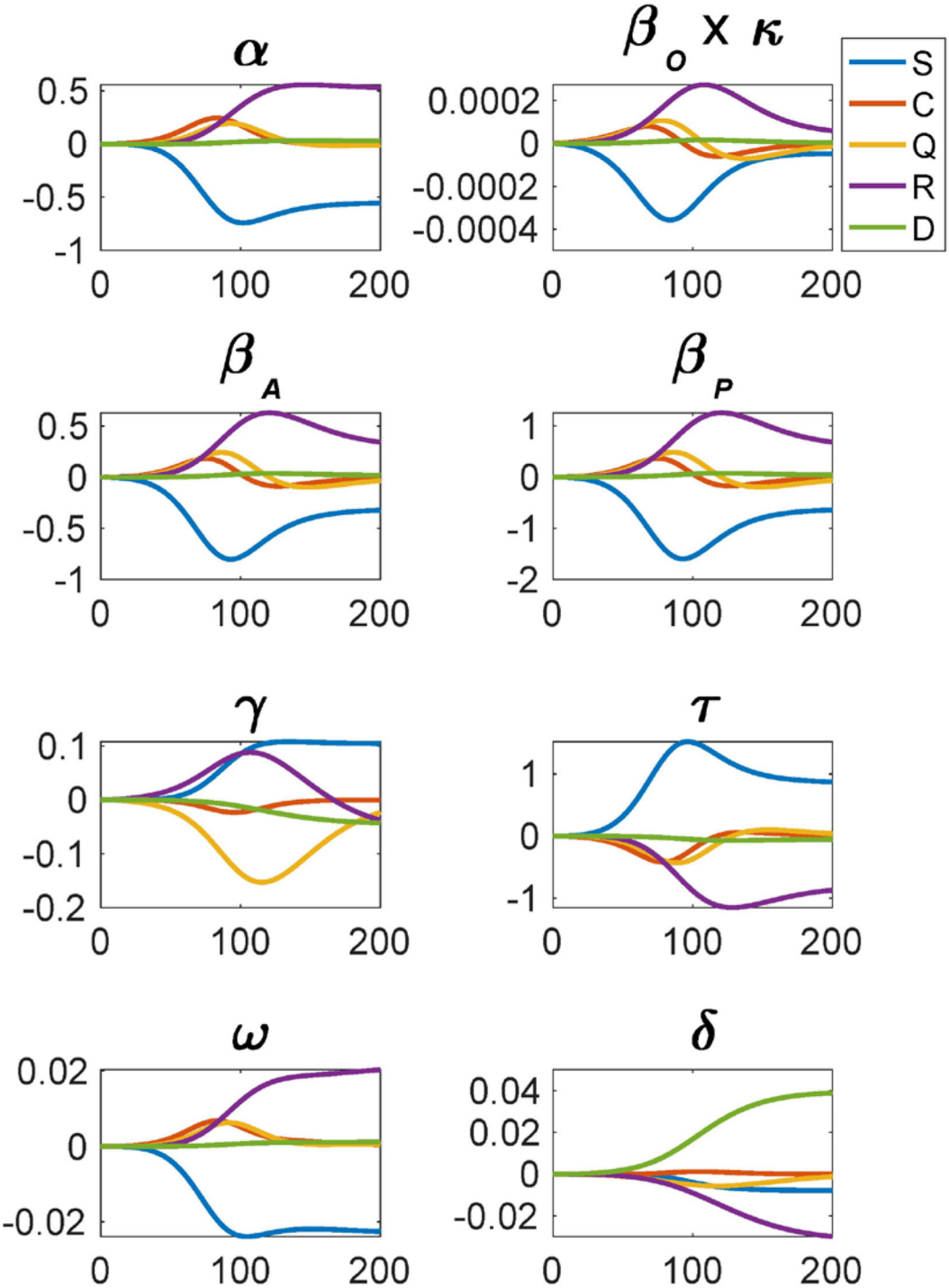


FIGURE 2 | Local sensitivity analysis of the transmission model, showing the changes in the model outcome variables corresponding to 1% increase in one parameter value (with the other seven parameter values fixed at their default values as in **Table 1**).

rigorously followed the guidelines and immediately reported their symptoms or infections once identified.

According to the study of Koo et al. [49], the practice of social distancing together with disease testing, reporting and quarantine policy (Guidelines 1, 4, and 6) could reduce the number of infections by 78.20% (IQR: 59.0–94.4%) compared with the baseline scenario when R_o was 2.5. That is, the transmission rates might drop to 1–78.2% = 21.80% of their

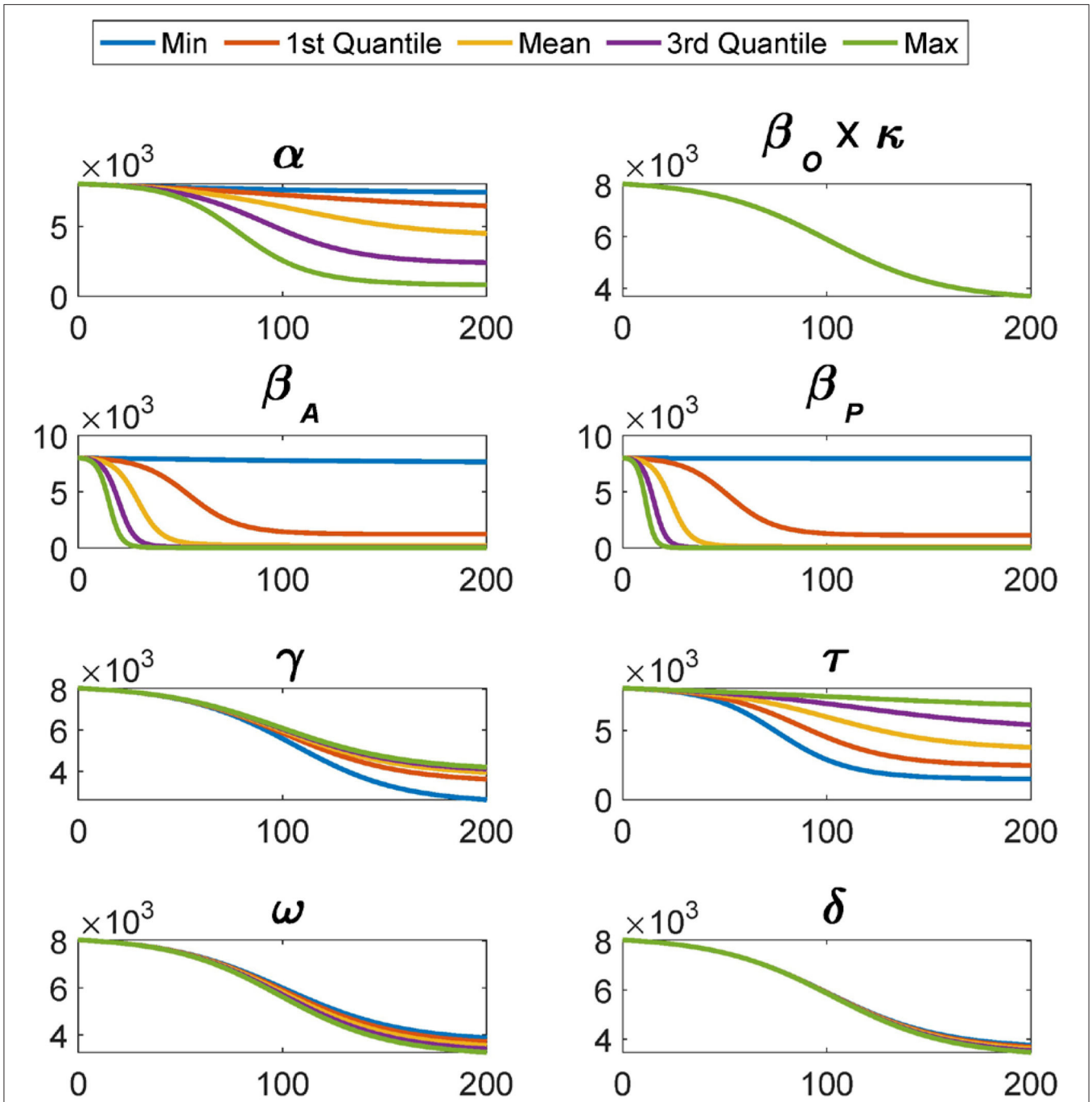


FIGURE 3 | The trajectories of S over the entire range of each parameter. For each parameter, the trajectories of S are generated corresponding to the minimum, the 1st quantile, the mean, the 3rd quantile, and the maximum parameter values.

baseline values after implementing Guidelines 1, 4, and 6. To quantify the effects of wearing mask, glove, goggle and hand washing (Guideline 2) on infection transmission, we adopted the results from Seto et al. [46] and Yin et al. [50] for severe acute respiratory syndrome (SARS). In their results, the odds ratio for mask wearing was 13 (95% CI: 3–60, attack rate $\frac{11}{83} = 13.25\%$), for glove wearing was 2 (95% CI: 0.6–7, attack rate $\frac{9}{133} = 6.77\%$), for hand-washing was 5 (95% CI: 1–19, attack rate $\frac{3}{17} = 17.65\%$), and for goggles wearing was $\frac{1}{0.2} = 5$ (95% CI: $2.44 - 10$, attack rate 61.50%). Correspondingly, the transmission rates might drop to 8.83% (95% CI: 1.92–36.56%) for mask wearing, 48.85% (95% CI: 15.17–100%) for glove wearing, 39.37% (95% CI: 22.40–64.33%) for goggle wearing, and 23.91% (95% CI: 6.32–100%) for hand washing. Thus, due to the implementation

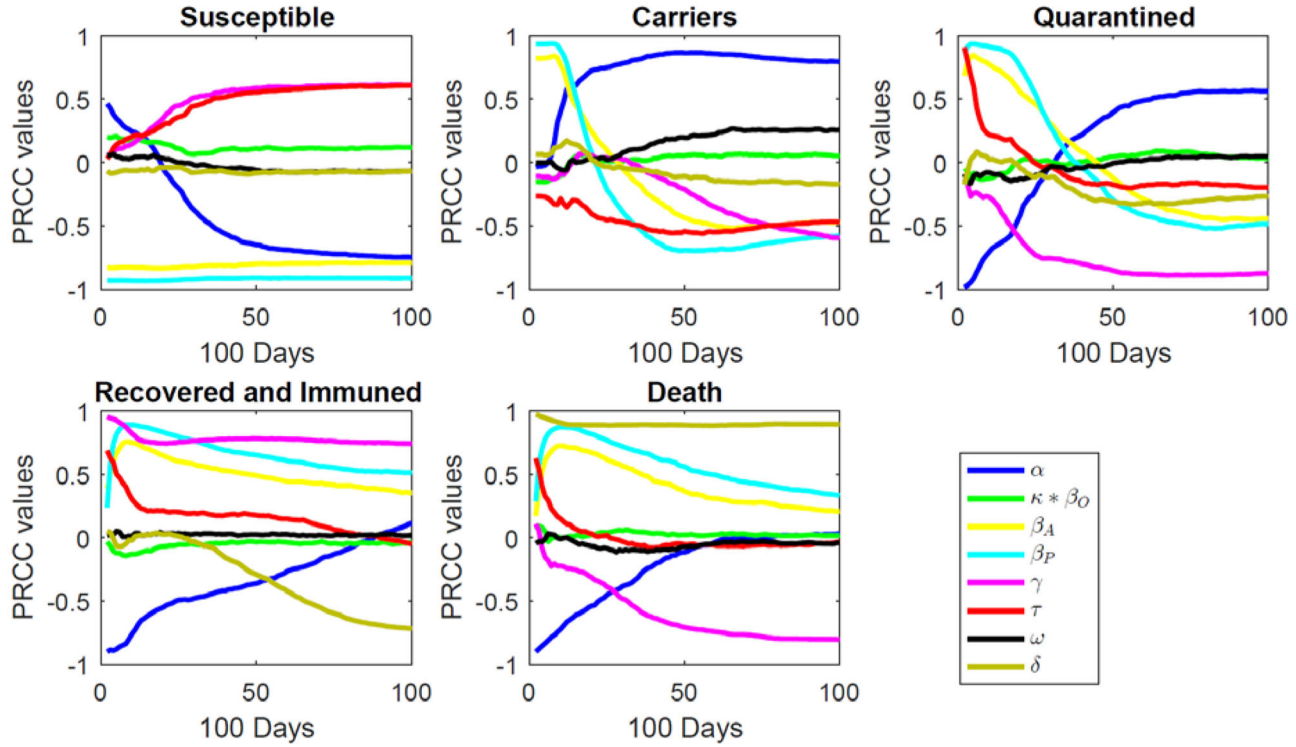


FIGURE 4 | Global sensitivity analysis (PRCC) of the transmission model.

of Guideline 2, the transmission rates might drop to 8.83% × 48.85% × 39.37% × 23.91% = 0.41% of their baseline values. Zhang et al. [51] investigated influenza A transmission in a student office setting, which showed that the infection risk could reduce by 2.14% with implementation of routine sanitization. Routine sanitation is expected to reduce the transmission rates to 1 − 2.14% = 97.86% of their baseline values after implementing Guideline 3. The implementation of Guideline 5 would improve the implementation of Guidelines 4 and 6 so its effect on infection transmission was not explicitly quantified in this study. For Guideline 7, we assumed a linear relationship between the number of working hours and the infection risk. The average working hours per day after reopening was assumed to be 70% of the regular working hours. Correspondingly, the transmission rates might drop to 70% (range 50–100%) with Guideline 7. For Guideline 8, according to the report of Mendell et al. [52], a 10–14% reduction in communicable respiratory infections might result from improved work environments. Taking 12% as the median, we assumed that the transmission rates might drop to 100% − 12% = 88% (range 86–90%) with implementing Guideline 8. Furthermore, the transmission rates among employees in contact-based positions were not affected by Guideline 9. Finally, in scenarios II, III, and IV, the transmission rates dropped to21.80% × 0.41% × 97.86% × 70% × 88% = 0.05%, 21.80% × 0.41% × 97.86% = 0.09% and 21.80% × 97.86% = 21.33% of their baseline values, respectively. See **Table 2** for the adjusted transmission rates.

All the simulation results in this section were generated using the same set of initial values for simplicity. Taking the DELL^{TM} center at Austin, Texas as an example, the total number of active workers after reopening could be around 14,000. The ratio between the on-site workers and the WFH workers was assumed as 2:1 (i.e., 9,333 on-site workers and 4,667 WFH workers). On the first day of reopening (day 0), the proportion of silent carriers among all the 14,000 employees was approximated by the proportion of infections in the general population of U.S. estimated using the number of reported cases as of May 5, 2020. The proportion of workers that had recovered from the infection and thus acquired immunity (referred as "recovered" in our model) were also estimated from the reported number of recoveries of the general U.S. population. The numbers of quarantined employees and death on day 0 were assumed to be 0. We conducted the simulations for 200 days for Scenario I and 100 days for Scenarios II–IV to verify the short-term feasibility of reopening. The outcome trajectories in the four scenarios were shown in **Figures 5**, **6** showed the reopening feasibility by plotting the net profit and the prevalence of infections in the workplace as well as in the general population (the left column), and the corresponding values of cost, profit, and net profit (the right column).

As shown in **Figures 5A–C**, without implementing any of the guidelines, reopening merely led to a large number of infections and deaths (230 deaths by the last day, 649 quarantined and 422 silent carriers at their peaks). The prevalence of infections

TABLE 2 | Parameters value adjustment in different scenarios.

		Scenario I (Baseline)	Scenario II	Scenario III	Scenario IV
Guidelines (*denotes implemented guidelines)	1. Social distancing		*	*	*
	2. PPE		*	*	
	3. Routine sanitization		*	*	*
	4. COVID-19 tests		*	*	*
	5. Non-contact sensors		*		
	6. Case reporting and quarantine		*	*	*
	7. Maximum duration of exposure		*		
	8. Equipment for aerosol transmission prevention		*		
	9. WFH if applicable		*	*	*
Parameter Values	$\beta_O \cdot \kappa$ $(\% \times day^{-1})$	3.100×10^{-7}	1.653×10^{-20}	2.684×10^{-20}	6.613×10^{-18}
	$\beta_A (day^{-1})$	0.099	5.280×10^{-05}	8.571×10^{-05}	2.112×10^{-02}
	β_P (day^{-1})	0.197	1.051×10^{-04}	1.706×10^{-04}	4.203×10^{-02}

within the business entity exceeded that in the general population throughout the 200-days window, and peaked at 114 infections per 1,000 people, which was 30 folds higher than that of the general population. Even by the end of the 200 days, the within-entity prevalence was 8 folds higher (32.9 cases per 1,000 persons) than the population average. Also, even after we extended the simulation time window to 200 days, the transmission model still did not reach its steady states, suggesting a less predictable risk of disease transmission. In short, despite that the net profit remained positive (**Figure 6B**), since the prevalence in workplace was constantly (much) higher than that in the general population, reopening turned out to be infeasible in Scenario I.

If a business entity strictly followed all guidelines 1–9 as in Scenario II, the number of infections and deaths reduced remarkably, as shown in **Figures 5D–F** (2 deaths among 9,333 onsite workers, at most 16 quarantined and 35 silent carriers). **Figure 5** also showed that the outcomes of the transmission model reached a plateau toward the end of the 100-days window. In other words, the infection could be contained in this scenario, with the number of carriers and the number of the quarantined reaching and staying at 0 by day 48 and day 92, respectively. According to **Figure 6C**, under the assumption that the prevalence of infections within the business entity was the same as in the general population at the beginning of reopening, the within-business prevalence dropped under the general population prevalence immediately (on day 2), and continuously decreased down to 0.035 infections per 1,000 people by the last day of simulation. This within-business prevalence was 104 folds lower than the general population prevalence reported in other studies. The business also attained higher and more stable net profit than in Scenario I (**Figure 6D**).

In Scenario III, guidelines 5, 7, 8 were skipped and the transmission rate values changed accordingly (see **Table 2**); however, the simulation results were surprisingly similar to those in Scenario II during the 100-day time window (**Figures 5G–I**, **6E,F**). Specifically, the numbers of deaths, the quarantined and silent carriers were nearly the same as those in Scenario II; also, it took approximately the same amount of time for the numbers of carriers and the quarantined to drop to 0 (**Figures 5G–I**). The nearly same predicted population trajectories resulted in a nearly same prediction on net profit (**Figures 6E,F**). To confirm such results, additional local sensitivity analysis was conducted at the parameter values in Scenarios II and III, suggesting that the simulation outputs were not sensitive to parameter value changes (e.g., a change of <0.1 in all the five output variables corresponding to 1% parameter value change).

After a business entity further dropped guideline 2 (Mask, glove, and goggle wearing while not alone) in Scenario IV, our model predicted that the spread of infections did not become uncontrollable within 100 days but associated with a higher cost (**Figures 5J–L**, **6G,H**). While the initial within-business prevalence was the same as the general population prevalence, the workplace quickly became a "hot spot" of infection spreading, and a prevalence much higher than the population average was reached (**Figure 6G**). The within-business prevalence continued to stay above the general population average for 11 days, and peaked at a level of 4 infections per 1,000 people on day 5. The results suggested that reopening should stop to prevent this entity from developing into a source of infection and posing significant risk on its workers as well as their close social contacts. Note that the within-business prevalence dropped to 0.09 infections per 1,000 people by the end of the 100-day time window (**Figure 6G**), and the numbers of deaths and the quarantined were controlled under 3 and 20 among 9,333 onsite workers, respectively. Business executives should not rely on such optimistic predictions and underestimate the infection risk for two reasons. First, constant parameter values were used in our simulations, which were not capable of capturing every possible time-varying characteristic of disease transmission over time (i.e., parameter values could be time-varying instead of

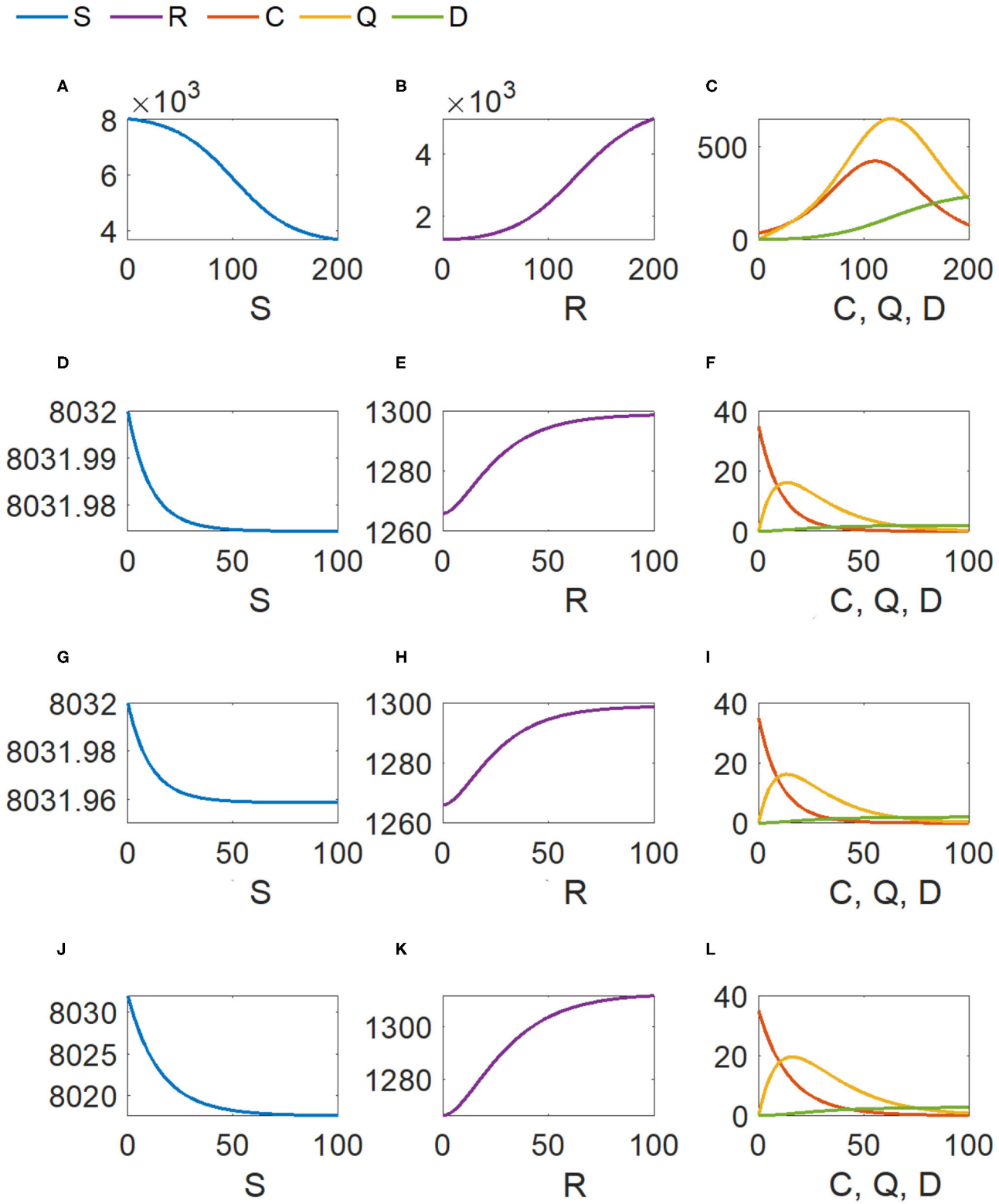


FIGURE 5 | Outcome trajectories for the four scenarios. For the same scenario, the five outcome variables of different orders of magnitude are visualized by three subfigures in one row. The scale of y-axis is adjusted to enable a clear view of each single trajectory over time. **(A,D,G,J)** Susceptible (S); **(B,E,H,K)** Recovered (R); **(C,F,I,L)** Carrier (C), Quarantined (Q), and Death (D).

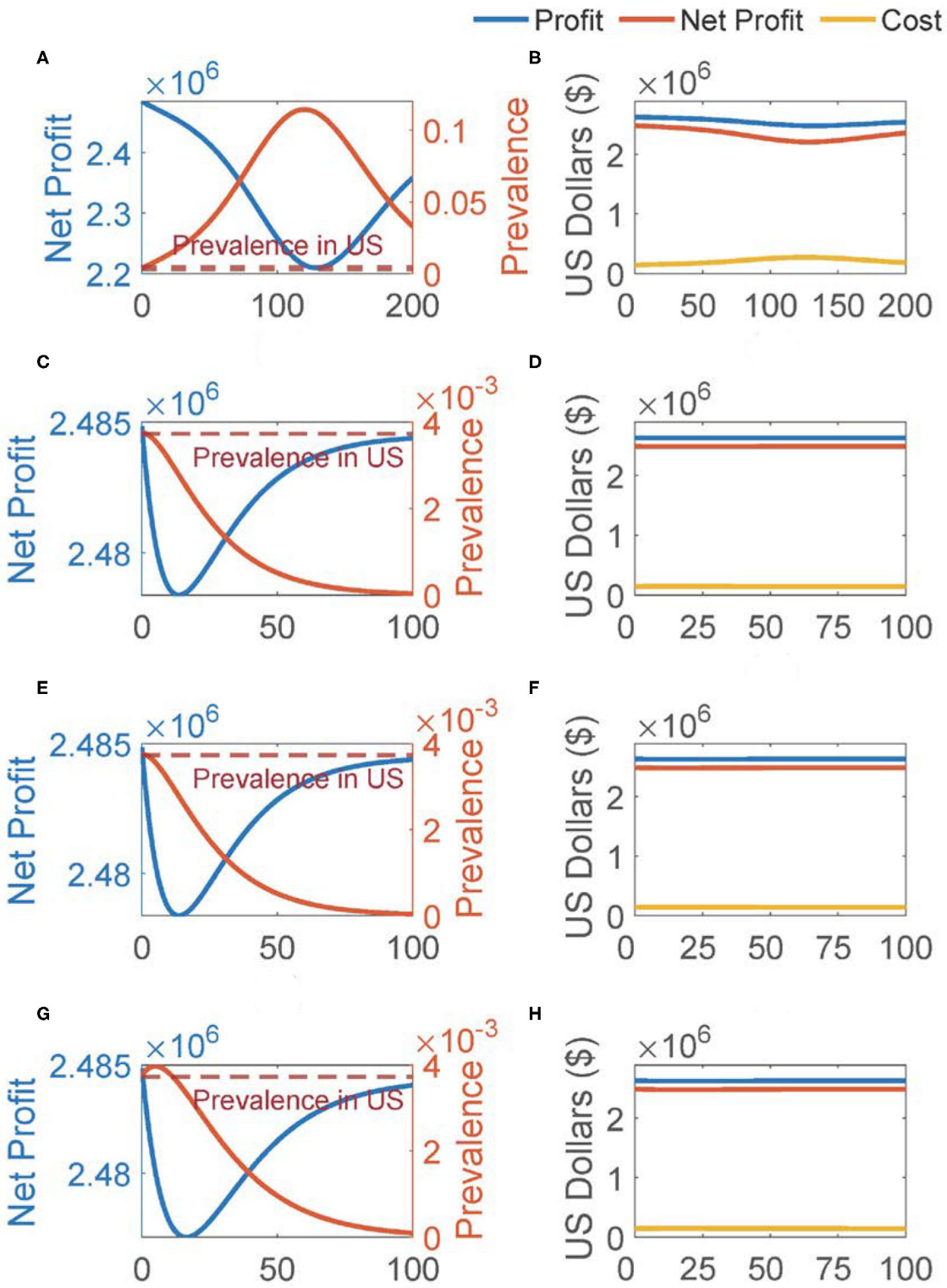


FIGURE 6 | Cost, profit, net profit, and infection prevalence for the four scenarios (each row for one scenario and Scenarios I–IV in the order of the top row to the bottom). For each scenario, the left subplot shows the reopening feasibility based on both net profit and infection prevalence in the workplace, with the prevalence in the population presented for comparison; and the right subplot shows the cost, profit and net profit. **(A,C,E,G)** Reopening feasibility, **(B,D,F,H)** cost and profit.

being constant). Second, our simulation was performed under the assumption that all workers strictly followed the selected guidelines. In reality, it is unlikely that every single employee would faithfully stick to such rules and guidelines daily and constantly over time. Moreover, **Figure 6G** suggested that the use of PPE was very important at the early stage of reopening. In Scenarios II and III, with the use of PPE, we did not observe the rapid increase of prevalence since the beginning of reopening; however, in Scenario IV without the protection of PPE, the within-business prevalence started to increase and exceed the population average on day one. This result was consistent with the recent study by Kai et al. [53], which demonstrated the significant effect of universal use of facial masks (e.g., at least 80% population wear masks) on impeding the spread of infections.

DISCUSSION

SARS-CoV-2 struck the whole world since 2019 and caused significant loss of human lives and economy. While various lockdown, quarantine, and isolation rules and polices worked effectively to control COVID infection transmission, numerous businesses were closed and tens of millions of people lost their jobs. As of May 20, 2020, all the states in the U.S. strategically moved toward gradual reopening to save the economy. However, given that neither effective drug treatments nor vaccines were available, the risk of infection spreading within a business entity became a key issue that any business executive had to consider. In this study, an ODE-based transmission model was developed together with a net profit equation to quantitatively evaluate the trade-off among profit, cost, and infection risk within a business entity after reopening. Model parameter values were calibrated from heterogeneous sources to enable computer simulations. Both local and global sensitivity analyses were conducted to understand our model behavior and result robustness. Finally, a case study assessed scenarios, in which different combinations of behavioral and social practice guidelines were implemented. The simulation results suggested that infection transmission was controllable within a business entity and a positive net profit could be generated after reopening only if a combination of selected guidelines were implemented. Also, our results suggested that the use of PPE could be significantly important at the early stage of reopening to prevent infection spreading.

Modeling business reopening is a novel and important problem during the COVID-19 pandemic or a similar situation. This study adopted a business entity's perspective instead of a social perspective (e.g., one can further consider social cost of death due to infection and model the interactions out of a workplace). However, a study from a business entity's point of view has its own added value. Note that reopening decision is not just the call of local governments but also business executives. For instance, even if a state or local government issues the reopening orders, a business entity may decide by itself whether it should reopen immediately or postpone the reopening after taking into consideration of both business-specific parameters and benefit-cost-risk trade-off. With that said, future researches from different perspectives are more than welcome to shed novel insights into the business reopening problem.

There are several issues worth of further discussions. First, while the proposed model did not explicitly introduce a term for the incompliance of employees on the use of PPE, we have investigated both Scenario III and Scenario IV in the Case Study Section. The only difference between the two scenarios is whether PPE is used; therefore, the results corresponding to partial or full incompliance by employees will fall between the results of Scenario III and those of Scenario IV. Second, our model implicitly accounted for the fact that a virus testing process might not be instant and completely error free. Recall that the outcome variable for "carriers" (C) in our model is the total number of asymptomatic, pre-symptomatic, and very-mildly-symptomatic (VMS) carriers (collectively known as silent spreaders), as defined in the Mathematical Model Section. It is reasonable to expect that any person with notable symptoms should practice self-quarantine or receive medical treatment even when waiting for virus test results or tested false negative; therefore, cases that are not immediately identified or quarantined are accounted for in C. Third, $\beta o \times \kappa$ was found not to play a very significant role in this study. Note that after reopening, employees will spend the majority of their 'gathering' time in the workplace so the workplace transmission would and should play a major role. Consequently, the more time employees spend on work-related activities, the less number of interactions they could have with others outside the workplace, suggesting a small value for $\beta o \times \kappa$. Finally, while mathematically it could be interesting to examine all the $2^9 = 512$ possible combinations of the nine guidelines to identify the optimal combination(s), it may not serve the practical purpose of this study. For instance, we can hardly imagine a worker who would strictly stick to the use of PPE but not practice any handwash; or a workplace gives up on the cost-effective measures such as social distancing and routine sanitization, while spending lots of budget on equipment for aerosol transmission prevention. More importantly, the infection protection measures considered in Scenario III are those strongly recommended by CDC due to their efficacy and cost-effectiveness. That is also why the optional protection measures were only considered in Scenario II to quantify the additional benefits such measures may provide.

We also recognize a number of limitations of this study. First, the mathematical model and the net profit equation were developed under multiple assumptions. While many of those assumptions are commonly adopted by related professional communities (e.g., epidemiology or mathematical modeling), some assumptions were introduced due to the lack of accurate and/or complete information (e.g., asymptomatic infection transmission rate). With additional efforts invested in future SARS-CoV-2 related research, we expect that informative and high-quality data will become available such that less assumptions are needed and more accurate results can be generated by our approach. Second, while we have compiled information from many different sources to calibrate the possible ranges of model parameter values, the parameter uncertainty may not be completely characterized by such parameter ranges due to, e.g., the heterogeneity in population demographics, health conditions, behavioral patterns, and social networks. Third, although the costs associated with guidelines #1-9 were included in the net profit calculation, our cost estimation is subject to

market fluctuations and many other (unpredictable) factors. It is thus suggested for users of our model to fine tune parameter values upon available business-specific data or information. Finally, in this study, herd immunity and vaccination strategies were not considered because they were not available as of the current moment. Considering the active research on SARS-CoV-2 vaccine and drug development, the availability of effective vaccines and medications can be expected and it should be taken into consideration in our modeling work at some point.

In summary, this modeling work provides a quantitative tool for decision-makers to explore and evaluate the business reopening option in the midst of COVID pandemics, and it can be extended to similar scenarios (e.g., outbreak of unknown or new virus) by re-calibrating related parameter values. We expect further research efforts in this direction to better prepare for possible strikes of infectious diseases in the future.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

HM designed and supervised this study, implemented computing code, and wrote the manuscript. QG, HF, CZ, PZ, and LW calibrated parameter values, implemented computing code, and contributed to manuscript writing. MS, XL, SD, DLai, CB, AP, LR, and DLair co-developed the modeling idea, provided parameter values, and contributed to manuscript editing.

FUNDING

This work was partially supported by NSF grant DMS-1620957 (HM), Weatherhead Foundation (DLai), and NSF grant DMS-1758290 (LR).

ACKNOWLEDGMENTS

Also, thank Dr. Eric Boerwinkle for motivating this study and sharing constructive comments and thank Drs. Wenyaw Chan and Hulin Wu for insightful discussions and feedbacks.

REFERENCES

1. Atkeson A. What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. *National Bureau of Economic Research Working Paper Series* No. 26867 (2020). doi: 10.3386/w26867
2. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet.* (2020) **395**:931–4. doi: 10.1016/S0140-6736(20)30567-5
3. McKibbin WJ, Fernando R. *The Global Macroeconomic Impacts of COVID-19: Seven Scenarios*. CAMA Working Paper No 19/2020 (2020). Available online at SSRN: https://ssrn.com/abstract=3547729.
4. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet.* (2020) **365**:1569–78. doi: 10.1016/S0140-6736(20)31022-9
5. Williamson BN, Feldmann F, Schwarz B, Meade-White K, Porter DP, Schulz J, et al. Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *bioRxiv.* (2020). doi: 10.1101/2020.04.15.043166. [Epub ahead of print].
6. Piguillem F, SHI L. Optimal Covid-19 quarantine and testing policies. *CEPR Discussion Paper No DP14613.* (2020). Available online at SSRN: https://ssrn.com/abstract=3594243.
7. Huremović D. Social Distancing, Quarantine, and Isolation. In: Huremović D, editor. *Psychiatry of Pandemics: A Mental Health Response to Infection Outbreak.* Cham: Springer International Publishing (2019). p. 85–94. doi: 10.1007/978-3-030-15346-5_8
8. Song B, Hei X. Models and strategies on reopening lockdown societies due to COVID-19. *OSF Preprints.* (2020). doi: 10.31219/osf.io/umtvh
9. Bhatia R, Klausner J. *A Step-by-Step Plan to Reopen California.* (2020). Available online at: https://theboldtaliccom/covid-19-next-steps-for-california-c01632c8e8b6 (accessed January 5, 2020).
10. Aboelkassem Y. COVID-19 pandemic: a Hill type mathematical model predicts the US death Number and the reopening date. *medRxiv.* (2020). doi: 10.1101/2020.04.12.20062893
11. Wang X, Tang S, Chen Y, Feng X, Xiao Y, Zongben X. When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemic? A data-driven network modeling. *Sci Sinica Math.* (2020) 50:969–78. doi: 10.1360/SSM-2020-0037
12. Thunstrom L, Newbold S, Finnoff D, Ashworth M, Shogren JF. The benefits and costs of using social distancing to flatten the curve for COVID-19. *Forthcoming J Benefit Cost Anal.* (2020) **11**:179–95. doi: 10.1017/bca.2020.12
13. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *N Engl J Med.* (2020) **382**:1199–207. doi: 10.1056/NEJMoa2001316
14. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance.* (2020) **25**:2000180. doi: 10.2807/1560-7917.ES.2020.25.10.2000180
15. Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA.* (2020) **323**:1406–7. doi: 10.1001/jama.2020.2565
16. Chisholm RH, Campbell PT, Wu Y, Tong SYC, McVernon J, Geard N. Implications of asymptomatic carriers for infectious disease transmission and control. *Royal Soc Open Sci.* (2018) **5**:172341. doi: 10.1098/rsos.172341
17. Dijkstra J, Hashimoto K. Expected immune recognition of COVID-19 virus by memory from earlier infections with common coronaviruses in a large part of the world population [version 1; peer review: awaiting peer review]. *F1000Research.* (2020) **9**:285. doi: 10.12688/f1000research.23458.2
18. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature.* (2020) **581**:465–9. doi: 10.1038/s41586-020-2196-x
19. Sun WW, Ling F, Pan JR, Cai J, Miao ZP, Liu SL, et al. Epidemiological characteristics of 2019 novel coronavirus family clustering in Zhejiang Province. *Zhonghua Yu Fang Yi Xue Za Zhi.* (2020) **54**:E027. doi: 10.3760/cma.j.cn112150-20200227-00199
20. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ.* (2020) **368**:m1165. doi: 10.1136/bmj.m1165
21. Kimball A, Hatfield KM, Arons M. Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility — King County, Washington. *MMWR Morb Mortal Wkly Rep.* (2020) **69**:377–81. doi: 10.15585/mmwr.mm6913e1
22. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung S, 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
23. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med.* (2020) **9**:462. doi: 10.3390/jcm9020462

24. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med.* (2020) **9**:538. doi: 10.3390/jcm9020538
25. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.* (2020) **368**:489–93. doi: 10.1126/science.abb3221
26. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med.* (2020) **27**:taaa021. doi: 10.1093/jtm/taaa021
27. Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR test results in patients recovered from cOVID-19. *JAMA.* (2020) **323**:1502–3. doi: 10.1001/jama.2020.2783
28. Xing Y, Mo P, Xiao Y, Zhao O, Zhang Y, Wang F. Post-discharge surveillance and positive virus detection in two medical staff recovered from coronavirus disease 2019 (COVID-19), China, January to February 2020. *Eurosurveillance.* (2020) **25**:2000191. doi: 10.2807/1560-7917.ES.2020.25.10.2000191
29. Omer SB, Malani P, del Rio C. The COVID-19 Pandemic in the US: a clinical update. *JAMA.* (2020) **323**:1767–8. doi: 10.1001/jama.2020.5788
30. Wang Q, Shi N, Huang J, Cui T, Yang L, Ai J, et al. Effectiveness and cost-effectiveness of public health measures to control COVID-19: a modelling study. *medRxiv.* (2020). doi: 10.1101/2020.03.20.20039644
31. Chen SC, Liao CM. Cost-effectiveness of influenza control measures: a dynamic transmission model-based analysis. *Epidemiol Infect.* (2013) **141**:2581–94. doi: 10.1017/S0950268813000423
32. White NM, Barnett AG, Hall L, Mitchell BG, Farrington A, Halton K, et al. Cost-effectiveness of an Environmental Cleaning Bundle for Reducing Healthcare-associated Infections. *Clin Infect Dis.* (2019) **70**:2461–3. doi: 10.1093/cid/ciz717
33. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* (2020) **382**:1708–20. doi: 10.1056/NEJMoa2002032
34. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the icelandic population. *N Engl J Med.* (2020) **382**:2302–15. doi: 10.1101/2020.03.26.20044446
35. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet.* (2020) **395**:689–97. doi: 10.1016/S0140-6736(20)30260-9
36. Shen M, Peng Z, Xiao Y, Zhang L. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. *bioRxiv.* (2020). doi: 10.1101/2020.01.23.916726
37. Liu T, Hu J, Kang M, Lin L, Zhong H, Xiao J, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *bioRxiv.* (2020). doi: 10.2139/ssrn.3526307
38. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv.* (2020). doi: 10.1101/2020.01.23.20018549
39. Majumder M, Mandl KD. *Early Transmissibility Assessment of a Novel Coronavirus in Wuhan, China*. Available online at SSRN: https://ssrncom/abstract=3524675 (accessed January 5, 2020).
40. Cao Z, Zhang Q, Lu X, Pfeiffer D, Jia Z, Song H, et al. Estimating the effective reproduction number of the 2019-nCoV in China. *medRxiv.* (2020). doi: 10.1101/2020.01.27.20018952
41. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis.* (2020) **92**:214–7. doi: 10.1101/2020.01.23.916395
42. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, et al. *Report 3: Transmissibility of 2019-nCov*. COVID 19 Resources, Imperial College London COVID-19 Reports (2020).
43. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance.* (2020) **25**:2000058. doi: 10.2807/1560-7917.ES.2020.25.4.2000058
44. 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. *Anna Int Med.* (2020) **172**:577–82. doi: 10.7326/M20-0504
45. 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.* (2020) 395:1054–62. doi: 10.1016/S0140-6736(20)30566-3
46. Seto WH, Tsang D, Yung RWH, Ching TY, Ng TK, Ho M, et al. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). *Lancet.* (2003) **361**:1519–20. doi: 10.1016/S0140-6736(03)13168-6
47. Marino S, Hogue IB, Ray CJ, Kirschner DE. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J Theoret Biol.* (2008) **254**:178–96. doi: 10.1016/j.jtbi.2008.04.011
48. Overholser BR, Sowinski KM. Biostatistics Primer: Part 2. *Nutrit Clin Pract.* (2008) **23**:76–84. doi: 10.1177/011542650802300176
49. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis.* (2020) **20**:678–88. doi: 10.1016/S1473-3099(20)30162-6
50. Yin WW, Gao LD, Lin WS, Du L, Zhang XC, Zou Q, et al. Effectiveness of personal protective measures in prevention of nosocomial transmission of severe acute respiratory syndrome. *Zhonghua Liu XIng Bing Xue Za Zhi.* (2004) **25**:18–22.
51. Zhang N, Li Y. Transmission of Influenza A in a student office based on realistic person-to-person contact and surface touch behaviour. *Int J Environ Res Public Health.* (2018) **15**:1699. doi: 10.3390/ijerph15081699
52. Mendell MJ, Fisk WJ, Kreiss K, Levin H, Alexander D, Cain WS, et al. Improving the health of workers in indoor environments: priority research needs for a national occupational research Agenda. *Am J Public Health.* (2002) 92:1430–40. doi: 10.2105/AJPH.92.9.1430
53. Kai D, Goldstein GP, Morgunov A, Nangalia V, Rotkirch A. Universal masking is urgent in the covid-19 pandemic: Seir and agent based models, empirical validation, policy recommendations. *arXiv [Preprint] arXiv*:200413553 (2020). doi: 10.13140/RG.2.2.21662.08001

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.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 14 August 2020
doi: 10.3389/fphy.2020.00311

Estimation of Infection Rate and Predictions of Disease Spreading Based on Initial Individuals Infected With COVID-19

*Seo Yoon Chae[1], KyoungEun Lee[2,3], Hyun Min Lee[1], Nam Jung[1], Quang Anh Le[1], Biseko Juma Mafwele[1], Tae Ho Lee[1], Doo Hwan Kim[1] and Jae Woo Lee[1,4,5]**

[1] Department of Physics, Inha University, Incheon, South Korea, [2] Ecology and Future Research Institute, Busan, South Korea, [3] National Institute of Ecology, Seocheon-gun, South Korea, [4] Institute of Natural Basic Sciences, Inha University, Incheon, South Korea, [5] Institute of Advanced Computational Sciences, Inha University, Incheon, South Korea

We consider the pandemic spread of COVID-19 in selected countries after the outbreak of the SARS-CoV-2 coronavirus in Wuhan City, China. We estimated the infection rate and the initial individuals infected with COVID-19 by using officially reported data from the early stages of the epidemic for a model of susceptible (S), infectible (I), quarantined (Q), and officially confirmed recovered (R_k) populations (the so-called SIQR$_k$ model). In the officially reported data, we know the number of quarantined cases and the officially reported number of recovered cases. We cannot know about recovered cases from asymptomatic patients. In the SIQR$_k$ model, we can estimate the parameters and the initial infections (confirmed cases + asymptomatic cases) from fitted values. We obtained an infection rate in the range $\beta = 0.233 \sim 0.462$, a basic reproduction number of $R_o = 1.8 \sim 3.5$, and the initial number of infected individuals, $I(0) = 10 \sim 8409$, for selected countries. By using fitting parameters, we estimated that the maximum time span of the infection was around 50 days in Germany when the government invoked the quarantine policy. The disease is expected to subside about 6 months after the first patients are found.

Keywords: coronavirus (2019-nCoV), epidemic model, SIR (Susceptible Infected-Recovered) model, quarantine, asymptomatic

OPEN ACCESS

Edited by:
Aristides (Aris) Moustakas, Natural History Museum of Crete, University of Crete, Greece

Reviewed by:
Antonio Cadilhe, Universidade Federal da Bahia, Brazil
Alexis Toda, University of California San Diego, United States

***Correspondence:**
Jae Woo Lee
jaewlee@inha.ac.kr

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *18 May 2020*
Accepted: *06 July 2020*
Published: *14 August 2020*

Citation:
Chae SY, Lee K, Lee HM, Jung N, Le QA, Mafwele BJ, Lee TH, Kim DH and Lee JW (2020) Estimation of Infection Rate and Predictions of Disease Spreading Based on Initial Individuals Infected With COVID-19. Front. Phys. 8:311. doi: 10.3389/fphy.2020.00311

INTRODUCTION

On December 31, 2019, Chinese authorities reported pneumonia from an unknown cause to the World Health Organization (WHO) in Wuhan City, Hubei province, China. On January 7, 2020, the virus was identified as a new coronavirus, first referred as 2019-nCov (SARS-CoV-2), which causes the disease named COVID-19. On January 11, 2020, China reported the first death from the novel coronavirus [1]. The victim was a 61-year-old man in Wuhan. On January 20, 2020, the WHO reported the first confirmed cases outside China (in Thailand, Japan, and South Korea [1]). The disease was spreading rapidly in Wuhan City, and cases were reported outside that city. On January 23, 2020, China placed Wuhan, a city of 11 million people, under quarantine. All transportation departures were canceled or suspended [1]. The president of the WHO declared COVID-19 a pandemic on March 11, 2020. After the first report on December 31, 2019, in Wuhan, COVID-19 was spreading very quickly all over the world [2] and is the first pandemic in the twenty first century.

Some states, such as the Republic of Korea, Taiwan, Singapore, and Hong Kong, have been controlling the disease successfully up to now. However, other countries, like the U.S.A., Italy, Spain, France, and the U.K., are suffering from the outbreak and from shortages of medical materials and overcrowded hospitals. Since the outbreak, scientists all over the world have been struggling to find a vaccine and drugs for treatment. In the highly connected societies, information and data on the disease are shared through the internet, social media, and mass media. We can obtain information from websites like worldometer[1] or livecornamap in South Korea [2].

A flood of articles and preprints is appearing on many journal and preprint websites. Recently, preprint websites like arXiv.org[3], bioRxiv[4], and medRxiv[5] are servicing a section with COVID-19 quick links. It is important to predict the spread of the disease in the early stages of the outbreak. Many epidemic models were proposed based on dynamic spreading models, agent-based models, the Monte Carlo model, and data-based spreading models [3–10].

The evolution of the virus was described with a modified susceptible (S), infectious (I), recovered (R) population, the so-called SIR model [3–7]. The prediction of COVID-19 evolution in Brazil was suggested by using the susceptible, infectious, quarantined, recovered (SIQR) model [5]. Numerical analysis provided an estimated basis reproduction number of $R_o = 5.25$, and a doubling time estimated at 2.72 days. The SIQR model includes a rate that quantifies the recovering of asymptomatic individuals for the evolution equation of the infection and the recovering population. Peng et al. introduced an epidemic model for COVID-19 including the exposed population [4]. A model by Carcione et al. is called the generalized susceptible, infectious, exposed, recovered (SEIR) model [6]. They introduced time-dependent parameters, such as mortality rate and protection rate. They applied the model to the situation in the Italian Region of Lombardy, and estimated a basic reproduction number of $R_o = 2.6$ in the early stages of the outbreak. Fanelli and Piazza analyzed and forecast COVID-19 spread by using the susceptible, infected, recovered, dead (SIRD) model in China, Italy, and France [3]. Pedersen and Meneghini quantified undetected COVID-19 cases and the effects of containment measures in Italy, introducing the SIQR model [7], which includes a rate for patients to become non-infectious.

There are some limitations in the SIR-type models because of the ignorance of the age-structure, spatial heterogeneities, activity types of the people, latent periods, the polices to prevent the spreading, etc. [8, 9]. In real situation the government of each country enforces the protection strategies such as physical distance, face masks, eye protection, wide testing, household quarantine, lockdown, and effects of media in the early phase of outbreak [10–16].

[1]Worldometer. Available online at: https://www.worldometers.info/coronavirus/
[2]Livecoronamap. Available online at: https://livecorona.co.kr/
[3]arXiv.org. https://arxiv.org/
[4]bioRxiv.org. https://www.biorxiv.org/
[5]medRxiv.org. https://www.medrxiv.org/

In this article, we consider a susceptible, infectible, quarantined, and confirmed recovered ($SIQR_K$) model based on only known data for active cases and recovered cases. In particular, we cannot know the recovered cases of the asymptomatic infected individuals. In our model we can predict the whole infected cases, quarantined cases, and recovered cased based on the early data for the known quarantined cases and the officially recovered cases. We estimate the parameters of the model from data on reported cases for selected countries. We obtained the infection rate and the initial number of infected individuals. From the fitting parameters, we estimated the basis reproduction number, and predict the maximum time span of the infection and the annihilation period of the disease.

EPIDEMIC MODEL

We consider an epidemic model for COVID-19 that is characterized by the variables $\{S(t), I(t), Q(t), R(t)\}$ denoting the susceptible population, the infected population, the quarantined population, and the recovered population, all at time t. The total population satisfies the constraint $N = S(t) + I(t) + Q(t) + R(t)$ where N is the total population. Let us define the recovered population as $R(t) = R_k(t) + R_a(t)$, where $R_k(t)$ is the known- or confirmed-as-recovered cases as reported officially, and $R_a(t)$ is the unknown or asymptomatic recovered population (infected, but not showing symptoms). Under the homogenous mixing postulate, we consider the so-called $SIQR_K$ model as follows:

$$\frac{dS(t)}{dt} = -\beta \frac{S(t) I(t)}{N} \tag{1}$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t) I(t)}{N} - (\alpha + \eta) I(t) \tag{2}$$

$$\frac{dQ(t)}{dt} = \eta I(t) - \gamma Q(t) \tag{3}$$

$$\frac{dR_k(t)}{dt} = \gamma Q(t) \tag{4}$$

$$\frac{dR_a(t)}{dt} = \alpha I(t) \tag{5}$$

In this model, the parameter β denotes the infection rate, and α is the rate at which patients become non-infectious by recovering without showing any symptoms. The parameter η is the rate of detection for newly infected people, and γ is the rate of recovery for quarantined cases. In the $SIQR_k$ model, the infected populations are divided into officially confirmed cases and asymptomatic cases. We only know the official number of quarantined cases and the official number of recovered cases. We do not know the actual number in the infected population owing to the unknown number of asymptomatic cases. Some asymptomatic individuals have recovered without any severe suffering from the disease. We propose that the parameter included in the dynamic equations should be the initial number of infected cases, which is the sum of officially known cases and the unknown population of asymptomatic cases. In this model we don't include the number of deaths. The number of deaths is included implicitly in the number of quarantined cases. If the

TABLE 1 | Date of reporting the first case, plus the state and location of COVID-19 outbreaks for selected countries.

Ranking	Country	First Report	State/Province	Location
1	U.S.A.	Jan. 21	Washington	Sonohomish
2	Spain	Jan. 31	Canary Islands	La Gomera
3	Italy	Jan. 31	Lazio (Latium)	Rome
5	Germany	Jan. 27	Babaria	Munich
6	U.K.	Jan. 31	Newcastle	York City
7	China	Dec. 31, 2019	Hubei	Wuhan
8	Turkey	Mar. 11	Ankara	Ankara
9	Iran	Feb. 19	Qom Province	Qom
11	Brazil	Feb. 25	São Paulo	São Paulo
13	Canada	Jan. 25	Ontario	Toronto
23	Japan	Jan. 16	Kanagawa Prefecture	Kanagawa
24	Korea	Jan. 20	Gyeonggi-do	Gimpo
32	Mexico	Feb. 28	Mexico City/Sinaloa	Mexico City

The ranking is based on total reported cases of the infection[1].

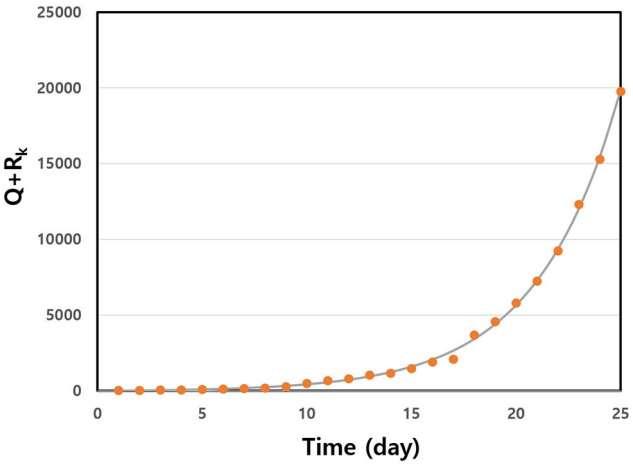


FIGURE 1 | $Q+R_K$ data were fitted to a non-linear least squares fit early in the outbreak as a function of time for Germany. The solid line is from fitted data, and the solid circles are the real data. We obtained fitting parameters $a = 9.62$ and $b = 0.25$.

death is occurring, it is quarantined indefinitely. From this idea we can predict some parameters from the officially provided data.

RESULTS

The outbreak of COVID-19 started around the world in January and February 2020, as summarized in **Table 1**. The disease was first reported in Wuhan City, Hubei province, China, on December 31, 2019. Some states, like the Republic of Korea, Taiwan, Hong Kong, etc., have controlled the disease well, up to now. They have executed massive inspections for the disease. When patients are found at a location, doctors and experts from the Korean Center for Disease Control and Prevention (KCDC) checked all people who had been in contact with those patients. All infected individuals were quarantined in hospitals or some remote places. Some persons suspected of infection would self-quarantine, and controllers checked on them frequently via cellphone app, the internet, and phone. However, many countries did not prepare to control and prevent the disease in the early stages, for example, the U.S.A. and Japan. Patients in these unprepared countries were incubating the disease in the early stages. Recently, these countries have suffered from abrupt outbreaks, and many people have died. We aggregate data set from worldometer website which is supported by the American Library Association[1].

In the reported data for each state, the active cases are transferred immediately to quarantined cases. Therefore, active cases correspond to quarantined cases, Q. Almost-recovered cases come from the isolation cases. From the reported data for Q and R_k, we can fit $(Q+R_k)$ as a function of time in the early stages of disease spread. $(Q+ R_k)$ is fitted by the exponential function $g(t) = \frac{a}{b}[e^{bt} - 1]$ in the early stages of disease spread (see the **Appendix**). From the obtained fitting parameters a and b, we estimated model parameters such as $a = \eta I(0)$ and $b = \beta - (\alpha + \eta)$, where $I(0)$ is the number of infected individuals at the outset. We have to determine four parameters: α, β, η, and $I(0)$. We determined rates α, η, and γ according to the method used in Bjornstad et al. [8].

Let ε denote the fraction of infectious individuals entering Q. There is a controversy over the ratio of asymptomatic cases for COVID-19 [17–22]. The reported rate of people testing positive for COVID-19, but being asymptomatic, in several instances ranged between 5 and 80%. We set the fraction as $\varepsilon = 1/3$ [15, 17–22]. The average incubation time is about 5 days [17, 18], and the duration for milder cases of the disease is about 5 to 6 days [19]. The average duration from infection to recovery or death in non-isolated patients is about 10 days, corresponding to a rate of 0.1/day [7]. Therefore, we obtained $\alpha = (1 - \varepsilon) \times 0.1/\text{day}$ and $\eta = \varepsilon \times 0.2/\text{day}$. Finally, we obtained $\alpha = \eta = 0.067/\text{day}$. Using these parameters and the fitting parameters a and b, we obtained parameters β and $I(0)$ from the fitting parameters and the predetermined rates. We summarize the results obtained from the data of each country.

We estimated infection rate β and the initial number of infected individuals, $I(0)$. The symptoms of COVID-19 do not appear in many cases. In **Figure 1**, we represent the non-linear least squares fit for $Q+R_K$ as a function of time in the early stages of disease spread in Germany. The early data fit the exponential function well. We give the fitted data for some selected countries in **Table 2**. We observed that there were large numbers of initially infected people. The infection rate shows a very high value in the range $0.233 \leq \beta \leq 0.462$ for the selected countries. We calculated the basic reproduction numbers of the estimated parameters for the countries. The R for many countries was >2. In particular, the basic reproduction number, R_o, for the U.S.A. shows a high value of $R_o = 3.45$. This high value induced the large number of infected people throughout many of the states in America.

TABLE 2 | Parameters of the $SIQR_K$ model.

Country	*a*	*b*	χ^2	β	I(0)	R_o	τ
U.S.A.	1.13	0.33	0.99	0.462	17	3.5	2.1
Spain	43.7	0.19	0.99	0.324	652	2.4	3.6
Italy	93.26	0.16	0.99	0.294	1,392	2.2	4.3
Germany	9.62	0.25	0.99	0.384	144	2.9	2.8
U.K.	6.6	0.21	0.99	0.344	99	2.6	3.3
China	563.38	0.23	0.93	0.364	8,409	2.7	4.3
Iran	138.46	0.10	0.98	0.233	2,077	1.8	6.9
Brazil	22.16	0.15	0.98	0.284	331	2.1	4.6
Canada	0.65	0.23	0.99	0.364	10	2.7	3.0
Japan	1.06	0.12	0.96	0.254	16	1.9	5.8
Korea	90.95	0.294	0.98	0.294	1,356	2.2	4.3
Mexico	12.84	0.12	0.99	0.254	192	1.9	5.8

We obtained fitting constants a and b. Then, we derived the infection rate, β, and the rate of detection, η, for new cases from the fitted values. We obtained recovery rate γ for quarantined individuals from the daily Q and R_K data. We obtained the basic reproduction number, R_o. Using fitting parameters, we numerically solved the $SIQR_K$ model and predicted the characteristic times.

We observed the high number of initially infected individuals, $I(0)$, from data fitting. In **Table 1**, we summarize the first official confirmed days for COVID-19 patients in many countries. Because of the incubation period and the asymptomatic cases in young, healthy people, we expect that there were many infected people when the health organizations of these countries reported their first cases. In China, we estimated $I(0) = 8,409$. The first confirmed time for the virus in China was a long time after the first case, because this is a new type of coronavirus. For the U.S.A., the number of initially infected people is a small value at $I(0) = 17$. In the U.S., the first patient was found in the state of Washington. However, late inspections and the delayed quarantine policy from the US Centers for Disease Control and Prevention (CDC) and the US federal government resulted in the huge outbreak in the USA. South Korea is one of the countries that excellently controlled this disease. In the early stages of the outbreak, the initial cases were estimated at $I(0) = 1,356$. In South Korea, a super-spreader was found in the metropolitan city of Daegu on Feb. 17, 2020, who attended worship services of a church gathering with a lot of people. Although the number of initially infected people is very big, the World Health Organization and the KCDC performed a wide range of inspections, imposed a strong policy for quarantines, and provided information on the people contacted by the confirmed patient. These strong protection policies have been preventing widespread infections of the disease up to now in South Korea.

We calculated the $SIQR_K$ model by using fitting parameters for the countries. **Figure 2** shows the predicted cases of susceptible (S), infecting (I), quarantined (Q), and individuals officially confirmed as accumulated recovered (R_k) for Germany. The maximum number of infected people was at around 50 days, when the government enforced quarantine on all infected persons. Of course, the maximum time and the lasting time of the disease depend on the fitting parameters and the number of initially infected people. For Germany, the disease subsides after 200 days. We need about 6 months to eradicate the disease, according to our model. We observed that the asymptomatic recovered population, $R_a = N - (S + I + Q + R_k)$, dramatically increased after the maximum time span of the infection, as shown in **Figure 2**. When we predict the evolution of the disease by some model, we need to use a confirmed data set, such as active cases, recovered cases, and terminal cases. In some cases, it is possible that data reported officially include any errors in the early phase of the outbreak. Then the prediction of the model also includes the uncertainty. However, in that case, our model can use observing the changing trend of the epidemic spreading within errors. When we give some variations of the initial conditions by the intrinsic errors in data, we can observe some varying patterns of the evolving disease.

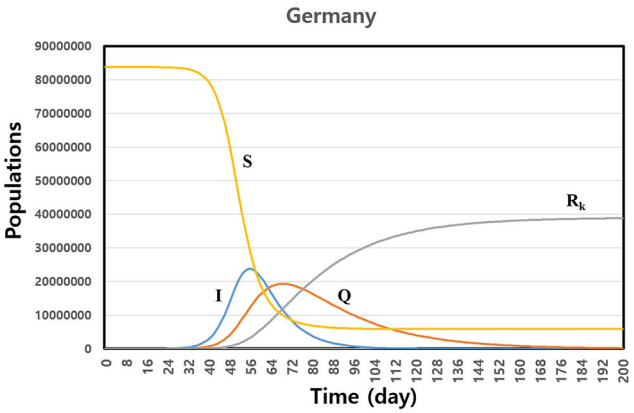


FIGURE 2 | Prediction of the susceptible (S), infecting (I), quarantined (Q), and confirmed accumulated recovered (R_k) individuals based on numerical integration using fitting parameters for Germany. The disease is lasting 200 days after first infecting a patient.

CONCLUSIONS

We consider a spreading epidemic model called the $SIQR_K$ model. In this model, we include a dynamic equation for quarantined individuals. We estimate the parameters of the dynamic evolution equation from the sum of quarantined cases and recovered cases. We obtained the parameters via non-linear least squares fit by using the set of reported data. When we consider the fitting of the model, it is important to use data for both types of confirmed cases, official and asymptomatic individuals in the model. In particular, there are officially reported recovered cases and asymptomatic recovered cases. We cannot know the asymptomatic recovered cases because we have no data. In this study we suggest a model to overcome this difficulty. The observed high value of the basic reproduction number indicates COVID-19 is a pandemic. We predict that the maximum time span of the infection is around 50 days to 2 months. The disease should last about 6 months when we quarantine infected individuals. We predict based on the model that the epidemics will last in some countries if the policy of the quarantine is not strict. In this model we don't include the number of deaths which are implicitly included in the quarantined cases. If we include the number of deaths, there are some mathematical difficulties to predict the parameters of the model. We will extend this model to include the death case in general.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.worldometers.info/coronavirus/.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (Grant No. NRF-2020R1A2C1005334).

REFERENCES

1. *Novel Coronavirus (2019-nCoV) Situation Report-1*. World Health Organization (2020).
2. *WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020* (2020). Available online at: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19
3. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals.* (2020) **134**:109761. doi: 10.1016/j.chaos.2020.109761
4. Peng P, Yang W, Zhang E, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv:2002.06563v1.* (2020) doi: 10.1101/2020.02.16.20023465
5. Crokidakis N. Data analysis and modeling of the evolution of COVID-19 in Brazil. *arXiv:2003.12150v1*. (2020).
6. Carcione JM, Santos J, Bagaini C, Ba J. A simulation of a COVID-19 epidemic based on a deterministic SEIR model. *arXiv:2004.035752v2.* (2020) doi: 10.1101/2020.04.20.20072272
7. Pedersen MG, Meneghini M. Quantifying undetected COVID-19 cases and effects of containment measures in Italy: predicting phase 2 dynamics. *[Preprint]*. (2020). doi: 10.13140/RG.2.2.11753.85600
8. Bjornstad ON, Shea K, Krzywinski M, Altman N. Modeling infectious epidemics. *Nature Methods.* (2020) **17**:453–6. doi: 10.1038/s41592-020-0822-z
9. Okell LC, Verity R, Watson OJ, Mishra S, Walker P, Whittaker C, et al. Have deaths from COVID-19 in Europe plateaued due to herd immunity? *Lancet.* (2020) **395**:e111. doi: 10.1016/S0140-6736(20)31357-X
10. Walker PGT, Wittaker C, Watson OJ, Beguelin M, Winskill P. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle- income countries. *Science.* (2020) **369**:413–22. doi: 10.1126/science.abc0035
11. Maclndtyre CR, Wang Q. Physical distance, face masks, and eye protection for prevention of COVID-19. *Lancet.* (2020) **27**:1950–1. doi: 10.1016/S0140-6736(20)31183-1
12. Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV2. *Science.* (2020) **368**:1422–4. doi: 10.1126/science.abc6197
13. Chu DK, Duda S, Solo K, Yaacoub S, Schunemann HJ. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS_CoV-2 and Covid-19: a systematic review and meta-analysis. *Lancet.* (2020) **395**:1973–87. doi: 10.1016/s0140-6736(20)31142-9
14. Tsallis C, Tirnakli U. Predicting COVID-19 peaks around the world. *Front Phys.* (2020) **8**:1–6. doi: 10.3389/fphy.2020.00217
15. Yang J, Wang G, Zhang S. Impact of household quarantine on SARS-Cov-2 infection in mainland China: a mean-field modelling approach. *Math Biosci Eng.* (2020) **17**:4500–12. doi: 10.3934/mbe.2020248
16. Zhou W, Wang A, Xia F, Xiao Y, Tang S. Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak. *Math Biosci Eng.* (2020) **17**:2693–707. doi: 10.3934/mbe.2020147
17. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Eng J Med.* (2020) **382**:1199–207. doi: 10.1056/NEJMoa2001316
18. 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**:577–582. doi: 10.7326/M20-0504
19. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA.* (2020) **323**:1406–7. doi: 10.1001/jama.2020.2565
20. Nishiura H, Kobayashi T, Suzuki A, Jung SM, Hayashi K, Kinoshita R, 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
21. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond princess cruise ship, Yokhama, Japan. *Euro Surveill.* (2020) **25**:1–5. doi: 10.2807/1560-7917.ES.2020.25.10.2000180
22. Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ.* (2020) **368**:m11165. doi: 10.1136/bmj.m1165

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.

APPENDIX

Let us examine the $SIQR_K$ model. In the early phase of disease spread, we expect that the susceptible population is similar to the total population, $S/N \approx 1$. Therefore, we can write a dynamic infection equation as found in Crokidakis [5] and Carcione et al. [6]:

$$\frac{dI(t)}{dt} = [\beta - (\alpha + \eta)]\, I(t) \tag{A1}$$

By integrating this equation with initial condition $I(0)$, we obtain the solution

$$I(t) = I(0)\, e^{[\beta-(\alpha+\eta)]t} \tag{A2}$$

The reproduction number, R_o, is given by

$$R_o = \frac{\beta}{\alpha + \eta} \tag{A3}$$

With COVID-19, the reproduction number is >1. The disease can spread easily through contact between individuals. The doubling time, τ, is given by $\tau = \frac{\ln 2}{[\beta-(\alpha+\eta)]} = \frac{\ln 2}{(\alpha+\eta)(R_o-1)}$. Infection rate β and the rate of detection of new cases, η, can be derived from the evolution time after early infection. Adding equations (3) and (4), we obtained a quantity such as

$$\frac{d(Q + R_k)}{dt} = \eta I(t) \tag{A4}$$

Therefore, we obtain the sum of quarantined cases and recovered cases as follows:

$$(Q + R_k)(t) = \frac{\eta I(0)}{\beta - (\alpha + \eta)} \left\{ e^{[\beta-(\alpha+\eta)]t} - 1 \right\} \tag{A5}$$

We calculated recovery rate γ obtained from the data set. The recovery rate is given by $\gamma = (R_{ki} - R_{ki-1})/Q_{i-1}$. The value of the recovery rate depends on time in the early stages, and converges to a constant value. We obtained a recovery rate of $\gamma = 0.036$/day.

frontiers in Physics

ORIGINAL RESEARCH
published: 01 September 2020
doi: 10.3389/fphy.2020.00336

Estimation of Local Novel Coronavirus (COVID-19) Cases in Wuhan, China from Off-Site Reported Cases and Population Flow Data from Different Sources

Zian Zhuang[1], Peihua Cao[2], Shi Zhao[3,4], Yijun Lou[1], Shu Yang[5], Weiming Wang[6], Lin Yang[7]* and Daihai He[1]**

[1] Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China, [2] Clinical Research Center, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [3] JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China, [4] Shenzhen Research Institute of Chinese University of Hong Kong, Shenzhen, China, [5] College of Medical Information Engineering, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [6] School of Mathematics and Statistics, Huaiyin Normal University, Huai'an, China, [7] School of Nursing, Hong Kong Polytechnic University, Hong Kong, China

OPEN ACCESS

Edited by:
Aristides (Aris) Moustakas,
Natural History Museum of Crete,
University of Crete, Greece

Reviewed by:
Ling Xue,
Harbin Engineering University, China
Martin Kröger,
ETH Zürich, Switzerland

***Correspondence:**
Weiming Wang
weimingwang2003@163.com
Lin Yang
l.yang@polyu.edu.hk
Daihai He
daihai.he@polyu.edu.hk

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: 16 May 2020
Accepted: 20 July 2020
Published: 01 September 2020

Citation:
Zhuang Z, Cao P, Zhao S, Lou Y, Yang S, Wang W, Yang L and He D (2020) Estimation of Local Novel Coronavirus (COVID-19) Cases in Wuhan, China from Off-Site Reported Cases and Population Flow Data from Different Sources. Front. Phys. 8:336. doi: 10.3389/fphy.2020.00336

In December 2019, novel coronavirus disease (COVID-19) hit Wuhan, Hubei Province, China and spread to the rest of China and overseas. The emergence of this virus coincided with the Spring Festival Travel Rush in China. It is possible to estimate the total number of COVID-19 cases in Wuhan, by 23 January 2020, given the cases reported in other cities/regions and population flow data between Wuhan and these cities/regions. We built a model to estimate the total number of COVID-19 cases in Wuhan by 23 January 2020, based on the number of cases detected outside Wuhan city in China, with the assumption that cases exported from Wuhan were less likely underreported in other cities/regions. We employed population flow data from different sources between Wuhan and other cities/regions by 23 January 2020. The number of total cases in Wuhan was determined by the maximum log likelihood estimation and Akaike Information Criterion (AIC) weight. We estimated 8 679 (95% CI: 7 701, 9 732) as total COVID-19 cases in Wuhan by 23 January 2020, based on combined source of data from Tencent and Baidu. Sources of population flow data impact the estimates of the total number of COVID-19 cases in Wuhan before city lockdown. We should make a comprehensive analysis based on different sources of data to overcome the bias from different sources.

Keywords: COVID-19, mobility, pneumonia, transportation, outbreaks

INTRODUCTION

In December 2019, a cluster of patients with pneumonia of unknown causes was reported in Wuhan, Hubei Province, China [1]. On 9 January 2020, a novel coronavirus, named SARS-CoV-2, was identified as the cause of this outbreak [2]. The emergence of this virus coincided with the Spring Festival Travel Rush in China. It was estimated that there would be around 3 billion trips made in China during the period of 10 January to 18 February 2020 [3]. Some researchers have pointed out the risk of the regional and global disease spreading during the Spring Festival

Travel Rush [4]. However, due to the small number of severe cases reported by mid-January and most cases were linked to the Huanan Seafood Market of Wuhan city, neither international nor regional traveling restrictions were implemented to Wuhan at the early stage of this outbreak. On 13 January 2020, the first case exported from Wuhan was reported in Thailand and the case numbers dramatically increased after the diagnostic kits became available in mid-January. As of 13 July 2020, there were 85 560 laboratory confirmed cases and 4 648 deaths (58.8 and 83.2% in Wuhan) [5]. In recognition of a wide-spreading outbreak, the government has suspended all public transportations inside Wuhan city since 23 January 2020, and some regional traveling restrictions were also implemented by other cities/regions [6].

METHODS

Objective

At the early stage of this outbreak, the cases might have been severely underreported due to the lack of diagnostic kits and insufficient screening for all suspected cases [7, 8]. Several efforts have been made to estimate the COVID-19 case numbers in Wuhan using different modeling approaches, and the estimates range from 4 000 to 75 815 during the period of 18–29 January [8–10].

In this study, we aimed to estimate the number of COVID-19 cases in Wuhan, based on the cases exported from Wuhan to other cities/regions in mainland China and different sources of the population flow data between Wuhan and these cities/regions. We tested the impact of different sources of population flow data on estimating cases in Wuhan before city lockdown and combined different sources of data to overcome the bias from different sources. The estimates were made by 23 January 2020 (before the suspension of public transportations in Wuhan). We assumed that the cases exported from Wuhan were less likely underreported in other cities/regions in mainland China, as stringent temperature screening was implemented at airports and railway stations.

Data

We obtained daily number of inbound and outbound domestic passengers traveling by air, train or road to/from Wuhan from two data sources:

(1) Tencent's LBS (location-based services) database (see: https://heat.qq.com/). According to location data of Tencent's mobile software users, population flow number during 10 December 2016 and 24 January 2017 was generated, between Wuhan and 24 cities/regions in China (Anhui, Beijing, Chongqing, Fujian, Gansu, Guangdong, Guangxi, Guizhou, Hainan, Hebei, Heilongjiang, Henan, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Ningxia, Shandong, Shanghai, Sichuan, Tianjin, Yunnan, Zhejiang). We assumed that the amount of population flow in 2017 is same as that in 2020.

(2) Baidu map database (see: https://qianxi.baidu.com/). According to location data of Baidu's mobile software users, population flow number from 1 to 20 January 2020 was generated, between Wuhan and 26 cities/regions (Shanxi, Shaanxi, and other cities/regions are the same as Tencent data).

We equally divided the population flow data from different sources separately to get average daily population flow number. **Figure 1** shows the geographical location of cities/regions which reported COVID-19 cases and the number of COVID-19 cases in each city/region.

As shown in **Table S1**, we collected total numbers of reported COVID-19 cases exported from Wuhan to other cities in China by 23 January 2020, and cases which were not exported from Wuhan (e.g., family or hospital clusters) were excluded from the analysis [11]. Thirteen cases were excluded due to the lack of traveling history to Hubei before illness onset. As for 161 cases that not specified the traveling history, we assumed that the probability of a single case being an exported case is θ, and each case is independent of each other. Then all of these unspecified cases follow a binomial distribution. As shown in Equation (1), θ represents the probability that a case is exported from Wuhan and n means the number of COVID-19 cases that not specified the traveling history, which is 161. P represents the probability that k out of n COVID-19 cases came from Wuhan. Since the most cases detected outside Wuhan are exported cases from Wuhan, by 23 January 2020 [12], we assumed that the probability of a case to be an exported case from Wuhan is based on a different level of θ (1, 0.9, 0.8). Then we obtained the expectation number of cases exported from Wuhan in city/region i.

$$P\{X = k\} = \binom{n}{k}\theta^k(1-\theta)^{n-k} \tag{1}$$

Models

Table S3 presents the general process of the whole method. The total number of COVID-19 cases exported from Wuhan and diagnosed in each city/region outside Wuhan by 23 January 2020 was assumed to follow a Binomial distribution [8], as in Equation (2), where λ is the total number of cases infected in Wuhan by 23 January and p_i is the probability of detecting any exported cases from Wuhan in city/region i outside Wuhan in China.

$$\begin{aligned}\text{Number of cases exported from Wuhan and detected in city/region } i \\ = Binomial(\lambda, p_i)\end{aligned} \tag{2}$$

The probability p_i can be derived from dividing daily outbound passengers of Wuhan to city/region i by the population size that the Wuhan airport, railway and road serves and multiplying by the mean time for patients from being infected to being detected, see Equation (3).

$$\begin{aligned}p_i = &\frac{\text{Daily outbound travellers from Wuhan to city/region } i}{\text{Catchement population of Wuhan airport}} \\ &\times \text{Mean time from infection to detection}\end{aligned} \tag{3}$$

Then, we used cases exported from Wuhan to estimate the total number of COVID-19 cases infected in Wuhan (λ). Based on the data obtained from each city/region, we obtained the λ by maximum likelihood estimation.

$$l(\lambda) = \log L(\lambda) = \sum_{i=1}^{k} \log f(n_i, p_i; \lambda) \tag{4}$$

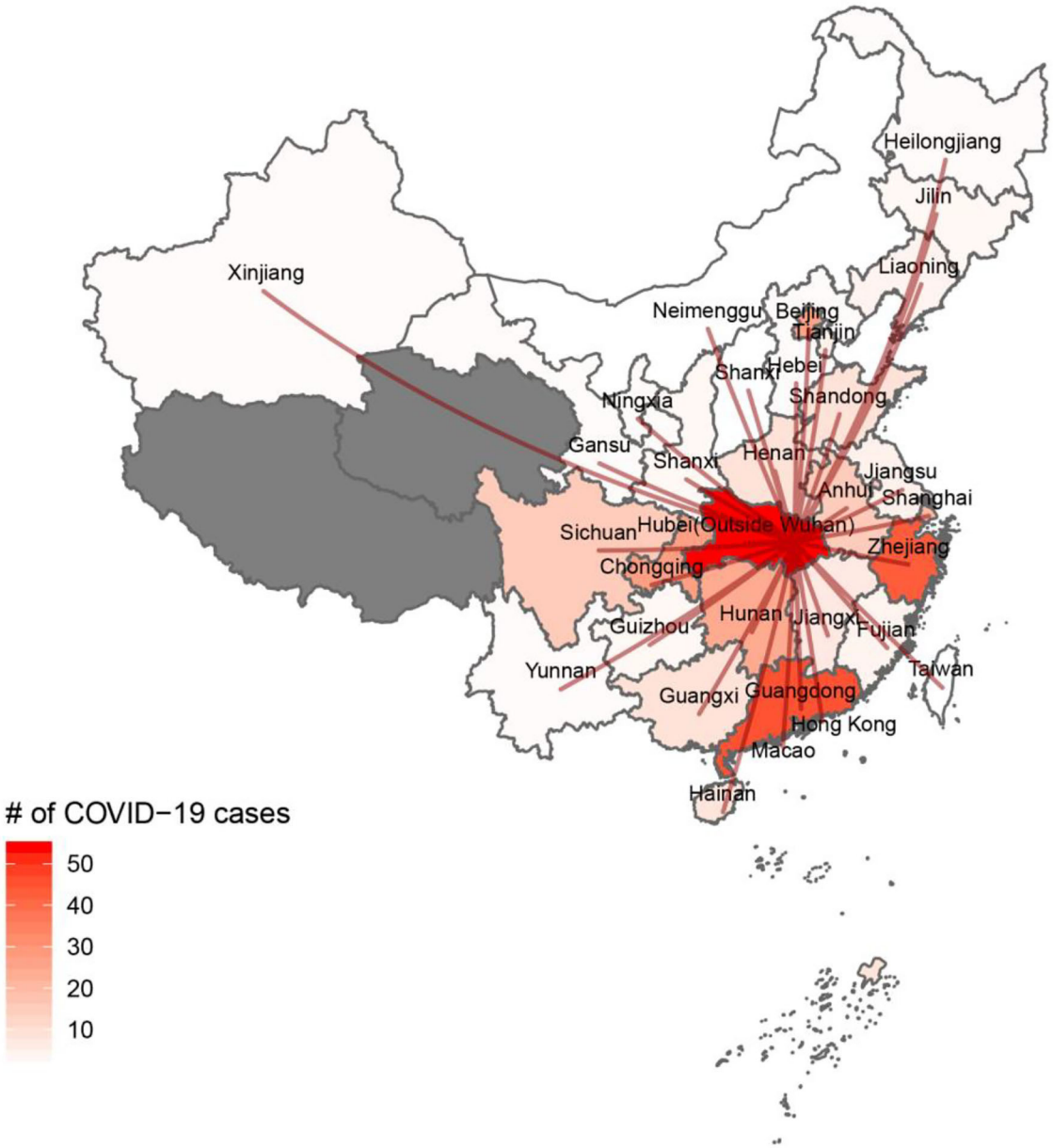


FIGURE 1 | The geographical distribution of exported COVID-19 cases in China. This figure reported number of reported COVID-19 cases in China, the dark gray area indicates the regions with zero COVID-19 cases as of 23 January 2020. Red paths show routes from Wuhan to other cities/regions.

In Equation (4), $l(\cdot)$ and $L(\cdot)$ are the total log-likelihood and the total likelihood, respectively. $f(\cdot)$ is the function for computing the value of the probability density function of the binomial distribution (Equation 2). The k represents the total estimated number of cities/regions. The n_i represents the number of cases exported from Wuhan and detected in city/region i, and p_i means the probability of finding any exported cases from Wuhan in in city/region i. The 95% confidence intervals (95% CI) of log-likelihood, l, can be calculated after obtaining λ, since residuals of log maximum likelihood estimation follow Chi-square distribution [13]. Then we can extrapolate a 95% CI about the total number of COVID-19 cases infected in Wuhan. As in Equation (5), by deducting the number of exported cases from the total number of cases infected in Wuhan, we got the final estimate of the total number of COVID-19 cases in Wuhan as of 23 January 2020.

$$\begin{aligned}\text{Cases in Wuhan} = {} & \text{Cases infected in Wuhan} \\ & - \text{Cases exported from Wuhan} \end{aligned} \quad (5)$$

We assumed a population of 19 million (catchment population) traveling through the airport, railway stations and highways in

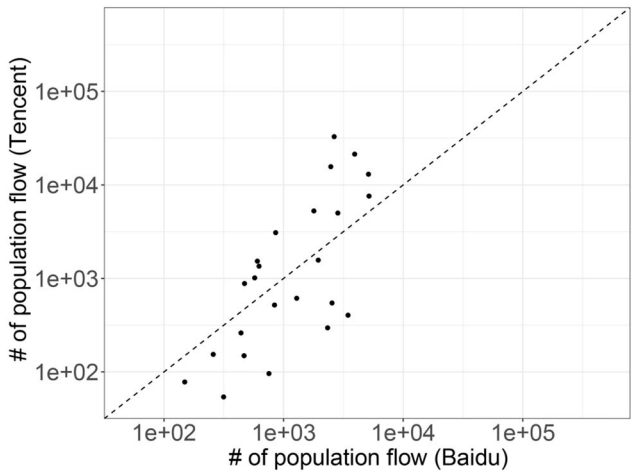


FIGURE 2 | Comparison between the number of population flow data from Baidu and Tencent. Y-axis presents the number of population flow between Wuhan and other cities/regions from Tencent data. X-axis presents the number of population flow between Wuhan and other cities/regions from Baidu data.

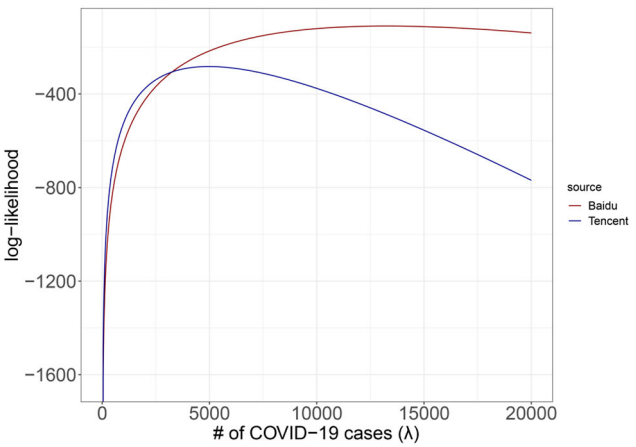


FIGURE 3 | Log maximum likelihood estimation for λ, based on data sourced from Baidu and Tencent.

Wuhan, and a 10-days delay on average, which accounted for the time interval reported between infection timing and case timing [8]. Since exported cases were much less than those in Wuhan as of 23 January 2020, it was assumed that all cases in other cities/regions outside Wuhan are detected. If cases in other cities/regions are missed, our estimate would underestimate the acute number of cases in Wuhan. In addition, we assumed that all of the passengers outflowed from Wuhan were equally likely to be infected, regardless of transfer passengers or local residents, as passengers may be a high risk of infection while traveling in trains and airplanes cabins.

To overcome the bias from different sources of data, we first evaluated the correlation between two datasets to determine whether there is an apparent inconsistency or discrepancy between different sources of data. We found that the Spearman's rank correlation coefficient of Baidu and Tencent data for the same 24 cities/regions is 0.75, which means that two sources of data are correlated under 99.99% confidential level. We assumed a linear relationship between the Baidu data and Tencent data (see **Figure 2**). For all observations, we assumed error terms are independent of each other. We also assumed all error terms follow a normal distribution and have the same variance. We then built the linear model (Equation 6) and tested the null hypothesis H_0 that $\alpha = 0$. In Equation (6), N_{Baidu} and $N_{Tencent}$ represent the number of population flow data from Baidu and Tencent.

$$N_{\text{Baidu}} = \alpha N_{\text{Tencent}} + \beta \tag{6}$$

We got the result that estimated coefficient α equals 0.10, β equals 1 272 and P-value for F-test was <0.01. Then we rejected the null hypothesis H_0 under 99% CI, which suggests that two sources of data are likely to have a linear relation. Since both sets of data is likely to be reasonable. We then applied Akaike Information Criterion (AIC) [14] to test the fitting result of number of cases exported from Wuhan and detected in city/region i, $n_{est<uscore>i}$, which follows a binomial distribution (Equation 2), based on Baidu and Tencent data, see Equation (7). To estimate the number of cases exported from Wuhan, the model used estimated the total number of COVID-19 cases infected in Wuhan (λ) from Equation (4). p_i is the probability that we will find any exported cases from Wuhan in city/region i outside Wuhan in China, which we have already obtained from Equation (3). Please note that we only included 24 cities/regions of which both Baidu and Tencent have population flow data in AIC weight calculation.

$$n_{\text{est\_}i} = p_i \lambda \tag{7}$$

Since Baidu and Tencent data show significant linear relationship, which confirmed with each other that the general pattern of data is rational, we weighted (Equation 8) and combined (Equation 9) the estimated number of cases from Baidu and Tencent based on AIC value to obtain the final estimate.

$$\mathrm{W}_s = 1 - \frac{\mathrm{AIC}_s}{\mathrm{AIC}_{\text{Baidu}} + \mathrm{AIC}_{\text{Tencent}}} \tag{8}$$

$$\lambda = \sum_{i=1}^{2} \mathrm{W}_s \lambda_s \tag{9}$$

In Equations (8) and (9), W_s and AIC_s represents the weight of estimated number of COVID-19 cases infected in Wuhan and AIC value for source s, respectively. λ_s is the estimate of the total number of cases from Equation (4), based on source s, and λ means the final estimate of total number of cases infected in Wuhan by 23 January 2020.

RESULT

Based on the data sourced from Tencent and Baidu, we estimated the total number of cases in Wuhan, λ (**Figure 3**). Then we estimated the 95% CI of the total number of COVID-19 cases. We estimated 4 672 (4 129, 5 257) and

TABLE 1 | AIC value and calculated Weight of final estimate for different sources of population data (under different assumptions of, θ, probability of an unspecified case reported in other cities/regions being an exported case from Wuhan).

	Tencent data			Baidu data		
	($\theta = 1$)	($\theta = 0.9$)	($\theta = 0.8$)	($\theta = 1$)	($\theta = 0.9$)	($\theta = 0.8$)
AIC value	137.8	137.5	136.6	147.1	146.7	144.6
Weight	51.6%	51.6%	51.4%	48.4%	48.4%	48.6%

TABLE 2 | Summary table of estimated total number of cases infected in Wuhan (including cases exported from Wuhan to other cities/regions) and number of cases in Wuhan (excluding cases exported from Wuhan to other cities/regions) by 23 January 2020, from different sources of data.

	Tencent Data			Baidu Data			Combined Data		
	Baseline ($\theta = 1$)	($\theta = 0.9$)	($\theta = 0.8$)	Baseline ($\theta = 1$)	($\theta = 0.9$)	($\theta = 0.8$)	Baseline ($\theta = 1$)	($\theta = 0.9$)	($\theta = 0.8$)
Estimated total number of cases infected in Wuhan (95% CI)	4 969 (4 426, 5 554)	4 819 (4 284, 5 395)	4 635 (4 111, 5 201)	13 251 (11 811, 14 803)	12 855 (11 437, 14 384)	12 371 (10 981, 13 872)	8 977 (8 000, 10 031)	8 708 (7 746, 9 746)	8 395 (7 450, 9 415)
Estimated number of cases in Wuhan by 23 January 2020 (95% CI)	4 672 (4 129, 5 257)	4 531 (3 996, 5 107)	4 358 (3 834, 4 924)	12 950 (11 510, 14 502)	12 563 (11 145, 14 092)	12 090 (10 700, 13 592)	8 679 (7 701, 9 732)	8 418 (7 456, 9 456)	8 116 (7 171, 9 137)

θ *represents the probability of an unspecified case reported in other cities/regions being an exported case from Wuhan.*

12 950 (11 510, 14 502) as total cases in Wuhan by 23 January 2020, based on Tencent and Baidu population flow data. In addition, based on the AIC weighting (**Table 1**), we combined results from Baidu and Tencent and estimated 8 679 (7 701, 9 732) as total cases in Wuhan. **Table 2** presents the estimates under different sources and different level of probability of an unspecified case reported in other cities/regions being an exported case from Wuhan. **Table S2** present the probability of finding any cases for each city/region outside Wuhan.

DISCUSSION

A recent study by Imai et al. estimated that a total of 4 000 (95% CI: 1 000–9 700) cases on 18 January 2020 [8]. Compared with the total number of confirmed cases provided by the government as of 23 January 2020, which is 495 [15], Imai et al. obtained around 8-fold of cases before 23 January [8]. This is partly because the screening effort targeting population from Wuhan in other cities is much more effective than the local screening effort in Wuhan due to the worsening situation. Estimates based on combination of Baidu data and Tencent data provided closer result by Imai et al.'s [8] than the official report [15]. Our model is mostly like Imai et al.'s [8]. The difference between the two models is that we estimated the total number of cases in Wuhan based on separated data of each city/region in China. In addition, we applied maximum likelihood estimation by calculating the log-likelihood value for each city/region. Imai et al. [8] obtained the estimate based on overall overseas data, applying maximum likelihood estimation by calculating the simple ratio. In the sensitivity analysis, **Table 2** shows that when the probability of an unspecified case reported in other cities/regions being an exported case from Wuhan is close to 1, slight fluctuations of the probability will have little impact on the estimation.

Estimates of the population outflow provided by Baidu and Tencent show substantial fluctuation, leading to results with significant differences. We found that Baidu and Tencent data show significant linear relation, which means that pattern of two sources of data is largely consistent. One possible reason for the phenomenon is that different institutions have a various definition of the number of people flow from one city to another. Methods include people who travel to other cities through Wuhan in the population flow may provide a much more significant figure than those that only calculate people who originally depart from Wuhan. At the same time, multiple round trips may also affect the count. Another possible reason is that Baidu and Tencent would fail to track the whole amount of population flow since not everyone uses mobile phone software from Baidu and Tencent.

Imai et al. suggested that by further improving the definition and testing of COVID-19 cases, and further expanding the scope of epidemic monitoring, the gap between the estimated number and official reported cases would be further narrowed. According to our results, statistics of population flow also play significant roles in estimation. At present, many researches use data from Baidu and Tencent platforms [10, 16–18]. Among them, Tian et al. [18] referred to different data sources to gain a more comprehensive measure of movement volume. According to the data presented in the article, Tian et al. [18] integrated the population flow data first, and then conducted relevant calculation and analysis. In this paper, we calculated estimates from each data source first and then weighted the results. Both of methods provide more reasonable results ranged between conclusions that generated by either data sources, contributing to overcome the bias from different sources.

CONCLUSIONS

Different sources of population flow data impact the estimates of the total number of COVID-19 cases in Wuhan before city lockdown. We built a model that could be reproduced to employ incompatible sets of population flow data to estimate the number of COVID-19 cases more reasonably. We estimated 8 679 (95% CI: 7 701, 9 732) as total COVID-19 cases in Wuhan by 23 January 2020, based on the combined source of data from Tencent and Baidu. What data source can be used to make the most reliable estimation is not clear yet, though estimates based on a single source of data are likely to be biased. A comprehensive analysis based on different statistics is need before we reach any conclusions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/Larryzza/COVID-19.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

DH was supported by General Research Fund (15205119) of Research Grants Council of Hong Kong and Alibaba (China) Co. Ltd. Collaborative Research grant. WW was supported by National Natural Science Foundation of China (Grant Number 61672013) and Huaian Key Laboratory for Infectious Diseases Control and Prevention (Grant Number HAP201704), Huaian, Jiangsu, China. PC was supported by National Natural Science Foundation of China (Grant Number 81903406).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00336/full#supplementary-material

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* (2020) **382**:727–33. doi: 10.1056/NEJMoa2001017
2. *WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China*. World Health Organization (WHO). Available online at: https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china (accessed January 9, 2020)
3. *In 2020, the National Passenger Volume of Spring Festival Transportation Will Reach About 3 Billion Person Times*. xinhuanet (in Chinese). Available online at: http://www.xinhuanet.com/2019-12/18/c_1125362460.htm (accessed December 28, 2019).
4. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K. Pneumonia of unknown etiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med.* (2020) 27:taaa008. doi: 10.1093/jtm/taaa008
5. *Real Time Epidemic Data.* Dingxiang doctor (in Chinese). Available online at: https://3g.dxy.cn/newh5/view/pneumonia (accessed July 13, 2020).
6. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. (2020) **395**:470-473. doi: 10.1016/S0140-6736(20)30185-9
7. Zhao S, Musa SS, Lin Q, Ran J, Yang G, Wang W, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in china in the first half of january 2020: a data-driven modelling analysis of the early outbreak. *J Clin Med.* (2020) **9**:388. doi: 10.3390/jcm9020388
8. Imai N, Dorigatti I, Cori A, Riley S, Ferguson NM. *Report 2: Estimating the Potential Total Number of Novel Coronavirus (2019-nCoV) Cases in Wuhan City, China. Imperial College London*. (2020). Available online at: https://spiral.imperial.ac.uk/bitstream/10044/1/77150/12/2020-01-22-COVID19-Report-2.pdf
9. Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, et al. The rate of underascertainment of novel coronavirus (2019-nCoV) infection: estimation using japanese passengers data on evacuation flights. *J Clin Med.* (2020) 9:419. doi: 10.3390/jcm9020419
10. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet.* (2020) **395**:689–97. doi: 10.1016/S0140-6736(20)30260-9
11. *Situation report of the Pneumonia Cases Caused by the Novel Coronavirus.* National Health Commission of Each Province of People's Republic of China (in Chinese). Available online at: http://www.nhc.gov.cn/ (accessed 23 January 2020).
12. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, et al. Risk for transportation of coronavirus disease from wuhan to other cities in China. *Emerg Infect Dis.* (2020) **26**:1049–52. doi: 10.3201/eid2605.200146
13. Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat.* (1938) **9**:60–2. doi: 10.1214/aoms/1177732360
14. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G, editors. *Selected Papers of Hirotugu Akaike.* New York, NY: Springer New York. (1998). p. 199–213. doi: 10.1007/978-1-4612-1694-0_15
15. *Report of Hubei Provincial Health Committee on Pneumonia Caused by New Coronavirus.* Health Commission of Hubei Province (in Chinese). Available online at: http://wjw.hubei.gov.cn/fbjd/dtyw/202001/t20200124_2014626.shtml] (accessed January 24, 2020).
16. Ai S, Zhu G, Tian F, Li H, Gao Y, Wu Y, et al. Population movement, city closure and spatial transmission of the 2019-nCoV infection in China. *medRxiv*. (2020). doi: 10.1101/2020.02.04.20020339. [Epub ahead of print].
17. Jin G, Yu J, Han L, Duan S. The impact of traffic isolation in Wuhan on the spread of 2019-nCov. *medRxiv*. (2020). doi: 10.1101/2020.02.04.20020438. [Epub ahead of print].
18. Tian H, Liu Y, Li Y, Wu C-H, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science.* (2020) **368**:638. doi: 10.1126/science.abb6105

Conflict of Interest: DH received a grant from Alibaba (China) Co. Ltd., Collaborative Research grant.

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

frontiers in Physics

ORIGINAL RESEARCH
published: 17 September 2020
doi: 10.3389/fphy.2020.00347

Estimating the Serial Interval of the Novel Coronavirus Disease (COVID-19): A Statistical Analysis Using the Public Data in Hong Kong From January 16 to February 15, 2020

Shi Zhao[1,2], Daozhou Gao[3], Zian Zhuang[4], Marc K. C. Chong[1,2], Yongli Cai[5], Jinjun Ran[6], Peihua Cao[7], Kai Wang[8], Yijun Lou[4], Weiming Wang[5]*, Lin Yang[9], Daihai He[4]* and Maggie H. Wang[1,2]*

[1] JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China, [2] CUHK Shenzhen Research Institute, Shenzhen, China, [3] Department of Mathematics, Shanghai Normal University, Shanghai, China, [4] Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China, [5] School of Mathematics and Statistics, Huaiyin Normal University, Huaian, China, [6] School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China, [7] Clinical Research Centre, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [8] Department of Medical Engineering and Technology, Xinjiang Medical University, Ürümqi, China, [9] School of Nursing, Hong Kong Polytechnic University, Hong Kong, China

Background: The emerging virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has caused a large outbreak of novel coronavirus disease (COVID-19) since the end of 2019. As of February 15, there were 56 COVID-19 cases confirmed in Hong Kong since the first case with symptom onset on January 23, 2020.

Methods: Based on the publicly available surveillance data in Hong Kong, we identified 21 transmission events as of February 15, 2020. An interval censored likelihood framework is adopted to fit three different distributions including Gamma, Weibull, and lognormal, that govern the serial interval (SI) of COVID-19. We selected the distribution according to the Akaike information criterion corrected for small sample size (AICc).

Findings: We found the lognormal distribution performed slightly better than the other two distributions in terms of the AICc. Assuming a lognormal distribution model, we estimated the mean of SI at 4.9 days (95% CI: 3.6–6.2) and SD of SI at 4.4 days (95% CI: 2.9–8.3) by using the information of all 21 transmission events.

Conclusion: The SI of COVID-19 may be shorter than the preliminary estimates in previous works. Given the likelihood that SI could be shorter than the incubation period, pre-symptomatic transmission may occur, and extra efforts on timely contact tracing and quarantine are crucially needed in combating the COVID-19 outbreak.

Keywords: COVID-19, serial interval, statistical analysis, Hong Kong, contact tracing

OPEN ACCESS

Edited by:
Lin Wang,
University of Cambridge,
United Kingdom

Reviewed by:
Zhanwei Du,
University of Texas at Austin,
United States
Sheikh Taslim Ali,
The University of Hong Kong,
Hong Kong
Haiyan Wang,
Arizona State University, United States

****Correspondence:***
Shi Zhao
zhaoshi.cmsa@gmail.com
Weiming Wang
weimingwang2003@163.com
Daihai He
daihai.he@polyu.edu.hk

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *04 June 2020*
Accepted: *23 July 2020*
Published: *17 September 2020*

Citation:
Zhao S, Gao D, Zhuang Z, Chong MKC, Cai Y, Ran J, Cao P, Wang K, Lou Y, Wang W, Yang L, He D and Wang MH (2020) Estimating the Serial Interval of the Novel Coronavirus Disease (COVID-19): A Statistical Analysis Using the Public Data in Hong Kong From January 16 to February 15, 2020. Front. Phys. 8:347. doi: 10.3389/fphy.2020.00347

INTRODUCTION

The coronavirus disease 2019 (COVID-19) is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, formerly known as the "2019-nCoV"), which has emerged at the end of 2019 [1–5]. COVID-19 cases were soon exported to many Chinese cities and overseas [6], and the travel-related risk of disease spread was suggested in previous studies [4, 7–9]. The risks of rapid spreading were evaluated based on the early surveillance data and also compared to other previous respiratory infectious diseases [5, 10–14]. Since the first confirmed imported case in Hong Kong on January 23 [15], the local government has implemented a series of control and prevention measures for COVID-19, including enhanced border screening and traffic restrictions [16, 17].

The COVID-19 pandemic has affected most of the regions around the world, including places with less developed healthcare systems. Hong Kong was the hardest hit region in the severe acute respiratory syndrome (SARS) outbreaks in 2003 [18, 19], and thus it is expected to be more prepared in mitigation of emerging infectious disease outbreaks [20]. The lesson in Hong Kong shall be an example for other regions, in particular those less developed places with poor settings [21–24]. As of February 15, there were 56 COVID-19 cases confirmed in Hong Kong [16], and local transmission was also recognized by the contact tracing investigation. Given the risk of human-to-human transmission, the serial interval (SI), which refers to the time interval from illness onset in a primary case (i.e., infector) to that in a secondary case (i.e., infectee) [25–28], was of interest to the iterative rate of transmission generations of COVID-19. SI could be used to assist strategic decision-making of public health policies and construct analytical frameworks for studying the transmission dynamics of SARS-CoV-2.

In this study, we examined the publicly available materials released by the Center for Health Protection (CHP) of Hong Kong. Adopting the case-ascertained design [29], we identified the transmission chain from index cases to secondary cases. We estimated the SI of COVID-19 based on 21 identified transmission chains from the surveillance data and contact tracing data in Hong Kong.

DATA AND METHODS

As of February 15, there were 56 confirmed COVID-19 cases in Hong Kong [16], which followed the case definition in official diagnostic protocol released by the World Health Organization (WHO) [30]. To identify the pairs of infector (i.e., primary case) and infectee (i.e., secondary case), we scanned all news press released by the CHP of Hong Kong between January 16 and February 15, 2020 [17]. The exact symptom onset dates of all individual patients were released by CHP [16], which were publicly available, and used to match each transmission chain. For those infectees associated with multiple infectors, we recorded the range of onset dates of all associated infectors, i.e., lower and upper bounds. With all publicly available information from CHP, we constructed the transmission events by subjectively screening the exposure link between consecutive COVID-19 infections. We identified 21 transmission events, including 12 infectees matched with only one infector, that were used for SI estimation. Note that all the 21 transmission events occurred in Hong Kong, and most of the cases involved Hong Kong residents.

Following previous studies [25], we adopted a distribution function with mean μ and standard deviation (SD) σ, denoted by $g(|\mu, \sigma)$, to govern the distribution of SI. We defined $g(|\mu, \sigma)$ as three different distributions; Gamma, Weibull, and lognormal distribution. The interval censored likelihood [31], denoted by L_0, of SI estimates is defined in Equation (1). It happens in the practical analyses of serial interval (as well as incubation period), observations are typically integer while the population mean can be a real value.

$$L_0\left(\mu, \sigma|\tau, T^{\text{low}}, T^{\text{up}}\right) = \prod_i \left[\int_{T_i^{\text{low}}}^{T_i^{\text{up}}} h\left(t|T_i^{\text{low}}, T_i^{\text{up}}\right) g(\tau_i - t|\mu, \sigma)\mathrm{d}t\right] \quad (1)$$

The $h(\cdot)$ was the probability density function (PDF) of exposure following a uniform distribution with a range from T^{low} to T^{up}. The terms T_i^{low} and T_i^{up} denoted the lower and upper bounds, respectively, for the range of onset dates of multiple infectors linked to the i-th infectee. The τ_i was the observed onset date of the i-th infectee. Hence, the likelihood function in Equation (1) can be interpreted as the probability of the SI being observed with uncertain onset dates of infectors but a fixed onset date of the infectee [25, 31]. We calculated the maximum likelihood estimates of μ and σ. Their 95% confidence interval (95% CI) were calculated by using the profile likelihood estimation framework with a cutoff threshold determined by a Chi-square quantile [32]. We select the distribution of $g(\cdot|\mu, \sigma)$ according to the Akaike information criterion corrected for small sample size, denoted by AICc. We employed both Pearson's correlation and coefficient of determination, i.e., R-squared, to measure the goodness-of-fit of the selected model.

In addition, as pointed out in [31], it was possible that the naive likelihood in Equation (1) underestimated the SI due to sampling biases. Hence, we adjusted for the right truncation observation bias due to isolation by using an alternative likelihood function, L, in Equation (2), which is based on the non-truncated version in Equation (1). The truncation scheme in Nishiura et al. [31], as well as adopted in Kwok et al. [21], is relying on prior knowledge of an additional parameter, i.e., the intrinsic growth rate of the epidemic, which is commonly assumed and fixed in the likelihood framework. The truncation scheme adopted in this work was previously discussed in Zhao [33], which considers both likelihood of occurrence and likelihood of being observed subjected to the implementation of isolation.

$$L\left(\mu, \sigma|\tau, T^{\text{low}}, T^{\text{up}}\right) = \prod_i \frac{\left[\int_{T_i^{\text{low}}}^{T_i^{\text{up}}} h\left(t|T_i^{\text{low}}, T_i^{\text{up}}\right) g(\tau_i - t|\mu, \sigma)\mathrm{d}t\right]}{G(d_i - T_i^{low}|\mu, \sigma)} \quad (2)$$

Here, the $G(\cdot)$ was the cumulative distribution function of $g(\cdot|\mu, \sigma)$. The d_i was the isolation date of the infector associated with the i-th infectee. All other notations were the same as those in Equation (1). The maximum likelihood estimates were calculated, and AICc was employed for model selection.

RESULTS AND DISCUSSION

The observed SIs of all 21 samples have a mean of 4.3 days, median of 4 days, interquartile range (IQR) between 2 and 5, and a range from 1 to 13 days. For the 12 "infector- infectee" pairs, the observed SIs have a mean of 3 days, median of 2 days, IQR between 2 and 4, and range from 1 to 8 days. **Figure 1** shows the likelihood profiles of varying SI with respect to μ and σ of SI. In

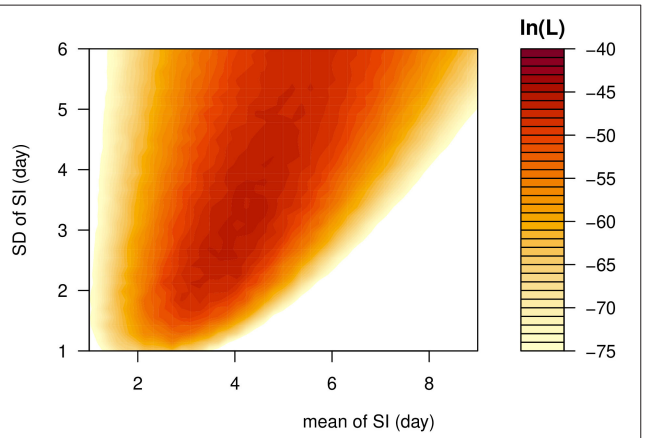


FIGURE 1 | The likelihood profile of the varying serial interval (SI) of COVID-19 by using all samples. The color scheme is shown on the right-hand side, and a darker color indicates a larger log-likelihood, i.e., ln(L), value.

Table 1, for the non-truncated scenario [i.e., using Equation [1]], we found the three distributions have almost equivalent fitting performance in terms of the AICc. The lognormal distribution has the lowest AICc, and thus it is presented as the main result for the SI estimation. By using all 21 samples, we estimated the mean of SI at 3.9 days (95% CI: 2.8–7.2) and SD of SI at 2.6 days (95% CI: 1.6–9.3). Between the observed and the fitted distributions, the Pearson's correlation is 0.98, and the R-squared is 0.97. These estimates largely matched the results in the existing literature [31, 34, 35]. Considering only the 12 "infector-infectee" pairs, we found the lognormal distribution also outperformed, and we estimated the mean of SI at 3.0 days (95% CI: 1.9–6.8) and SD of SI at 2.0 days (95% CI: 1.0–10.5). In this case, the Pearson's correlation is 0.96, and the R-squared is 0.92. The fitted lognormal distributions were shown in **Figure 2**.

For the right-truncated scenario [i.e., using Equation (2)], the lognormal distribution also outperformed in terms of the AICc, see **Table 1**. By using all 21 samples, we estimated the mean of SI at 4.9 days (95% CI: 3.6–6.2) and SD of SI at 4.4 days (95% CI: 2.9–8.3). By only using the 12 "infector-infectee" pairs, we estimated the mean of SI at 3.0 days (95% CI: 2.1–3.9) and SD of SI at 2.0 days (95% CI: 1.2–4.6). The Pearson's correlation and coefficient of determination were no longer applicable here since the likelihood function was adjusted and thus not solely depended on the SI observations.

Comparing to the SI of SARS with a mean of 8.4 days and SD of 3.4 days [36], the estimated 4.9-day SI for COVID-19 indicated rapid cycles of generation replacement in the transmission chain. Hence, highly efficient public health control measures, including contact tracing, isolation, and screening, were strongly recommended to mitigate the epidemic size. The timely supply and delivery of healthcare resources, e.g., facemasks, alcohol sterilizer, and manpower and equipment for treatment, were required in response to the rapid growing incidences of COVID-19 [4, 37]. In the places with less

TABLE 1 | Summary of the estimates of the serial interval (SI) mean and standard deviation (SD) from three different distributions.

Truncation	Dataset	Distribution	Serial interval (day)		AICc
			Mean	SD	
Not truncated	All transmission events ($n = 21$)	Gamma	4.0 (2.9, 5.9)	2.4 (1.6, 4.5)	95.6
		Weibull	4.0 (2.8, 5.8)	2.4 (1.8, 4.5)	96.2
		Lognormal	3.9 (2.8, 7.2)	2.6 (1.6, 9.3)	95.5
	"Infector-infectee" pairs ($n = 12$)	Gamma	3.0 (1.9, 5.4)	1.8 (1.0, 4.6)	49.9
		Weibull	3.0 (1.8, 5.5)	1.9 (1.3, 5.1)	51.0
		Lognormal	3.0 (1.9, 6.8)	2.0 (1.0, 10.5)	49.0
Truncated	All transmission events ($n = 21$)	Gamma	4.4 (3.2, 5.6)	3.0 (2.1, 4.8)	100.7
		Weibull	4.4 (3.1, 5.8)	2.9 (2.0, 5.0)	101.4
		Lognormal	4.9 (3.6, 6.2)	4.4 (2.9, 8.3)	99.5
	"Infector-infectee" pairs ($n = 12$)	Gamma	3.0 (2.0, 4.0)	1.8 (1.2, 3.3)	49.9
		Weibull	3.0 (1.9, 4.2)	1.8 (1.3, 3.6)	51.0
		Lognormal	3.0 (2.1, 3.9)	2.0 (1.2, 4.6)	48.9

The highlight estimates are considered as the main results.

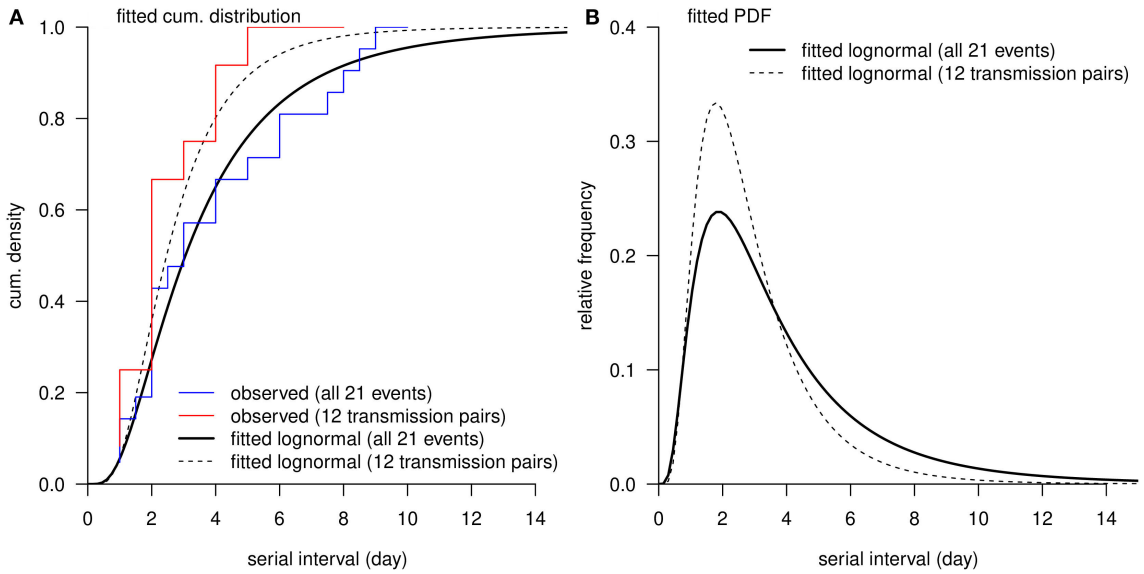


FIGURE 2 | The distribution of serial interval (SI). **(A)**, the red curve is the observed cumulative distribution of SI from 12 transmission pairs, and the blue curve is the observed cumulative distribution of SI from all 21 transmission events. **(A,B)**, the black bold curve is the fitted lognormal distribution using all 21 transmission events, and the dashed thin curve is the fitted lognormal distribution using 12 transmission pairs without truncation.

developed healthcare systems and limited medical resources, such a rapid growth of the epidemic may cause a burden to the public health system. Therefore, preparedness and cautiousness for the risk of COVID-19 are crucial to minimize impacts [38, 39].

As also pointed out by recent works [31, 34, 35, 40], the mean of SI at 4.9 days is slightly smaller than the mean incubation period, roughly 5 days, estimated by many previous studies [41–44]. The pre-symptomatic transmission may occur when the SI is shorter than the incubation period. If isolation can be conducted immediately after the symptom onset, the pre-symptomatic transmission is likely to contribute to most of SARS-CoV-2 infections. This situation has been recognized by a recent epidemiological investigation [45], and has been implemented in the mechanistic modeling studies of the COVID-19 epidemic [4, 46], where the pre-symptomatic cases were contagious. As such, merely isolating the symptomatic cases will lead to a considerable proportion of secondary cases, and thus contact tracing and immediately quarantine were crucial to reduce the risk of infection. In addition, we would like to point out that minor negative SI observations were reported in recent studies [34, 35, 47–49]. The negativity in the SI may occur when the incubation period is short with a large variance. However, negative value was not observed in our dataset, which may be due to the small sample size. We further remark that this is unlikely to bias the estimation of mean SI, but may lead to a slight underestimation of the SD of SI. The purpose of estimating SI is to approximate the generation interval (time lag of infections of successive cases) which is strictly positive. Caution should be taken when dealing with negative SI.

A recently epidemiological study used 5 "infector-infectee" pairs from contact tracing data in Wuhan, China during the early outbreak to estimate the mean SI at 7.5 days (95% CI: 5.3–19.0) [42], which appeared larger than our SI estimate of 4.9 days. Although the 95% CIs of SI estimates in this study, consistent with previous studies [21, 31, 33–35], and those in Li et al. [42] were not significantly separated, the difference in the SI estimates might exist. If this difference was not due to sampling chance, one of the possible explanations could be enhanced public awareness and swift control measures including the contact tracing and isolation implemented in Hong Kong. Since Hong Kong was the hardest hit in the SARS outbreaks in 2003 [18, 19], the local public health control was one of the most effective in the world. In the initial phase of the outbreak in Wuhan, the transmission occurred without sufficient awareness and effective intervention, thus the SI estimate in Li et al. [42] may be regarded as the intrinsic (wild) SI, as defined in Champredon et al. [50], of COVID-19. Whereas, the SI estimate in Hong Kong may be regarded as the effective SI, in more practical situations where timely action (quarantining cases and their close contacts) is in place [23], such that one case could be isolated before having the chance to further infect others. If timely action was not in place, infections of longer serial interval may occur. Thus, shorter SI observations might be an outcome of effectiveness in control in a location. The practice in Hong Kong is an example for other regions, including less developed countries.

The SI estimate can benefit from larger sample size. The estimates in our study were based on 21 identified transmission events including 12 "infector-infectee" pairs. Although the sample size was smaller than 28 transmission events in

Nishiura et al. [31], 71 in You et al. [35] and 468 in Du et al. [34], the advantage of this analysis is that all the 21 transmission events were identified in Hong Kong. Hence, the surveillance data were under consistent reporting and recording standards, which further reduced the heterogenicity in the observations. Our analysis can be improved if larger records on the local transmission events can be produced. Furthermore, a comparison between different localities is important, which sheds light on the effects of different external factors on SI.

Accurate and consistent records on dates of illness onset were essential to the estimation of the SI. All samples used in this analysis were identified in Hong Kong and collected consistently from the CHP [16, 17]. Hence, the reporting criteria were most likely to be the same for all COVID-19 cases, which potentially made our findings more robust.

Clusters of cases can occur by person-to-person transmission within a cluster, e.g.,

- scenario (**I**): person A infected B, C, and D; or
- scenario (**II**): A to B to C to D; or
- scenario (**III**): a mixture of (**I**) and (**II**), e.g., A to B, B to C and D;

or they can occur through common exposure to an unrecognized source of infection, e.g.,

- scenario (**IV**): an unknown person X infected A, B, C, and D; or
- scenario (**V**): a mixture of (**IV**) and (**I**) or (**II**), e.g., X to A and B, B to C and D.

The lack of information in the publicly available dataset made it difficult to disentangle such complicated situations. The scenarios (**I**) and (**II**) can be covered by a pair of "infector-infectee" such that we could identify the link between two unique consecutive infections. Under the scenario (**III**), we cannot clearly identify the pairwise match between the infector and infectee, which means there were multiple candidates for the infector of one infectee. As such, we employed the PDF $h(\cdot)$ in Equation (1) to account for the possible time of exposure ranging from T^{low} to T^{up}. There is no information available on the SI for scenarios (**IV**) as well as (**V**) due to the onset date of person X being unknown, and thus our analysis was limited in the scenarios (**I**)-(**III**). We note that we should be extra cautious in interpreting the clusters of cases because of this potential limitation. Although we used interval censoring likelihood to deal with the multiple-infector matching issue, more detailed information of the exposure history and clue on "who acquires infection from whom" (WAIFW) would improve our estimates.

Longer SI might be difficult to confirm in reality due to the isolation of confirmed infections [51, 52], or to identify and link together due to the less accurate information associated with memory error which occurred in the backward contact tracing exercise [34]. The issue associated with isolation could possibly bias the SI estimates and lead to an underestimated result [31]. It is possible that at the initial stage the SI is longer than later when strict isolation takes place [23]. Nevertheless, a comparison of estimated SI for SARS and COVID-19 in Hong Kong is still meaningful. We found that the estimated SI of COVID-19 appears shorter than that of SARS. It would be hard to imagine that isolation is responsible for the difference. It is unlikely that the isolation is more rapid in cases of COVID-19 than cases in SARS in Hong Kong, as well as other limitations, which would have happened for both. Thus, the difference we observed for COVID-19 and SARS is likely intrinsic. Given the rapid spreading of COVID-19, effective contact tracing and quarantine/isolation were even more crucial for successful control.

CONCLUSION

Together with the basic reproduction number, the serial interval is one of the most important epidemiological parameters, although is difficult to estimate and garners less attention than the former. Here, we found that the SI of COVID-19 may be shorter than the preliminary estimates seen in previous works. Since SI could be shorter than the incubation period among some cases, pre-symptomatic transmission may occur, and extra efforts on timely contact tracing and quarantine are crucially needed in combating the COVID-19 outbreak.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR'S NOTE

We note the preprint version of this study is posted in the medRxiv with doi: 10.1101/2020.02.21.20026559 [53], and in the ResearchSquare with doi: 10.21203/rs.3.rs-18805/v3 [54].

AUTHOR CONTRIBUTIONS

SZ conceived the study and carried out the analysis. SZ and DH drafted the first manuscript. All authors discussed the results, critically read, revised the manuscript, and gave final approval for publication.

FUNDING

DH was supported by General Research Fund (grant number: 15205119) of the Research Grants Council (RGC) of Hong Kong, China and Alibaba (China) Co. Ltd. Collaborative Research project. WW was supported by National Natural Science Foundation of China (grant number: 61672013) and Huaian Key Laboratory for Infectious Diseases Control and Prevention (grant number: HAP201704), Huaian, Jiangsu, China.

REFERENCES

1. World Health Organization. 'Pneumonia of unknown cause – China', Emergencies preparedness, response, Disease outbreak news, World Health Organization (WHO). (2020). Available online at: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/ (accessed March 8, 2020).
2. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. (2020) **368**:489–93. doi: 10.1126/science.abb3221
3. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digital Health*. (2020) **2**:e201–8. doi: 10.1016/S2589-7500(20)30026-1
4. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. (2020) **395**:689–97. doi: 10.1016/S0140-6736(20)30260-9
5. Zhao S, Musa SS, Lin Q, Ran J, Yang G, Wang W, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of january 2020: a data-driven modelling analysis of the early outbreak. *J Clin Med*. (2020) **9**:388. doi: 10.3390/jcm9020388
6. Wells CR, Sah P, Moghadas SM, Pandey A, Shoukat A, Wang Y, et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc Natl Acad Sci USA*. (2020) **117**:7504–9. doi: 10.1073/pnas.2002616117
7. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K. Pneumonia of unknown etiology in Wuhan, China: potential for international spread via commercial air travel. *J Travel Med*. (2020) **27**:taaa008. doi: 10.1093/jtm/taaa008
8. Zhao S, Zhuang Z, Cao P, Ran J, Gao D, Lou Y, et al. Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis. *J Travel Med*. (2020) **27**:taaa022. doi: 10.1093/jtm/taaa022
9. Zhao S, Zhuang Z, Ran J, Lin J, Yang G, Yang L, et al. The association between domestic train transportation and novel coronavirus outbreak in China, from 2019 to 2020: a data-driven correlational report. *Travel Med Infect Dis*. (2020) **33**:101568. doi: 10.1016/j.tmaid.2020.101568
10. Kang D, Choi H, Kim J-H, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int J Infect Dis*. (2020) **94**:96–102. doi: 10.1016/j.ijid.2020.03.076
11. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis*. (2020) **20**:553–8. doi: 10.1016/S1473-3099(20)30144-4
12. Liang K. Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infect Genet Evol*. (2020) **82**:104306. doi: 10.1016/j.meegid.2020.104306
13. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis*. (2020). **92**:214–7. doi: 10.1016/j.ijid.2020.01.050
14. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*. (2020) **25**:2000058. doi: 10.2807/1560-7917.ES.2020.25.4.2000058
15. Kwok KO, Wong V, Wei VWI, Wong SYS, Tang JW-T. Novel coronavirus (2019-nCoV) cases in Hong Kong and implications for further spread. *J Infect*. (2020) **80**:671–93. doi: 10.1016/j.jinf.2020.02.002
16. Centre for Health Protection. *Summary of data and outbreak situation of the Severe Respiratory Disease associated with a Novel Infectious Agent, Centre for Health Protection, the government of Hong Kong*. (2020). Available online at: https://www.chp.gov.hk/en/features/102465.html (accessed March 8, 2020).
17. Centre for Health Protection. *The collection of Press Releases by the Centre for Health Protection (CHP) of Hong Kong*. (2020). Available online at: https://www.chp.gov.hk/en/media/116/index.html (accessed March 8, 2020).
18. Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. *Epidemiology*. (2005) **16**:791–801. doi: 10.1097/01.ede.0000181633.80269.4c
19. Leung GM, Hedley AJ, Ho L-M, Chau P, Wong IOL, Thach TQ, et al. The epidemiology of severe acute respiratory syndrome in the 2003 Hong Kong epidemic: an analysis of all 1755 patients. *Ann Intern Med*. (2004) **141**:662–73. doi: 10.7326/0003-4819-141-9-200411020-00006
20. Ran J, Zhao S, Zhuang Z, Chong MKC, Cai Y, Cao P, et al. Quantifying the improvement in confirmation efficiency of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) during the early phase of outbreak in Hong Kong in 2020. *Int J Infect Dis*. (2020) **96**:284–7. doi: 10.1016/j.ijid.2020.05.015
21. Kwok KO, Wong VWY, Wei WI, Wong SYS, Tang JW-T. Epidemiological characteristics of the first 53 laboratory-confirmed cases of COVID-19 epidemic in Hong Kong, 13 February 2020. *Eurosurveillance*. (2020) **25**:2000155. doi: 10.2807/1560-7917.ES.2020.25.16.2000155
22. Leung GM, Cowling BJ, Wu JT. From a sprint to a marathon in Hong Kong. *N Engl J Med*. (2020) **382**:e45. doi: 10.1056/NEJMc2009790
23. Zhao S, Cao P, Chong MKC, Gao D, Lou Y, Ran J, et al. COVID-19 and gender-specific difference: Analysis of public surveillance data in Hong Kong and Shenzhen, China, from January 10 to February 15, 2020. *Infect Control Hosp Epidemiol*. (2020) **41**:750–1. doi: 10.1017/ice.2020.64
24. Wong SYS, Kwok KO, Chan FKL. What can countries learn from Hong Kong's response to the COVID-19 pandemic? *CMAJ*. (2020) **192**:E511–5. doi: 10.1503/cmaj.200563
25. Cowling BJ, Fang VJ, Riley S, Peiris JM, Leung GM. Estimation of the serial interval of influenza. *Epidemiology*. (2009) **20**:344–7. doi: 10.1097/EDE.0b013e31819d1092
26. Fine PEM. The interval between successive cases of an infectious disease. *Am J Epidemiol*. (2003) **158**:1039–47. doi: 10.1093/aje/kwg251
27. Nishiura H, Chowell G, Heesterbeek H, Wallinga J. The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. *J R Soc Interf*. (2010) **7**:297–307. doi: 10.1098/rsif.2009.0153
28. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B: Biol Sci*. (2007) **274**:599–604. doi: 10.1098/rspb.2006.3754
29. Yang Y, Longini I, Halloran ME. Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. *J R Stat Soc Ser C-Appl Stat*. (2006) **55**:317–30. doi: 10.1111/j.1467-9876.2006.00539.x
30. World Health Organization. *Laboratory Testing for 2019 Novel Coronavirus (2019-nCoV) in Suspected Human Cases, World Health Organization* (WHO). (2020). Available online at: https://www.who.int/health-topics/coronavirus/laboratory-diagnostics-for-novel-coronavirus (accessed March 8, 2020).
31. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis*. (2020) **93**:284–6. doi: 10.1016/j.ijid.2020.02.060
32. Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*. (2005) **11**:1031–57. doi: 10.3150/bj/1137421639
33. Zhao S. Estimating the time interval between transmission generations when negative values occur in the serial interval data: using COVID-19 as an example. *Math Biosci Eng*. (2020) **17**:3512–9. doi: 10.2807/1560-7917.ES.2020.25.17.2000257
34. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg Infect Dis J*. (2020) **26**:1341–43. doi: 10.3201/eid2606.200357
35. You, C., Deng, Y., Hu, W., Sun, J., Lin, Q., Zhou F, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int J Hygiene Environ Health*. 228:113555. doi: 10.1016/j.ijheh.2020.113555
36. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*. (2003) **300**:1966–70. doi: 10.1126/science.1086616
37. Zhao S, Stone L, Gao D, Musa SS, Chong MKC, He D, et al. Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to 2020. *Ann Transl Med*. (2020) **8**:488. doi: 10.21037/atm.2020.03.168

38. Chong KC, Cheng W, Zhao S, Ling F, Mohammad KN, Wang MH, et al. Monitoring disease transmissibility of 2019 novel coronavirus disease in Zhejiang, China. *Int J Inf Dis*. (2020) 96:128–130. doi: 10.1016/j.ijid.2020.04.036
39. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infect Dis*. (2020) **93**:211–6. doi: 10.1016/j.ijid.2020.02.058
40. Wang K, Zhao S, Liao Y, Zhao T, Wang X, Zhang X, et al. Estimating the serial interval of the novel coronavirus disease (COVID-19) based on the public surveillance data in Shenzhen, China from January 19 to February 22, 2020. *Transbound Emerg Dis*. (2020). doi: 10.1111/tbed.13647. [Epub ahead of print].
41. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*. (2020) **25**:2000062. doi: 10.2807/1560-7917.ES.2020.25.5.2000062
42. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*. (2020) **382**:1199–207. doi: 10.1056/NEJMoa2001316
43. Linton MN, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov RA, Jung S-M, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med*. (2020) **9**:538. doi: 10.3390/jcm9020538
44. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith H, et al. The incubation period of 2019-nCoV from publicly reported confirmed cases: estimation and application. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.02.02.20020016
45. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med*. (2020) **382**:970–1. doi: 10.1056/NEJMc2001468
46. Chowell G, Dhillon R, Srikrishna D. Getting to zero quickly in the 2019-nCov epidemic with vaccines or rapid testing. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.02.03.20020271
47. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. (2020) **368**:eabb6936. doi: 10.1126/science.abb6936
48. 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
49. Ma S, Zhang J, Zeng M, Yun Q, Guo W, Zheng Y, et al. Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. *medRxiv [Preprin]*. (2020). doi: 10.1101/2020.03.21.20040329
50. Champredon D, Dushoff J. Intrinsic and realized generation intervals in infectious-disease transmission. *Proc R Soc B: Biol Sci*. (2015) **282**:20152026. doi: 10.1098/rspb.2015.2026
51. Park SW, Champredon D, Dushoff J. Inferring generation-interval distributions from contact-tracing data. *J R Soc Interf*. (2020) **17**:20190719. doi: 10.1098/rsif.2019.0719
52. Yan P. Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *J Theoret Biol*. (2008) **251**:238–52. doi: 10.1016/j.jtbi.2007.11.027
53. Zhao S, Gao D, Zhuang Z, Chong M, Cai Y, Ran J, et al. Estimating the serial interval of the novel coronavirus disease (COVID-19): a statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.02.21.20026559
54. Zhao S, Gao D, Zhuang Z, Chong M, Cai Y, Ran J, et al. Estimating the serial interval of the novel coronavirus disease (COVID-19): a statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. *Res Square*. (2020). doi: 10.21203/rs.3.rs-18805/v1. [Epub ahead of print].

Conflict of Interest: DH received funding from Alibaba (China) Co. Ltd. Collaborative Research project.

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

The reviewer STA declared a shared affiliation with one of the authors, JR to the handling editor at time of review.

frontiers in Physics

ORIGINAL RESEARCH
published: 29 September 2020
doi: 10.3389/fphy.2020.564698

Epidemic Characteristics of COVID-19 in Africa

Lijun Wang[1,2†], Shengwen Dong[3†], Ying Zhao[4†], Yulei Gao[1], Jian Wang[2,5], Muming Yu[1], Fei Xu[6] and Yanfen Chai[1*]*

[1] Department of Emergency Medicine, Tianjin Medical University General Hospital, Tianjin, China, [2] Hôpital de l'amitié Sino-Congolaise, Brazzaville, Congo, [3] School of Nursing, Tianjin Medical University, Tianjin, China, [4] Department of Neonatology, Tianjin Central Hospital of Gynecology Obstetrics, Tianjin, China, [5] International Diagnosis and Treatment Center, Tianjin First Central Hospital, Tianjin, China, [6] Educational Technology and Laboratory Management Centre, Tianjin Foreign Studies University, Tianjin, China

Objectives: The aim of this study was to investigate coronavirus disease 2019 (COVID-19) epidemic characteristics in Africa and provide some references to prevent pandemic.

Methods: We collect information of the laboratory-confirmed case of COVID-19 that has been reported since April 28, 2020 in Africa, describe the epidemic characteristics in different regions of Africa, and predict the spread trend.

Results: (1) There are 35,309 COVID-19 cases and 1,522 deaths in 54 African countries. Both case amount and death toll in Northern Africa are higher than those in other African regions, whereas the case fatality rate (CFR) in Africa is lower than that of the world. (2) Among the 54 countries in Africa, 39 of them (72.22%) report confirmed cases between 1 and 499. (3) In total, 596,146 tests have been conducted in 35 countries, and 28,271 patients have been confirmed, with a positive rate of 4.74%. However, there are only 335 tests (interquartile range: 145.5–874) for every 1 million people in Africa. (4) The number of infection cases will increase to around 150,000 at the end of May according to our model.

Conclusion: In Africa, COVID-19 spreads faster while with a relatively lower CFR, and it may be related to higher young people ratio. In addition, test intensity of COVID-19 is obviously insufficient in Africa.

Keywords: COVID-19, Africa, pandemic, prediction, test intensity

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Zhanwei Du,
University of Texas at Austin, United States
Chao Gao,
Southwest University, China
Wen-xuan Wang,
Beijing University of Posts and Telecommunications (BUPT), China

***Correspondence:**
Yanfen Chai
chaiyanfen2012@126.com
Lijun Wang
wanglijun2005icu@foxmail.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *22 May 2020*
Accepted: *12 August 2020*
Published: *29 September 2020*

Citation:
Wang L, Dong S, Zhao Y, Gao Y, Wang J, Yu M, Xu F and Chai Y (2020) Epidemic Characteristics of COVID-19 in Africa. Front. Phys. 8:564698. doi: 10.3389/fphy.2020.564698

INTRODUCTION

Since the first coronavirus disease 2019 (COVID-19) case was reported in Wuhan, China, a dramatically increasing number of cases had been identified, although its exact origin remained unknown until now [1]. As of April 28, 2020, there were a total of 3,136,508 cases from 210 countries and territories, and 217,813 cases died. Africa reported the first case of COVID-19 on February 14; currently, a total of 35,309 confirmed cases and 1,522 related deaths have been reported in 54 African countries [2].

Although African cases only account for 1.13% in the world, due to the unique demographic and geographic characteristics, COVID-19 spreads much faster in Africa than that in other regions. In addition, the African relatively higher ratio of young population and the fragile public health system have exacerbated the disease epidemic [3]. Hence, Africa is suffering from the huge challenge of the COVID-19 pandemic. This article introduces some epidemic characteristics of COVID-19

in Africa, such as the case number and distribution, the case fatality rate (CFR) and its regional difference, the total tests and test intensity, and the transmission type. In addition, as the COVID-19 is at the early stage, we use the Matlab Curve Fitting toolbox to fit the model with the outbreak data and make the forecasting. We hope that this article would make the government pay more attention to the epidemic and provide some references for effective prevention and control in Africa and other countries.

MATERIALS AND METHODS

Materials

Data Sources and Definition

Population characteristics, such as population density (= people number/km^2), urban population rate (= current urban people number/total population amount × 100%), age structure, median age (index by dividing a population into two numerically equally sized groups), and current population amount in each country, are searched on the website "Worldometer-Population" [4] and "Central Intelligence Agency" [5].

Data of total confirmed COVID-19 cases, total COVID-19 deaths, total test samples, test intensity (= test numbers/1 million population), the country reporting the first case, transmission type (imported or local), and amount are derived from the "Worldometer-Coronavirus" [2], "African Centers for Disease Control (CDC)" [6], and "COVID-19 Global Pandemic Real-time Report" [7]. The age and gender distributions of confirmed COVID-19 cases originated from the "World Health Organization" [8]. All data in this article are updated to April 28, 2020, unless otherwise stated.

Different organizations have different divisions of African regions. In this article, the regional classification of Eastern, Southern, Western, Northern, and Central Africa is based on the "African CDC" [6].

Methods

The total number of COVID-19 cases and deaths in African countries has been recorded each day since February 14. The total case number and the CFR in different African regions are calculated according to the regional division [6]. The global total number of COVID-19 cases, CFR, total test amount, and intensity are searched online [2]. At the same time, population characteristics of Africa and related 16 countries are investigated on the Internet.

We predict the required time that Africa arrives at the indicated case number (100, 500, 1,000, 5,000, 10,000, 50,000, and 100,000) using formulas (1) and (2):

$$\mathrm{Day_a} = \mathrm{Date}_\alpha - \mathrm{Date}_\beta \quad (1)$$

$$\mathrm{Date_b} = \mathrm{Date}_\gamma + \mathrm{Day}_\delta \quad (2)$$

notes: a = required in a country; α = case arrival indicated number in the country; β = the case first reported in the country; b = predicative case arrival indicated number in Africa; γ = the case first reported in Africa; δ = median $\mathrm{Day_a}$ among 16 countries.

Data we use to calculate the median day is from 16 countries, including USA, Italy, Germany, and China, etc. (see **Table 1**). The test intensity in this article is defined as "test numbers/1 million population," and test positive rate is calculated by formula (3):

$$\text{Test positive rate}(\%) = \frac{\text{confirmed case number}}{\text{total tests}} \times 100. \quad (3)$$

In addition, we use the Matlab Curve Fitting toolbox to fit the model with the outbreak data and make the forecasting. The number of cumulative number of cases at time interval t, $C(t)$, is modeled by the following equation:

$$C(t) = \left(r\left(1-p\right)t + A\right)^{\frac{1}{1-p}}$$

where $p \in [0, 1]$ is a "deceleration of growth" parameter and r is a positive parameter denoting the growth rate.

Figure 1 shows the flowchart in this article.

Statistical Methods

Categorical data are presented as frequencies and percentages, whereas continuous data without normal distribution are expressed as median and interquartile range (IQR). A chi-square test is used to compare the differences in CFR of COVID-19 or test positive rate among different regions. The Wilcoxon rank-sum test is applied to test intensity between the globe and Africa. Multiple linear regression analysis (stepwise linear regression) is used to explore the relationship between population characteristics and COVID-19 disease transmission speed in countries. Statistical analyses are performed using IBM SPSS Statistics for Windows (version 23.0, IBM Corp., Armonk, NY, USA), and $p < 0.05$ is considered statistically significant.

RESULTS

Case Number and Distribution

Figure 2A shows the trend of COVID-19 cases in Africa: the daily increase was <100 cases before March 19, whereas it reached up to 100–500 during March 19–24, and it exceeded 500 cases/day after March 24. Similar to the above-mentioned, there were only 13 countries that reported COVID-19 cases before March 13, whereas the entire African countries discovered the case on April 6 (**Figure 2B**). **Figure 2C** demonstrates the COVID-19 distribution in African countries on April 28.

Regarding the case distribution, there were 13,986 cases coming from Northern Africa, taking account for 40.59%, followed by Western Africa (n = 8,614, 25.00%), Southern Africa (n = 5,364, 15.57%), Eastern Africa (n = 3,474, 10.08%), and Central Africa (n = 3,020, 8.76%). **Figure 2D** indicates that the case number in Northern Africa grew faster than that in other African regions.

The five countries in Africa with the highest cumulative number of cases (proportion of reported cases in Africa) were

Abbreviations: COVID-19, coronavirus disease 2019; CDC, Centers for Disease Control; CFR: case fatality rate; IQR, interquartile range; HCQ, hydroxychloroquine; rRT-PCR, (real-time) reverse transcription polymerase chain reaction; WHO, World Health Organization.

TABLE 1 | Population characteristics and transmission dynamics of COVID-19 in some countries.

	Population density (P/km^2)	Urban population (%)	Median age (year)	Current population (million)	Reported 1st case[a]	Days required for case growth						
						100	500	1 k	5 k	10 k	50 k	100 k
USA	36	82.8	38.3	330.52	Jan 10	52	58	61	67	69	74	77
Italy	206	69.5	47.3	60.48	Jan 29	25	29	31	38	41	52	61
Germany	240	76.3	45.7	83.78	Jan 26	35	39	42	49	52	61	70
Spain	94	80.3	44.9	46.75	Jan 30	32	37	39	43	47	56	62
France	119	81.5	42.3	65.28	Jan 23	37	43	45	52	56	68	75
UK	281	83.2	40.5	67.8	Jan 30	35	42	44	51	56	67	77
China	153	60.8	38.4	1,439.32	Dec 1	48	52	54	58	61	73	
Iran	52	75.5	32.0	83.72	Feb 18	8	11	13	18	23	44	
Turkey	110	75.7	31.5	84.11	Mar 9	9	11	13	18	21	33	
Russia	9	73.7	39.6	145.92	Mar 2	15	23	25	34	38	50	
Switzerland	219	74.1	43.1	8.64	Feb 26	8	14	16	23	28		
Belgium	383	98.3	41.9	11.58	Feb 3	32	39	42	52	55		
Netherlands	508	92.5	43.3	17.13	Feb 24	11	18	21	29	36		
Austria	109	57.3	43.5	8.99	Jan 24	44	49	52	60	67		
S. Korea	527	81.8	43.7	51.26	Jan 19	32	35	38	44	75		
Brazil	25	87.6	33.5	212.18	Feb 24	18	23	26	36	40		
Africa	45	43.8	19.7	1,332.38	Feb 14	25	33	37	46	53		
Median (day)*	153	80.3	42.3	65.28		**32**	**37**	**39**	**44**	**52**	**58.5**	**72.5**

**African data not included.*

[a] The first case was reported in 2019 in China; the other countries were in 2020.

Bold values mean the median day required for case growth to indicated number among those 16 countries.

Egypt ($n = 5{,}042$, 14.28%), followed by South Africa ($n = 4{,}996$, 14.15%), Morocco ($n = 4{,}252$, 12.04%), Algeria ($n = 3{,}649$, 10.33%), and Cameroon ($n = 1{,}705$, 4.83%). These five countries accounted for 55.63% in entire Africa, whereas these ratios in Europe and China were 69.16 and 86.88%, respectively ($\chi^2 = 10{,}526.70$, $p < 0.001$). In addition, 39 (72.22%) countries reported cases between 1 and 499 (**Figure 2E**). All those data suggest that the case distribution in Africa is relatively scattered.

According to the available data [8] on the age and gender distributions of confirmed COVID-19 cases in the World Health Organization (WHO) African region ($n = 4{,}639$) from February 25 to May 5, the male to female ratio among confirmed cases was 1.8, and the median age was 42 years (range: 0–105). Males (62%) in the 31–39 and 40–49 age groups were more affected than females (38%) across the same age groups.

Case Fatality Rate and Regional Difference

The COVID-19 CFR in the globe gradually increased over time. However, the CFR in Africa increased from February 14, the peak value was on April 11 (5.24%), and it began to slowly decrease after that day. Moreover, there were 1,522 death tolls as of April 28 with a CFR of 4.31% in Africa, which was lower than the global CFR of 6.94% ($\chi^2 = 113.96$, $p < 0.001$; **Figure 3A**). The countries with the highest CFR in Africa were Algeria, followed by Liberia, Egypt, Burkina Faso, and Democratic Republic of the Congo (**Figure 3B**).

Figures 3C,D indicate that the COVID-19 death toll in Northern Africa has increased dramatically since March 28, and its CFR was higher than that of other regions in Africa (7.18 vs. 1.99–3.34%, $\chi^2 = 228.18$, $p < 0.001$).

Total Tests and Test Intensity

As of April 28, a total of 30,832,243 COVID-19 tests were conducted in 174 countries, and 3,070,343 patients were confirmed, with a positive rate of 9.96%. In these countries, there were 5,307 tests (IQR: 837.5–17,152.75) for every 1 million people. Regarding Africa, a total of 596,146 tests were conducted in 35 countries, and 28,271 patients were confirmed, with a positive rate of 4.74%. However, there were only 335 tests (IQR: 145.5–874) for every 1 million people in Africa. Both test intensity and test positive rate in Africa were lower than those in the globe (χ^2 was 27.239 and 15,041.85, respectively, all $p < 0.001$). Moreover, the test positive rate and test intensity in African regions were different. **Figure 4** demonstrates that the test intensity in Northern Africa was higher than that in other regions, whereas the test positive rate in Central Africa seemed better.

Transmission Type

The transmission type is divided into two ways: imported or local transmission. Just like in Tianjin, China, imported cases predominated at the early stage. However, local transmission patients became more and more over time [9]. In Africa, there were only 33.68% local transmission cases from February 14 to March 10, whereas this ratio was up to 99.06% in the first week in April (**Table 2**). However, the duration of imported transmission in Africa was longer than that in Tianjin (24 vs. 7 days).

A total of 105 cases were confirmed from February 14 to March 10. Among them, 63 imported cases were reported by nine countries. Forty-five cases (71.43%) came from an Egypt cruise ship (12 crew members, 14 Egyptians, and 19 foreigners),

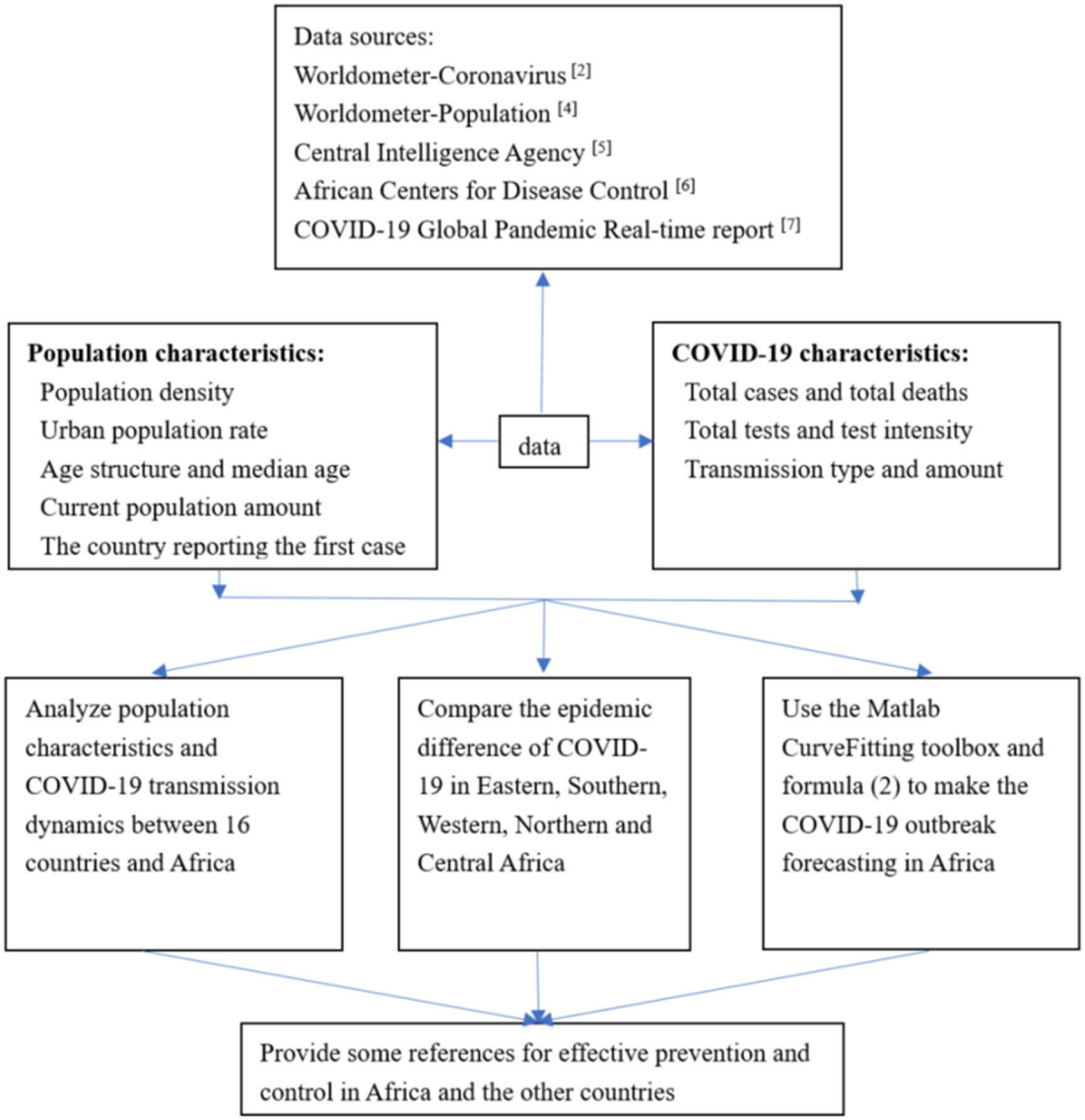


FIGURE 1 | Flowchart of this article.

followed by France ($n = 9$), Italy ($n = 5$), Kazakh ($n = 2$), United Kingdom ($n = 1$), and Canada ($n = 1$) [6].

COVID-19 Case Dynamic Prediction for Africa

We analyze the population characteristics and COVID-19 transmission dynamics in 16 countries from different continents [2, 10] (**Table 1**). The median days required for arrival to 100 and 500 cases of COVID-19 in these countries are 32 and 37, respectively. However, it only requires 25 and 33 days in Africa, respectively. The age structure in Africa and other countries was shown in **Figure 2F**. It indicates that the ratio of 0–24 years old in Africa was significantly higher than that in the globe (50.06 vs. 40.75%, $\chi^2 = 46{,}613{,}552.9$, $p < 0.001$). Moreover, multivariate linear regression analysis shows that when the confirmed case is below 1,000, the transmission speed in these countries is only related to the median age ($B = 2.044$, $t = 2.285$, $p = 0.043$), whereas the other three variables, including population density, total population, and urban population rate, are not significantly different (p-value was 0.311, 0.051, and 0.414, respectively). According to the speed of virus transmission in these 16 countries, we predict that the African COVID-19 cases will reach 50,000 and 100,000 cases on April 12 and 26, respectively (**Figure 5**).

Moreover, in order to estimate the future situation of COVID-19 infection in Africa for the upcoming intervention policy planning, we analyze and forecast the number of infection cases in Africa. As the COVID-19 outbreak in Africa was at the early stage, we adopt the early epidemic growth model [10]

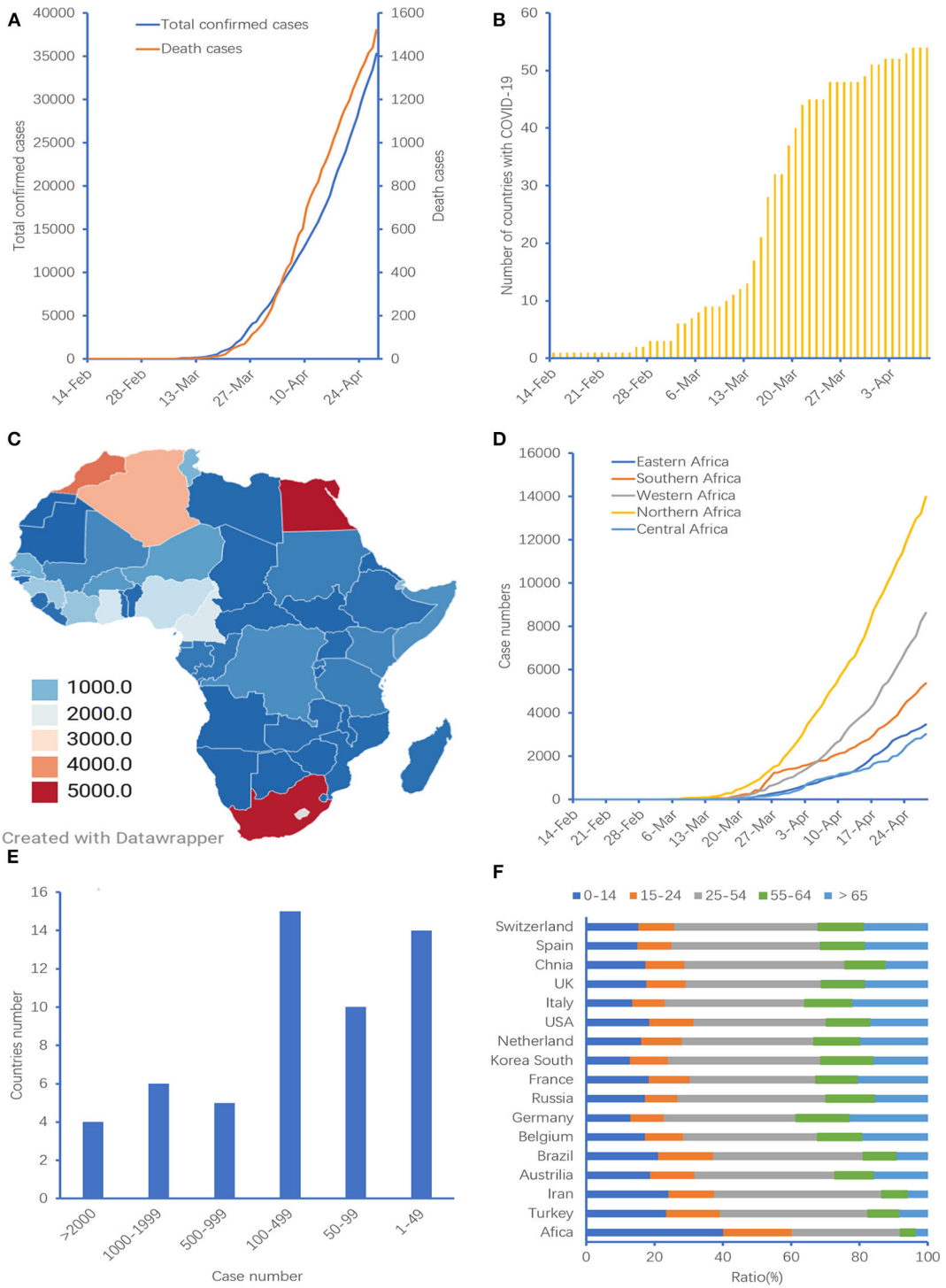


FIGURE 2 | Case number and distribution in Africa. **(A)** Confirmed COVID-19 and death cases in Africa over time, **(B)** number of countries with COVID-19 over time, **(C)** case distribution in Africa, **(D)** confirmed COVID-19 cases in different African regions. Three countries of Sahrawi Arab Democratic Republic (Northern Africa), Union of the Comoros (Eastern Africa), and Kingdom of Lesotho (Southern Africa) are not included, **(E)** distribution of COVID-19 cases in different countries in Africa, **(F)** age structure in Africa and other countries. Two countries of Mayotte and Reunion are not included.

developed to analyze the COVID-19 transmission pattern in Africa. **Figure 6** (upper) shows the outbreak data of all countries in Africa. **Figure 6** (lower) shows the fitting result and the forecast result until May 31, 2020. As shown in **Figure 6** (lower), the early epidemic growth model well-characterizes COVID-19 transmission dynamics in Africa. According to the model, the

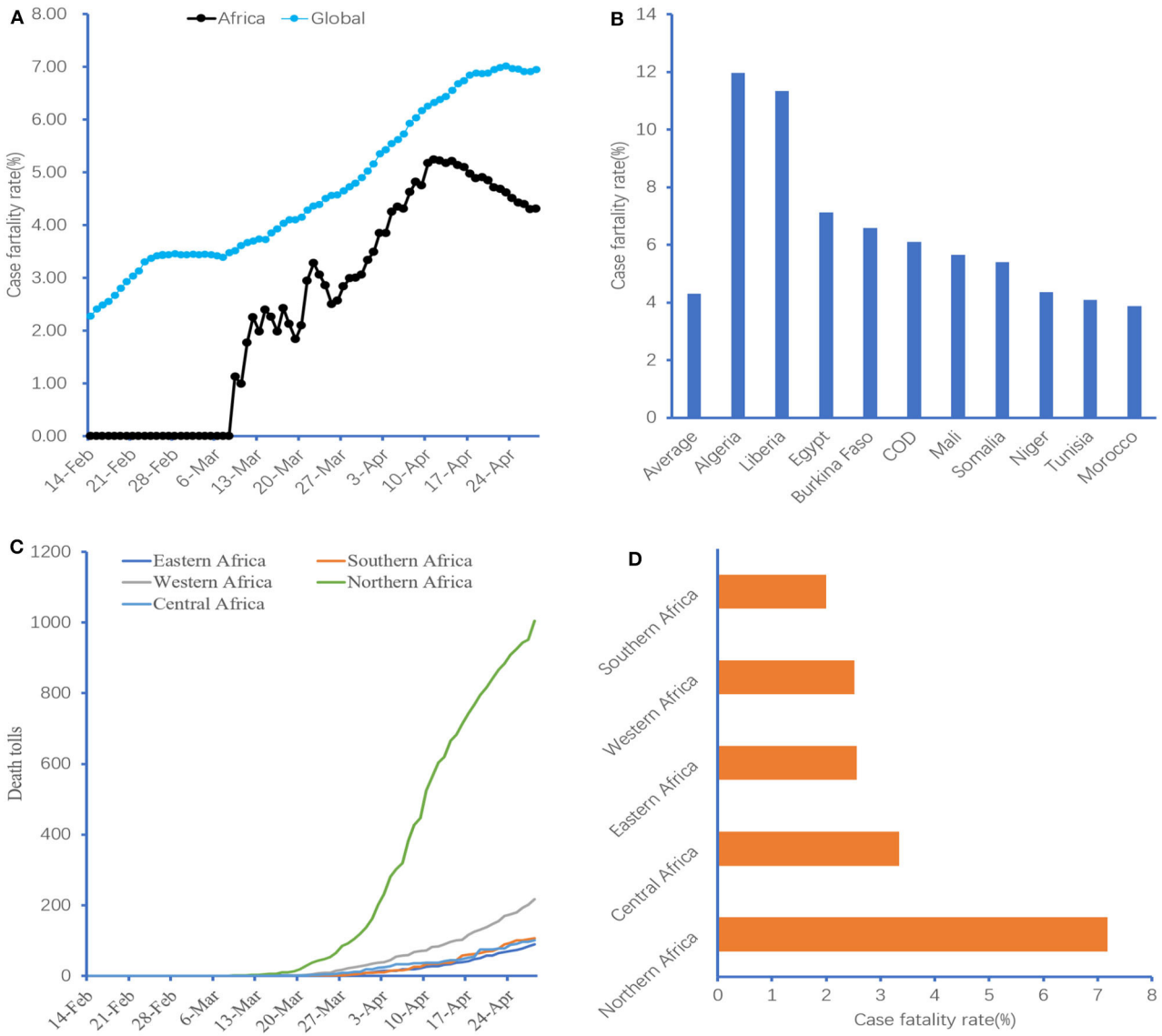


FIGURE 3 | Case fatality rate and region difference. **(A)** CFR of COVID-19 in Africa, **(B)** the top 10 African countries with COVID-19 CFR. COD, The Democratic Republic of the Congo; CFR, case fatality rate. **(C,D)** Comparison of COVID-19 CFR in different African regions. Three countries of Sahrawi Arab Democratic Republic (Northern Africa), Union of the Comoros (Eastern Africa), and Kingdom of Lesotho (Southern Africa) are not included.

number of infection cases will increase to around 150,000 at the end of May.

DISCUSSION

In Africa, COVID-19 has spread to only 13 countries from February 14 to March 13, whereas up to the entire 54 countries on April 6. The virus in Africa spreads significantly faster at the early stage than that in other regions, as demonstrated by **Figure 5**. Our results show that the transmission speed of COVID-19 is related to the median age. The younger the population, the faster the virus spreads. **Table 1** shows that the virus spreads significantly faster in Iran (median age = 32 years) or Turkey (median age = 31.5 years) than in other countries. Data analyzed about the first 425 confirmed cases in China suggest that the epidemic doubled in size every 7.4 days at the early stage, whereas it needs <4 days in Africa (105 confirmed cases in 24 days) [6]. The median age of those patients is 59 years, and more than 80% of patients are ≥45 years in that Chinese study [11]. However, Africa is a high concentration of young people, with a median age of 19.7 years, which may lead to the faster spread of COVID-19.

In addition, cases distribution in Africa is relatively scattered. The top five countries with the largest cases account for 55.63% of all African cases, which is obviously lower than Europe (69.16%), Asia (68.62%), or North America (99.51%) [7]. On the other hand, although imported cases are predominant at the

TABLE 2 | Transmission type of COVID-19 in Africa.

Date	Transmission type			Ratio of local transmission (%)
	Local transmission	Imported	Under investigation	
Feb 14–Mar 10	32	63	Not mentioned	33.68
Mar 11–Mar 17	263	23	Not mentioned	91.96
Mar 18–Mar 24	1,434	142	Not mentioned	90.99
Mar 25–Mar 31	3,334	134	Not mentioned	96.14
Apr 1–Apr 7	4,763	40	5	99.06

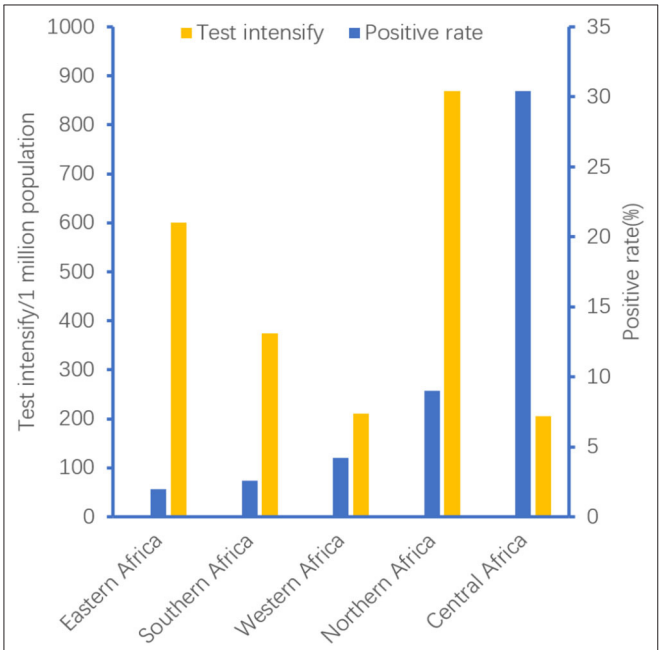


FIGURE 4 | Test positive rate and test intensity in different Africa regions.

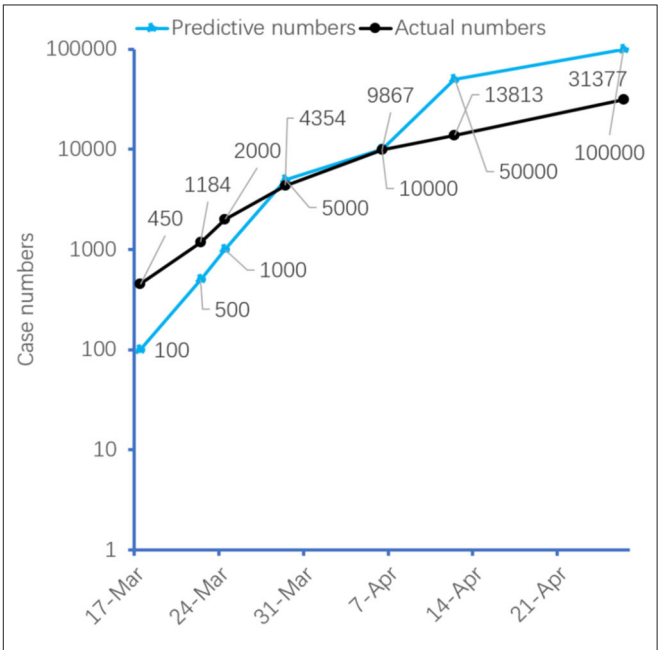


FIGURE 5 | Predictive curve of transmission dynamics of COVID-19 in Africa (1).

early stage of the African outbreak, community-transmitted cases have recently occurred in more and more countries over time. There are only 1–499 confirmed COVID-19 cases in 39 (72.22%) countries among Africa (**Figure 2B**); however, it will lead to the rapid spread of the epidemic, just like a single spark could start a prairie fire.

Furthermore, many people in Africa have serious health issues, such as malaria, HIV, and tuberculosis, that make them vulnerable to COVID-19 [12]. Sub-Saharan Africa has one of the highest levels of child malnutrition globally. Recent research finds that the stunting and underweight rates in Burundi are 57.7 and 28.8%, respectively, whereas the wasting rate is 18.0% in Niger, followed by Burkina Faso (15.50%). COVID-19 may put malnourished children at significant risk [13, 14].

Moreover, COVID-19 shares some of the highly similar symptoms with malaria, such as fever, difficulty in breathing, fatigue, and headaches of acute onset. What is more, both two diseases could affect similar organs of the body. Thus, a COVID-19 case may be misclassified as malaria if symptoms alone are used to define a case during this emergency period [15]. Hence, the actual case numbers at the early stage are much smaller than prediction, as shown in **Figure 5**. For example, predictive cases would reach 50,000, whereas actual cases were only 13,813 on April 12. Although the WHO has already supported Africa with early detection by providing thousands of COVID-19 testing kits to limit the widespread transmission, health officials in many African countries concede that the case amount is far higher than those confirmed [16, 17]. Meanwhile, the shortage of testing kits in Africa has induced that test intensity only accounts for 6.31% of the world (median: 335 tests vs. 5,307 tests/1 million population) [18].

The lower CFR of COVID-19 in Africa may relate to a higher proportion of younger population. Data from Italy indicate a much higher CFR among the elderly [19]. Another data also find that COVID-19 kills an estimated 13.4% of patients of 80 years old or older, compared with 1.25% of those in their 50's and 0.3% of those in their 40's [20]. In addition, over 80% of the deaths in China occur to people of 60 years old or older

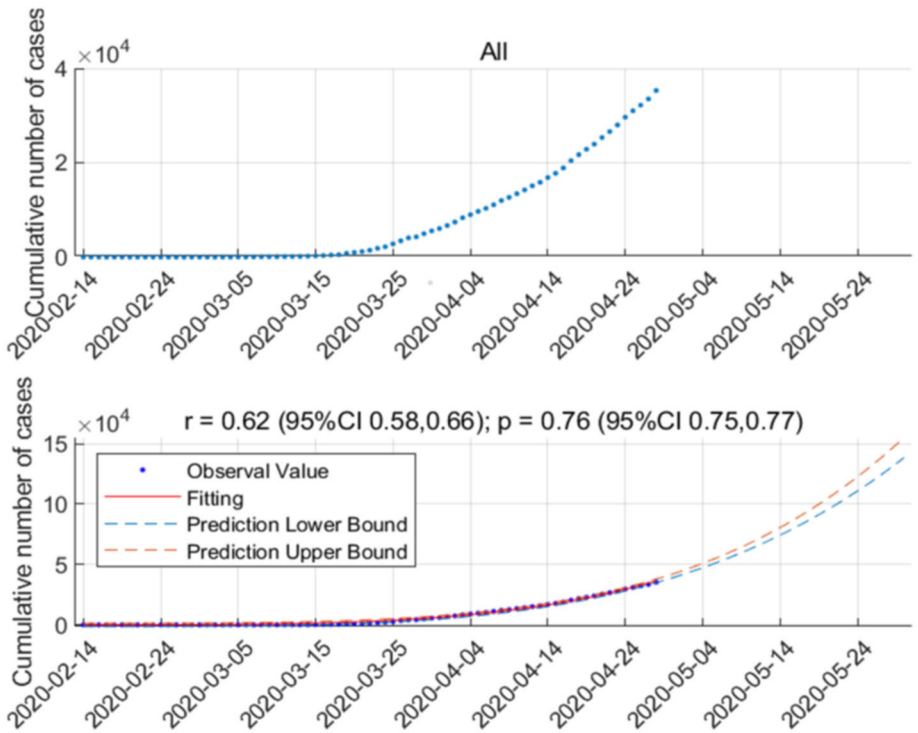


FIGURE 6 | Predictive curve of transmission dynamics of COVID-19 in Africa (2).

at the early stage of COVID-19 [21]. People in that age group account for 16% of the population, whereas in Africa, they only make up just 5% of the population [12]. Africa's relatively young population could help buffer the continent against CFR. Similarly, the relatively higher CFR in Northern Africa may relate to the lower ratio of younger people. For example, the median ages in Egypt, Algeria, and Morocco are 24.6, 28.5, and 29.5 years, respectively, whereas the median age in Africa is only 19.7 years [4]. On the other hand, Africa has a high incidence area of malaria, and chloroquine or hydroxychloroquine (HCQ) is the first-line drug for the treatment. Recently, research shows that chloroquine or HCQ is also a potential treatment for COVID-19 [22]. Is the lower CFR related to some changes in the body caused by repeated use of chloroquine or HCQ by Africans? Until now, however, there is no sufficient evidence whether the lower COVID-19 CFR in Africa is related to the exposure of chloroquine or HCQ.

The reason for the relatively lower test positive rate in Africa remains unclear until now. The current gold standard for the etiological diagnosis of COVID-19 is (real-time) reverse transcription polymerase chain reaction (rRT-PCR), which is affected by many factors, such as inadequate procedures for collection, handling, transport, and storage of the swabs and collection of inappropriate or inadequate material, as well as sample contamination [23]. On the other hand, a high proportion of young people in Africa may be another reason. Liu et al. [24] tested 4,880 cases using quantitative RT-PCR and found that a positive rate of 18–29 years in patients is only 24.90%, whereas it is ≥70 years with 61.81%. A single negative RT-PCR should not exclude COVID-19, especially if clinical suspicion is high. Chest CT has higher sensitivity for the diagnosis of COVID-19 than initial RT-PCR from swab samples [25].

According to the volume of air travel between China and Africa, a modeling study about African countries imported COVID-19 cases demonstrates that countries with the highest importation risk are Egypt, Algeria, and South Africa [26]. However, data from the African CDC indicate that the origin of imported cases in Africa is mainly from Europe (i.e., France, Italy, and United Kingdom), although COVID-19 is first reported in China [6, 27]. In addition, among these 105 patients reported as of March 10, 80 (76.19%) were distributed in Northern Africa, and most of them were imported from Europe. This may be the reason why the number of COVID-19 cases in Northern Africa is higher than that in other African regions.

Although accounting for 1.13% of the world cases, Africa also needs a response immediately to prevent its pandemic. However, most African countries remain woefully unprepared for what is coming [28]. So far, quarantine is the most effective measure to prevent virus transmission, but in African cities, particularly in tightly packed informal settlements, keeping people apart is very difficult because most people need daily income to survive. People are relatively dispersed in rural areas, while communicating information about social distancing will also be challenging [12, 17].

After closing borders and mitigated measures, including restrictions on movement, public gatherings, and schools, some

countries have seen a reduction in the average daily case growth, whereas it will undoubtedly depress economics in Africa [27]. Whether lockdown is suitable for Africa remains unknown. However, Africa could work together to prevent the spread of COVID-19 and amplify their collaboration by sharing data, capabilities, and coordinating strategies. Global or regional unions (the WHO, African CDC, or West Africa Health Organization) can be powerful drivers of that cooperation [12, 27, 29]. In addition, regional workshops need to conduct to strengthen the capacity for enhanced surveillance at points of entry, infection prevention and control, risk communication, and clinical case management [27]. Countries in Africa can apply a combination of containment and mitigation measures to delay a surge in cases that could overwhelm the availability of hospital beds, while protecting the medically vulnerable, such as the elderly and those with comorbidities [27]. What is more, African countries should raise response fund to strengthen test capacity, effective surveillance and contact tracing, and isolation. For example, the Jack Ma Foundation has provided medical supplies, including diagnostics, and equipment to each of the 55 countries in Africa [30].

There are some limitations in this article. Firstly, some information, such as basic reproduction number (R0), incubation, clinical presentation, and treatment strategies are not collected in this study due to unavailable public data. Secondly, this article only considers population density, age, urban rate, and total population as factors affecting the spread of COVID-19. In fact, some factors, such as poverty and education level, may also affect the spread, and these factors will be researched in the future. Lastly, the epidemic characteristics of COVID-19 in each country are not the same, and the accuracy of the data provided by 16 countries is uncertain. Using the epidemiological characteristics of these countries to predict development trends in Africa may be biased. There are some new methods, such as multi-scale asynchronous belief percolation model, temporal networks, and dynamical clustering in electronic commerce systems [31–33], that may improve the limitations in this article. We will use these methods in the future research.

In total, COVID-19 spreads faster in Africa, with a relatively lower CFR, and it may be related to higher young people ratio. Test intensity of COVID-19 is obviously insufficient in Africa. Africa could work together to prevent the spread of COVID-19.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

LW and YC wrote the original draft and edited and critically revised the manuscript. SD, YZ, FX, and JW collected the data. YG, JW, and MY analyzed and interpreted the data. All authors substantially contributed to the conception, writing, and revision of the work and approved the final content of the manuscript.

ACKNOWLEDGMENTS

We thank professor Cui Shuzhang and Shou Songtao of the Emergency Department of Tianjin Medical University General Hospital for the article review.

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* (2020) **382**:727–33. doi: 10.1056/NEJMoa2001017
2. *COVID-19 Coronavirus Pandemic*. (2020). Available online at: https://www.worldometers.info/coronavirus/ (accessed April 29, 2020).
3. Azevedo MJ. The state of health system(s) in Africa: challenges and opportunities. In: *Historical Perspectives on the State of Health and Health Systems in Africa, Volume II. African Histories and Modernities.* Cham: Palgrave Macmillan. (2017). p. 1–73. doi: 10.1007/978-3-319-32564-4_1
4. World Population. *Population.* (2020) Available online at: https://www.worldometers.info/population/ (accessed April 29, 2020).
5. *Central Intelligence Agency.* (2020). Available online at: https://www.cia.gov/library/publications/the-world-factbook/fields/341.html (accessed April 29, 2020).
6. *African Centers for Disease Control.* (2020) Available online at: https://africacdc.org/ (accessed April 29, 2020).
7. *COVID-19 Global Pandemic Real-time Report.* (2020) Available online at: https://ncov.dxy.cn/ncovh5/view/en_pneumonia?from=dxy&source=&link=&share= (accessed April 29, 2020)
8. WHO. *COVID-19 Situation update for the WHO African Region, External Situation Report 10.* (2020). Available online at: https://reliefweb.int/report/south-africa/covid-19-situation-update-who-african-region-external-situation-report (accessed May 5, 2020).
9. Wang L, Zhang H, Lu B, Wang J, Dong Q, Shou S, et al. Effects and strategies on Emergency department Caused by patients with 2019 novel coronavirus infection in Tianjin. *Chin J Emerg Med.* (2020) **29**:515–7. doi: 10.3760/cma.j.issn.1671-0282-2020022.012
10. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: a review. *Phys Life Rev.* (2016) **18**: 66–97. doi: 10.1016/j.plrev.2016.07.005
11. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med.* (2020) **382**:1199–207. doi: 10.1056/NEJMoa2001316
12. *Fighting COVID-19 in Africa Will Be Different.* (2020). Available online at: https://www.bcg.com/publications/2020/fighting-covid-in-africa.aspx (accessed March 26, 2020).
13. Ginsburg AS, Izadnegahdar R, Berkley JA, Walson JL, Rollins N, Klugman KP. Undernutrition and pneumonia mortality. *Lancet Global Health.* (2015) **3**:e735–6. doi: 10.1016/S2214-109X(15)00222-3
14. Chanda-Kapata P, Kapata N, Zumala A. COVID-19 and malaria A symptom screening challenge for malaria endemic countries. *Int J Infect Dis.* (2020) **94**:151–3. doi: 10.1016/j.ijid.2020.04.007
15. Akombi BJ, Agho KE, Merom D, Renzaho AM, Hall JJ. Child malnutrition in sub-Saharan Africa: A meta-analysis of demographic and health surveys (2006-2016). *PLoS ONE.* (2017) **12**:e0177338. doi: 10.1371/journal.pone.0177338

16. *African Countries Move From COVID-19 Readiness to Response as Many Confirm Cases.* (2020). Available online at: https://www.afro.who.int/health-topics/coronavirus-covid-19 (accessed March 25, 2020).
17. *The Last Continent to Face up COVID-19, Africa 'Needs to Wake up.* (2020) Available online at: https://www.euractiv.com/section/coronavirus/news/the-last-continent-to-face-up-covid-19-africa-needs-to-wake-up/?_ga=2.188869291.144338809.1585666547-435116306.1585666547 (accessed March 23, 2020).
18. *African Nations Try to Overcome Shortage of COVID-19 Test Kits.* (2020). Available online at: https://www.voanews.com/africa/african-nations-try-overcome-shortage-covid-19-test-kits (accessed March 20, 2020).
19. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA.* (2020) **323**:1775–6. doi: 10.1001/jama.2020.4683
20. *What Explains Covid-19's Lethality for the Elderly? Scientists Look to 'Twilight' of the Immune System.* (2020). Available online at: https://www.statnews.com/2020/03/30/what-explains-coronavirus-lethality-for-elderly/ (accessed March 30, 2020).
21. China CCDC. *The Epidemiological Characteristics of an Outbreak of COVID19 China 2020.* (2020) Available online at: http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51 (accessed February 17, 2020).
22. Colson P, Rolain JM, Lagier JC, Brouqui P, Raoult D. Chloroquine and hydroxychloroquine as available weapons to fight COVID-19. *Int J Antimicro Agents.* (2020) **55**:105932. doi: 10.1016/j.ijantimicag.2020.105932
23. Lippi G, Simundic AM, Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). *Clin Chem Lab Med.* (2020) **58**:1070–6. doi: 10.1515/cclm-2020-0285
24. Liu R, Han H, Liu F, Lv Z, Liu Y, Feng Y, et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clin Chim Acta.* (2020) **505**:172–5. doi: 10.1016/j.cca.2020.03.009
25. Ai T, Yang ZL, Hou HY, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology.* (2020) **296**:E32–40. doi: 10.1148/radiol.2020200642
26. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle PY, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *Lancet.* (2020) **395**:871–7. doi: 10.1016/S0140-6736(20)30411-6
27. Loembé MM, Tshangela A, Salyer SJ, Varma JK, Ouma AEO, Nkengasong JN. Covid-19 in Africa: the spread and response. *Nat Med.* (2020) **26**: 996–1008. doi: 10.1038/s41591-020-0961-x
28. El-Sadr WM, Justman J. Africa in the Path of Covid-19. *N Engl J Med.* (2020) **383**:e11. doi: 10.1056/NEJMp2008193
29. Fairhead J, Leach M. One *Size Fits all? Why Lockdowns Might not be Africa's Best Bet.* (2020). Available online at: https://africanarguments.org/2020/04/22/one-size-fits-all-why-lockdowns-might-not-be-africa-best-bet/ (accessed April 22, 2020).
30. Africa CDC. (2020). Available online at: https://africacdc.org/news-item/africa-cdc-receives-third-donation-of-medical-supplies-from-jack-ma-foundation-co-hosts-global-medixchange-webinar-on-covid-19/ (accessed April 27, 2020).
31. Li HJ, Wang L. Multi-scale asynchronous belief percolation model on multiplex networks. *N J Phys.* (2019) **21**:015005. doi: 10.1088/1367-2630/aaf775
32. Bu ZH, Wang YY, Li HJ, Jiang JC, Wu ZA, Cao J. Link prediction in temporal networks: Integrating survival analysis and game theory. *Inform Sci.* (2019) **498**:41–6. doi: 10.1016/j.ins.2019.05.050
33. Li HJ, Bu ZH, Wang ZH, Cao J. Dynamical clustering in electronic commerce systems via optimization and leadership expansion. *IEEE Trans Industr Inform.* (2020) **16**:5327–34. doi: 10.1109/TII.2019.2960835

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 30 September 2020
doi: 10.3389/fphy.2020.00371

Stochastic Filtrate of Essential Workers to Reactivate the World Economy Safely

Elias Manjarrez[1]*, *Roberto Olmos*[2], *Lilia Cedillo*[3] and *Humberto A. Salazar-Ibarguen*[2]

[1] Instituto de Fisiología, Benemérita Universidad Autónoma de Puebla, Puebla, México, [2] Laboratorio Nacional de Supercómputo del Sureste de México, Benemérita Universidad Autónoma de Puebla, Puebla, México, [3] Centro de Detección Biomolecular, Benemérita Universidad Autónoma de Puebla, Puebla, México

Because there is a rapidly emerging risk that essential supply chains could collapse during the COVID-19 pandemic, new strategies are urgently needed by governments to protect workers regardless of whether or not they have symptoms. We propose a "stochastic filtrate" of such workers to reduce the risk of physical interactions with workers infected with SARS-CoV-2. Here, we suggest the random real-time-RT-PCR test of SARS-CoV-2 as a filtrating agent, although other more useful tests developed in the future to detect SARS-CoV-2 could also be used instead. Such a focused strategy, when combined with other preventive measures, could be successfully replicated in many countries to reactivate the world's economy safely. Our stochastic filtrate concept includes a mathematical framework and conceptual model. The simulations of this stochastic filtrate process support its viability.

Keywords: epidemic control, stochastic filtrate, stochastic processes, complex systems, social physics, collective phenomena, essential supply chains, COVID-19

OPEN ACCESS

Edited by:
Chengyi Xia,
Tianjin University of Technology, China

Reviewed by:
Xiao Han,
University of California, Davis, United States
Satyam Mukherjee,
Indian Institute of Management Udaipur, India

***Correspondence:**
Elias Manjarrez
eliasmanjarrez@gmail.com;
elias.manjarrez@correo.buap.mx

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *10 June 2020*
Accepted: *31 July 2020*
Published: *30 September 2020*

Citation:
Manjarrez E, Olmos R, Cedillo L and Salazar-Ibarguen HA (2020) Stochastic Filtrate of Essential Workers to Reactivate the World Economy Safely. Front. Phys. 8:371. doi: 10.3389/fphy.2020.00371

INTRODUCTION

Real-time reverse transcription-polymerase chain reaction (RT-PCR) testing is a method commonly employed in molecular biology laboratories. It is one of the most widely used laboratory techniques for detecting the severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2). The highly contagious SARS-CoV-2 has resulted in the COVID-19 pandemic, which is affecting the economy in an unprecedented way, despite extensive efforts to use real-time-RT-PCR testing of SARS-CoV-2 to mitigate virus propagation [1]. A possible reason for this failure is that many countries are only administering these tests to patients presenting COVID-19-like symptoms, or to those who were in close contact with such patients, disregarding the asymptomatic SARS-CoV-2 infected population [2]. Given the essential nature of supply chains in meeting society's needs, "focused" strategies of RT-PCR testing should be implemented to test workers needed to maintain those supply chains. Here we propose an approach, termed "stochastic filtrate" with a filtrating agent, which could be employed to timely identify and isolate, under quarantine, essential workers infected with SARS-CoV-2. We propose focused real-time-RT-PCR-testing as an optimal filtrating agent (though other future tests for early detection of SARS-CoV-2 could also be employed). Our method takes advantage of the random nature of stochastic processes, as in randomized controlled trials [3]. Such random sampling has been found to be effective in providing a "quick count" in elections, in which a random sample of polling places is used to identify the "winner" without having to count all the votes. In other words, instead of waiting for the healthcare sector to carry out extensive testing on the whole population, tests would be given to workers employed in essential businesses using a random-sampling based on the stochastic filtrate. Because workers could be infectious several days

before showing symptoms, it will be vital to apply the stochastic filtrate to all workers regardless of symptoms. Therefore, we propose that our approach provides a strategy that governments could employ to optimize and speed up testing to protect workers in essential supply chains all over the world. In this article, we present a mathematical framework and simulation to demonstrate that the stochastic filtrate of essential workers could contribute to a safe and productive workforce during the pandemic.

METHODS

Stochastic Filtrate

Briefly, to define the concept of "stochastic filtrate," we first define a "filtrating agent." It is an abstract mathematical-instrument that removes (isolates) only one particular distinctive element from a set of Z elements:

$$Z = \sum_{j=0}^{R} W_j^v + \sum_{j=0}^{F} W_j \tag{1}$$

$$W_j^v = W_j = 1 \tag{2}$$

where, W_j^v are elements to be removed and W_j represent elements to be filtrated (**Figure 1**). Here, R is the number of elements to be removed, and F is the number of elements to be filtrated (Z = R + F, is the total number of elements in the set). Thus, the filtrating agent can be compared with a "sieve," a device able to remove a distinctive component from a set of elements. In chemistry, the removed element is called the "filtride," whereas the filtrated element is the "filtrate" (as a noun). In our definition of "stochastic filtrate," we use "filtrate" as a verb.

Specifically, we define "stochastic filtrate" as a process in which a filter of N non-filtrating agents and M filtrating agents, randomly allocated, remove a particular number of distinctive elements W_j^v in a set of size Z, while the rest of the elements W_j^v and W_j from the set pass through such a filter. The following relationships combine the number of elements in the set and the stochastic filter:

$$Z \leq N + M \tag{3}$$

$$R + F \leq N + M \tag{4}$$

The stochastic filtrate is a process related to a stochastic filtrating agent. For clarity, in **Figure 1**, we illustrate the filtrating agent with a sieve symbol, and the non-filtrating agent with white circles, whereas each filtrating instrument is depicted by a vertical gray rectangle.

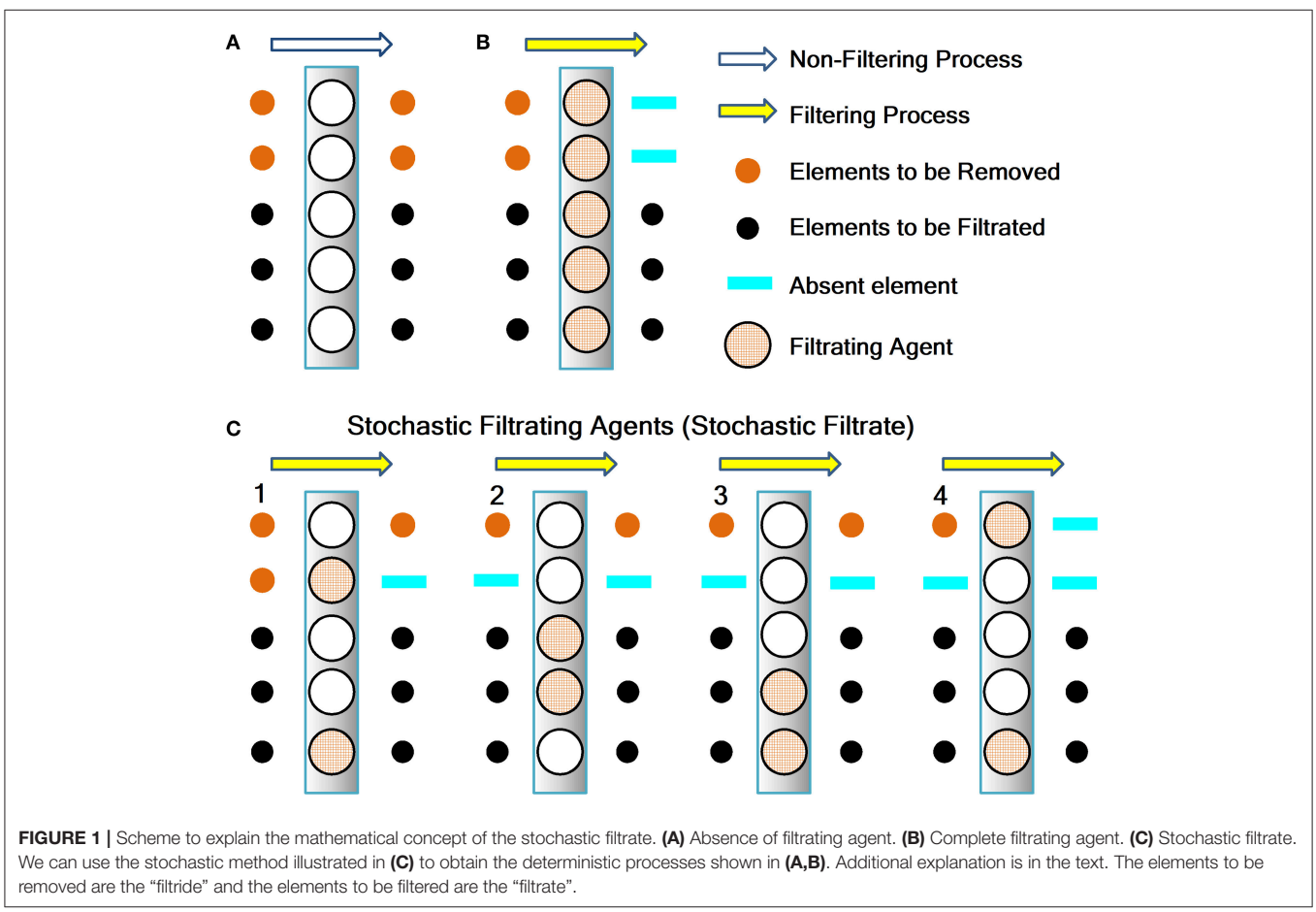


FIGURE 1 | Scheme to explain the mathematical concept of the stochastic filtrate. **(A)** Absence of filtrating agent. **(B)** Complete filtrating agent. **(C)** Stochastic filtrate. We can use the stochastic method illustrated in **(C)** to obtain the deterministic processes shown in **(A,B)**. Additional explanation is in the text. The elements to be removed are the "filtride" and the elements to be filtered are the "filtrate".

There are three types of possible filters, depending on values N and M:

1. When N = Z and M = 0, we obtain a filter with the "absence of a filtrating agent." In this case, there is no filtering process, and the distinct elements (filtrides) remain mixed without separation from the original set (**Figure 1A**).
2. When N = 0 and M = Z, we obtain a filter with a "complete filtrating agent." In this case, there is a separation of all distinctive elements (filtrides) from the set in a single trial (**Figure 1B**).
3. When $N \neq 0$ and $M \neq 0$, we obtain a filter with a "stochastic filtrating agent" (stochastic filtrate). In this case, M filtrating agents randomly remove one or several distinctive elements (filtrides) from the set. We define M as the "random sampling size" (RSS) of our stochastic filtrate. For instance, if the random sample size is two, then up to two elements from the set can be randomly separated (**Figure 1C**). A characteristic of this stochastic filtrate is that it requires several trials to remove all distinctive elements from the set. The number of trials needed for the complete extinction of unique elements is inversely proportional to M. Moreover, when M = 0, the number of trials needed for complete removal is indefinite. Additionally, when M is equal to the number of elements in the set, then the number of trials necessary for complete disappearance is one.

Note that the abovementioned stochastic filtrate in the third type of filter is a general class of filtering that reproduces the first and second types of filters, depending on the value of parameter RSS. Thus, it is interesting that these two *deterministic* filters derive from a *stochastic* one, with the level of uncertainty in the stochastic filtrate being a function of RSS. Furthermore, if we set RSS to the "random sampling size" M for the stochastic filtrate, then:

$$\mathrm{RSS} = \mathrm{M} \tag{5}$$

$$RSS \geq \sum_{j=0}^{R} W_j^{\nu} + \sum_{j=0}^{F} W_j - N \tag{6}$$

To our knowledge, this is the first time that this formalism and method of "stochastic filtrate" is introduced. In the next section, we will explain how to employ this approach, as a strategy to opportunely identify and isolate, under quarantine, essential workers infected with SARS-CoV-2. For such a purpose, we will suggest the use of the focused real-time-RT-PCR-testing as a filtrating agent.

RESULTS

Stochastic Filtrate of Workers From an Essential Company

We adapted the general conceptual framework illustrated in **Figure 1**, to SARS-CoV-2-infected and non-infected workers (see **Figure 2**). **Figure 2A** shows the scenario in which a focused filtrating agent is lacking for workers from a particular company. Such a scenario is similar to the current world situation of unfocused usage of real-time RT-PCR testing, as these tests are performed only on patients with COVID-19 symptoms, or those persons who were in close contact with such patients. Current testing is not focused on the population sector without COVID-19 symptoms, including essential supply chain workers.

The scheme in **Figure 2A** illustrates five consecutive vertical rectangles with white circles, each depicting a lack of filtrating agents. The space around the vertical boxes represents the scenario of contagion between workers each successive day (see numbers) after they come into the company. Note that an infection occurred on day 2 resulting from the close interaction between workers on day 1 (see the new orange circle). On day 3, another contagion occurred and there was the addition of another infected worker outside the company (orange circle highlighted with a ring). Note that in this scenario, there is a collapse of the company on day 4.

In contrast, with the use of a complete filtrating agent from days 1 to 5 (i.e., periodic testing using focused real-time-RT-PCR for all workers; see **Figure 2B**), the contagion is controlled inside the company, and the workforce is not affected. Although this approach is intuitive, it could be unpractical and expensive for companies with thousands of workers. However, there is a more feasible approach, as illustrated in **Figure 2C**. With this approach, the filtrating agent (real-time-RT-PCR test) is adequate to mitigate the contagion and maintain the workforce, as expected from the theory of "stochastic filtrate." In this last scenario, the number of real-time-RT-PCR tests (i.e., RSS) could be empirically adjusted depending on the number of workers. This would save money and the human resources needed to perform such focused and periodical tests. Note that the effectiveness of the strategy described in **Figure 2C** is between that of the procedures depicted in **Figures 2A,B**.

It is clear that the filtrating agent randomly changes its position across trials, from days 1 to 5 (see vertical rectangles with circles, and the horizontal yellow line in **Figure 2C**). When the filtrating agent detects an infected worker (an orange circle), then it "sieves" such an infected worker from the rest of the workers and isolates it under quarantine (horizontal blue lines). The "sieve" (i.e., the filtrating agent) changes its position in each trial according to a stochastic process. In this way, there is a chance that in repeated tests, any infected worker could be identified and timely separated. Note that the filtrating agent does not separate non-infected workers (black circles), who continue working in the company installations. In this example, there are three workers who are infected, but two others continue working (workforce), maintaining the company economically active. This is an advantage compared with the scenario illustrated in **Figure 2A**, in which all workers quickly become infected, possibly leading to the company's collapse. After quarantine, all workers could return to the company, with the exception of those with potential long-term health problems associated with COVID-19.

It is clear that the cases shown in **Figures 2B,C**, the complete and the stochastic filtrate, respectively, were useful in avoiding the company's collapse, allowing a safe economic reactivation maintained by non-infected workers. The model introduced in **Figure 2** could be generalized to any number of workers and random-sampling size across trials.

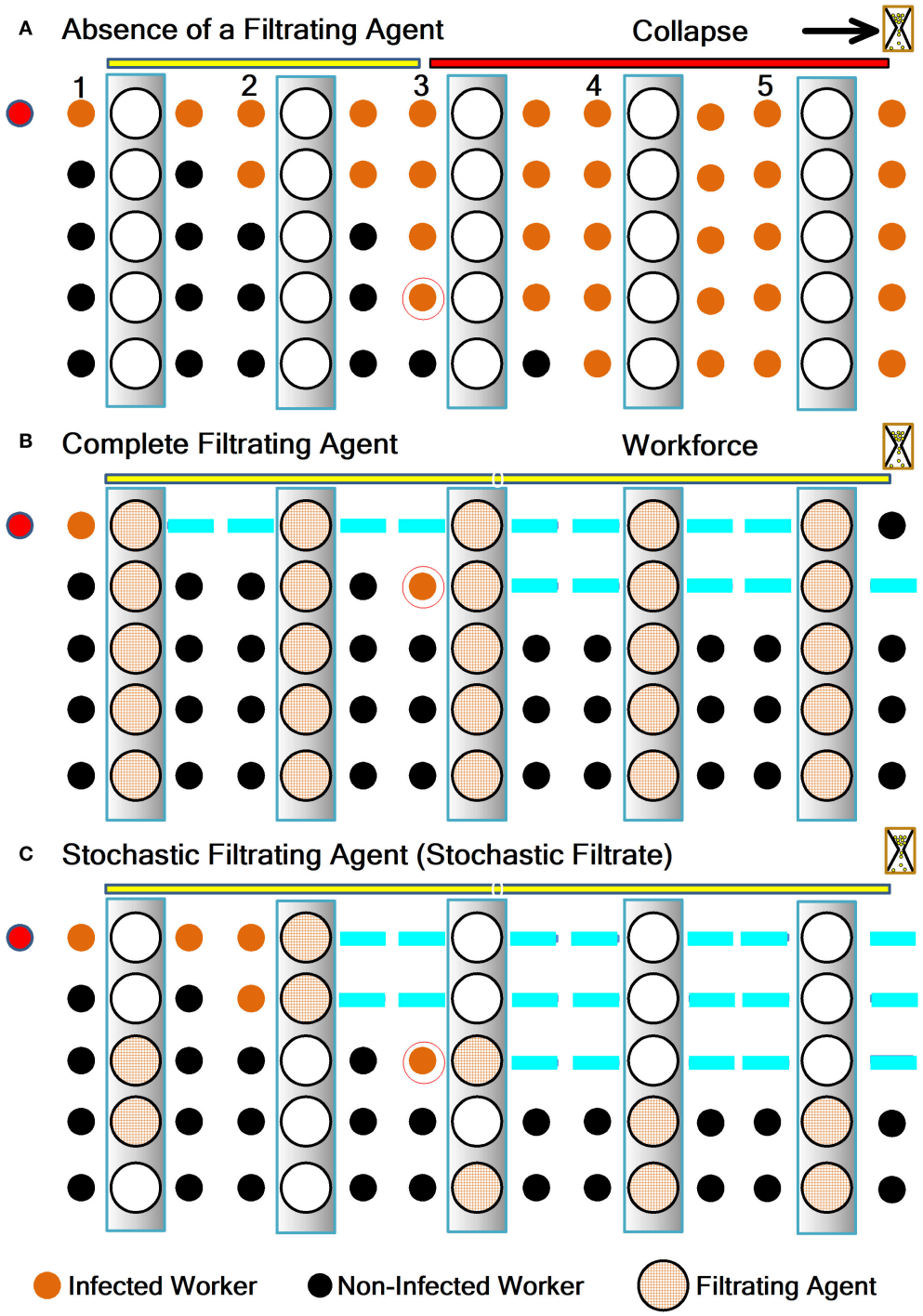


FIGURE 2 | Strategies with real-time-RT-PCR as a filtrating agent, including the stochastic filtrate. **(A)** There is no filtration process when a filtrating agent is absent. **(B)** The filtrate is highly effective when there is a complete filtrating agent (real-time-RT-PCR testing for all workers). **(C)** The case of the stochastic filtrate. Here the infected workers are the "filtride," and the non-infected workers are the "filtrate." Note that **(C)** is similar to **Figure 1C**, but for infection between workers. The blue line indicates that an infected worker is in quarantine. The filtrating agent randomly selects subjects (see vertical rectangles containing white circles and sieve circles). For this example, the random sampling size (RSS, number of filtrating agents) is two in a population of five complete filtrating agents (i.e., RSS = 40%). Interestingly, when the RSS is equal to the number of complete filtrating agents (100%), then we obtain the scenario illustrated in **(B)**. However, when the RSS is zero %, we get the situation depicted in **(A)**. Thus, our method shown in **(C)** reproduces the three scenarios. RSS is expressed as percentage of the total number of filtrating agents; i.e., when the filter is in the "complete filtrating agent" condition (N = 0 and M = Z, see Methods).

Conceptual Model of a Stochastic Process

To validate the proposal that stochastic filtrate is useful in mitigating COVID-19 contagion, we developed a conceptual model of a stochastic process for risky physical interactions between workers in the three scenarios explained in **Figure 2**. We hypothesize that asymptomatic workers carrying the virus "v" can be timely identified and isolated in focused tests with the complete filtrating agent and stochastic filtrate with real-time-RT-PCR testing.

Briefly, our conceptual model consists of two compartments in Cartesian space: "company" and "street." We also included a restrictive criterion for entering the company (i.e., when a worker tests positive). With these assumptions, we generated the three scenarios mentioned above for a random walk of infected and non-infected workers:

$$W_j^v\left(x^v, y^v, t\right), j = 1, \ldots, R \tag{7}$$

$$W_j\left(x, y, t\right), j = 1, \ldots, F \tag{8}$$

These are matricial representations for worker "j" as a function of Cartesian coordinates "xy" and time "t," where the superscript "v" indicates a "j" worker infected with the virus SARS-CoV-2. Here R is the number of workers put under quarantine and F is the number of non-infected workers to be filtrated and allowed to work at the company building.

We generated a stochastic process by assigning random numbers to the xy coordinates. Then we obtained random walk-like trajectories for each subject, with the possibility of emulating a more realistic walking behavior using filters to smooth such paths. We included the option of persons walking on the street infected with SARS-CoV-2, who are represented by:

$$P_j^v\left(x^v, y^v, t\right), j = 1, \ldots, C \tag{9}$$

We assumed that some "C" citizens could potentially infect workers in the street. To simplify, we defined a risky interaction of non-infected workers with infected workers and persons walking on the road as the case when the first crossing between trajectories occurs in proximity, and one of the workers carries the SARS-CoV-2.

$$\left(x^v, y^v, t\right) - \left(x, y, t\right) = \Delta\mu \tag{10}$$

Here, $\Delta\mu$ is a proximity value of social distancing. We employed trajectory crossings to estimate the incidence of infected workers in the three scenarios. We also considered that the trajectory crossing represents infection with a determined probability. This last point is necessary because not all persons who are walking in proximity will be infected; it will depend on various preventive measures (see Discussion section).

Furthermore, RSS is the number of real-time-RT-PCR tests performed during the stochastic filtrate of workers from a particular company, where:

$$RSS = \#\ of\ ''Real\_Time\_RT\_PCR\_test'' \tag{11}$$

$$RSS \geq \sum_{j=0}^{R} W_j^v\left(x^v, y^v, t\right) + \sum_{j=0}^{F} W_j\left(x, y, t\right) - N \tag{12}$$

$$N = \#\ of\ workers - RSS \tag{13}$$

$$Workforce\ (D) = \sum_{day=1}^{D} [\sum_{j=0}^{F} W_j\ \left(x, y, t\right)]_{day} \tag{14}$$

Here the workforce is defined as the number of non-infected workers to testing day D, who are still working after the successive stochastic filtrate. In our model, RSS can be expressed as the percentage of the total number of filtrating agents; i.e., when the filter is in the "complete filtrating agent" condition (N = 0 and M = Z, see Methods).

Python Simulation

To illustrate the viability of the conceptual model, we employed Python to simulate the three scenarios with a different number of workers (infected and non-infected) and infected-persons walking on the street. The software can emulate the number of workers in companies around the world. The simulation was based on the mathematical framework and conceptual model described in the previous section. **Figures 3A–C** show representative elements of the simulation. Black traces represent the random walk of non-infected workers, and orange and red marks indicate the random walk of infected workers and persons walking on the street, respectively. The green circles indicate the Cartesian coordinates of the occurrence of a contagion, and blue rectangles show those infected workers in quarantine after a positive result of a real-time-RT-PCR test. As an example, **Figures 3D–F** show simulated trajectories for Z = 49 workers on a working day. The initial conditions are 48 workers not infected ($F = 48$), one worker infected ($R = 1$), and one infected person ($C = 1$) walking on the street. This means that for these simulations, we used the following set of parameters and initial conditions: Z = 49, $F = 48$, $R = 1$, and $C = 1$. In **Figures 3D–F**, we also considered a random sampling size (RSS) = 40%. In the computer simulation RSS was expressed as a percentage of the total number of filtrating agents, i.e., when the filter is in the "complete filtrating agent" condition (N = 0 and M = Z, see Methods). We also included that the probability of contagion by proximity is = 20% inside and outside the company. In the previous section, we defined the parameters: F, R, C, and RSS.

To analyze our proposed concept in more detail, we ran the simulation 15 times to compute the infection incidence for 15 different companies with Z = 49 workers. We assigned an RSS = 40% for ten companies and RSS = 75% for another five companies. Then we calculated the cumulative number of cases of infected workers with SARS-CoV-2 for these companies. We obtained a graph of the grand average for the cumulative number of cases (expressed as a percentage of the total number of workers) in the three scenarios for 5 days (**Figure 4A**). Such a graph shows that the number of infected workers in the first

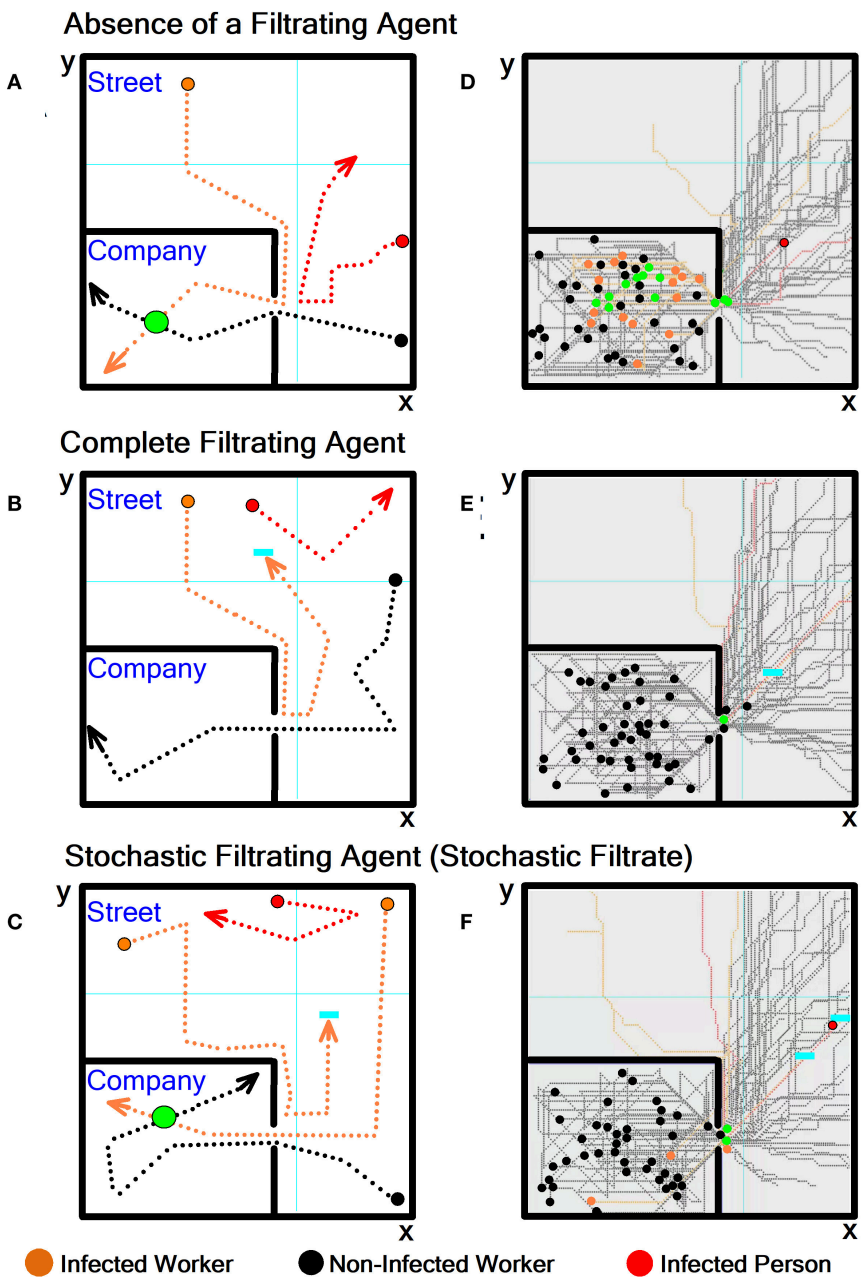


FIGURE 3 | Simulated strategies with real-time-RT-PCR as a filtrating agent. **(A–C)** Scheme of trajectories for two workers, one infected and another non-infected, and for one infected-person walking on the street. **(D–F)** The same as **(A–C)**, but for superimposed trajectories of 48 non-infected workers and one infected worker as initial conditions. In this case, there is also one infected-person walking on the street (more explanations in the text). Note that each approach yields different scenarios, with more contagions (orange circles) in situation **(D)** than in scenarios **(E,F)**. In the **Supplementary Material**, we included two movies and the Python code with detailed explanations (for purposes of reproducing or extending the analysis). The first movie simulates in sequence the three scenarios during 5 working days in real time: **(D)** Absence of a filtrating agent with RSS = 0%, **(E)** Complete filtrating agent RSS = 100%, and **(F)** Stochastic filtrating agent, where RSS = 40%. The second movie is similar to **(D–F)**, but for RSS = 75%. RSS is expressed as percentage of the total number of filtrating agents; i.e., when the filter is in the "complete filtrating agent" condition (see **Figure 2**).

scenario was significantly higher than the number of infected workers in the second and third scenarios. We also calculated the level of uncertainty (95% confidence intervals) in our simulation, utilizing the percentage of infected workers at the end of 5 days vs. four different RSS values (**Figure 4B**). This graph shows that at intermediate levels of RSS, there is an intermediate

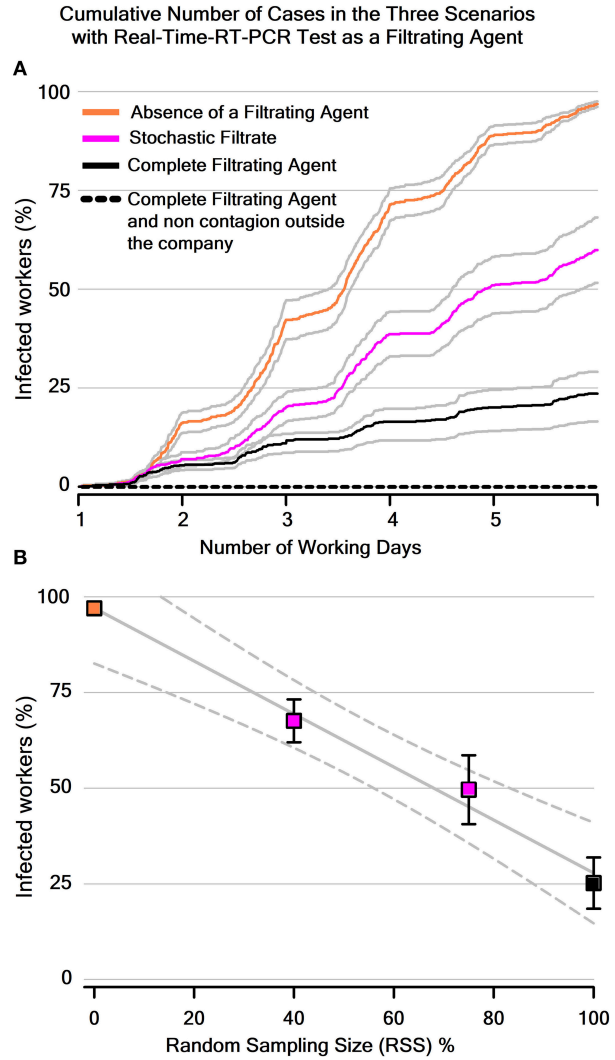


FIGURE 4 | Cumulative mean number of cases of infected workers in the three scenarios: absence of a filtrating agent, complete filtrating, and stochastic filtrate. Quantitative results from the Python simulation for 15 different companies. **(A)** Superimposed graphs show the number of infected workers (as a percentage of the total number of workers) in the absence of a filtrating agent, complete filtrating, and stochastic filtrate. Note the effectivity of the stochastic filtrate of workers compared with the lack of a filtrating agent. Gray lines indicate the standard error of the mean. In the real situation, the time between "working days" could be defined empirically. These are simulated strategies with real-time-RT-PCR as a filtrating agent, but other filtrate agents detecting SARS-CoV-2 could be employed. **(B)** Number of infected workers (as a percentage of the total number of workers) vs. random sampling size (RSS in %) employed to generate the three scenarios in the Python simulation. Note that the number of infected workers is inversely proportional to RSS. The graph also shows that the stochastic process is between two deterministic situations: "absence of a filtrate agent" and "complete filtrate agent," with an uncertainty range to quantify the number of infected workers for intermediate values of RSS (with 95% confidence intervals, dashed lines). Vertical bars represent the standard error.

uncertainty-level to calculate the number of infected workers. The graph in **Figure 4B** also illustrates a significant tendency with a low number of infected workers for high values of RSS.

Finally, we defined the workforce as the number of non-infected workers, and we obtained the graph of the workforce vs. the number of working days (**Figure 5A**). In a real context, the periodicity of such "working days" could be defined empirically (see Discussion section). Note that the workforce is 100% for the ideal case of a complete filtrating agent and no contagion from outside the company. This result indicates that both the complete filtrating agent and the stochastic filtrate strategies could help to maintain an active workforce in essential companies. We also found that the mean workforce for the three scenarios was: Absence of a Filtrating Agent = 3.0 ± 0.6%, Complete Filtrating Agent = 74.8 ± 6.7%, Stochastic Filtrate = 32.4 ± 5.6% for RSS = 40% and Stochastic Filtrate = 50.4 ± 9% for RSS = 75% (Mean ± SE). **Figure 5B** shows this workforce vs. RSS.

The results clearly illustrate how the stochastic filtrate strategy could be useful for mitigating the risk of COVID-19 infection in simulated data, thus supporting our proposal. In the **Supplementary Material**, we included the source code of the simulation software and two movies illustrating the computer simulation in operation.

DISCUSSION

Here, we introduced for the first time the concept of "stochastic filtrate," in which a mathematical instrument of non-filtrating and filtrating agents, randomly allocated, removes a particular number of distinctive elements in a set. In contrast, the rest of the items from the collection pass through such a filter. We defined "filtrating agent" as an abstract mathematical-instrument that isolates only one particular distinctive element from a group. We extended this mathematical framework to show in a model that the stochastic filtrate of workers could be employed to mitigate SARS-CoV-2 propagation in essential supply chains. For this purpose, we used focused real-time-RT-PCR-testing as an optimal filtrating agent, although other more effective tests or procedures developed in the future could be implemented instead. We compared three scenarios: "absence of a filtrating agent," "complete filtrating agent," and "stochastic filtrate." The first scenario represented the case of "unfocused," RT-PCR testing to any person in a country, but not specifically targeting workers of essential companies. The second scenario illustrated the case of focused RT-PCR testing for all workers at an essential company. Finally, the third scenario involved random focused-RT-PCR testing. Our results support the hypothesis, and thus our proposed concept that both the "complete filtrating agent" and "stochastic filtrate" could be useful to mitigate the risk of SARS-CoV-2 contagion in simulated data.

Advantages of the Stochastic Filtrate

An advantage of our concept is that it does not require the RT-PCR testing, and it could be implemented with other clinically approved tests for SARS-CoV-2 that will be available after this publication. Another advantage is that governments, companies, and citizens would most certainly benefit from this concept, given the current emergency and the long time before

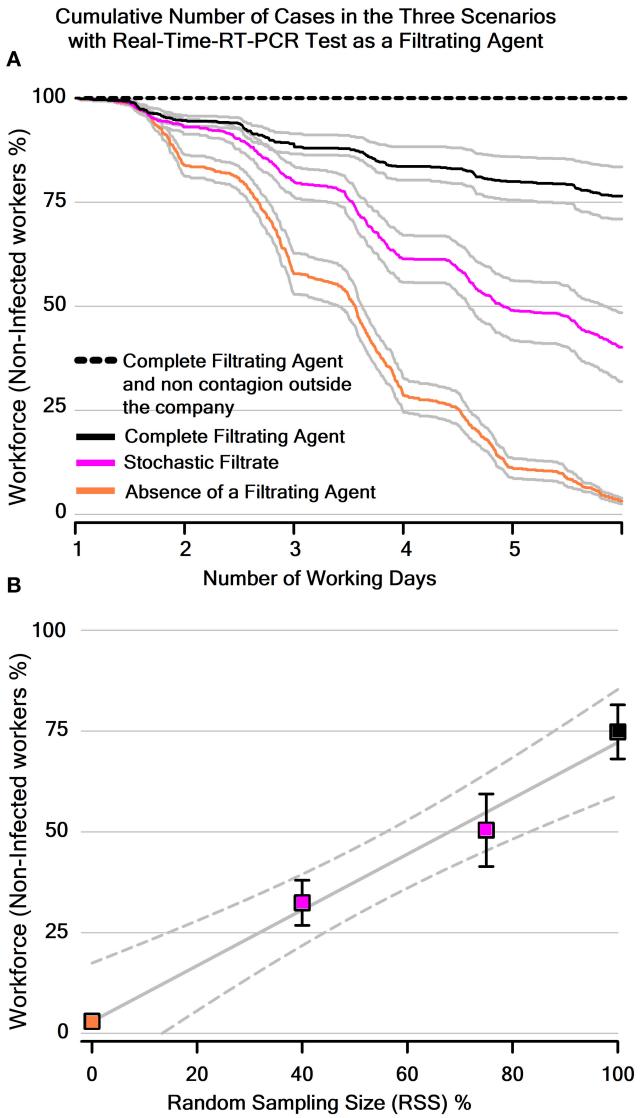


FIGURE 5 | Cumulative mean number of cases of non-infected workers ("workforce") in the three scenarios: absence of a filtrating agent, complete filtrating, and stochastic filtrate. **(A)** The same as in **Figure 4**, but for the workforce (as a percentage of the total number of workers). **(B)** Workforce vs. random sampling size (RSS in %) employed to generate the three scenarios in the Python simulation. Note that a higher RSS ensures a higher workforce (with 95% confidence intervals, dashed lines). Vertical bars represent the standard error. The graph also shows that the stochastic process is between two deterministic situations: "absence of a filtrate agent" and "complete filtrate agent." The filtrate agent could be the real-time-RT-PCR.

a vaccine for COVID-19 is available [4]. Another advantage is the random nature of the sampling, necessitating only a few RT-PCR assays per day (depending on the RSS). Such a strategy is economically convenient, given that supplies for RT-PCR analyses are scarce and overpriced in some countries. Therefore, it would be ideal if governments could freely provide, as a human right, this type of periodic test to workers involved in essential supply chains during the COVID-19 pandemic. Then, the proposed strategy of random focused-RT-PCR testing, when combined with other preventive measures against COVID-19 [5–7] and future vaccines, could be successfully implemented. Some additional precautionary measures are frequent handwashing, the combined use of medical face shields and masks [8], social distancing, body temperature sensors, and targeted isolation [9]. Because many experts warn that SARS-CoV-2 spreads through the air in respiratory aerosols (see references in the last paragraph of this section), then air-conditioning systems should be temporarily shut down and replaced by natural ventilation through open windows and doors. Another logical preventive measure is the implementation of monitoring stations outside the company doors to detect a quick response (QR) code for every worker.

Such a QR code could contain timely information about the results of RT-PCR testing of each worker, as well as of their possible risky physical interactions with others within company installations. You Only Look Once (YOLO) based human action recognition and localization software [10] combined with a QR code, for instance, could be employed to analyze walking trajectories. Alternative technologies are available from many private companies and universities, such as social distancing software [11] with contact data tracing and a variety of monitoring devices to mitigate the COVID-19 pandemic. Recently entries to the market include wearable badge holders based on proximity sensors with an alarm and contact tracing with mapping software to be used inside the workplace [12]. Also, there has been a proposal to employ prediction methods to forecast the number of COVID-19 cases, which could be useful to government decision-making on the measures to impose [13, 14]. All of these new technological devices and prevention methods could be implemented together with the stochastic filtrate for essential worker populations. We hypothesize that the combined and rigorous implementation of such measures could keep the number of infected workers low in order to maintain a highly productive workforce, as illustrated by the horizontal dashed lines in **Figures 4A**, **5A**. Another significant advantage of our method is that it could allow us to choose, as a filtrating agent, another simpler and quicker testing-technique in the future to detect SARS-CoV-2 with the best precision.

Limitations of the Stochastic Filtering With a Filtrating Agent

The first limitation of our proposal is that empirical research is needed to determine the number of days between the working days illustrated in **Figures 4**, **5**. A possibility could be to employ a 10-4 cyclic strategy, i.e., a cyclic schedule of a 4-day work period and 10-day lockdown, or similar variants, as suggested in recent models [7]. Another limitation is that some workers will not agree to the privacy and data protection policy necessary for QR information, or the complementary use of wearable proximity sensors and video surveillance in offices and other workplace environments. The contributions of psychologists and social scientists would be very helpful in addressing these

challenges. Governments could also assist by fostering empathy campaigns during the "new reality" in the COVID-19 pandemic. Another concerning limitation is that some governments may disregard the proposal described here due to political issues or conflicts of interest with the selective distribution of funds to protect essential workers with random focused-RT-PCR testing. However, it would be crucial to highlight that the proportion of resources necessary for such tests would be only a small part of the total funds used for all citizens. This means that the use of random focused-RT-PCR assays for the rest of the population would still be available.

Accuracy of the filtrating agent (real-time-RT-PCR test) is another relevant issue of concern. False-negative cases could affect the valuable utility of RT-PCR assays to correctly identify and isolate infected workers. There is evidence that 29% of patients infected with SARS-CoV-2 could receive an initial RT-PCR false-negative result [15]. Therefore, such findings reinforce the need for repeated testing in persons with suspicion of SARS-CoV-2 [16]. Another alternative is the combined use of other types of available tests to detect SARS-CoV-2, or IgM antibodies against the virus, thus raising the quality of diagnostic accuracy. Importantly, correct sampling procedures with high-quality extraction of viral RNA and the real-time RT-PCR kit performed by experts in biomolecular assays could also improve the efficiency and reduce inaccurate results [16].

Simulation Limitations, Advantages, and Challenges

A limitation of our Python simulation is that it is simple, due to the use of only two compartments and one restriction parameter. However, simplicity in simulation also offers advantages. For example, simple simulations are easy to understand, and they motivate more sophisticated uses. Nevertheless, other simulation studies of our conceptual model and method could include more compartments of different sizes within the company boundaries, a different number of workers in different working spaces, and more detailed probabilities of incorrect use of masks or incomplete handwashing, quantified by YOLO-like approaches. Other more sophisticated parameters could include decision making in emulated workers via artificial intelligence procedures, big data strategies, and 3D avatar-based representations. In this form, walking trajectories and physical interactions among several emulated workers may be more realistic. Inclusion of "decision making" with avatars, based on social networks with positive and negative prevention information, could also be useful to quantify a more realistic epidemic propagation [17]. It would also be interesting to extend our simulation by using epidemiologic models of two-layer multiplex networks, where the upper layer represents the awareness diffusion regarding epidemics, and the lower layer expresses the epidemic propagation [18]. On the other hand, future simulations should consider preventive measures to eliminate respiratory aerosols with SARS-CoV-2, a form of virulent particulate matter (PM) 2.5 in the air [19]. The computer simulation of these respiratory aerosols is relevant, given that essential workers very often share close spaces. It would be convenient to simulate far-UVC light [20] and electrostatic precipitators [21, 22] to eliminate SARS-CoV-2 in virulent PM 2.5 to control the air quality in such spaces. This new realistic computer simulation could help to capture real-world situations and thus examine more effective measures to prevent and control infectious spreading among essential workers during the COVID-19 pandemic.

CONCLUSION

The stochastic filtrate of essential workers could undoubtedly contribute to protect and efficiently improve each country's economy during the different stages of the "new reality" in the COVID-19 pandemic.

DATA AVAILABILITY STATEMENT

All data is available in the main text or the **Supplementary Material**. All data, code, and materials used in the analysis are open to any researcher for purposes of reproducing or extending the analysis, and they are available at the **Supplementary Material** section.

AUTHOR CONTRIBUTIONS

EM conceived the article, mathematical framework, and conceptual model of stochastic filtrate and wrote the paper. Coauthors RO and HAS-I implemented the Python simulation, and LC provided information about the real-time-RT-PCR test, SARS-CoV-2, and COVID-19. EM supervised all aspects of the work. All the authors revised and approved the manuscript.

FUNDING

CONACyT Fronteras de la Ciencia #536 (EM), Cátedra Marcos Moshinsky (EM), and VIEP-PIFI-FOMES-PROMEP-BUAP-Puebla (EM), Comité de Internacionalización and VIEP BUAP (HAS-I, EM), México.

ACKNOWLEDGMENTS

We thank Dr. Judith Percino for encouraging the publication of this research and Dr. David Pinto for comments on the manuscript. We thank John Reid and Paul Kretchmer for proofreading the English document.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00371/full#supplementary-material

Supplementary Video 1 | Video 1.MP4. Movie of the Python simulation for RSS = 40%. The first stage of this video shows part of a code and starts with simulated action after 4 s. The first five trials reproduce scenarios as illustrated in **Figure 3D**; whereas the second group of five trials, reproduce the situation of

Figure 3E. The last five trials reproduce a scene as illustrated in **Figure 3F**. Green symbols indicate the moment of occurrence of new infections and orange circles represent the infected workers. The blue rectangles represent the infected workers in quarantine. We considered the same parameters mentioned in the Results section.

Supplementary Video 2 | Video 2.MP4. Movie of the Python simulation for RSS = 75%. Note that the stochastic filtrating agent could be most efficient for RSS = 75% than for RSS = 40%.

Simulation Code | SimulationCode-40percentRSS.py (to reproduce random scenarios similar to those illustrated in Video 1.MP4) (Here we used porTest = 40 in the main Python code).
SimulationCode-75percentRSS.py (to reproduce random scenarios similar to those illustrated in Video 2.MP4) (Here we used porTest = 75 in the main Python code). actors.py positions.py. The code is in the Data Sheet 1.ZIP file and is open to any researcher for purposes of reproducing or extending the analysis.

REFERENCES

1. Liu R, Han H, Liu F, Lv Z, Wu K, Liu Y, et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clin Chim Acta.* (2020) **505**:172–5. doi: 10.1016/j.cca.2020.03.009
2. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health.* (2020) **8**:e488–96. doi: 10.1016/S2214-109X(20)30074-7
3. Haushofer J, Metcalf CJE. Which interventions work best in a pandemic? *Science.* (2020) **368**:1063–5. doi: 10.1126/science.abb6144
4. Corey BL, Mascola JR, Fauci AS, Collins FS. A strategic approach to COVID-19 vaccine R&D. *Science.* (2020) **368**:948–50. doi: 10.1126/science.abc5312
5. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health.* (2020) **5**:e261–70. doi: 10.1016/S2468-2667(20)30073-6
6. Balachandar V, Mahalaxmi I, Kaavya J, Vivekanandhan G, Ajithkumar S, Arul N, et al. COVID-19: emerging protective measures. *Eur Rev Med Pharmacol Sci.* (2020) **24**:3422–5. doi: 10.26355/eurrev_202003_20713
7. Karin O, Bar-On YM, Milo T, Katzir I, Mayo A, Korem Y, et al. Adaptive cyclic exit strategies from lockdown to suppress COVID-19 and allow economic activity. *medRxiv [Preprint].* (2020) 2020.04.04.20053579.
8. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect Dis Model.* (2020) **5**:293–308. doi: 10.1016/j.idm.2020.04.001
9. Neufeld Z, Khataee H, Czirok A. Targeted adaptive isolation strategy for Covid-19 pandemic. *medRxiv [Preprint].* (2020) 2020.03.23.20041897.
10. Shinde S, Kothari A, Gupta V. YOLO based human action recognition and localization. *Proc Comput Sci.* (2018) **113**:831–8. doi: 10.1016/j.procs.2018.07.112
11. Braithwaite I, Callender T, Bullock M, Aldridge R. Automated and semi-automated contact tracing: protocol for a rapid review of available evidence and current challenges to inform the control of COVID-19. *medRxiv [Preprint].* (2020) 2020.04.14.20063636.
12. Nguyen CT, Saputra YM, Huynh NV, Nguyen NT, Khoa TV, Tuan BM, et al. Enabling and emerging technologies for social distancing: a comprehensive survey. *arXiv [Preprint].* arXiv:2005.02816 (2020)
13. Perc M, Gorišek Miksić N, Slavinec M, StoŽer A. Forecasting COVID-19. *Front Phys.* (2020) **8**:127. doi: 10.3389/fphy.2020.00127
14. Tsallis C, Tirnakli U. Predicting COVID-19 peaks around the world. *Front Phys.* (2020) **8**:217. doi: 10.3389/fphy.2020.00217
15. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, Del Campo R, Ciapponi A, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *medRxiv [Preprint].* (2020) 2020.04.16.20066787.
16. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev.Mol Diagn.* (2020) 20:453. doi: 10.1080/14737159.2020.1757437
17. Wang ZS, Xia CY, Chen ZQ, Chen GR. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern.* (2020). doi: 10.1109/TCYB.2019.2960605
18. Wang ZS, Guo QT, Sun SW, Xia CY. The impact of awareness diffusion on SIR-like epidemics in multiplex networks. *Appl Math Comput.* (2019) **349**:134–47. doi: 10.1016/j.amc.2018.12.045
19. 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: 10.1073/pnas.2009637117
20. Buonanno M, Welch D, Shuryak I, Brenner DJ. Far-UVC light (222 nm) efficiently and safely inactivates airborne human coronaviruses. *Sci Rep.* (2020) **10**:10285. doi: 10.1038/s41598-020-67211-2
21. Kettleson EM, Ramaswami B, Hogan CJ, Lee MH, Statyukha GA, Biswas P, et al. Airborne virus capture and inactivation by an electrostatic particle collector. *Environ Sci Technol.* (2009) **43**:5940–6. doi: 10.1021/es803289w
22. Kettleson EM, Schriewer JM, Buller RML, Biswas P. Soft-X-ray-enhanced electrostatic precipitation for protection against inhalable allergens, ultrafine particles, and microbial infections. *Appl Environ Microbiol.* (2013) **79**:1333–41. doi: 10.1128/AEM.02897-12

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 09 October 2020
doi: 10.3389/fphy.2020.574514

Conditions for a Second Wave of COVID-19 Due to Interactions Between Disease Dynamics and Social Processes

Sansao A. Pedro[1], *Frank T. Ndjomatchoua*[2], *Peter Jentsch*[3,4], *Jean M. Tchuenche*[5], *Madhur Anand*[3] *and Chris T. Bauch*[4]*

[1] *Departamento de Matemática e Informática, Universidade Eduardo Mondlane, Maputo, Mozambique,* [2] *Sustainable Impact Platform, Geospatial Science and Modelling Cluster, International Rice Research Institute, Metro Manila, Philippines,* [3] *School of Environmental Sciences, University of Guelph, Guelph, ON, Canada,* [4] *Department of Applied Mathematics, University of Waterloo, Waterloo, ON, Canada,* [5] *Avenir Health, Glastonbury, CT, United States*

OPEN ACCESS

Edited by:
Zhanwei Du,
University of Texas at Austin,
United States

Reviewed by:
Sen Pei,
Columbia University, United States
Chao Gao,
Southwest University, China

****Correspondence:***
Chris T. Bauch
cbauch@uwaterloo.ca

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *20 June 2020*
Accepted: *27 August 2020*
Published: *09 October 2020*

Citation:
Pedro SA, Ndjomatchoua FT, Jentsch P, Tchuenche JM, Anand M and Bauch CT (2020) Conditions for a Second Wave of COVID-19 Due to Interactions Between Disease Dynamics and Social Processes. Front. Phys. 8:574514. doi: 10.3389/fphy.2020.574514

In May 2020, many jurisdictions around the world began lifting physical distancing restrictions against the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This gave rise to concerns about a possible second wave of coronavirus disease 2019 (COVID-19). These restrictions were imposed in response to the presence of COVID-19 in populations, usually with the broad support of affected populations. However, the lifting of restrictions is also a population response to the accumulating socio-economic impacts of restrictions, and lifting of restrictions is expected to increase the number of COVID-19 cases, in turn. This suggests that the COVID-19 pandemic exemplifies a coupled behavior-disease system where disease dynamics and social dynamics are locked in a mutual feedback loop. Here we develop a minimal mathematical model of the interaction between social support for school and workplace closure and the transmission dynamics of SARS-CoV-2. We find that a second wave of COVID-19 occurs across a broad range of plausible model input parameters governing epidemiological and social conditions, on account of instabilities generated by behavior-disease interactions. The second wave tends to have a higher peak than the first wave when the efficacy of restrictions is greater than 40% and when the basic reproduction number R_0 is less than 2.4. Surprisingly, we also found that a lower R_0 value makes a second wave more likely, on account of behavioral feedback (although a lower R_0 does not necessarily cause more infections, in total). We conclude that second waves of COVID-19 can be interpreted as the outcome of non-linear interactions between disease dynamics and social behavior. We also suggest that further development of mathematical models exploring behavior-disease interactions could help us better understand how social and epidemiological conditions together determine how pandemics unfold.

Keywords: COVID-19, epidemic model, behavioral fatigue, coupled behavior-disease system, SARS-CoV-2, evolutionary game theory, imitation dynamics, social learning

1. INTRODUCTION

The COVID-19 pandemic has given rise to an "epidemic of models" [1]. Diverse mathematical models of SARS-CoV-2 transmission (the virus that causes COVID-19 disease) have been instrumental in capturing infection dynamics and informing public health control efforts to mitigate the COVID-19 pandemic and reduce the mortality rate. The concept of "flattening the curve" comes from model outputs that show how reducing the transmission rate through efforts such as contact tracing and physical distancing can lower and delay the epidemic peak [2].

On account of limited options for pharmaceutical interventions such as vaccines, and inadequate testing capacity in many jurisdictions, the COVID-19 pandemic has also been characterized by large-scale physical distancing efforts–including school and workplace closure–being adopted by entire populations despite heavy economic costs. Mathematical models of SARS-CoV-2 transmission and control show that physical distancing can mitigate the pandemic [2–5] and this has subsequently been backed up by empirical analyses of case notification data. These analyses show how mitigation measures have reduced the effective reproduction number of SARS-CoV-2 below one, meaning that each infected case infects less than one person on average [6–8]. However, the population's willingness to support school and workplace closures could wane over time, as the economic costs of closure accumulate [9]. This has given rise to the possibility of a second wave of COVID-19 in many populations.

The large role played by physical distancing during the COVID-19 pandemic exemplifies a coupled behavior-disease system, in which human behavior influences infectious disease transmission and *vice versa* [10–16]. These systems are part of a broader class of pervasive systems in which human behavior both influences, and responds to, the dynamics of our environment. Hence, one might better speak of a single, coupled human-environment system, instead of just human systems or environmental systems in isolation from one another [17–19].

The interactions between disease dynamics and behavioral dynamics in COVID-19 are emphasized by research showing that the perceived risk of SARS-CoV-2 infection is a predictor of adherence to physical distancing measures [20] and moreover that individuals respond to the presence of COVID-19 cases in their population by increasing their physical distancing efforts [21]. In turn, physical distancing has been shown to reduce the number of cases [6], completing the loop of coupled behavior-disease dynamics. Some models have already begun exploring this interaction between disease dynamics and individual behavior and/or public health policy decision-making for COVID-19 [5, 16, 22–26]. The emergence of a second wave of COVID-19 on account of population attitudes to physical distancing has also been explored in mathematical models [27].

The social aspects of behavior-disease interactions seem to be relevant for COVID-19 decision-making. Individuals do not necessarily make the best possible (most rational) response to the presence of COVID-19 cases in their population. Instead, it has been found that political leaders can be influential in convincing individuals to change their physical distancing efforts [23]. Additionally, jurisdictions experiencing outbreaks that start relatively late appear to learn from the experiences of jurisdictions that were affected earlier [28]. Meanwhile, other research emphasizes a need for more work on the socio-economic aspects of the pandemic [29]. These findings suggest that imitation and social learning processes are important for understanding interactions between disease dynamics and decision-making for COVID-19, which ultimately determine the epidemic curve.

Here we model the coupled behavior-disease dynamics of SARS-CoV-2 transmission and population support for school and workplace closure, using a simplified theoretical model. We opted for a simple model that avoids heterogeneities because our objective is to gain insights into potential interactions between social and behavioral dynamics. Public opinion evolves according to social learning rules [10, 18], and public opinion in support of closure depends both on COVID-19 case incidence and accumulated socio-economic losses due to school and workplace closure. A central decision-maker chooses a time to initially close schools and workplaces when the outbreak begins, but may subsequently open and close them again depending on how public opinion ebbs and flows. Meanwhile, disease dynamics are described by a compartmental epidemic model [30]. The details of our mathematical model are described in the section 2. We analyze the model to characterize the conditions that give rise to a second wave of COVID-19 in the population.

2. METHODS

2.1. Model Equations

Transmission dynamics are given by an SEIR model, modified to take physical distancing into account,

$$\frac{dS}{dt} = -\beta(1 - C(t))SI, \frac{dE}{dt} = \beta(1 - C(t))SI - \sigma E,$$
$$\frac{dI}{dt} = \sigma E - \gamma I, \frac{dR}{dt} = \gamma I, \quad (1)$$

where S is the proportion of susceptible individuals ("susceptible"), E is the proportion of individuals who have been infected but are not yet infectious ("exposed"), I is the proportion of individuals who are both infected and infectious ("infectious"), and R is the proportion of individuals who are no longer infectious ("removed"). The time-varying parameter $C(t)$ captures the impact of school and workplace closure on the transmission of SARS-CoV-2. β is the baseline transmission rate in the absence of school/workplace closure, σ is the time rate at which an exposed person becomes infections, and γ is the time rate at which an infectious person recovers. We use an SEIR model since they are appropriate for population-level modeling epidemics of acute, self-limiting infections that confer natural immunity [30, 31]. Since our focus is on physical distancing and lockdown, we do not include compartments for testing, contact tracing and asymptomatic transmission.

The decision-maker decides to "turn on" closure at some time t_{close}, and then decides to "turn off" closure when population

support for closure, $x(t)$ drops below 50%. Hence $C(t)$ is given by:

$$C(t) = \begin{cases} 0 & t < t_{close} \text{ or } x < 50\%, \\ C_0 & t \geq t_{close} \text{ and } x \geq 50\%, \end{cases}$$

where C_0 is a combined measure of how many workplaces are closed (the remainder being essential workplaces such as hospitals) as well as the effectiveness of physical distancing in those workplaces that remain open.

Approaches to modeling human opinions and decision-making vary greatly and include agent-based network models based on complex systems science [32–34], evolutionary game theory (imitation dynamics) [10, 11, 13], mathematical models based on psychological theory [35] and other approaches. For our study system, the imitation dynamic approach is suitable because (1) imitation dynamics are sufficient to describe the population-level opinion dynamics when individuals learn from one another and response to changes in their utility functions, as epidemic and socio-economic conditions evolve [36]; (2) differential equations are usually easier to analyze (either rigorously or through numerical analysis) than agent-based models. The percentage of the population that supports school and workplace closure, x, evolves according to an imitation dynamic as:

$$\frac{dx}{dt} = \kappa x(1-x)(\omega I - \epsilon L), \tag{2}$$

where κ is the social learning rate, ω is sensitivity to infection prevalence, and ϵ is sensitivity to accumulated socio-economic losses L. Support for closure goes up when the prevalence of infection goes up, but it declines when the accumulated socio-economic losses, L, become too large. The quadratic term $x(1-x)$ represents a social learning dynamics where individuals sample others at some rate, and they change opinion based on the utility difference $\omega I - \epsilon L$ (A full derivation of this type of differential equation appears in [18]). The social learning rate κ represents how often individuals discuss opinions about lockdown. The parameter ω control how the population opinion about lockdown reacts to the prevalence of infection, and it is influenced by the perceived risk of the severity of infection (frequency of severe cases, hospitalizations, and deaths). The parameter ϵ controls how population opinion about lockdown reacts to socio-economic losses and is influenced by the perception of how severe the socio-economic losses as a function of media coverage, for instance.

We can absorb ω into κ, yielding:

$$\frac{dx}{dt} = \kappa\omega \cdot x(1-x)(I - \epsilon/\omega \cdot L) \tag{3}$$

and then, setting $\kappa' = \kappa\omega$ and $\epsilon' = \epsilon/\omega$ and dropping the primes for simplicity we obtain

$$\frac{dx}{dt} = \kappa x(1-x)(I - \epsilon L). \tag{4}$$

Finally, the variable L is a phenomenological representation of accumulated socio-economic losses obeying

$$\frac{dL}{dt} = \alpha C(t) - \delta L, \tag{5}$$

where α controls the rate at which school and workplace closure impacts socio-economic health of the population, and δ is a decay rate that represents adjustment to baseline losses. These two parameters represent the sum effect of multiple processes. For instance, α is influenced by what proportion of workers are affected by lockdown through loss of employment or working hours; household savings and debt; and economic stimulus packages. Similarly, δ is determined by economic discounting, the ability of individuals to adjust to new economic circumstances (for instance by offering new products and services to meet demand in a pandemic market), and other factors.

2.2. Parameterization

A full list of parameter definitions, baseline values, and literature sources appears in **Table 1**. The transition rates σ and γ, were set based on COVID-19 epidemiological literature [37–39, 44, 45], while the transmission rate β was estimated. Note that the last compartment of the model, R, does not correspond to a stage of illness preceding recovery but rather a stage of infectiousness [31], which wanes quickly after the imposition of case isolation, in addition to the decline in viral shedding after the first 5 days [46]. Moreover, the infectious stages is preceded by a latent stage in which the virus is still replicating inside its new host until it can reach a level where the host can transmit the infection to others. These features of COVID-19 disease history guided our choice of γ and σ. Since the prevalence $I(t)$ as used in the model is different from case incidence as appears in daily lab-confirmed case reports, a new state

TABLE 1 | Parameter values, baseline values, and literature sources.

Parameter	Meaning	Baseline value	Range	Source
$1/\sigma$	Latent period	2.5 days	2.0–3.0	[37]
$1/\gamma$	Infectious period	5 days	3.0–7.0	[37–39],
β	Transmission rate	0.54/day	0.42–0.54	[40], Calibrated
C_0	Closure efficacy	0.63	0–0.67	$C_0 = 1 - 1/R_0$, [30]
α	Rate of socio-economic loss	0.0657/day	0.00273–0.0822	Calibrated
δ	Discounting rate	0.0033/day	0.00014–0.0041	$\delta = 0.05\alpha$, [41]
κ	Social learning rate	1.5/day	1–500	Calibrated; [42, 43]
ϵ	Sensitivity to infection	0.005	0.001–0.01	Calibrated

variable denoting daily infection incidence, I_i was defined as the difference between accumulated health outcomes. That is, $I_i(t) = H(t) - H(t-1)$, where $H(t)$ is the cumulative number of cases at time t:

$$\frac{dH}{dt} = \sigma E(t). \quad (6)$$

The state variable $I_i(t)$ was used when fitting the model to daily case notifications was required.

Because ϵ, κ, and α and ω encapsulate many different factors, we take the approach of inferring parameter values by fitting the model to data. (We also explore the impact of variation in these parameter values in sensitivity analysis through parameter planes). To avoid over-fitting, δ is fixed based on commonly used discounting rates. The social parameters κ and ϵ were calibrated. While κ dictates how quickly population opinion changes, ϵ dictates how sensitive the population is to changes in case reports relative to socio-economic losses. κ was estimated from behavior early in the epidemic when socio-economic losses are small as $dx/dt \approx \kappa x(1-x)I$. We derived $I(t)$ from reports of confirmed positive cases from the early stages of the United States epidemic, adjusted by a case under-ascertainment factor of 8.7 in the United States [42]. We used 21 January 2020 as the initial date of the epidemic, when the first case of COVID-19 was reported in the United States. Most populations rapidly adopted physical distancing measures against COVID-19. Gallup polls indicate that 59%/79%/92% of the United States public avoided going to events with large crowds, as of 13-15 March/16-18 March/20-22 March 2020 respectively [43]. Similarly, 30%/54%/72% avoided public places, and 23/46/68% avoided small gatherings [43]. Taking the average of these responses across the three question types, we obtain that $x(52) = 0.373$, $x(59) = 0.597$, $x(66) = 0.773$ where time is measured in days since January 21. Finally, we shifted these points forward 14 days as physical distancing at time t will not be reflected in infection data until $t + 14$ due to the delays in testing and symptom recognition. We assumed $x(0) = 0.25$ when fitting to these three data points using least-squares error minimization for the κ estimation.

To estimate β we used least-squares error minimization to fit the modeled $I_i(t)$ to lab-confirmed daily case reports in the United States from the early epidemic [40]. The fitted infection trajectory, I_i is in good agreement with reported cases in the US during the initial phase of the epidemic leading up to 4 April 2020 (**Supplementary Figure 1b**). Our inferred estimate $\beta = 0.54$/day yields a basic reproduction number $R_0 \approx \beta/\gamma = 2.7$ [30], which agrees well with published estimates of R_0 for SARS-CoV-2 [47]. For the special case where there is an absence of any control measures, the model predicts that about 80% of the population becomes infected by the end of the outbreak (**Supplementary Figure 1a**).

The remaining two parameters, ϵ and α, were calibrated to obtain the result that x remains high after the initial surge in support for closure, but begins to drop after 2 months. This period of time was based on the observation that 2 months that have elapsed since the declaration of the national emergency in the United States on 13 March 2020, and the process of re-opening state economies that has unfolded over the month of May 2020. These two parameters control the timescale of lifting school and workplace closure based on its socio-economic impacts. Finally, the parameter δ was set such that $\delta = 0.05\alpha$ on the basis of commonly used discounting rates in economics and assuming that economic losses accumulated through the $\alpha C(t)$ term would be discounted at a rate of 5% per year [41].

In order to illustrate curve flattening and show that the model has the expected response to reduction in the transmission rate due to closure, we generated model timeseries of $I(t)$ for the special case where closure is applied throughout the entire outbreak. The epidemic curve for different values of the closure efficacy C_0 is shown, ranging from $C_0 = 0$ (no intervention) to $C_0 = 0.6$ (**Supplementary Figure 1c**). The timeseries show that the epidemic curve is flattened and delayed as closure becomes more efficacious, which reduces peak demand for intensive care beds and buys time for developing pharmaceutical interventions like vaccines and antiviral drugs, improving testing capacity, and establishing novel approaches to patient care. For the remainder of our analysis, to determine C_0 it was assumed that C_0 should be large enough to bring the effective reproduction number R_{eff} below 1, reflecting the observed success in multiple jurisdictions where physical distancing and closure have maintained $R_{eff} < 1$ [6–8]. Hence we chose $C_0 = 1 - 1/R_0$ based on the elimination threshold for the SEIR model [30]. We also assumed $t_{close} = 20$ days but in practice, our second requirement that $x \geq 50\%$ was not reached until after 20 days in all of the model simulations. We analyzed numerical simulations of our model to determine the conditions where one or more waves of SARS-CoV-2 infections could occur. A wave was defined as a local maximum in the prevalence $I(t)$, and the simulation time horizon was 730 days (2 years).

3. RESULTS

3.1. Mechanisms Causing a Second Wave

At our baseline parameter values, time series of infection prevalence $I(t)$ and support for closure $x(t)$ exhibit non-trivial time evolution, including a second wave of COVID-19 infections (**Figure 1**). These results illustrate the basic mechanisms underlying the model dynamics. As infection prevalence grows, support for closure rises and eventually crosses the 50% threshold by $t = 80$ days. After this, infection prevalence peaks and begins to decline. Support remains at a high plateau for a period of 2 months. After this period, infection prevalence remains small while the socio-economic impacts of lockdown continued to mount. This causes waning of support for lockdown, and hence restrictions are lifted by $t = 160$ days. Shortly thereafter, prevalence begins to rise again. Support for closure correspondingly rises again, but not quickly enough to prevent a second wave of COVID-19 with a peak at $t = 240$ days that is higher than the peak of the first wave.

3.2. Epidemiological Conditions for a Second Wave

Our goal was to gain insight into the conditions that generate a second wave of COVID-19, and to test the robustness of the predicted result at our baseline parameter values illustrate in **Figure 1**. Hence we explored the model dynamics in the neighborhood of our baseline parameter values (**Table 1**) using parameter planes that show how the dynamical regimes of the model vary with changes in two model parameters (one on each axis) around the baseline values. We explored two dynamical outcomes: the number of waves in the course of the entire pandemic, and the ratio of the peak height of the second wave to the peak height of the first wave.

We start by exploring the effects of variation in the epidemiological parameters β (transmission rate) and γ (inverse of the average duration of infectiousness) in **Figure 2**, while keeping the rest of the parameters at baseline values. The results show that one, two, or three waves are possible under variation in these parameter values. A second wave characterizes most of the $\beta - \gamma$ plane, however. A third wave appears when $1/\gamma$ is between 5 and 7 days and β is up to 0.4/day, corresponding to $R_0 < 2$ (we note that most estimates place $R_0 > 2$ for SARS-CoV-2 [47, 48]). The second peak may be higher or lower than the first peak, depending on the β and γ parameter combinations. For $R_0 > 2.4$, the second peak tends to be lower than the first, while for $R_0 < 2.4$ is higher.

The increase in the number of waves with a decrease in β (or equivalently, a decrease in the basic reproduction number R_0) is notable and surprising. The classical SEIR model without behavior shows the opposite effect: increasing R_0 when the endemic equilibrium is stable causes damped oscillations to be sustained for a longer period before the oscillations die down. Our model shows different dynamics than the classical SEIR model on account of strong behavioral feedback. When R_0 is sufficiently high, the infection can pass through the population rapidly and cause a large amount of herd immunity to build up before the population response causes a late dampening of the epidemic curve. As a result of herd immunity, a second wave is not possible. But when R_0 is smaller, the spread of SARS-CoV-2 is slow enough to allow a timely population response that flattens the curve and ends the first wave. After the first wave, cases are low, but so is herd immunity. In the meantime, the economic consequences of lockdown continue to build, causing a waning of support for continued lockdown, which in turns sparks a second wave among the remaining susceptible individuals. This process can be repeated in third and subsequent waves for some parameter values. But we emphasize that multiple waves do not necessarily correspond to more COVID-19 cases overall.

Changes in the duration of infectiousness $1/\gamma$ and the duration of the latent stage $1/\sigma$ around baseline values do not change the number of peaks: a second wave is still observed across the range we explored (**Supplementary Figure 2**). However, the second peak is higher than the first when $1/\gamma$ is between 3 to 5 days, while out of this range the second peak is lower. The lack of dependence of dynamics on σ is expected. When $1/\gamma < 3$ days, the second peak is less severe because R_0 drops below levels

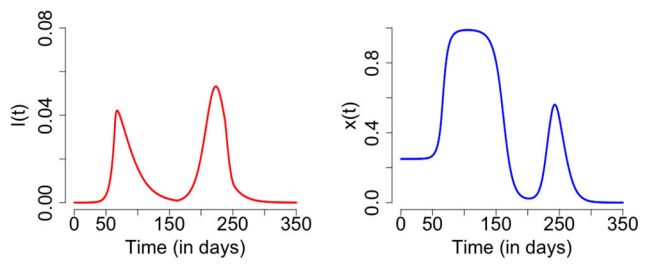


FIGURE 1 | Second wave of COVID-19: time series plot of the number of infected individuals **(left)** and proportion of individuals practicing physical distancing **(right)**. The results are obtained for $C_0 = 0.63$, $R_0 = 2.675$, $1/\sigma = 2.5$, $\gamma = 1/5$, $\kappa = 5$, $\epsilon = 0.008$, $\alpha = 1.0 * 24/365$, and $\delta = 0.05 * 24.0/365$ and initial conditions $S(0) = 1 - 1/1000000$, $E(0) = 0$, $I(0) = 1/1000000$, $R(0) = 0$, $x(0) = 0.25, L(0) = 0$, and $H(0) = 0$.

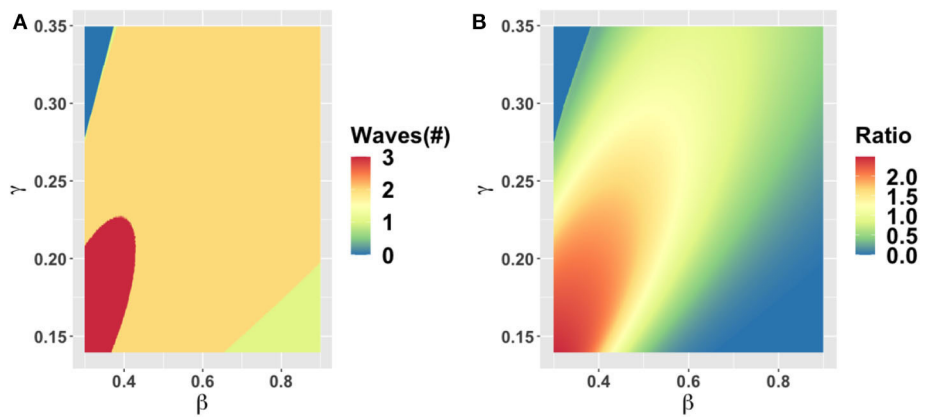


FIGURE 2 | Parameter plane showing model dynamics as they vary with changes in the transmission rate β and the rate γ at which the infectious period ends, for the number of COVID-19 waves **(A)** and the ratio of peak height of the second wave to the first wave peak **(B)**. Other parameters are at baseline values (**Table 1**).

that are feasible for continued transmission in the population. In contrast, when $1/\gamma > 5$ days the second peak is less severe because a heightened R_0 causes rapid build-up of herd immunity in the first wave of infection.

3.3. Socio-Economic and Intervention Conditions for a Second Wave

Next we explored the effect of variation in intervention, economic and social parameters. The parameter plane for α, the rate at which economic losses due to closure accumulate, and δ, the discounting rate for losses, shows little variation in these values across the ranges we explored (**Supplementary Figure 3**). Two waves are predicted and the peak of the second wave is higher than the first wave for almost all parameter combinations. The only exception is that when α is very small, only a single wave occurs because the population is willing to tolerate economic losses indefinitely. As a result, x remains high over the entire time horizon of the simulation and COVID-19 is effectively controlled throughout this period.

The behavioral parameter κ is a measure of how quickly novel social behavior spreads through a population as disease cases are reported. It has a large influence on the model dynamics, as represented in the $\kappa - \alpha$, $\kappa - \epsilon$, and $\kappa - C_0$ parameter planes (**Figures 3–5**). Higher values of κ indicate that individuals imitate more quickly. At our baseline value $\kappa = 5$/day, we observe a second wave. As the value of κ increases from this baseline value, the number of waves increases from two to six or seven in all three parameter planes, unless the effectiveness of closure (C_0) is so low that the population experiences a single large epidemic that rapidly confers herd immunity to everyone (**Figures 3–5**). As κ is reduced sufficiently from its baseline value, the second wave is lost as expected, since we enter a parameter regime where the population responds with an unrealistic slowness to the presence of COVID-19, and it

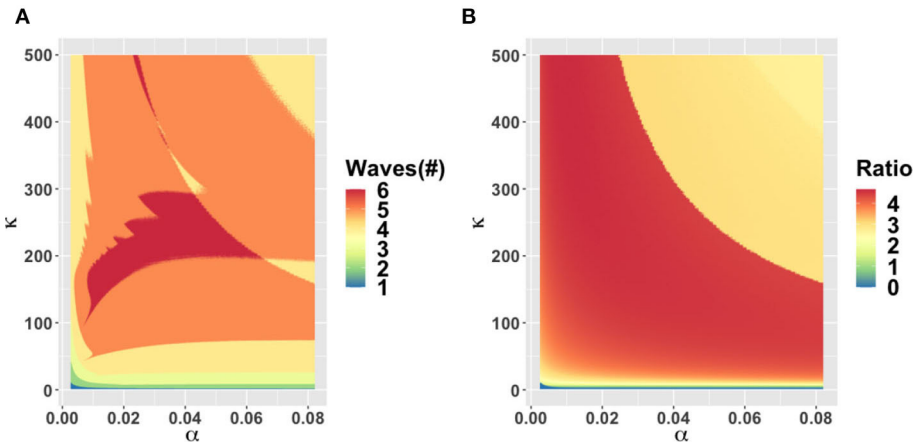


FIGURE 3 | Parameter plane showing model dynamics as they vary with changes in the rate α at which socio-economic losses accrue and the rate κ that controls the social learning rate, for the number of COVID-19 waves **(A)** and the ratio of peak height of the second wave to the first wave peak **(B)**. Other parameters are at baseline values (**Table 1**).

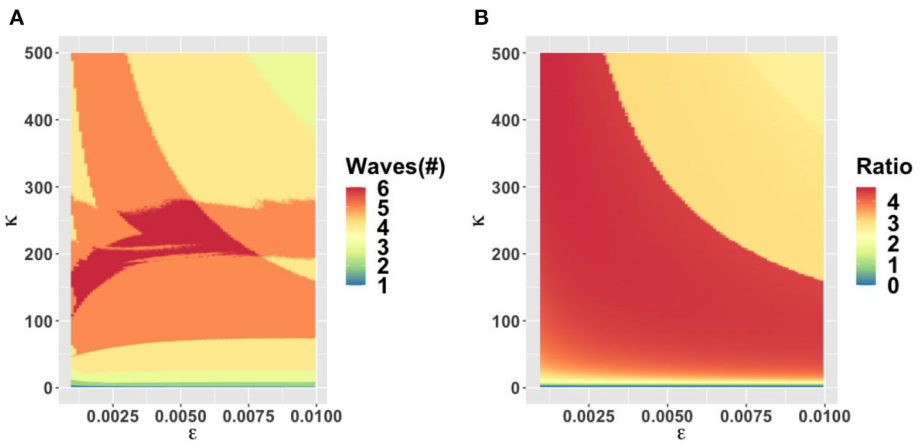


FIGURE 4 | Parameter plane showing model dynamics as they vary with changes in the rate κ that controls the social learning rate and the parameter ϵ which controls how sensitive the population is to economic losses relative to infection prevalence, for the number of COVID-19 waves **(A)** and the ratio of peak height of the second wave to the first wave peak **(B)**. Other parameters are at baseline values (**Table 1**).

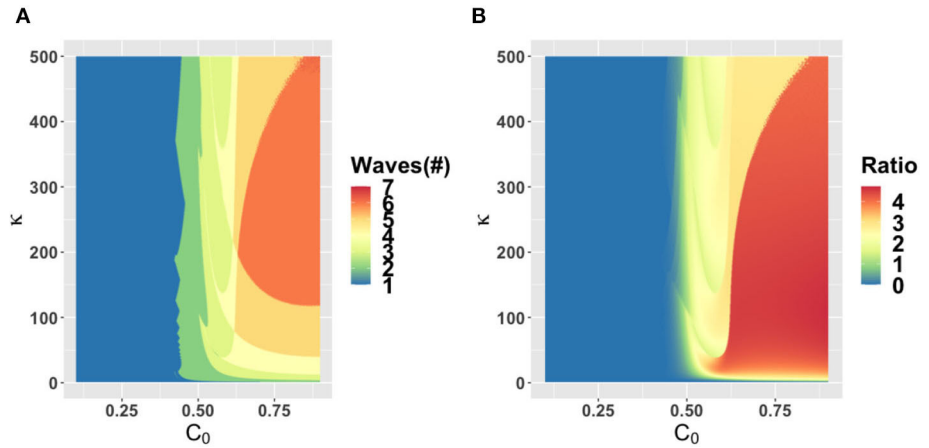


FIGURE 5 | Parameter plane showing model dynamics as they vary with changes in the rate κ that controls the social learning rate and the parameter C_0 which controls how effective closure is, for the number of COVID-19 waves **(A)** and the ratio of peak height of the second wave to the first wave peak **(B)**. Other parameters are at baseline values (**Table 1**).

experiences a single, rapid pandemic wave that rapidly confers herd immunity. The second peak is higher than the first peak in all three parameter planes, except again when C_0 is too low to effectively flatten the curve, and a single large outbreak results. Some examples of model outcomes for three or more waves are shown in **Supplementary Figure 4**. In these extreme scenarios, the second wave can either dominate the first and third waves, or it is also possible that the peaks of successive epidemic waves increase over time until it reaches a maximum peak in the fourth wave.

4. DISCUSSION

A second wave of COVID-19 is widely feared in 2020 as many jurisdictions around the world begin lifting restrictions that have held viral transmission in check. To address this issue, we analyzed a simple theoretical model of the interplay between SARS-CoV-2 transmission dynamics and social dynamics concerning public support for physical distancing and school and workplace closure. We found that a second wave of COVID-19 (and sometimes also a third wave) was likely across a broad range of epidemiological and behavioral parameters. In some cases, the second peak was higher than the first peak, while for other parameter combinations it was lower.

Our prediction of a second wave driven by behavior-disease interactions is plausible, given past and recent experience with novel emerging pathogens. One of the first affected countries in the COVID-19 pandemic (Iran) is now experiencing a large second wave on account of lifting restrictions in April 2020 [49]. During the 2003 SARS-CoV-1 epidemic in Toronto, premature relaxation of control measures resulted in a second outbreak that was as large as the first outbreak [50]. Finally, behavioral responses to disease dynamics appear to have played a role in shaping the three waves that some populations experienced during the "Spanish flu" pandemic in 1918 [51].

Our model makes simplifying assumptions that could influence its projections. For instance, our model assumes that populations respond to infection prevalence $I(t)$ but in fact, populations observe reported cases and deaths, both of which are delayed compared to time of actual infection. Time delays tend to destabilize dynamics in epidemic models [31] and hence we suspect that a model extension including a response to lagged outcomes like reported cases and deaths would exacerbate the severity of second waves in our model.

On the other hand, adding real-world spatial and demographic heterogeneities to our model could stabilize the dynamics and make the predicted oscillations less extreme, even if they do not remove them completely [52–56]. Similarly, on the behavioral modeling side, we suggest that the extreme oscillations observed in this model could also be stabilized if individuals use past and/or projected future states in their decision-making, instead of just the current prevalence, as we assumed [57, 58]. Alternatively, if individuals learn socially from other populations at differing stages of COVID-19 outbreaks [28], and not just their local population, this might also dampen the oscillations we observed in the model.

In summary, we speculate that incorporating social and spatial heterogeneities into the model would not completely remove the possibility of a second wave, although it could dampen the cycles [52–56] and give rise to epidemic curves more closely resembling that observed in the second wave in Iran [49]. Moreover, our prediction of a second wave was relatively robust across parameter space. Hence, we conclude that a second wave of COVID-19 on account of the coupled behavior-disease feedbacks we explore in this model will characterize many populations. Because interactions between the dynamics of disease spread and social processes will play a major role in shaping the pandemic, more effort in transmission modeling of COVID-19 should be devoted to accounting for them.

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

CB, MA, and JT conceived the study. SP and FN conducted the research. All authors developed the model and analysis and wrote the manuscript.

FUNDING

This research was funded by NSERC Discovery Grants and an Ontario Ministry of College and Universities grant to CB and MA. SP was partially supported by Program UEM-SIDA 2017-2022 under the Subprogramme Nr 1.4.2: Capacity Building in Mathematics, Statistics and Its Applications.

ACKNOWLEDGMENTS

The authors are grateful to comments from two reviewers. This manuscript has been released as a pre-print at medRxiv [59].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.574514/full#supplementary-material

REFERENCES

1. Yan P. Non-identifiables and invariant quantities in infectious disease models. In: *Mathematical Understanding of Infectious Disease Dynamics*. Singapore: World Scientific (2009). p. 167-229. doi: 10.1142/9789812834836_0005
2. Tuite AR, Fisman DN, Greer AL. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. *CMAJ*. (2020) 192:E497–505. doi: 10.1101/2020.03.24.20042705
3. Kucharski AJ, Klepac P, Conlan A, Kissler SM, Tang M, Fry H, et al. Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of SARS-CoV-2 in different settings. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.04.23.20077024
4. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infect Dis Modell*. (2020) **5**:256–63. doi: 10.1016/j.idm.2020.02.002
5. Karatayev, VA, Anand M, Bauch, CT. "Local lockdowns outperform global lockdown on the far side of the COVID-19 epidemic curve." In *Proceedings of the National Academy of Sciences*. (2020)
6. Dreher N, Spiera Z, McAuley FM, Kuohn L, Durbin JR, Marayati NF, et al. Impact of policy interventions and social distancing on SARS-CoV-2 transmission in the United States. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.05.01.20088179
7. McGrail DJ, Dai J, McAndrews KM, Kalluri R. Enacting national social distancing policies corresponds with dramatic reduction in COVID-19 infection rates. *medRxiv*. (2020) **15**:e0236619. doi: 10.1101/2020.04.23.20077271
8. Anderson SC, Edwards AM, Yerlanov M, Mulberry N, Stockdale J, Iyaniwura SA, et al. Estimating the impact of COVID-19 control measures using a Bayesian model of physical distancing. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.04.17.20070086
9. Brodeur A, Clark AE, Fléche S, Powdthavee N. *COVID-19, Lockdowns and Well-Being: Evidence from Google Trends*. Institute of Labor Economics (IZA) (2020).
10. Bauch CT. Imitation dynamics predict vaccinating behaviour. *Proc R Soc B Biol Sci*. (2005) **272**:1669–75. doi: 10.1098/rspb.2005.3153
11. Reluga TC. Game theory of social distancing in response to an epidemic. *PLoS Comput Biol*. (2010) **6**:e1000793. doi: 10.1371/journal.pcbi.1000793
12. Funk S, Salathé M, Jansen VA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J R Soc Interface*. (2010) **7**:1247–56. doi: 10.1098/rsif.2010.0142
13. Fu F, Rosenbloom DI, Wang L, Nowak MA. Imitation dynamics of vaccination behaviour on social networks. *Proc R Soc B Biol Sci*. (2011) **278**:42–9. doi: 10.1098/rspb.2010.1107
14. Mbah MLN, Liu J, Bauch CT, Tekel YI, Medlock J, Meyers LA, et al. The impact of imitation on vaccination behavior in social contact networks. *PLoS Comput Biol*. (2012) **8**:e1002469. doi: 10.1371/journal.pcbi.1002469
15. Bauch CT, Galvani AP. Social factors in epidemiology. *Science*. (2013) **342**:47–9. doi: 10.1126/science.1244492
16. Zhao S, Stone L, Gao D, Musa SS, Chong MK, He D, et al. Imitation dynamics in the mitigation of the novel coronavirus disease (COVID-19) outbreak in Wuhan, China from 2019 to 2020. *Ann Transl Med*. (2020) **8**:448. doi: 10.21037/atm.2020.03.168
17. Turner BL, Matson PA, McCarthy JJ, Corell RW, Christensen L, Eckley N, et al. Illustrating the coupled human-environment system for vulnerability analysis: three case studies. *Proc Natl Acad Sci USA*. (2003) **100**:8080–5. doi: 10.1073/pnas.1231334100
18. Thampi VA, Anand M, Bauch CT. Socio-ecological dynamics of Caribbean coral reef ecosystems and conservation opinion propagation. *Sci Rep*. (2018) **8**:1–11. doi: 10.1038/s41598-018-33573-x
19. Henderson KA, Bauch CT, Anand M. Alternative stable states and the sustainability of forests, grasslands, and agriculture. *Proc Natl Acad Sci USA*. (2016) **113**:14552–9. doi: 10.1073/pnas.1604987113
20. Pollak Y, Dayan H, Shoham R, Berger I. Predictors of adherence to public health instructions during the COVID-19 pandemic. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.04.24.20076620
21. Yan Y, Malik AA, Bayham J, Fenichel EP, Couzens C, Omer SB. Measuring voluntary social distancing behavior during the COVID-19 pandemic. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.05.01.20087874
22. Rossberg AG, Knell RJ. How will this continue? Modelling interactions between the COVID-19 pandemic and policy responses. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.03.30.20047597
23. Ajzenman N, Cavalcanti T, Da Mata D. More than words: leaders? Speech and risky behavior during a pandemic. *SSRN*. (2020). doi: 10.2139/ssrn.3582908. [Epub ahead of print].
24. Kochańczyk M, Grabowski F, Lipniacki T. Dynamics of COVID-19 pandemic at constant and time-dependent contact rates. *Math Modell Nat Phenomena*. (2020) **15**:28. doi: 10.1051/mmnp/2020011
25. Rowlett J, Karlsson CJ. Decisions and disease: the evolution of cooperation in a pandemic. *arXiv[Preprint].arXiv:200412446*. (2020) doi: 10.1038/s41598-020-69546-2
26. Steinegger B, Arenas A, Gómez-Gardenes J, Granell C. Pulsating campaigns of human prophylaxis driven by risk perception palliate oscillations of direct contact transmitted diseases. *Phys Rev Res*. (2020) **2**:023181. doi: 10.1103/PhysRevResearch.2.023181
27. Johnston MD, Pell B. A dynamical framework for modeling fear of infection and frustration with social distancing in COVID-19 spread. *arXiv[Preprint].arXiv:200806023*. (2020).
28. Tellis GJ, Sood N, Sood A. Why did US governors delay lockdowns against COVID-19? Disease Science vs Learning, Cascades, and Political Polarization. *SSRN*. (2020). doi: 10.2139/ssrn.3575004. [Epub ahead of print].
29. Hossain MM. Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping. *SSRN*. (2020). doi: 10.2139/ssrn.3547824. [Epub ahead of print].
30. Hethcote HW. The mathematics of infectious diseases. *SIAM Rev*. (2000) **42**:599–653. doi: 10.1137/S0036144500371907
31. Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control. Oxford: Oxford University Press (1992).

32. Gao C, Liu J. Modeling and restraining mobile virus propagation. *IEEE Trans. Mobile Comput.* (2012) **12**:529–41. doi: 10.1109/TMC.2012.29
33. Gao C, Liu J. Network-based modeling for characterizing human collective behaviors during extreme events. *IEEE Trans. Syst. Man Cybernet. Syst.* (2016) **47**:171–83. doi: 10.1109/TSMC.2016.2608658
34. Wang Z, Bauch CT, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, et al. Statistical physics of vaccination. *Phys Rep.* (2016) **664**:1–13. doi: 10.1016/j.physrep.2016.10.006
35. Beckage B, Gross LJ, Lacasse K, Carr E, Metcalf SS, Winter JM, et al. Linking models of human behaviour and climate alters projected climate change. *Nat Clim Change.* (2018) **8**:79–84. doi: 10.1038/s41558-017-0031-7
36. Bauch CT, Bhattacharyya S. Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS Comput Biol.* (2012) **8**:e1002452. doi: 10.1371/journal.pcbi.1002452
37. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med.* (2020) **9**:538. doi: 10.3390/jcm9020538
38. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* (2020) **20**:553–8. doi: 10.1101/2020.01.31.20019901
39. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med.* (2020) **382**:1199-207. doi: 10.1056/NEJMoa2001316
40. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* (2020) **20**:533–4. doi: 10.1016/S1473-3099(20)30120-1
41. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines, similarities, differences and some implications. *Value Health.* (2001) **4**:225–50. doi: 10.1046/j.1524-4733.2001.43040.x
42. Lachmann A. Correcting under-reported COVID-19 case numbers. *medRxiv [Preprint].* (2020). doi: 10.1101/2020.03.14.20036178
43. Saad L. Americans Step Up Their Social Distancing Even Further. (2020). Available from: https://news.gallup.com/opinion/gallup/298310/americans-step-social-distancing-even-further.aspx
44. Park SW, Cornforth DM, Dushoff J, Weitz JS. The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak. *medRxiv.* (2020). doi: 10.1101/2020.03.09.20033514
45. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health.* (2020). doi: 10.1101/2020.02.08.20021162
46. He X, Lau EH, 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/s41591-020-1016-z
47. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* (2020) **27**:taaa021. doi: 10.1093/jtm/taaa021
48. Hilton J, Keeling MJ. Estimation of country-level basic reproductive ratios for novel Coronavirus (COVID-19) using synthetic contact matrices. *medRxiv [Preprint].* (2020). doi: 10.1101/2020.02.26.20028167
49. Johns Hopkins Coronavirus Resource Center. (2020). Available online at: https://coronavirus.jhu.edu/ (accessed May 21, 2020).
50. News From the Centers for Disease Control and Prevention. Update: severe acute respiratory syndrome-Toronto, Canada, 2003. *MMWR Morbid Mortal Weekly Rep.* (2003) **52**:547.
51. He D, Dushoff J, Day T, Ma J, Earn DJ. Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proc R Soc B Biol Sci.* (2013) **280**:20131345. doi: 10.1098/rspb.2013.1345
52. Grenfell BT, Kleczkowski A, Gilligan C, Bolker B. Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases. *Stat Methods Med Res.* (1995) **4**:160–183. doi: 10.1177/096228029500400205
53. Lloyd AL, May RM. Spatial heterogeneity in epidemic models. *J Theor Biol.* (1996) **179**:1–11. doi: 10.1006/jtbi.1996.0042
54. Earn DJ, Rohani P, Grenfell BT. Persistence, chaos and synchrony in ecology and epidemiology. *Proc R Soc Lond Ser B Biol Sci.* (1998) **265**:7–10. doi: 10.1098/rspb.1998.0256
55. Viboud C, Bjornstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science.* (2006) **312**:447–51. doi: 10.1126/science.1125237
56. Metcalf CJE, Hampson K, Tatem AJ, Grenfell BT, Bjornstad ON. Persistence in epidemic metapopulations: quantifying the rescue effects for measles, mumps, rubella and whooping cough. *PLoS ONE.* (2013) **8**:e74696. doi: 10.1371/journal.pone.0074696
57. Wells CR, Bauch CT. The impact of personal experiences with infection and vaccination on behaviour-incidence dynamics of seasonal influenza. *Epidemics.* (2012) **4**:139–51. doi: 10.1016/j.epidem.2012.06.002
58. Bury TM, Bauch CT, Anand M. Charting pathways to climate change mitigation in a coupled socio-climate model. *PLoS Comput Biol.* (2019) **15**:e1007000. doi: 10.1371/journal.pcbi.1007000
59. Pedro SA, Ndjomatchoua FT, Jentsch P, Tcheunche JM, Anand M, Bauch CT. Conditions for a second wave of COVID-19 due to interactions between disease dynamics and social processes. *medRxiv [Preprint].* (2020). doi: 10.1101/2020.05.22.20110502

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 20 October 2020
doi: 10.3389/fphy.2020.00353

COVID-19 Reverse Prediction and Assessment on the *Diamond Princess* Cruise Ship

Juan Zhang[1,2], *Gui-Quan Sun*[1,2], *Mingtao Li*[3], *Rui Gao*[1,4], *Huarong Ren*[1,4], *Xin Pei*[3] and *Zhen Jin*[1,2]*

[1] Complex Systems Research Center, Shanxi University, Taiyuan, China, [2] Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Taiyuan, China, [3] College of Mathematics, Taiyuan University of Technology, Taiyuan, China, [4] School of Mathematical Sciences, Shanxi University, Taiyuan, China

As of July 21, 2020, the coronavirus SARS-CoV-2 had spread to almost all countries around the world and caused more than 14.8 million confirmed cases, owing to its high transmissibility and fast rate of spread. Of the infected locations, the *Diamond Princess* cruise ship is special in that it is an isolated system with a population highly concentrated in a limited space, providing particularly favorable conditions for the transmission of the novel coronavirus-associated pneumonia, COVID-19. The Japanese government's emergency measures for controlling the spread of COVID-19 on the cruise ship have also been questioned. In this paper we develop a homogeneous mixed difference system to describe the mechanism of transmission of COVID-19 on the cruise ship, reverse-predict the epidemic transmission trend from January 20 to February 20, 2020, including the daily number of infected people and the peak time of infection, estimate the range of the basic reproduction number of virus transmission on the cruise ship, and assess the effects of prevention and control measures. It is concluded that the isolation of people, along with rapid and comprehensive detection of infections, play an important role in controlling the epidemic. In fact, the Japanese government's emergency measures did have a certain effect on limiting the spread of COVID-19, but the number of infected people could have been reduced by at least 60% if all personnel on the cruise ship had been tested and isolated promptly as early as February 5.

Keywords: COVID-19, *Diamond Princess*, difference equation, reverse prediction, reproduction number, measures assessment, infection scale

OPEN ACCESS

Edited by:
Chengyi Xia,
Tianjin University of Technology, China

Reviewed by:
Lin Wang,
University of Cambridge, United Kingdom
Sanling Yuan,
University of Shanghai for Science and Technology, China

***Correspondence:**
Zhen Jin
jinzhn@263.net

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *28 May 2020*
Accepted: *24 July 2020*
Published: *20 October 2020*

Citation:
Zhang J, Sun G-Q, Li M, Gao R, Ren H, Pei X and Jin Z (2020) COVID-19 Reverse Prediction and Assessment on the Diamond Princess Cruise Ship. Front. Phys. 8:353. doi: 10.3389/fphy.2020.00353

1. INTRODUCTION

Since December 2019 a new coronavirus pneumonia, named COVID-19, has spread menacingly quickly and invaded more than 200 countries, infecting over 14.8 million people and causing more than 600,000 deaths by July 21, 2020. Among the infected places, the *Diamond Princess* cruise ship is a special case; the course of development of the epidemic on the ship is shown in **Figure 1**. On January 20, the *Diamond Princess* departed from Yokohama, Japan, and a passenger from Hong Kong embarked. The passenger disembarked in Hong Kong on January 25 and was confirmed to be infected with SARS-CoV-2 on February 1. The *Diamond Princess*, which was scheduled to return to Yokohama on February 5, arrived ahead of time on February 4 and was subjected to quarantine and inspection. According to requirements of the Ministry of Health, Labor and Welfare of Japan,

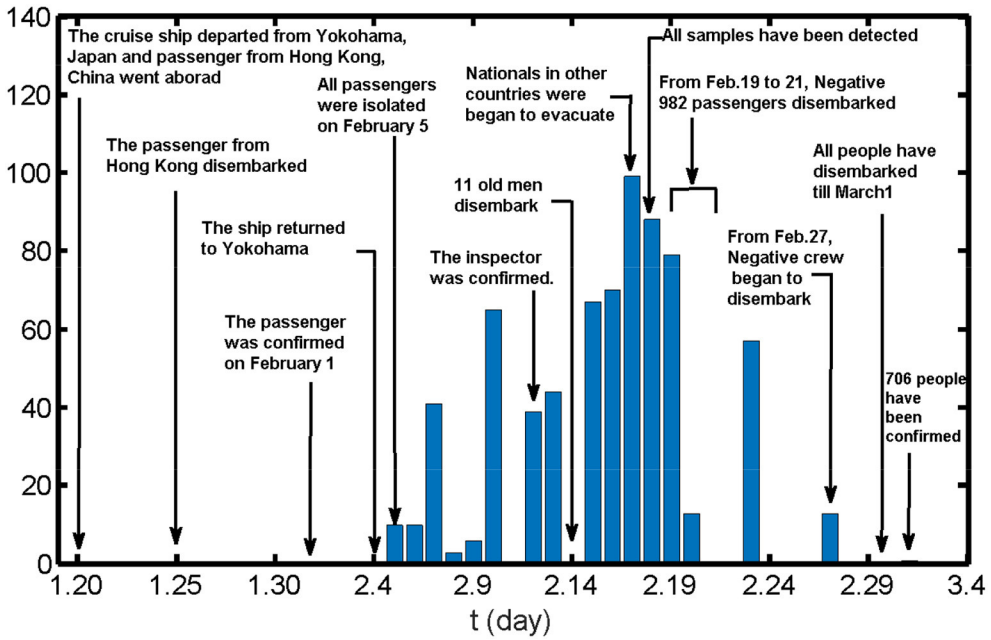


FIGURE 1 | The course of development of the COVID-19 epidemic on the *Diamond Princess* cruise ship.

from February 5 all passengers on the *Diamond Princess* had to be isolated in their cabins for 14 days. During the quarantine period, passengers and crew were tested in batches irregularly, and individuals confirmed to be positive for COVID-19 were to disembark for treatment at designated institutions in Japan. With the comprehensive implementation of systematic detection starting from February 14, all testing work was completed on February 18. Meanwhile, from February 17, other countries began to evacuate their nationals on chartered flights. From February 19, passengers who tested negative and in whose room there were no confirmed cases were allowed to disembark. From February 27, the crew of the cruise ship began to disembark and were sent to Saitama prefecture and Hikari city in Japan to be isolated for 14 days. By March 1, all people had disembarked from the cruise ship. Eventually, a total of 721 on the *Diamond Princess* were reported to be infected with SARS-CoV-2.

During the initial phase of infection on the cruise ship, the Japanese government did not pay enough attention to the outbreak and did not test, survey, evacuate, or isolate people in a timely fashion, but only asked passengers to isolate themselves in their cabins, and continued to allow air to be circulated on the ship through the central air-conditioning system. Siqi Sun, a lecturer at Shanghai Maritime University and the author of the book *Cruise Tourism Law Thesis*, has said that this is the most complex crisis in the history of cruises since the *Titanic* event in 1909 [1]. In the present work, we aim to use mathematical methods [2] to elucidate the mechanism of COVID-19 transmission on the *Diamond Princess* cruise ship, and thus provide a theoretical basis for guidance on epidemic prevention and control in similar closed systems.

At the early stage of the epidemic on the *Diamond Princess*, information about the outbreak was limited to daily detected and confirmed cases. However, the sampling method and the proportion of inspected samples were irregular, so it is difficult to capture the pattern of the epidemic during the early stage, let alone use it to predict the later development of the epidemic. However, by May 2020 all cases of infection had been confirmed. Therefore, based on the final cumulative number of confirmed cases, the transmission trend of the epidemic from January 20 to February 20 can be reverse-predicted and evaluated. For this purpose, considering that the mode of contact between people on the ship is close to uniform mixing, we derive a homogeneous mixing difference equation to describe the transmission mechanism of COVID-19 on the ship, with the goal of understanding the characteristics of epidemic transmission in a confined space. In addition, we estimate the number of people infected by the Hong Kong passenger from January 20 to January 25, the number of people who had been infected by February 5 when isolation measures were put in place, and the peak time and peak value of the epidemic on the cruise ship. Moreover, we estimate the basic reproduction number of the spread of COVID-19 on the ship. Finally, we assess the effects of the isolation measures taken by the Japanese government.

2. MATERIALS AND METHODS

2.1. Materials

We study the infection status of 3,711 people on the cruise ship, including 2,666 passengers and 1,045 crew members, from January 20 to February 20. The numbers of detected and confirmed cases are shown in **Table 1** [1, 3–5], where detected

TABLE 1 | Detected and confirmed cases out of 3,711 people on the *Diamond Princess*.

Date	Cumulative number of detected cases	Cumulative number of confirmed cases	Number of people disembarking
February 5	31	10	10
February 6	102	20	20
February 7	202	61	41
February 8	208	64	3
February 9	336	70	6
February 10	439	135	65
February 11	439	135	0
February 12	492	174	39
February 13	713	218	44
February 14	714	218	11
February 15	930	285	67
February 16	1,219	355	70
February 17	1,723	454	437
February 18	2,404	542	88
February 19	3,011	621	522
February 20	/	634	522
February 27	/	705	/
March 3	/	706	/
	/	721 (final number)	/

cases refer to individuals who had taken a nucleic acid test of which the result is uncertain (may be negative or positive), and confirmed cases refer to individuals whose tests came back positive.

2.2. Prerequisites

We assume that the first case on the cruise ship is the passenger from Hong Kong, who boarded the ship on January 20 and disembarked on January 25.

We also assume that everyone on the cruise ship has the same contact number per day, denoted by $C(t)$. Since the Japanese government began to test and isolate all individuals on the cruise ship from February 5, the model is divided into two stages: (i) January 20 to February 4, and (ii) February 5 to February 20. The average number of close contacts per person per day is denoted by C_1 during the first stage and by C_2 during the second stage.

2.3. Dynamical Model

We take January 20 to be the initial time, denoted by $t = 1$. At time t, the number of infected but not confirmed (INC for short) is denoted by $I(t)$. As given in **Figure 2**, the variation with time is given by

number of INC at time $t+1$ = number of INC at time t
+ number infected by the Hong Kong passenger at time t
+ number infected by other passengers at time t
− number of newly confirmed cases at time t.

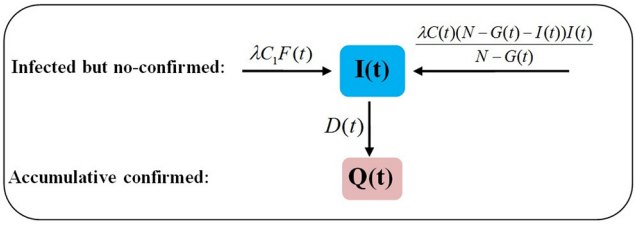


FIGURE 2 | Transfer diagram for the dynamical transmission of COVID-19 among subpopulations.

Let $F(t)$ be the number of imported infected persons at time t. It is assumed that the Hong Kong passenger is the only imported case. Then $F(t) = 1$ from January 20 to January 25, i.e., for $1 \le t \le 6$, and $F(t) = 0$ for $t > 6$. The daily number of close contacts of the Hong Kong passenger is C_1, and let λ be the probability of another individual being infected by the passenger after close contact. Then the number of people infected by the Hong Kong passenger at time t is $\lambda C_1 F(t)$.

At time t, for all infected individuals except the Hong Kong passenger, the close contact number per person per day is $C(t)$. The proportion of uninfected (i.e., susceptible) people is $\frac{N-G(t)-I(t)}{N-G(t)}$, where N is the total number of people on the cruise ship on February 5, $G(t)$ is the cumulative number of people who have disembarked, and $N - G(t)$ is the total number of people remaining on the cruise ship at time t. Hence, the number of infected people caused by one infected individual at time t is $\frac{\lambda C(t)(N-G(t)-I(t))}{N-G(t)}$, and the number of infected people caused by all infected individuals at time t is $\frac{\lambda C(t)(N-G(t)-I(t))}{N-G(t)} I(t)$. The number of newly confirmed cases at time t is denoted by $D(t)$. The cumulative number of confirmed cases at time t is denoted by $Q(t)$. The cumulative number of infected cases at time t is denoted by $A(t)$.

Hence we have the system of equations

$$\begin{aligned} I(t+1) &= I(t) + \lambda C_1 F(t) + \frac{\lambda C(t)(N - G(t) - I(t))I(t)}{N - G(t)} - D(t), \\ A(t+1) &= A(t) + \lambda C_1 F(t) + \frac{\lambda C(t)(N - G(t) - I(t))I(t)}{N - G(t)}, \\ Q(t+1) &= Q(t) + D(t), \end{aligned} \tag{1}$$

where

$$F(t) = \begin{cases} 1 & \text{for } t \le 6, \\ 0 & \text{for } t > 6 \end{cases}$$

and

$$C(t) = \begin{cases} C_1 & \text{for } t < 17, \\ C_2 & \text{for } t \ge 17. \end{cases}$$

2.4. Parameter Values

In this subsection, we use real data (see **Table 1**) to estimate the parameter values in model (1), which are given in

TABLE 2 | Values of parameters and variables in model (1).

Parameter	Meaning	Value	Confidence interval	Source
$D(t)$	Daily new number of confirmed cases at time t	See **Table 1**	/	Real data
$G(t)$	Daily cumulative number of people disembarking at time t	See **Table 1**	/	Real data
λ	Transmission probability per contact	0.05	/	(a)
C_1	Contact number per person per day before February 5	9.2472 (Case I)	[8.7184, 9.7760]	(b)
C_2	Contact number per person per day after February 5	1.5013 (Case I)	[0.8685, 2.1341]	(b)
C_1	Contact number per person per day before February 5	6.4386 (Case II)	[6.4301, 6.4472]	(b)
C_2	Contact number per person per day after February 5	6.4386 (Case II)	[6.4301, 6.4472]	(b)
N	Total number of staff on ship	3,711	/	Real data
$I(0)$	Initial number of infected individuals	0	/	Real data
$A(0)$	Initial number of cumulative infected individuals	0	/	Real data
$Q(0)$	Initial number of cumulative confirmed individuals	0	/	Real data
$A(32)$	Theoretical number of cumulative infected individuals on February 20	/		
$\hat{A}(32)$	Actual number of cumulative infected individuals on February 20	/	[720, 760]	(c)

Table 2. We describe the process of parameter estimation, which is implemented using the function `fminsearch` in the optimization toolbox of MATLAB.

(a) The meaning of λ is the transmission probability per contact. We estimate its value by combining information from references [6–8], which studied the spread of the epidemic in Wuhan city, in Shanxi province, and on the *Diamond Princess*, respectively. In Zhang et al. [6] the transmission probability per contact was found to be 0.0149; in Xue et al. [7] the values 0.01597 and 0.04644 were given; and Liu et al. [8] provided a range of 0.001–0.2 for the transmission probability per contact. Based on these estimates, here we take the value of λ to be 0.05.

(b) We use the least-squares estimation method to obtain the values of C_1 and C_2 in Case I and the value of $C_1 = C_2$ in Case II (see section 3 for description of the cases) by fitting the theoretical value of $A(32)$ in model (1) to the actual cumulative number of infected individuals, denoted by $\hat{A}(32)$. Individuals on the cruise ship were isolated, with each passenger staying in their own cabin. Each cabin on the *Diamond Princess* can accommodate four people [9], so we suppose that the average number of people in each cabin is 3. As for the crew, from February 5 they took meals in the canteen in turn, and the dining room table can accommodate 12 people. Therefore, we assume that the average contact number for each crew member is 12. The proportions of passengers and crew on the ship were 71.8 and 28.2%, respectively, so we calculate that the average contact number per person per day for passengers and crew is between 5 and 6.

(c) According to Japan's health ministry bulletin, all individuals on the ship had been tested and had disembarked by March 1, and the total number of officially reported confirmed cases is 721 [5]. Besides those, 38 passengers tested negative when disembarking but were later confirmed to be positive [10, 11]. Considering that some of these people may have become infected on their way back home, we estimate that on February 20 the actual cumulative number of infected individuals, $\hat{A}(32)$, was between 720 and 760.

3. RESULTS

The isolation measures taken by the Japanese government have been controversial, and the actual infection situations before and after the measures were imposed are not clear. In this section, we reverse-predict the infection on the cruise ship in two scenarios.

3.1. Case I

In this scenario it is assumed that the isolation measure is effective by itself, and that circulating air through the central air-conditioning system does not cause spread of the virus. Under isolation, close contacts of passengers were restricted to their cabin mates, and close contacts of crew members were limited to those dining together. Based on the model and parameter values, by calculating $\frac{\lambda C(t)(N-G(t)-I(t))}{N-G(t)} I(t)$ and $A(t)$ we can estimate the new daily number and cumulative number of infected people from January 20 to February 20; these values are reported in **Figure 3** and **Table 3**.

From **Figure 3A** we observe that the peak time of the epidemic in this case is from February 5 to February 15. As of February 20, there were still 100 infected but not confirmed individuals on the ship. By comparing the red dots with the black solid line, we can see that the number of confirmed cases was far less than the number of infected people before February 14. The comprehensive and systematic testing from February 14 led to a rapid fall in the number of infected persons, and the cumulative number of infected people and the total number of confirmed cases gradually became closer. So we can see that isolating people and conducting comprehensive testing play important roles in epidemic control. From **Table 3** we also see that there would have been about 258 ([188, 328]) infected individuals on February 5 when the cruise ship was isolated. It can be concluded that because the Japanese government did not pay enough attention to the epidemic at the initial stage and did not carry out comprehensive testing on all people onboard in time, the final proportion of infected reached nearly 20%. If measures had been taken for the *Diamond Princess* as promptly as for the *Costa Selena* cruise ship, which carried 4806 passengers

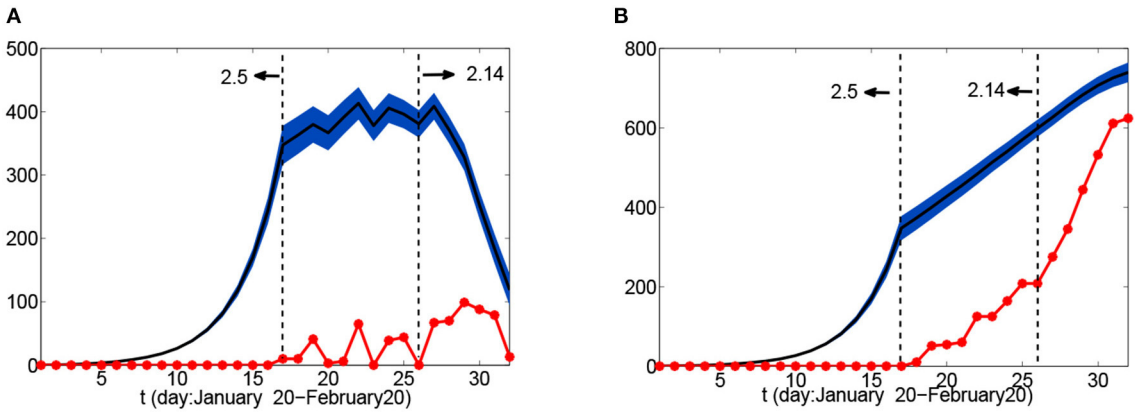


FIGURE 3 | Number of infected people as a function of time from January 20 to February 20 in Case I: (A) daily new number of infected cases; (B) daily cumulative number of infected cases. Red dots represent actual confirmed numbers; black dotted lines indicate critical time nodes; the black solid line represents the theoretical mean value of the number of infected, $\frac{\lambda C(t)(N-G(t)-I(t))}{N-G(t)} I(t)$ in (A) and $A(t)$ in (B); the blue area represents the interval of 10,000 times deduction results.

TABLE 3 | Reverse-predicted infection size and peak in Cases I and II.

Case	Number of people infected by Hong Kong passenger	Infected population on February 5	Infected population on February 20	Peak time of infection
I	2.7742 ([2.6155,2.9328])	257.9222 ([187.9917,327.8528])	109.0139 ([96.4057,121.6221])	February 5 to February 15
II	1.9316 ([1.9290,1.9341])	52.9066 ([52.6305,53.1826])	119.5218 ([94.9011,144.1424])	February 15 to February 16

and crew members and for which the Chinese government spent 1 h formulating the response plan, 4 h conducting the sampling work, 5 h producing the test results, and 24 h overall completing the detection, evacuation, and resettlement work [12], the number of infected on the *Diamond Princess* could have been reduced by at least 60–70%. This shows that the effect of non-pharmaceutical interventions on controlling the spread of COVID-19 is very considerable [13].

The reproductive number (including the basic reproductive number, or the effective reproductive number) is an index that measures the number of infected people caused by a single infected individual during the infection period. In Case I, the basic reproductive number can be estimated based on the Hong Kong passenger infecting about 3 (2.7742) persons from January 20 to January 25 by the sum $\sum_{t=1}^{6} \lambda C_1 F(t)$, which can be taken as the reproductive number of the Hong Kong passenger. Based on reference [14], the incubation period of COVID-19 is 3–5 days, so we assume that it will take 3–5 days from showing clinical symptoms to becoming a confirmed case on the cruise ship. Then the infection period is assumed to be 6–10 days. Besides, according to reports of the National Institute of Infectious Diseases of Japan, as of February 18 the percentage of infected but symptomless samples among 2,404 respiratory tract specimens was as high as 48%. Assume that the average time between being infected and being confirmed for asymptomatic infection is 10–15 days. Combining symptomatic and asymptomatic infections, we estimate the basic reproduction number of the cruise ship to be 3.66–5.54; see **Table 4**. For Wuhan city in Hubei province, China, which also became an isolated system after sealing, the basic reproduction number is 1.4–4.0 [15, 17, 19, 20]. For China as a whole, the basic reproduction number is 3.51–4.05 [21]. The basic reproduction number of the cruise ship is larger, which may be due to the higher population density, a more enclosed space, a higher percentage of asymptomatic infections, and delayed testing for all individuals.

TABLE 4 | Comparison of the basic reproduction number in different regions.

Region	Basic reproduction number	Source
Diamond Princess cruise ship	3.58–5.63	Case I
	2.54–3.99	Case II
Wuhan city	1.5–2.5	WHO
	1.47–2.59	[12]
	2.47–2.86	[14]
	3.6–4.0	[15]
	1.16–1.48	[16]
	2.2	[17]
Hunan province	1.34	[18]
Shandong province	1.71	[18]
Shenzhen city	1.08	[18]
China	3.51–4.05	[19]

3.2. Case II

Japan's emergency measures, which kept passengers isolated in cruise cabins but continued to circulate air through central air-conditioning, was questioned by many experts. Assume

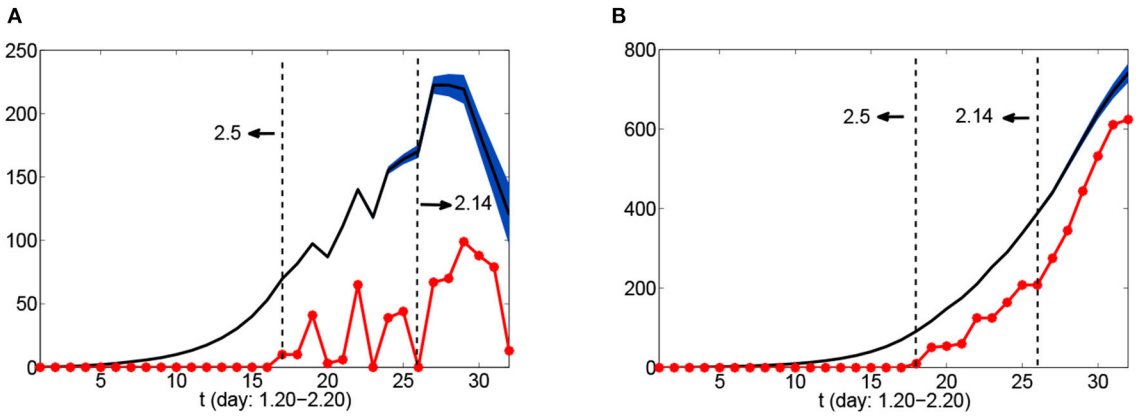


FIGURE 4 | Number of infected people as a function of time from January 20 to February 20 in Case II: (A) daily new number of infected cases; (B) daily cumulative number of infected cases. Red dots represent actual confirmed numbers; black dotted lines indicate critical time nodes; the black solid line represents the theoretical mean value of the number of infected, $\frac{\lambda C(t)(N-G(t)-I(t))}{N-G(t)}I(t)$ in (A) and $A(t)$ in (B); the blue area represents the interval of 10,000 times deduction results.

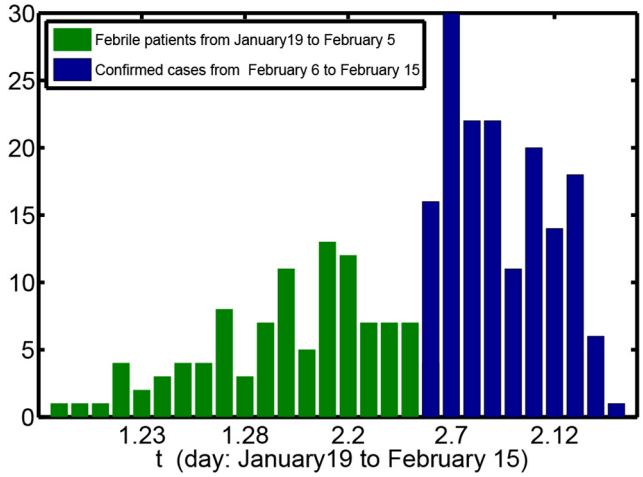


FIGURE 5 | Number of cases plotted against time. The green bars show the number of febrile patients and the blue bars the number of confirmed COVID-19 cases with reported onset times.

that after isolation, the close contact number of each person does not change due to circulation of air in the central air-conditioning system, that is, $C_1 = C_2$. Similarly to Case I, based on the model and parameter values we can calculate the new number and cumulative number of infected people from January 20 to February 20, as shown in **Figure 4** and **Table 3**. By inspection of **Figure 4A**, the infection peak time is from February 15 to February 16. Although the quarantine measure was imposed on February 5, the upward trend in the number of infections remained largely unchanged. Comparing the red dots with the black solid line in **Figure 4A**, we see that the comprehensive and systematic testing which started on February 14 resulted in a rapid decline in the number of infected people after February 17. We also find that there were about 53 infected individuals on February 5 when the cruise ship was isolated. If the Japanese government had completed response measures, such as detection, evacuation, and resettlement work as quickly as was done for the *Costa Selena*, the number of infected on the *Diamond Princess* could have been reduced by at least 90%. In Case II, it can be estimated that the Hong Kong passenger infected about two persons (1.9316) from January 20 to January 25. In this case, then, we estimate the basic reproduction number on the cruise ship to be 2.55–3.99.

4. DISCUSSION AND CONCLUSION

In this article we have used a mathematical dynamical model to reversely deduce, in two hypothetical cases, the number of people infected with COVID-19 on the *Diamond Princess* cruise ship from January 20 to February 20, 2020, the peak time of infection, the number of the infections caused by the Hong Kong passenger, and the basic reproduction number, information which can provide the theoretical basis for understanding the spread of the epidemic and developing effective control measures. The results reported in this paper demonstrate that a dynamical model which incorporates a transmission mechanism can be a very useful method for predicting the spread of a disease, especially when actual data are scarce [22–25].

Although the premises of the two cases are different, it can be deduced that quarantining populations and conducting comprehensive testing play an important role in control of the epidemic, and that detection of cases needs to be timely.

According to the report published by the National Institute of Infectious Diseases of Japan, as of February 18, 2,404 samples on the *Diamond Princess* cruise ship had been tested (including samples for double detection), and 531 cases were confirmed to be positive for COVID-19. Of these, 33 cases showed symptoms before February 6, 151 showed symptoms on or after February 6, and 255 (48%) were asymptomatic. Combining the time of

diagnosis of 79 patients with fever before February 6 (only 33 of which were confirmed cases) with the onset time of 151 confirmed cases after February 6, we obtain **Figure 5**. As can be seen from the figure, the infection peak was reached on February 7. Comparing with **Figures 3A**, **4A**, Case I seems closer to the actual situation, indicating that the role of central air-conditioning in transmission of the disease is relatively small, and that the isolation intervention adopted by the Japanese government had a certain impact on reducing the transmission of the virus from one person to another.

By our analysis, we estimate the reproduction number associated with the Hong Kong passenger to be between 1.9290 and 2.9328. In fact, the time period between the date of infection and date of confirmation for other persons on the cruise ship is irregular and cannot be determined, so the period of infection cannot be determined and hence the basic reproduction number cannot be found accurately. Nevertheless, we can give a general range [2.54, 5.63]. If the Japanese government had completed response measures for the *Diamond Princess*—including detection, evacuation, and resettlement work—as promptly as was done for the *Costa Selena* cruise ship (within 24 h), the number of infections on the *Diamond Princess* could have been reduced by at least 60%. In this respect, the response measures taken for the *Costa Selena* in Tianjin Port provide a positive paradigm. Moreover, 38 *Diamond Princess* passengers who tested negative on the ship were confirmed to have the disease after they disembarked, indicating that the nucleic acid test had certain errors and led to a proportion of infected people not being detected and thus becoming a potential source of risk for other regions. Therefore, the accuracy of the testing method should also be improved.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

JZ, G-QS, ML, and ZJ conceived and designed the experiments. JZ and G-QS performed the experiments. JZ, G-QS, and ML developed the dynamical model. RG, XP, and HR collected and analyzed the data. JZ, HR, XP, and RG analyzed the data. JZ, G-QS, and ZJ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the National Natural Science Foundation of China under Grant (11601292, 61873154, 11671241, 11701348, 11801398, and 11971278), General Youth Fund project in Shanxi Province (201901D211158), the National key research and development project of China (2019YFC1200501), the 1331 Engineering Project of Shanxi Province, Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi Province (2019L0114), and the Key Research and Development Project in Shanxi Province (202003D31011/GZ).

REFERENCES

1. *Adventure of "Virus Cruise": Infection Development of 621 Individuals on Princess Diamond Cruise Ship*. Available online at: http://news.caijingmobile.com/article/detail/412655?source_id=40 (accessed February 20, 2020).
2. Huang S, Peng Z, Jin Z. Studies of the strategies for controlling the COVID-19 epidemic in China: estimation of control efficacy and suggestions for policy makers. *Sci Sin Math.* (2020) **50**:1–14. doi: 10.1360/SSM-2020-0043
3. *31 Days on the Diamond Princess Cruise Ship*. Available online at: https://baijiahao.baidu.com/s?id=1659091075715805538&wfr=spider&for=pc (accessed February 21, 2020).
4. *About 150 Crew Members Have Not Yet Disembarked From the Diamond Princess After All Its Passengers Disembarked*. Available online at: http://news.eastday.com/eastday/13news/auto/news/world/20200227/u7ai9122386.html (accessed February 27, 2020).
5. *A New Crew Member of the Diamond Princess Cruise Ship Has Been Confirmed To Be Infected With COVID 19*. Available online at: https://baijiahao.baidu.com/s?id=1660062409001457350&wfr=spider&for=pc (accessed March 2, 2020).
6. Zhang J, Sun G, Li M, Zhang W, Chang L, Jia Z, et al. COVID-19 propagation prediction and assessment method with imported cases and infection generations: Shanxi Province as a case. *Res Square.* (2020). doi: 10.21203/rs.3.rs-16298/v1
7. Xue L, Jing S, Miller JC, Sun W, Li H, Estrado-Franco JG, et al. A data-driven network model for the emerging COVID-19 epidemics in Wuhan, Toronto and Italy. *Math Bioenc.* (2020) **326**:108391. doi: 10.1016/j.mbs.2020.108391
8. Liu F, Li X, Zhu G. Using the contact network model and Metropolis-Hastings sampling to reconstruct the COVID-19 spread on the Diamond Princess. *Sci Bull.* (2020) **65**:1297–305. doi: 10.1016/j.scib.2020.04.043
9. *Diamond Princess*. Baidubaike. Available online at: https://baike.baidu.com/item/%E9%92%BB%E7%9F%B3%E5%85%AC%E4%B8%BB%E5%8F%B7 (accessed May 17, 2020).
10. *Diamond Princess Passengers Returning to Hong Kong Confirmed Case Hong Kong, and Media: No Surprise*. Available online at: https://news.163.com/20/0224/08/F64TCTC20001899O.html (accessed February 24, 2020).
11. *4 People That Were Tested To Be Negative When Disembarking From "Diamond princess" Cruise Ship Were Confirmed*. Available online at: http://n.eastday.com/pvideo/1582783658015276 (accessed February 27, 2020).
12. *China's Version of the Horror Cruise Ship Incident*. Available online at: https://www.sohu.com/a/376232368_653731 (accessed February 27, 2020).
13. Ali S, Wang L, Lau E, Xu X, Du Z, Wu Y, et al. Evolution of effective serial interval of SARS-CoV-2 by non-pharmaceutical interventions. *Res Sq.* (2020) rs.3.rs-32486. doi: 10.21203/rs.3.rs-32486/v1
14. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of 2019 novel coronavirus infection in China. *medRxiv.* (2020). doi: 10.1101/2020.02.06.20020974
15. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, et al. Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China. *Emerg Infect Dis.* (2020) **26**:1049–52. doi: 10.3201/eid2605. 200146
16. Du Z, Wang L, Xu X, Wu Y, Cowling B, Meyers L. The serial interval of COVID-19 from publicly reported confirmed cases. *Emerg Infect Dis.* (2020) **26**:1341–3. doi: 10.1101/2020.02.19.20025452
17. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating

in Wuhan, China: a modelling study. *Lancet.* (2020) **395**:689–97. doi: 10.1016/S0140-6736 (20)30260-9

18. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis.* (2020) **20**:793–802. doi: 10.1016/S1473-3099(20)30230-9
19. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis.* (2020) **92**:214–7. doi: 10.1016/j.ijid.2020.01.050
20. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet.* (2020) **395**:1–21. doi: 10.1016/S0140-6736(20)30154-9
21. Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *medRxiv.* (2020). doi: 10.1101/2020.02.10.20021675
22. Wang Z, Xia C, Chen Z, Chen G. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern.* (2020) 1–9. doi: 10.1109/TCYB.2019.2960605
23. Xia C, Wang Z, Zheng C, Guo Q, Shi Y, Dehmer M, et al. A new coupled disease-awareness spreading model with mass media on multiplex networks. *Inform Sci.* (2019) **471**:185–200. doi: 10.1016/j.ins.2018.08.050
24. Zhang J, Jin Z, Yuan Y. Assessing the spread of foot and mouth disease in mainland china by dynamical switching model. *J Theor Biol.* (2019) **460**:209–19. doi: 10.1016/j.jtbi.2018.09.027
25. Zhang J, Jin Z, Sun G, Sun X, Wang Y, Huang B. Determination of original infection source of H7N9 avian influenza by dynamical model. *Sci Rep.* (2014) **4**:4846. doi: 10.1038/srep04846

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.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 28 October 2020
doi: 10.3389/fphy.2020.00383

How to Optimize the Supply and Allocation of Medical Emergency Resources During Public Health Emergencies

Chunyu Wang[1], *Yue Deng*[1], *Ziheng Yuan*[2], *Chijun Zhang*[3], *Fan Zhang*[1], *Qing Cai*[4], *Chao Gao*[5,1]* *and Jurgen Kurths*[6,7]

[1] College of Computer and Information Science, Southwest University, Chongqing, China, [2] Faculty of Humanities, Chang'an University, Xi'an, China, [3] College of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun, China, [4] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore, Singapore, [5] Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources of the People's Republic of China, Shenzhen, China, [6] Potsdam Institute for Climate Impact Research, Potsdam, Germany, [7] Nizhny Novgorod State University, Nizhny Novgorod, Russia

The solutions to the supply and allocation of medical emergency resources during public health emergencies greatly affect the efficiency of epidemic prevention and control. Currently, the main problem in computational epidemiology is how the allocation scheme should be adjusted in accordance with epidemic trends to satisfy the needs of population coverage, epidemic propagation prevention, and the social allocation balance. More specifically, the metropolitan demand for medical emergency resources varies depending on different local epidemic situations. It is therefore difficult to satisfy all objectives at the same time in real applications. In this paper, a data-driven multi-objective optimization method, called as GA-PSO, is proposed to address such problem. It adopts the one-way crossover and mutation operations to modify the particle updating framework in order to escape the local optimum. Taking the megacity Shenzhen in China as an example, experiments show that GA-PSO effectively balances different objectives and generates a feasible allocation strategy. Such a strategy does not only support the decision-making process of the Shenzhen center in terms of disease control and prevention, but it also enables us to control the potential propagation of COVID-19 and other epidemics.

Keywords: COVID-19, computational epidemiology, epidemic propagation, emergence management, multi-objective optimization, medical emergency resources

OPEN ACCESS

Edited by:
Lin Wang,
University of Cambridge,
United Kingdom

Reviewed by:
Peican Zhu,
Northwestern Polytechnical University,
China
Jinzhuo Liu,
Yunnan University, China

***Correspondence:**
Chao Gao
cgao@swu.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *04 July 2020*
Accepted: *07 August 2020*
Published: *28 October 2020*

Citation:
Wang C, Deng Y, Yuan Z, Zhang C, Zhang F, Cai Q, Gao C and Kurths J (2020) How to Optimize the Supply and Allocation of Medical Emergency Resources During Public Health Emergencies. Front. Phys. 8:383. doi: 10.3389/fphy.2020.00383

1. INTRODUCTION

In December 2019, a few patients with atypical pneumonia were first diagnosed in Wuhan, China, and this was reported to the World Health Organization (WHO) [1]. Then, its etiology was identified as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It is obvious that the disease was spreading continuously from person to person at the end of January 2020 [2]. The confirmed and imported cases in the Hubei province, China, brought about severe coronavirus outbreaks in many cities across China. More seriously, the disease, called COVID-19, has since spread rapidly across China and to the rest of the world [1]. By the end of June 30, 2020, there were 10,185,374 confirmed cases and 503,862 deaths in 210 countries [3].

Currently, vaccines and antiviral medicines for COVID-19 are still under development. Besides, other pharmaceutical treatments (e.g., plasma therapy) are not likely to be implemented on a large scale. Therefore, the use of a personal protective equipment (PPE) (e.g., hand sanitizers, surgical face masks, and protective clothing) becomes one of the major non-pharmaceutical measures to reduce the infectious hazard [4]. The sudden outbreak throughout China has exacerbated an unprecedented PPE shortage, causing austere challenges to epidemic control [5]. Due to the limited stock of a PPE during the early phase of COVID-19 epidemic, the local health organization suffered great challenges. It has been an open problem for us to efficiently allocate a PPE during an epidemic propagation.

Currently, various models (e.g., SIS, SIR, and SVIR) are considered to model the spreading process for the purpose of estimating the affecting scope from the perspective of theoretical analysis [6–10]. It is also possible to introduce evolutionary game theory during the resource allocation process [11]. However, balancing the allocation of a PPE in different regions is still a tough task due to multiple risk factors and high interest. Two key factors are a high-risk population and the severity of infection [12]. As a result, this paper develops an improved particle swarm optimization algorithm to reconcile this decision making problem. It balances the heterogeneity of a high-risk population and the severity of infection by incorporating the one-way crossover and mutation operations into the particle swarm optimization framework and eventually generates a reasonable allocation strategy of a PPE.

We apply this algorithm for balancing the allocation of PPE in Shenzhen, which is one of the largest megacities [with a population of more than 11 million [13]] in Guangdong province, China. During the COVID-19 epidemic in China, Shenzhen reported 195 confirmed COVID-19 cases as of February 1, 2020, which was higher than all other cities except Hebei province at that time. Therefore, improving the medical emergency system has been one of the highest priorities in Shenzhen [14]. More specifically, the supply and allocation of PPE is an urgent matter in Shenzhen.

2. METHODS

Shenzhen is divided into ten administrative districts. All parameters are set in **Table 1**. Let $X = (x_1, ..., x_{10})$ denote a PPE

TABLE 1 | Parameters are used in the experiments of the GA-PSO algorithm.

Parameters	Source
N_i	Population size in the *ith* district in 2016 [13]
O_i	Population size of people aged 65 or more in the *ith* district in 2015 [15]
I_i	The number of confirmed cases in the *ith* district by February 1, 2020 [16]
A_i	The number of impacted individuals in per confirmed case which is assumed as 1,000
X	The PPE allocation over 10 districts where $X = (x_1, ..., x_{10})$
G	The total allocation in Shenzhen where $G = \sum_i x_i$

TABLE 2 | The demographic information of 10 districts in Shenzhen and the predicated PPE allocation of the highest risk and the greatest severity objectives.

District	The predicted allocation of the largest high-risk	The predicted allocation of the greatest severity	Population in 2016	Age 65 in 2015	Cumulative cases on february 1, 2020
Bao'an	3	21	2,863,300	51,938	22
Dapeng	36	36	135,600	4,499	2
Futian	84	705	1,440,600	80,605	36
Guangming	4	3	531,200	8,620	5
Luohu	13	61	975,600	44,210	11
Longhua	48	4	1,511,500	23,434	24
Longgang	61	25	2,052,400	44,620	41
Nanshan	705	84	1,291,200	57,477	46
Pingshan	25	13	356,100	5,935	5
Yantian	21	48	221,200	8,173	3

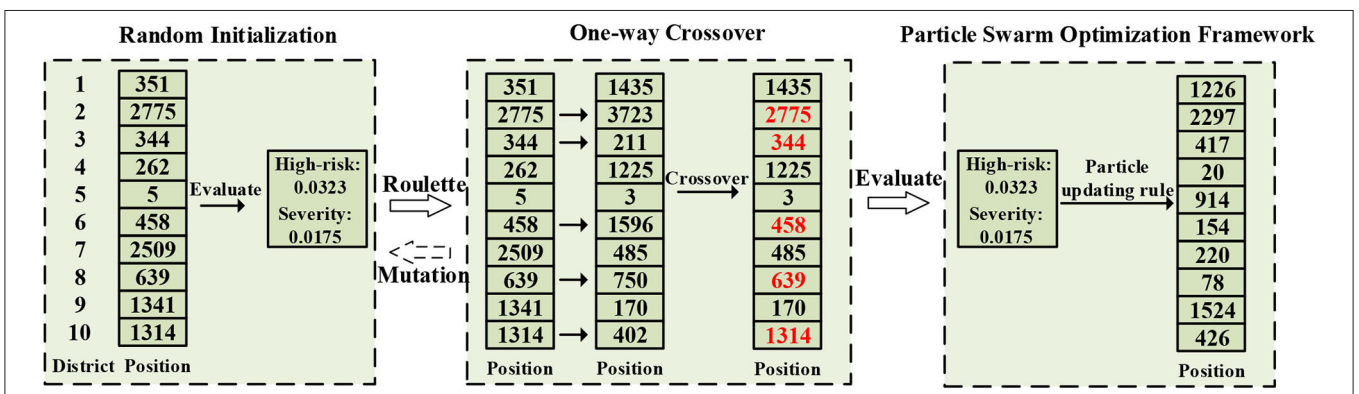


FIGURE 1 | The overview of the GA-PSO algorithm, which comprises three components, i.e., the random initialization, the genetic operations, and the particle swarm optimization framework. The initial particles are selected by the roulette method. When a random number is less than the mutation rate, the particle is executed the initialization operation.

allocation over 10 districts, and $G = \sum_i x_i$ represent the total allocation. The optimization strategy is based on the following assumptions and objectives:

- **High-risk population**: Due to multiple complications and a high fatality rate of people over 65, we select people over 65 [17]. The *ith* district with O_i people aged 65 or more is allocated x_i sets of PPE defined in Equation (1), where N_i is the population size in the *ith* district.

$$\sum_i x_i \times O_i/N_i \tag{1}$$

- **The severity of infection**: The *ith* district with I_i confirmed cases is allocated x_i sets of PPE defined in Equation (2). A_i is the number of impacted individuals per confirmed case, which is assumed to be 1,000 for all districts.

$$\sum_i x_i \times A_i \times I_i/N_i \tag{2}$$

Based on the above two objectives, this paper proposes a novel multi-objective optimization algorithm, named as GA-PSO, developed on the basis of MODPSO [18]. More specifically, GA-PSO is partially modified to simultaneously optimize the high-risk population and the severity of infection. The overview of GA-PSO is shown in **Figure 1**. It adopts one-way crossover and mutation operations to modify MODPSO in order to enhance the quality of solutions and escape the local optimum [19, 20].

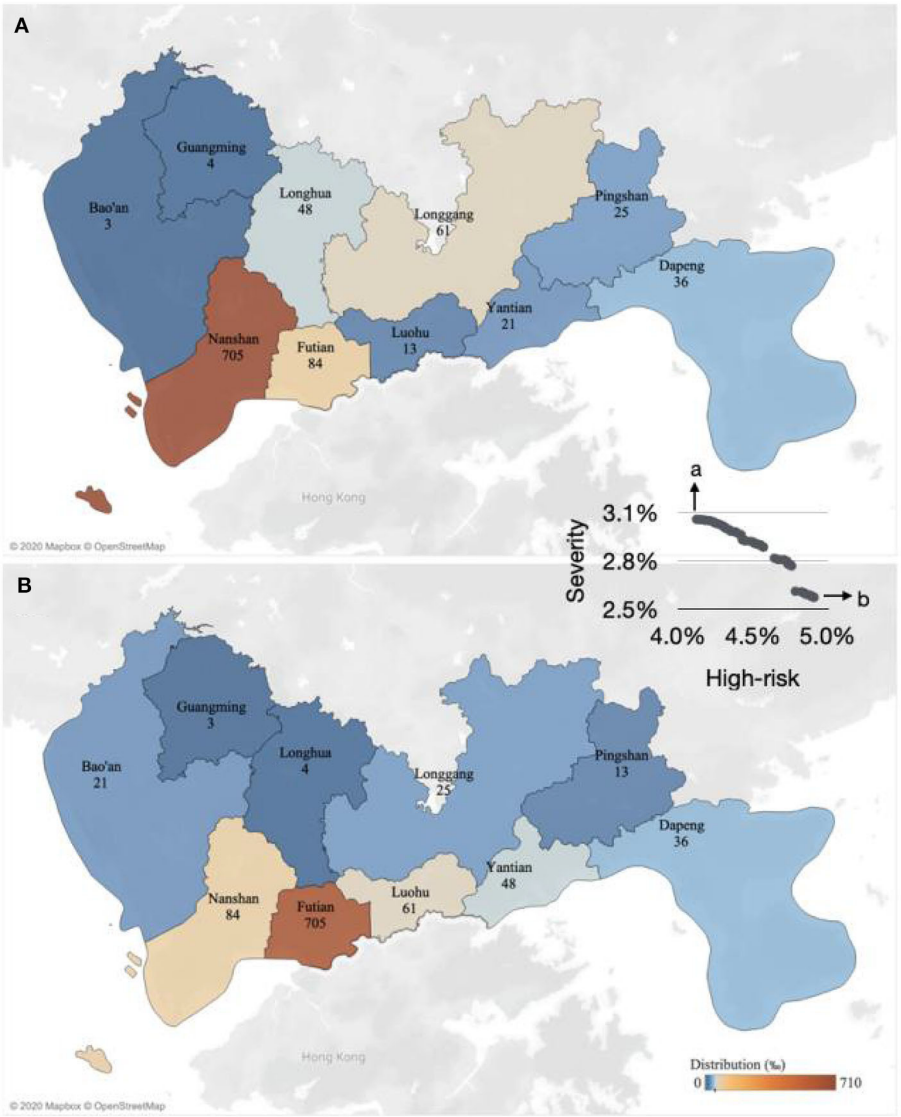


FIGURE 2 | Estimated PPE optimized allocation in the 10 districts of Shenzhen on February 14, 2020. The Pareto frontier uses the two objectives of high-risk and severity priorities in the middle right inset. The predicted allocation of the greatest high-risk priority in **(A)** and the predicted allocation of the greatest severity priority in **(B)**.

Every population is composed of a set of particles which is a feasible solution (i.e., X). Firstly, a random population of particles is initialized. X is a random integer vector by $G = \sum_i x_i$. Then, X is evaluated by the high-risk population and the severity of infection. It adopts the roulette method to select the preferable fitness particles. After that, it executes the one-way crossover operation. When a random number is less than the mutation value, it generates random particles. Finally, it iteratively generates new populations using particle status updating rules. During each iteration, the algorithm updates the Pareto solutions. When the number of iterations reaches the set value, the currently obtained Pareto solutions are the final solutions for the optimization problem.

3. RESULTS

The Shenzhen dataset is shown in **Table 2** and includes the population size in 2016 [13], the number of people aged 65 or more in 2015 [15], and the cumulative cases on February 1, 2020 [16]. Besides, the PPE optimized allocation is calculated by GA-PSO as shown in **Table 2**. **Figure 2** describes that the GA-PSO detects the Pareto front solutions for the PPE optimized allocation over 10 districts in Shenzhen. Every Pareto front solution indicates a feasible allocation. Different solutions can lead to different objective function values as quantified by the high-risk and severity priorities. The value of high-risk priority has ranged from 4.1% to 5.0%, and the value of the severity priority has ranged from 2.5% to 3.1% in **Figure 2**.

In summary, two objectives (i.e., the high-risk population and the severity of inflection) are added to obtain the PPE accessibility. The maximum value of the PPE accessibility is 7.53% for older or impacted individuals. When only the maximal high-risk population is the objective, the optimized allocation is estimated across 10 districts in Shenzhen, ranging from 3‰ in suburban Bao'an to 705‰ in central Nanshan. If the severity is the solitary objective, the allocation is from 3‰ in suburban Guangming to 705‰ in central Futian. The central districts of Futian and Nanshan top the PPE priority because of the maximal older people and confirmed cases.

4. DISCUSSION

Metropols (e.g., New York, London, Paris, and San Paulo) have been experiencing an urgent demand for medical emergency resources since the early phase of the COVID-19 epidemic. The use of a personal protective equipment (PPE) (e.g., hand sanitizers, surgical face masks, and protective clothing) has become one of the major non-pharmaceutical measures with which to reduce the infectious hazard. Due to the limited stock of PPE in particular during the early phase of the COVID-19 epidemic, the local health organization has suffered greatly. It is therefore a challenge to balance the supply and allocation of PPE during the available time. To overcome this challenge for practical applications, the multi-objective optimization is a promising method.

A multi-objective optimization algorithm generates a set of Pareto solutions each of which is a tradeoff between multiple objectives. Consequently, the multi-objective optimization facilitates the decision making problem. In addition to PPE, the pharmaceutical prophylactics (e.g., antibiotics, vaccines, antitoxins, and other critical medical supplies) are also faced with the decision making problem because of the limited supply in the early phase. Our optimization algorithm for PPE can therefore be extended to other pharmaceutical prophylactics to narrow the scope of a disease spread and to avoid catastrophic economic and social consequences.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

CW, CG, YD, ZY, and QC performed the experiments. CW, CG, and FZ analyzed the data. CW, CG, YD, and JK wrote the paper. CW, CG, ZY, CZ, and JK planned the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (No. 61976181), Natural Science Foundation of Chongqing (No. cstc2018jcyjAX0274), the Fundamental Research Funds for the Central Universities (No. XDJK2020D022), the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation (No. MNR KF-2019-04-034), Chongqing Graduate Student Research Innovation Project (No. CYS20116), and the Russian Ministry of Science and Education Agreement (No. 13.1902.21.0026).

REFERENCES

1. Chen S, Yang J, Yang W, Wang C, Barnighausen T. COVID-19 control in China during mass population movements at New Year. *Lancet.* (2020) **395**:764–6. doi: 10.1016/S0140-6736(20)30421-9
2. Wu P, Hao X, Lau EHY, Wong JY, Leung KSM, Wu JT, et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Eurosurveillance.* (2020) **25**:2000044. doi: 10.2807/1560-7917.ES.2020.25.3.2000044
3. World Health Organization. *Coronavirus Disease 2019 (COVID-19): Situation Report - 162* (2020). Available online at: https://www.who.int/docs/default-source/coronaviruse/20200630-covid-19-sitrep-162.pdf?sfvrsn=e00a5466_2.
4. Leung NH, Chu DK, Shiu EY, Chan KH, McDevitt JJ, Hau BJ, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med.* (2020) **26**:676–80. doi: 10.1038/s41591-020-0843-2
5. Ranney ML, Griffeth V, Jha AK. Critical supply shortages-the need for ventilators and personal protective equipment during the covid-19

pandemic. *N Engl J Med.* (2020) 382:e41. doi: 10.1056/NEJMp2006141
6. Zhu P, Wang X, Zhi Q, Ma J, Guo Y. Analysis of epidemic spreading process in multi-communities. *Chaos Solitons Fractals.* (2018) **109**:231–7.
7. Zhu P, Wang X, Li S, Guo Y, Wang Z. Investigation of epidemic spreading process on multiplex networks by incorporating fatal properties. *Appl Math Comput.* (2019) **359**:512–24. doi: 10.1016/j.amc.2019.02.049
8. Du Z, Holme P. Coupling the circadian rhythms of population movement and the immune system in infectious disease modeling. *PLoS ONE.* (2020) **15**:e0234619. doi: 10.1371/journal.pone.0234619
9. Du Z, Nugent C, Galvani AP, Krug RM, Meyers LA. Modeling mitigation of influenza epidemics by baloxavir. *Nat Commun.* (2020) **11**:1–6. doi: 10.1038/s41467-020-16585-y
10. Du Z, Xu X, Wang L, Fox SJ, Cowling BJ, Galvani AP, et al. Proactive social distancing mitigates COVID-19 outbreaks within a month across 58 mainland China cities. *Emerg Infect Dis.* (2020) 26. doi: 10.1101/2020.04.22.20075762
11. Zhu P, Wang X, Jia D, Guo Y, Li S, Chu C. Investigating the co-evolution of node reputation and edge-strategy in prisoners dilemma game. *Appl Math Comput.* (2020) **386**:125474. doi: 10.1016/j.amc.2020.125474
12. Bansal S, Pourbohloul B, Meyers LA. A comparative analysis of influenza vaccination programs. *PLoS Med.* (2006) **3**:e387. doi: 10.1371/journal.pmed.0030387
13. Yang X, Xie Z. *Shenzhen Statistical Yearbook.* Shenzhen.
14. Zhang D, Mou J, Cheng J, Griffiths S. Public health services in Shenzhen: a case study. *Public Health.* (2011) **125**:15–19. doi: 10.1016/j.puhe.2010.10.007
15. Futian Government. *Study on Evaluation of Population Aging in Futian District*. Shenzhen (2017). Available online at: http://www.szft.gov.cn/wap/ftzw/tjxx/qttjgb_41536/content/post_3321769.html
16. Shenzhen Health Commission. *Distribution of districts in Shenzhen on February 1, 2020*. Tencent (2020). Available online at: https://xw.qq.com/cmsid/20200202A04J5Y00
17. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. *The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19)–China, 2020* (2020). Available online at: http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51
18. Cai Q, Gong M, Ruan S, Miao Q, Du H. Network structural balance based on evolutionary multiobjective optimization: a two-step approach. *IEEE Trans Evol Comput.* (2015) **19**:903–16. doi: 10.1109/TEVC.2015.2424081
19. Wang C, Deng Y, Li X, Chen J, Gao C. Dynamic community detection based on a label-based swarm intelligence. *IEEE Access.* (2019) **7**:161641–53. doi: 10.1109/ACCESS.2019.2951527
20. Wang C, Deng Y, Li X, Xin Y, Gao C. A label-based nature heuristic algorithm for dynamic community detection. In: *PRICAI 2019: Trends in Artificial Intelligence.* Fiji (2019). p. 621–632.

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.

frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fams.2020.565336

Estimation and Monitoring of COVID-19's Transmissibility From Publicly Available Data

Antonio Silveira and Antonio Pereira Jr.*

Department of Electrical and Biomedical Engineering, Federal University of Pará, Belém, Brazil

The 2019 coronavirus disease (COVID-19) pandemic began in the city of Wuhan, China, at the end of 2019 and quickly spread worldwide. The disease is caused by contact with the SARS-CoV-2 virus, which probably jumped from an animal host to humans. SARS-CoV-2 infects various tissues in the body, notably the lungs, and patients usually die from respiratory complications. Mathematical models of the disease have been instrumental to guide the implementation of mitigation strategies aimed at slowing the spread of the disease. One of the key parameters of mathematical models is the basic reproduction ratio R_0, which measures the degree of infectivity of affected individuals. The goal of mitigation is to reduce R_0 as close or below 1 as possible, as it means that new infections are in decline. In this work, we use the recursive least-squares algorithm to establish the stochastic variability of a time-varying $R_0(t)$ from eight different countries: Argentina, Belgium, Brazil, Germany, Italy, New Zealand, Spain, and the United States. The proposed system can be implemented as an online tracking application providing information about the dynamics of the pandemic to health officials and the public at large.

Keywords: COVID-19, epidemic spreading, pattern recognition, mathematical modeling, transmission dynamics, disease prediction

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Wen-Xuan Wang,
Beijing University of Posts and Telecommunications (BUPT), China
Liangli Yang,
University of Science and Technology Beijing, China
Jun Hu,
Central University of Finance and Economics, China

****Correspondence:***
Antonio Silveira
asilveira@ufpa.br

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

Received: *24 May 2020*
Accepted: *12 August 2020*
Published: *05 November 2020*

Citation:
Silveira A and Pereira A Jr (2020) Estimation and Monitoring of COVID-19's Transmissibility From Publicly Available Data. Front. Appl. Math. Stat. 6:565336. doi: 10.3389/fams.2020.565336

1. INTRODUCTION

On March 11, 2020, the World Health Organization (WHO) declared the 2019 coronavirus disease (COVID-19) a global pandemic [1]. COVID-19 is caused by the SARS-CoV-2 coronavirus and was first reported in Wuhan, China, in December 2019 [2]. Since then, COVID-19 has spread globally with millions of laboratory-confirmed cases and hundreds of thousands of deaths [3]. The median incubation period of COVID-19 is 5.1 days, and nearly all infected persons who have symptoms will do so within 12 days of infection [4]. However, an unprecedented characteristic of COVID-19 is its capacity for asymptomatic transmission [5], which contributes to increase the probability of transmission [6]. So far, there is no specific treatment for the disease, and many research teams are currently working on a vaccine that, optimistically, will only be available in 2021. Even then, it will take some time to inoculate a significant share of the population. Meanwhile, hospital structures around the globe (e.g., intensive care units' beds, ventilators, and so on) are becoming overwhelmed with new patients, and the increasing caseload will prove most catastrophic for poor countries, which lack the adequate healthcare capacity to deal with the unparalleled demand posed by COVID-19 [7, 8].

So far, efforts to control the spread of the disease have focused on the adoption of non-pharmaceutical interventions (NPI) based on population behavioral change and social distancing, such as banning large gatherings, enforcing the use of face masks, washing hands, or imposing severe lockdowns [9, 10]. Due to significant uncertainties regarding the transmissibility of SARS-CoV-2 as well as other political, social, and economic considerations, it is necessary to delineate effective social distancing policies that are able to alleviate COVID-19's burden on healthcare structure and borrow time for the development of a vaccine or drug candidates while also simultaneously reducing the socioeconomic strain of living in a locked-down, confined society [11]. The effective monitoring of the epidemic's dynamics plays a crucial role in the ongoing containment efforts but will also continue to do so for some time ahead when social distancing control measures are eventually relaxed, and the first wave of the pandemic is followed some months later by second or third waves of infection that may be more severe than the first [12].

Mathematical models, by providing a quantitative framework for hypothesis evaluation and the estimation of changes in transmission of infectious diseases over time and space, can indicate whether containment measures are having a measurable effect while guiding the design of alternative interventions [13]. Mathematical models vary in many aspects, including complexity in terms of the number of variables and parameters used, spatial and temporal resolution (e.g., discrete vs. continuous time), and design (e.g., deterministic or stochastic) [14]. Mechanistic models of the susceptible–infected–recovered (SIR) type [15, 16] are the standard framework for a wide array of infectious diseases, including COVID-19 (see, for example, [17, 18]). However, parameter estimates for a given model are subject to two major sources of uncertainty: the noise intrinsic to the data and the *ad hoc* assumptions used for ascertaining parameter estimates [14].

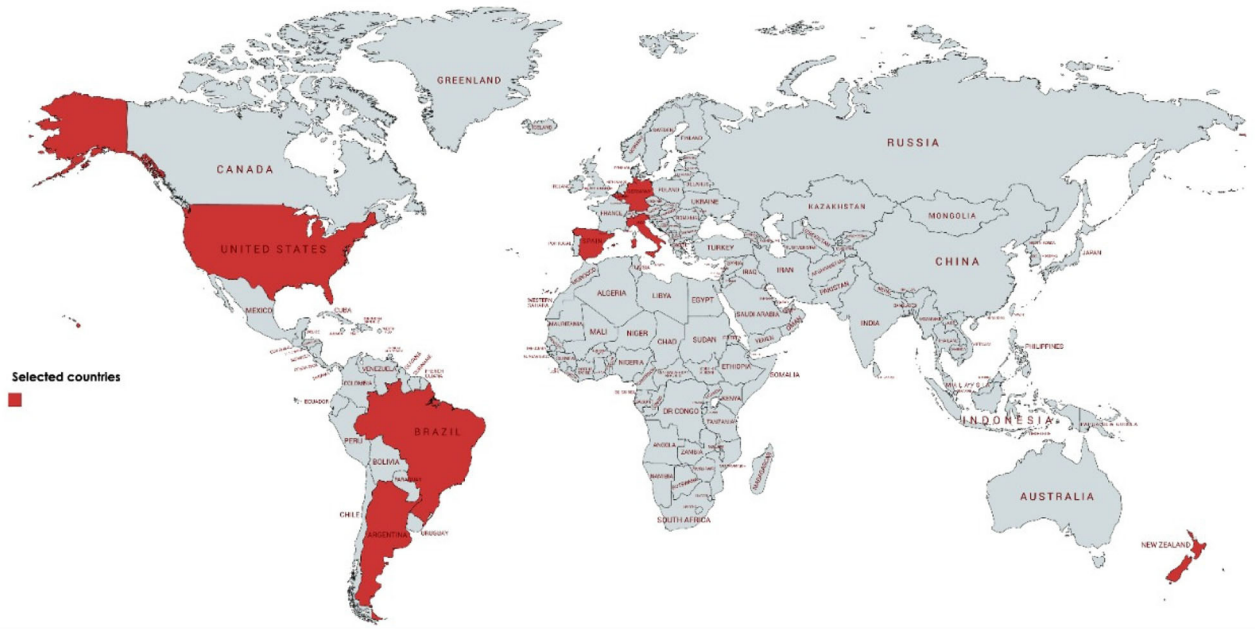
The basic reproduction number [19], R_0, or the average number of new infections caused by an infectious individual [20], is widely used to characterize the dynamics of infectious outbreaks and guide policy planning. For instance, the average R_0 at the start of the SARS pandemic in 2003 was estimated to be around 2.75 and was later reduced to 1 due to intervention strategies, including isolation and quarantine activities [21]. R_0 is an imprecise estimate that is rarely measured directly [it is the product of disease parameters: duration, opportunity, transmission probability, and susceptibility (DOTS)] and rests on particular model structures and assumptions [22]. Modelers face many challenges when trying to provide robust estimations of R_0 in the current pandemics, such as the existence of superspreaders, the fact that SARS-CoV-2 can also be spread by asymptomatic individuals and the scarce availability of testing supplies [17]. Several methods have been proposed to track trends in R_0 during the course of an epidemic [23–27]. The access to reliable estimates of R_0 could provide useful information about the efficacy of containment measures and allow their effective management in order to keep hospitalization rates within a desired approximate range [28].

In the present study, we model the transmission dynamic of COVID-19 in eight countries (**Figure 1**) using least-squares algorithm (LSA) techniques. The criteria used to select the eight countries were due to their geographical location in either the northern/southern hemisphere or due to specificities regarding their first response to the pandemic threat as reported in the news media. Some (e.g., New Zealand and Argentina) were very effective in responding rapidly to the pandemic by implementing lockdown measures, whereas others (e.g., USA and Brazil) initially downplayed the virus' threat and took longer to initiate containment protocols. Our goal is to contribute to understanding the spread of SARS-CoV-2 and compare R_0 uncertainty arising from noise in the time series data gathered from public online sources. We used machine learning algorithms to optimally estimate a time-varying $R_0(t)$, which allows the monitoring of the ongoing pandemic in almost real time. We compared the daily country reports for R_0 to those we estimated in the present work in order to assess the reliability of official data. Besides, LSA-based techniques were used to reveal clues regarding the dynamics of $R_0(t)$ and its stochasticity in terms of the linear power of the estimation error and provide information on how such stochastic behavior is correlated to the outcome of the ongoing pandemics.

The LSA is one of the most popular estimation methods in machine learning and has been used in many scientific and engineering applications [30–32], including epidemiology, for calibrating mathematical models' parameters based on time series data while also generating disease forecasts in the near or long terms [14, 33–35]. While it has been used for centuries as a classic curve-fitting technique [31, 32], it is still a basic tool in modern data science because of its least-squares Euclidean ℓ_2-norm minimization that is advantageous over other norms and metrics, such as the ℓ_∞ and ℓ_1 norms, granting reduced sensibility to outliers due to the squared error [32, 36]. Higher-order norms and the use of more generalized cost functions, both linear and non-linear, often require gradient descent solutions, which are the foundation of deep learning algorithms with none, many, or infinite solutions. In comparison, LSA is computationally inexpensive and can even be solved analytically.

LSA estimation is based on the least mean squares (LMS) algorithm, a special case of Bayesian estimation and the foundation of classical optimal estimation theory, where its applicability is commonly attributed to offline *batch processing*, that is, the whole data set must be *a priori* available and be processed at once "in one single step into the estimate" [36], involving complex algebraic procedures, such as the inversion of high-order matrices, with dimension equal to the length of the data set. Its offshoot, the recursive least squares (RLS) algorithm, has been used for real-time estimation applications in diverse areas, such as signal and data processing, communications, and control systems [37, 38], since it benefits from the recursive method and avoids matrix inversion by working one sample at a time, both speeding up processing and avoiding a possible ill-conditioned (non-invertible) information matrix formation.

Some of the advantages of RLS algorithm over LMS algorithm, and other more complex gradient descent-based estimation methods, are as follows: its recursive or sequential processing, which requires less memory over a single iteration step; the possibility to capture the dynamics of non-linear and time-varying systems; its native discrete-time synthesis to



Country	Population	GDP	HDI	Covid-19		Date of 100th case	Day control started (in days since 100th case)			
				Cases	Deaths		L1	L2	L3	L4
Argentina	44,938,712	9,887	0.830	9,918	416	03/19/20	prior	prior	prior	prior
Belgium	11,589,623	46,724	0.919	56,511	9,212	03/06/20	4	6	11	-
Brazil	210,147,125	8,955	0.761	312,000	20,112	03/13/20	6	-	-	-
Germany	83,149,300	46,653	0.939	179,000	8,320	03/01/20	10	12	13	21
Italy	60,317,116	32,947	0.883	228,000	32,486	02/23/20	7	10	17	27
New Zealand	4,984,340	41,616	0.921	1,154	21	03/19/20	prior	2	4	6
Spain	47,007,367	30,774	0.893	235,000	28,628	03/2/20	8	13	13	13
USA	328,239,523	67,426	0.920	1,610,000	92,513	03/8/20	4	12	12	12

FIGURE 1 | Sociodemographic and COVID-19 data for the eight countries targeted in this study. COVID-19 data accessed on 22/05/2020 on https://coronavirus.jhu.edu/map.html. Intervention data gathered from reference [29]. GDP per capita in US$. Interventions are categorized into Alert Levels 1 to 4 according to the New Zealand framework (L1, Level 1, Prepare; L2, Level 2, Reduce; L3, Level 3, Restrict; L4, Level 4, Eliminate) [29]. Map from mapchart.net.

deal with discrete-time signals; and its dead-beat convergence characteristics, that is, convergence in minimal time [31]. However, one important drawback is the possibility of RLS getting stuck in local minima and becoming unstable when tracking certain classes of signals, as remarked in reference [31]. Thankfully, the exponential functions used in epidemic modeling are within the stable scope of convergence of the RLS method.

In this work, we compare the performance of both the RLS and LMS algorithms on estimating $R_0(t)$, in terms of processing speed and accuracy. While both are LSA-based methods apparently differing just by the sequential-recursive and the batch-processing forms of implementation, we are dealing with a general case problem of estimating a random variable $R_0(t)$ given random variables as well in a maximum-likelihood estimation problem that "implies ignorance of any statistics of the estimated variable" [36], $R_0(t)$. Thus, if the available measurements are independent, RLS and LMS algorithms should achieve the same accuracy performance. Otherwise, the RLS algorithm should perform better.

Our modeling approach is based on a discrete-time multiple-input, multiple-output (MIMO) setup that uses a well-established concept in the design of aerospace navigation systems called sensor fusion, which is summarized by the following statement, "one always gains by adding a new measurement in terms of navigation error, and this no matter how bad the additional measurement is" [39, Remark 4.8]. In this work, we combine data from the number of susceptibles, infections, recoveries, deaths, and individual parameters of three coupled differential equations in order to improve the estimation of $R_0(t)$.

2. METHODS

2.1. Data Sources

The COVID-19 data used in this report are publicly available from The Center for Systems Science and Engineering of the Johns Hopkins University (JSU CCSE) [40], which maintains a Repository on Github [41].

2.2. Procedures

The transmission dynamics of the COVID-19 outbreak is usually described by a compartmental model SEIR, where S denotes susceptible, E denotes exposed, I denotes infected, and R denotes removed [42, 43]. In a closed population of P_n individuals, the transitions between the compartments (cf. **Figure 2**) are described through the following set of differential equations:

$$\frac{d}{dt}s(t) = -\bar{\lambda}s(t)i(t) \tag{1}$$

$$\frac{d}{dt}e(t) = \bar{\lambda}s(t)i(t) - \kappa e(t) \tag{2}$$

$$\frac{d}{dt}i(t) = \kappa e(t) - \gamma i(t) \tag{3}$$

$$\frac{d}{dt}r(t) = \gamma i(t) \tag{4}$$

where $\bar{\lambda} = \lambda/P_n$, λ is the infection rate, γ is the remove rate, and κ is the incubation rate. From these parameters, it is possible to calculate the basic reproduction ratio, $R_0 = \lambda/\gamma$. Thus, R_0 is not solely dependent on the infection rate but also on the frequency of removals due to death or recoveries.

Assuming that the incubation period of the disease is instantaneous, and the duration of infectivity is the same as the length of the disease, we can consider both groups E and I as contagious and $E(t) := I(t)$. Also, according to the Akaike information criterion (AIC), the simpler SIR model performs much better than an SEIR model in representing the information contained in the confirmed-case data available for COVID-19 [44]. The basic SIR model is described by the set of Kermack–McKendrick equations [45]:

$$\frac{d}{dt}s(t) = -\frac{\lambda}{P_n}s(t)i(t) \tag{5}$$

$$\frac{d}{dt}i(t) = \frac{\lambda}{P_n}s(t)i(t) - \gamma i(t) \tag{6}$$

$$\frac{d}{dt}r(t) = \gamma i(t) \tag{7}$$

2.3. Discrete-Time Susceptible–Infected–Recovered System Parametric Estimation in Real Time

Traditionally, mathematical epidemiology models have been approached with a continuous-time perspective, due in part to the fact that these are more tractable mathematically [46]. However, in order to use machine learning techniques, there is a need for a discrete-time equivalent realization to cope with daily-sampled data [47]. Due to the slow dynamics of the pandemics, a first-order continuous to discrete Euler approximation can be applied to the Kermack–McKendrick equations.

For a general $f(t)$ function, a backward discrete-time derivative approximation is given as:

$$\frac{d}{dt}f(t) := \frac{f(k+1) - f(k)}{T_s} = \Delta f(k+1)/T_s \tag{8}$$

where $T_s = 1$ is the sampling interval in days, and $\Delta = 1 - q^{-1}$ is the discrete difference operator, defined in q^{-1}, the backward shift operator domain. The discrete-time approximations of Equations (5)–(7) are given by the following difference equations, respectively:

$$s(k) = s(k-1) - \lambda(k-1)\frac{s(k-1)i(k-1)}{P_n} \tag{9}$$

$$i(k) = i(k-1) + \lambda(k-1)\frac{s(k-1)i(k-1)}{P_n} - \gamma(k-1)i(k-1) \tag{10}$$

$$r(k) = r(k-1) + \gamma(k-1)i(k-1) \tag{11}$$

The discrete-time SIR system described above considers time-varying parameters in order to continuously adapt the model as new data become available. Using the time series of infections and removals (due to death or recovery), Equations (9)–(11) can be used to estimate the model parameters.

Since Equation (11) has an exclusive dependence with $\gamma(k)$, this poses a direct estimation problem that can be stated as "for N registered samples, minimize the following quadratic cost function":

$$J_r = \frac{1}{2}\sum_{0}^{N} e_r^2(k) = \frac{1}{2}\sum_{k=0}^{N}\left[r(k) - \hat{r}(k)\right]^2 \tag{12}$$

Equation (12) is based on the estimation error of $r(k)$. By applying the RLS method to minimize J_r, it is possible to optimally estimate $\hat{\gamma}(k)$ using the following equation:

$$\hat{r}(k) = r(k-1) + \hat{\gamma}(k-1)i(k-1) \tag{13}$$

We assume that the estimation error is Gaussian, $e_r(k) \sim (0, \sigma_{e_r}^2)$, with zero mean and variance $\sigma_{e_r}^2$, such that $r(k) = \hat{r}(k) + e_r(k)$. The estimator gain, the parametric estimation, and error covariance minimization are solved recursively as follows:

$$L_r(k) = \frac{p_r(k-1)i(k-1)}{1 + i(k-1)p_r(k-1)i(k-1)} \tag{14}$$

$$\hat{\gamma}(k) = \hat{\gamma}(k-1) + L_r(k)\left[r(k) - \hat{r}(k)\right] \tag{15}$$

$$p_r(k) = [1 - L_r(k)i(k-1)]p_r(k-1) \tag{16}$$

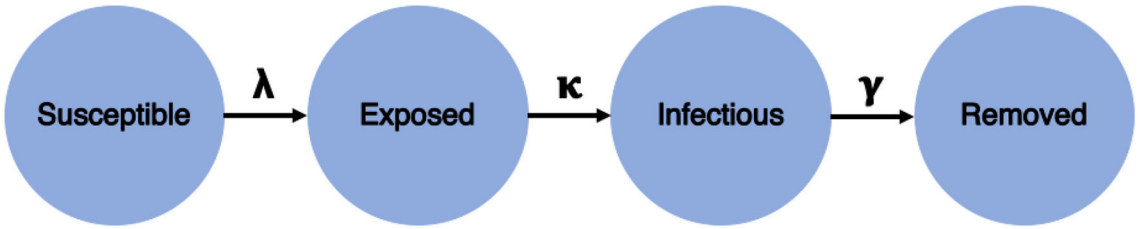


FIGURE 2 | SEIR (S denotes susceptible, E denotes exposed, I denotes infected, and R denotes removed) model structure.

where the error covariance matrix $p_r(k)$ can be reset periodically to prioritize more recent data. Specifically in this work, $p_r(k)$ is a scalar, since only a single parameter is being estimated. Choices to initialize the covariance matrix may vary, depending on prior available covariance information or other positive definite matrix. The higher its magnitude, the higher the estimator gain in the transitory dynamical stage of the estimation procedure.

In regard to Equation (11) and the estimation of $\hat{\gamma}(k)$, since the infection numbers increase before any removal report is available during the first stages of the pandemic, during this period $\hat{\gamma}(k)$ tends to zero and eventually makes the time-varying estimated reproduction number tend to infinity:

$$\hat{R}_0(k) = \frac{\hat{\lambda}(k)}{\hat{\gamma}(k) \to 0} = \infty \tag{17}$$

As a consequence of Equation (17), the estimation method proposed in this work cannot be applied when the number of recovered is not available.

Since Equation (10) depends on both SIR parameters, due to its stronger dependence on $\hat{\gamma}(k)$ and Equation (11), its estimated value is substituted into Equation (10), and another RLS problem is constructed to estimate $\hat{\lambda}(k)$, using

$$\hat{i}(k) = i(k-1) + \hat{\lambda}(k-1)\frac{\hat{s}(k-1)i(k-1)}{P_n} - \hat{\gamma}(k)i(k-1) \tag{18}$$

The solution is akin to minimizing the estimation error $e_i(k) = i(k) - \hat{i}(k)$, using the following equations for estimator gain, parametric estimation update, and error covariance minimization, respectively:

$$L_i(k) = \frac{p_i(k-1)\hat{s}(k-1)i(k-1)/P_n}{1 + \left[\hat{s}(k-1)i(k-1)p_i(k-1)\hat{s}(k-1)i(k-1)\right]/P_n^2} \tag{19}$$

$$\hat{\lambda}(k) = \hat{\lambda}(k-1) + L_i(k)\left[i(k) - \hat{i}(k)\right] \tag{20}$$

$$p_i(k) = \left\{1 - L_i(k)\left[\frac{\hat{s}(k-1)i(k-1)}{P_n}\right]\right\} p_i(k-1) \tag{21}$$

Time-series data for Equation (9) is not available, and the evolution of the susceptible compartment in time is in fact estimated based on the known initial condition (i.e., the population P_n) and on the estimated $\hat{\lambda}(k)$. Thus, it is always estimated and fed back to Equation (18), such that the correct form to represent it, within this time-varying SIR model, is by rewriting Equation (9) based on the estimated susceptible:

$$\hat{s}(k) = \hat{s}(k-1) - \hat{\lambda}(k)\frac{\hat{s}(k-1)i(k-1)}{P_n} \tag{22}$$

The estimation of the time-varying reproduction number, based on the derived discrete-time SIR model, is given by:

$$\hat{R}_0(k) = \frac{\hat{\lambda}(k)}{\hat{\gamma}(k)} \tag{23}$$

We also adopted two modifications to the nominal $\hat{R}_0(k)$ equation: a moving 4-days average to compensate for the randomness of daily updates on incidence data, as seen in the German Daily Situation Report of the Robert Koch Institute on COVID-19 [48], and the proportion of susceptible individuals in the population, known as the effective reproduction number [14]:

$$\bar{R}_0(k) = \frac{\hat{s}(k)}{P_n}\left[\frac{\hat{R}_0(k) + \hat{R}_0(k-1) + \hat{R}_0(k-2) + \hat{R}_0(k-3)}{4}\right] \tag{24}$$

Figure 3 shows a block diagram of the proposed $\bar{R}_0(k)$ estimator. This diagram presents a clearer view of the coupled multivariate dynamics and closed-loop characteristics of the proposed estimation approach.

With this formulation, it is possible to analyze the transmission ratio of the pandemics on a daily basis, as with a sensor. Besides, estimations of the transmission ratio produce a dynamic representation from the perspective of the time series of $\bar{R}_0(t)$ [henceforth designated simply as $\hat{R}_0(t)$], which allows the modeling of its dynamics and its randomness in order to assess stochastic properties correlated to the time-varying reproduction number, which might reflect how health authorities have been handling the challenges posed by the pandemics in each country considered in this work.

2.4. Modeling of R_0

Henceforth, we assume that we are able to estimate the reproduction number on a daily basis and it is thus possible to consider it as another output of the proposed pandemic model. Thus, by relying on the time series of $\hat{R}_0(k)$ and knowing it is

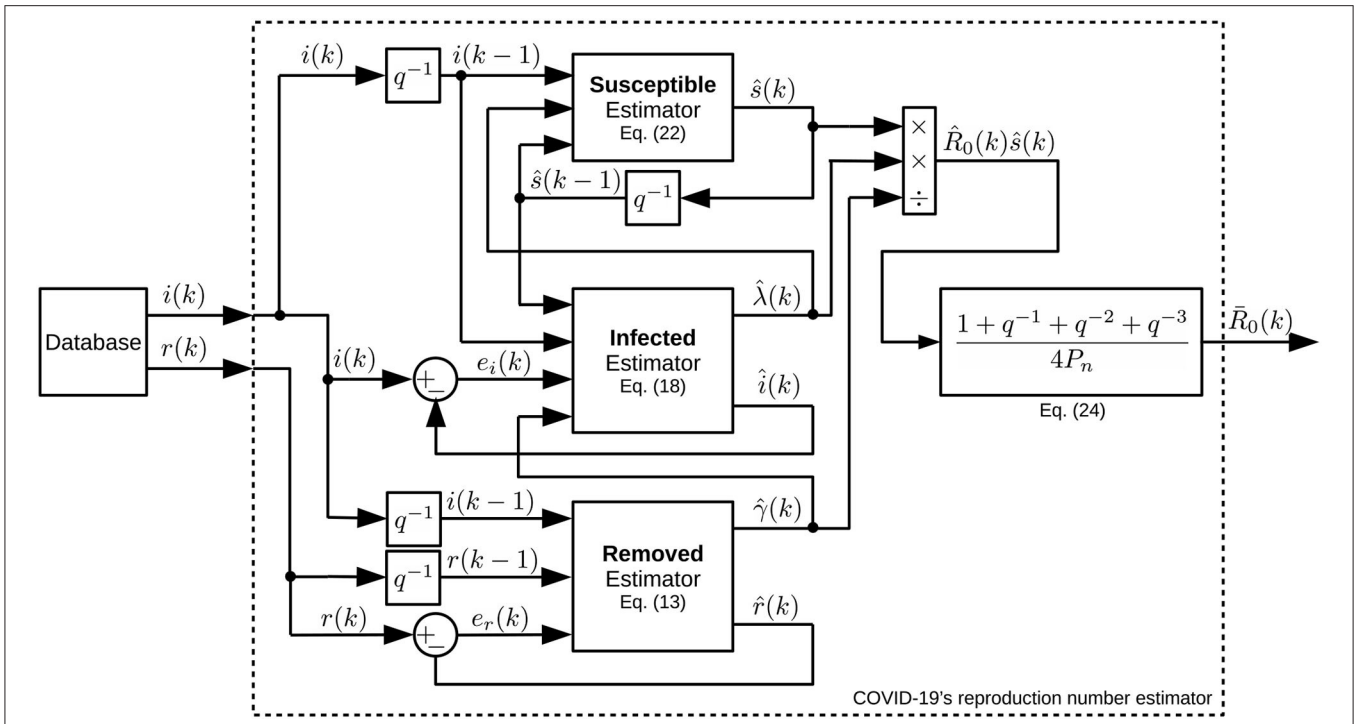


FIGURE 3 | Block diagram of the estimation system. Notice how its subsystems are interconnected using closed-loop strategies and data fusion.

correlated with the number of infected and removed individuals, we deploy machine learning techniques to identify a dynamic system that fits the data.

Different from the real-time monitoring/sensing procedure adopted to estimate $\hat{R}_0(k)$, now, we are interested in obtaining a general model that can describe the dynamics of a system, $\hat{R}_0(q^{-1})$, for a certain period of interest. In order to model such a system, we used the LMS algorithm, which is a non-RLS estimation technique [31, 32], and propose a black-box linearized polynomial model:

$$\hat{R}_0(q^{-1}) = \frac{\hat{B}(q^{-1})I(q^{-1}) + \hat{C}(q^{-1})R(q^{-1}) + e_{R_0}(q^{-1})}{\hat{A}(q^{-1})} \tag{25}$$

$$\begin{aligned} \hat{A}(q^{-1}) &= 1 + \hat{a}_1 q^{-1} + \hat{a}_2 q^{-2} \\ \hat{B}(q^{-1}) &= \hat{b}_0 q^{-1} + \hat{b}_1 q^{-2} \\ \hat{C}(q^{-1}) &= \hat{c}_0 q^{-1} + \hat{c}_1 q^{-2} \end{aligned} \tag{26}$$

where $e_{R_0}(q^{-1})$ is the Gaussian process based on the estimation error $e_{R_0}(k)$ of estimated polynomials shown in Equation (26).

This second-order autoregressive with exogenous input (ARX)-based model structure is assumed considering the fundamental simplicity of the SIR model, in which the infected and removed systems together form a second-order system.

The non-RLS estimator is a batch-processing technique used to optimally estimate the set of parameters that minimizes a quadratic performance index as the one shown in Equation (12), but using a vector–matrix form of error, $\mathbf{e}_{R_0} = \hat{\mathbf{R}}_\mathbf{0} - \boldsymbol{\Phi}\hat{\boldsymbol{\theta}}$. This vector–matrix system is defined as:

$$\mathbf{e}_{R_0} = \left[e(0) \ \cdots \ e(N)\right]^T \tag{27}$$

$$\hat{\mathbf{R}}_\mathbf{0} = \left[\hat{R}_0(0) \ \cdots \ \hat{R}_0(N)\right]^T \tag{28}$$

$$\hat{\boldsymbol{\theta}}^\mathbf{T} = \left[\hat{a}_1 \ \hat{a}_2 \ \hat{b}_0 \ \hat{b}_1 \ \hat{c}_0 \ \hat{c}_1\right] \tag{29}$$

$$\boldsymbol{\Phi}^\mathbf{T} = \left[\phi^T(0) \ \cdots \ \phi^T(N)\right] \tag{30}$$

The above equations represent, respectively, the vector of errors, the vector of observed outputs, the estimated parameter vector, and the matrix of regressors. The latter is based on the vectors of regressors formed up to N registered samples, with such vectors defined as:

$$\phi^T(k) = \left[\hat{r}_0(k-1) \ \hat{r}_0(k-2) \ i(k-1) \ i(k-2) \ r(k-1) \ r(k-2)\right] \tag{31}$$

The solution to obtain the estimated parameters is straightforward and given by

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{\Phi}^\mathbf{T}\boldsymbol{\Phi}\right)^{-\mathbf{1}}\boldsymbol{\Phi}^\mathbf{T}\hat{\mathbf{R}}_\mathbf{0} \tag{32}$$

By assuming an ARX linear model, it is possible to evaluate the pandemics from the perspective of linear stochastic systems theory, assessing how the reproduction number decays linearly in time in different countries and how the random nature of events associated with the pandemics affects the model's uncertainties, that is, $e_{R_0}(k) \sim (0, \sigma^2_{e_{R_0}})$. This calculated uncertainty can give us some clues regarding the effectiveness of pandemic control measures.

We propose that by analyzing the linear power of the $\hat{R}_0(q^{-1})$ estimation error, that is, $\sigma^2_{e_{R_0}}$, we can use this stochastic property to generate *stochastic ratio* curves based on the estimated linearized model of Equation (25). The higher the variability associated with the stochastic ratio, the higher we expect the variance or linear power of the error to be, and consequently, the more uncertain is the official COVID-19 data reported by health authorities.

3. RESULTS

We used publicly available data to validate the algorithms and the estimated time-varying SIR model parameters. This section is organized as follows: we first present the estimated results based on the number of infectious and removals available from different countries. Then, we present time series of daily estimates on $R_0(k)$ based on the RLS-estimated $\hat{\lambda}(k)$ and $\hat{\gamma}(k)$ and compare both processing time and accuracy results to the classical LMS method to verify the efficiency of the proposed technique.

These results are followed by the presentation of linearized estimated outputs generated by the ARX-based $\hat{R}(q^{-1})$ models [shown in Equation (25)] using non-recursive or batch processing of 30-days of $R_0(k)$ estimates. The modeling residuals were assumed to be Gaussian (zero mean). Estimated error variances were used to produce 200 discrete Gaussian sequences as surrogates to additive white noises, depicted by $e_{R_0}(k)$ in Equation (25). These 200 additive noise sequences were used to generate, for every analyzed country, 200 stochastic reproduction ratio trajectories, shown together with the linearized ratio and the real-time estimated ratio. It must be remarked, however, that this value of 200 Monte Carlo simulations were selected heuristically in order to give the necessary visual information in our figures, such that the readers could verify by themselves how the modeling uncertainties of the pandemics could mislead our judgment of the probable effective reproduction number of COVID-19 during the studied period. The number of Monte Carlo simulations thus affects only a post-modeling phase and do not interfere with the estimation of the reproduction number.

3.1. Discrete-Time Susceptible–Infected–Recovered Model Estimation Results

The Achilles' heel of the RLS algorithm for parametric estimation may be the setup of initial conditions and whether they are optimal or not. However, this is a major problem only if the estimated parameters are to be used in a real-time adaptive control system where the closed-loop stability must be guaranteed [31]. This is not the case in the present work, which is interested only in the modeling question itself. Thus, the initial RLS parameters can be either arbitrarily set or set at zero since, theoretically, the RLS estimator is dead-beat and converges in the minimum possible number of iterations [32, 36].

The discrete SIR model proposed in this work is described by three coupled differential equations, each based on a single recursive regression and thus forming a third-order system. Then, theoretically, as a dead-beat estimator, the RLS estimator would take three iterations to converge and estimate optimal parameters. However, we used plenty more iterations than the theoretical requirements by commencing to process data from March 22, 2020 onward but evaluating the results after April 23, 2020, thus giving more than 30 days/iterations for the RLS to converge to an optimal set of parameters by April 23, 2020, when we started our analysis.

The estimators were implemented with arbitrary initial parameters of $\hat{\gamma}(0) = \hat{\lambda}(0) = 1$ and the initial guess for $\hat{R}_0(0) = 1$. The magnitudes of the estimation error covariance matrices were initialized as $p_r(0) = p_i(0) = 1$, considering that the initial error is large. Both $p_r(k)$ and $p_i(k)$ were reset to 1 every 7 days to prioritize more recent data [49]. The selection of the reset period considered not only the number of days required to wash-out outliers but also to allow weekly changes in the parameters' dynamical behavior, such as due to lockdowns or quarantine relaxation. We also adopted a moving 4-days average for $\hat{R}_0(k)$, shown in Equation (24), to compensate for daily random effects [48]. We only used data from March 22, 2020 onward, when all eight countries already had more than 100 infections reported.

Figures 4, **5** show the dynamics of both infected and removed cases using the SIR model. It must be remarked that the infected curve in the United States was downscaled by a factor of 4 in order to fit in the graph along with the other studied countries. It is evident from **Figure 4** that the RLS method provide parametric estimates that fit the reported data. The same goodness of fit cannot be observed in **Figure 5**, as the number of removals for the United States, for example, poses some difficulties for the RLS estimator, as depicted by the estimated removals (dotted lines) showing a certain dispersion from the real data (continuous lines). However, the inclusion of the removed data set, even with bad measurements, leads to a better estimate of $\hat{R}_0(k)$.

Figure 6 presents the $\hat{R}_0(k)$ of the eight different countries during a period of 2 weeks. Despite the large variability of both Argentina and the United States, there are other countries that are already reducing the number of new infections and where the frequency of removals has increased, such as in Germany, Italy, and New Zealand (see **Table 1**).

The trajectory of the curves shown in **Figure 6** can also be associated with some extraordinary events that occurred during the same period. For example, after May 3, 2020, when some U.S. states had relaxed social distancing guidelines, it is possible to observe a corresponding phasic increase in the basic reproduction ratio of the United States, which was also reported in the Washington Post on May 9, 2020 [50].

One interesting trajectory in **Figure 6** regards Argentina. For most of the time, the estimated reproduction ratio of Argentina was one of the highest and comparable only to the United States, despite its low number of infected individuals (cf. **Figure 4**).

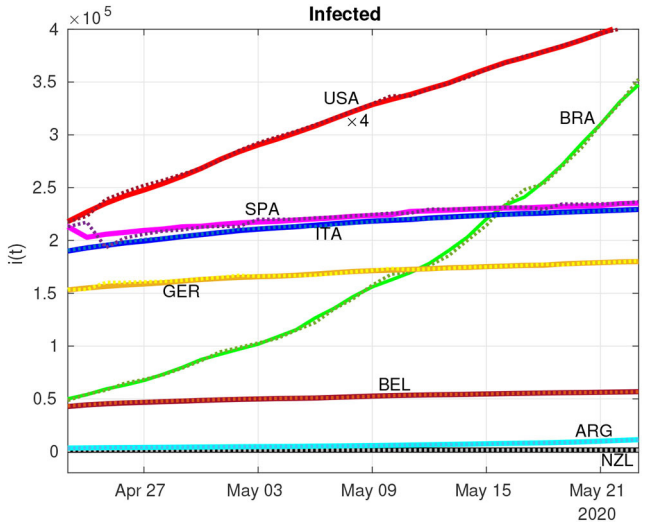


FIGURE 4 | Estimation (dotted lines) of infected individuals (continuous lines) using the recursive least-squares (RLS) technique (the U.S. data was downscaled four times to fit in the graph).

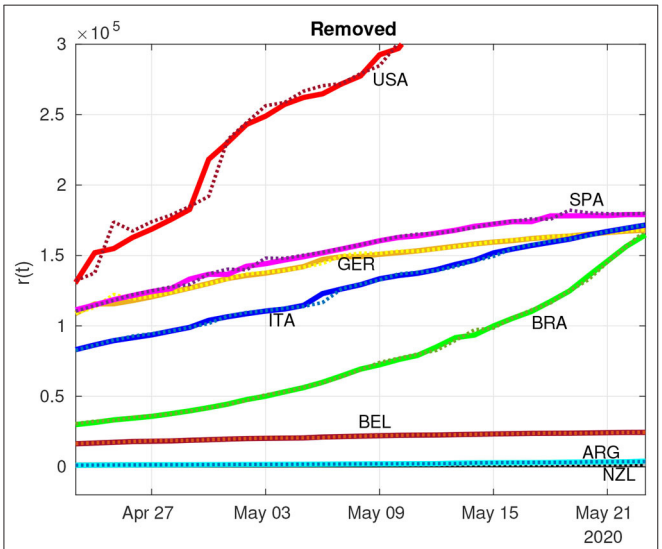


FIGURE 5 | Estimation (dotted lines) of removed individuals (continuous lines) using the RLS technique.

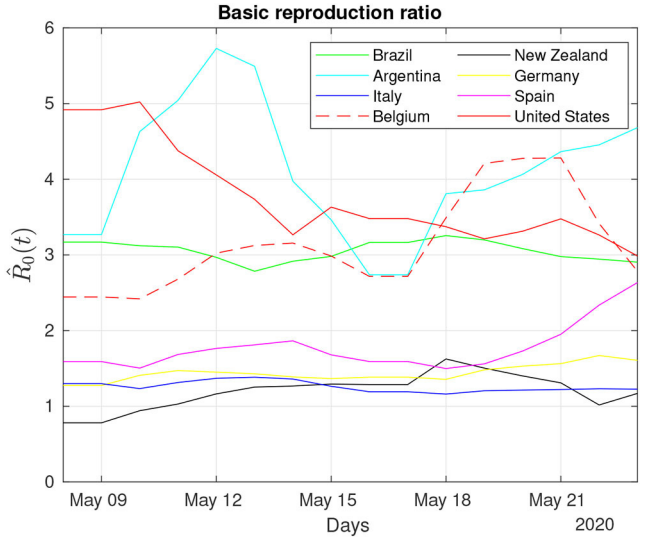


FIGURE 6 | Daily monitoring of the COVID-19 pandemic: Dynamic behavior of the estimated reproduction number for 2 weeks in May 2020.

TABLE 1 | Recoveries (May 21, 2020).

Country	Recoveries (%)
Argentina	30.5
Belgium	26.6
Brazil	40.5
Germany	88.2
Italy	59.0
New Zealand	96.6
Spain	70.3
United States	23.5

This apparent contradiction is related to the low number of removals at the beginning of the pandemic that tends to raise the $\hat{R}_0(k)$ (see Equation 17). Recoveries in Argentina, based on the data of May 16 (cf. **Table 1**), are ∼30.5% of its total infected, close to Belgium with 26.6% and whose transmission ratio is below Argentina's ratio, reinforcing the notion that the transmission ratio is not a static parameter and is best approached by dynamical systems theory.

The results displayed in **Figure 4** seem at odds with a recent report that claimed that Brazil's R_0 had recently dropped to 1.4 [51]. However, on the same day a Brazilian newspaper quoted this report, the country had a record number of new infections and deaths [52]. This reinforces the notion that machine learning methods might give better and faster clues regarding the severity of the pandemic in terms of $\hat{R}_0(k)$.

We compared the performance of the proposed RLS estimator with its most common counterpart, the LMS algorithm. In this comparison, we considered the average processing time and the normalized relative estimation error based on the mean-square-based cost function shown in Equation (12). The average processing time results for the RLS and LMS algorithms were, respectively, 14 and 20 ms. We used a computer with a fourth-generation Intel Core i5-4200U CPU at 1.6 GHz, 4 GB of RAM, running Ubuntu 18.04 LTS and Matlab R2018a (Mathworks, Inc., Natick, MA, United States). The accuracy results expressed by normalized relative errors are summarized in **Figure 7**. Notice that the RLS method outperformed the LMS method in all cases, except for Argentina.

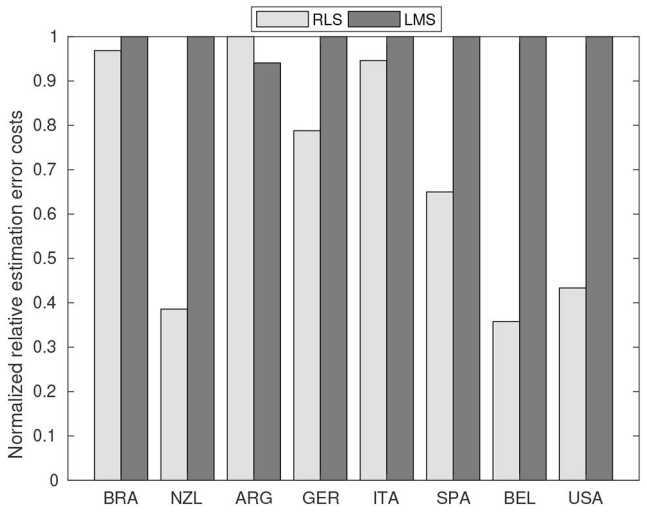


FIGURE 7 | Comparison of normalized relative estimation error costs between the proposed RLS-based technique and the least mean squares (LMS) estimation algorithm. The average processing time of RLS and LMS algorithms was, respectively, 14 and 20 ms.

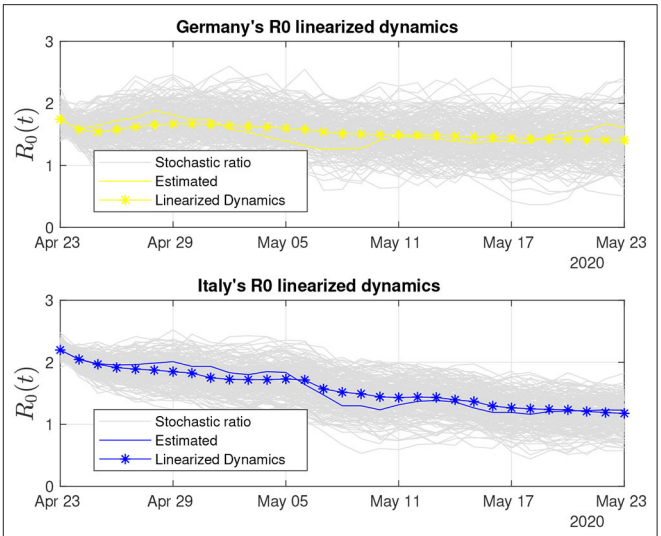


FIGURE 8 | Thirty days of linearized R_0 dynamics for Germany and Italy. Both countries display low variability in the stochastic ratio curve.

3.2. Assessing the Pandemics Through R_0 Dynamics

Figures 8–11 show $\hat{R}_0(k)$ for a 30-days period for the eight investigated countries. These figures show linearized R_0 together with the respective real-time estimated data that originated this second-stage estimation. The variance of the estimation error is then used to generate Gaussian sequences that are superimposed on the linearized R_0 estimate, the stochastic ratio, which synthetically reproduces the stochasticity and uncertainties of R_0 estimates. Both Germany and Italy (**Figure 8**) are through a period of consistently decaying $R_0(t)$. Despite the hard way COVID-19 hit Italy before, its stochastic ratio currently is among the lowest.

Brazil and the United States (**Figure 9**) are the two most populous countries of our sample and the most hard-hit by COVID-19 among them. Both countries also have been struggling with their uneven response to the pandemics [53]. This outcome is captured by our R_0 sensors, with the Brazilian stochastic ratio being in decrease, as recovered data become more available (see **Figure 9**).

In **Figure 10**, we present linearized R_0 estimates for Spain and Belgium. Spain's estimates have suffered the influence of annotation errors (by subtracting infected individuals on April 24, 2020), which eventually provoked oscillations in the real-time ratio estimate, making it zero-cross on April 27. Such an error was washed out by the linearized estimate and has certainly contributed to its increased degree of uncertainty (see **Table 2**).

Belgium's degree of uncertainty was the worst among all the countries, at least during the period we analyzed. One possible clue to understand why Belgium's stochastic ratio became so variable is to consider the impossibility to linearize its dynamics and the associated increase in error. However, our estimation

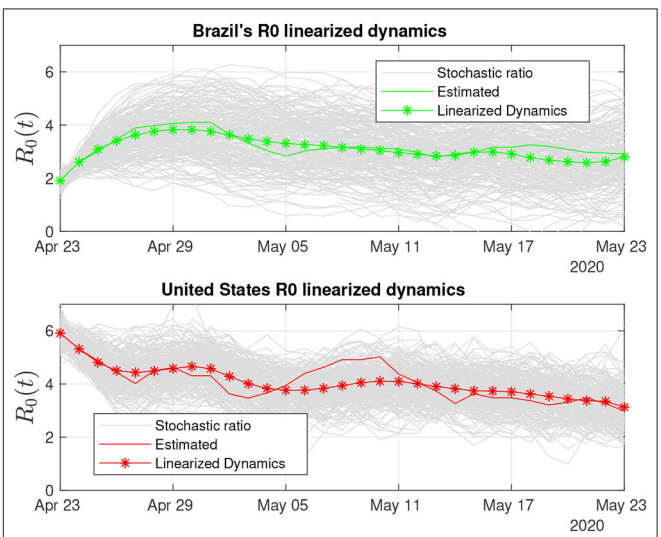


FIGURE 9 | Brazil and the United States display large variability in the stochastic ratio curve, which probably reflects data uncertainty and difficulties to control COVID-19's spread.

procedure considers the error as Gaussian, and its mean value in fact approaches zero. Thus, the linearized estimate based on 30 days of data has a high probability to be close to the values shown.

Belgium has also shown an increased R_0 during the analyzed period. The low number of recoveries (see **Table 1**) influenced the frequency of removals and the basic reproduction number, which is the ratio between the frequency of infections and the frequency of removals (see Equation 23). Even though a recent report [54] proposed that Belgium's R_0 at the beginning of May 2020 was 0.8, according to our adaptive SIR-based R_0 estimations, the current situation in Belgium is uncertain with a probable R_0

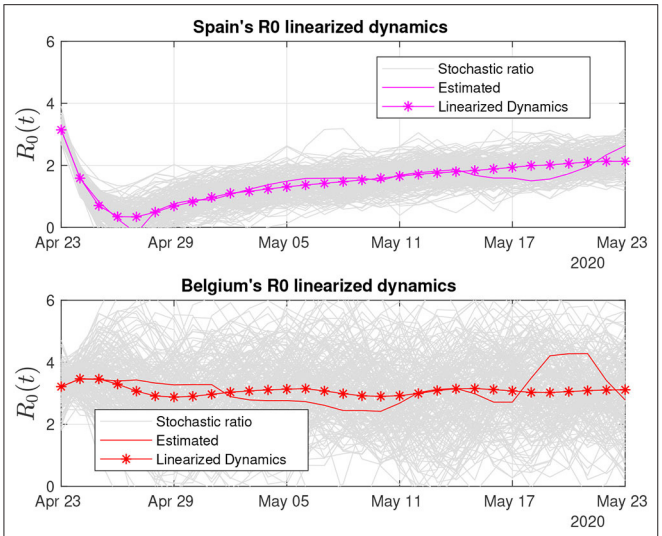


FIGURE 10 | Thirty days of linearized R_0 dynamics for Spain and Belgium. Both countries exhibit high variability of the stochastic ratio, which is suggestive of problems with either data compilation or difficulties to stabilize the pandemic.

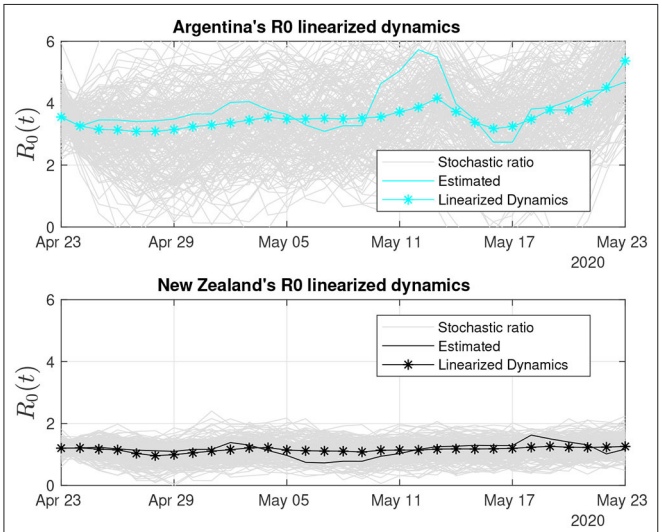


FIGURE 11 | Thirty days of linearized R_0 dynamics for Argentina and New Zealand. New Zealand has been able to control its COVID-19 infection rate, which is reflected on the low variability of the stochastic ratio curve.

close to 3.0, which is the mean value of 30 days of linearized dynamics (see **Table 3**).

Figure 11 shows the trends for Argentina and New Zealand. The R_0 estimates for the two countries, which have the lowest number of infections of the group, suggest that several recent R_0 reports are considering a ratio solely based on the number of infections and population, without estimating and taking into account the frequency of removals, which makes R_0 increase.

We compiled a table of the uncertainty of the estimated stochastic ratios by country (see **Table 2**), where the percentage

TABLE 2 | Pandemic's uncertainty (up to May 23, 2020).

Country	Power	Uncertainty (%)
Argentina	0.3099	30.9
Belgium	0.2426	24.2
Brazil	0.0546	5.4
Germany	0.0220	2.2
Italy	0.0113	1.1
New Zealand	0.0349	3.4
Spain	0.0521	5.2
United States	0.1835	18.3

TABLE 3 | R_0 estimates (on/up to May 23, 2020).

Country	Real-time	Linearized	30-days mean
Argentina	4.6	5.3	3.5
Belgium	2.7	3.1	3.0
Brazil	2.9	2.8	3.0
Germany	1.6	1.4	1.5
Italy	1.2	1.1	1.5
New Zealand	1.1	1.2	1.1
Spain	2.6	2.1	1.4
United States	2.9	3.1	4.0

of uncertainty is proportional to the linear power of the Gaussian estimation error. **Table 3** shows the estimated R_0 estimates by country and demonstrates the usefulness of our approach to provide a real-time picture of the pandemic that can be used to support decision making.

4. DISCUSSION

Almost 2 months after the WHO declared COVID-19 a global pandemic, health complications due to SARS-CoV-2 have caused many deaths and upended the routines of billions of people around the world. Research efforts for the development of a vaccine are being accelerated, but it is still a distant target [55]. At this moment, however, the most effective interventions are NPIs aimed at reducing SARS-CoV-2 transmission rates in order to increase the fraction of severe cases having access to scarce medical resources, such as mechanical ventilation [56].

Mathematical modeling is a valuable instrument to gauge the epidemics' dynamics and evaluate the effects of interventions aimed to control its spread. A crucial parameter is R_0, the basic reproduction number, which is closely followed by health officials and the public alike. As the world hopefully transitions to a gradual release from social distancing measures, many questions still remain about the SARS-CoV-2 virus, and there is all but the inevitability of secondary waves of infection ahead. Thus, the continuing use of mathematical models to track the disease will remain a necessity. However, the utility of models depends on the quality of the data they are fed, and there are many uncertainties regarding publicly available data on COVID-19

cases. For instance, due to the lack of widespread testing and subnotifications on the cause of deaths due to the disease, it is almost impossible to have a definitive picture of transitions between the compartments used to model the disease. Thus, in this work, we provide a system that takes into consideration the degree of uncertainty of the results presented by epidemiological SIR-type models.

The NPIs aimed to control the spread of the COVID-19 pandemic are implemented in a way similar to feedback control systems, with health authorities implementing social restriction measures in response to real-time evaluation of the number of infected and removed cases. Therefore, as in control systems, we are interfering with COVID-19's dynamics, affecting its behavior and parameters, and even discussing the development of real-time monitoring techniques using automation and control systems technologies. However, such closed-loop, control systems still need humans in the loop, and some data annotation mistakes may occur, such as seasonal dynamic changes related to workers shifts, weekend reduced reports, and even possible data manipulation of the data log may compromise the stability of estimators.

We showed that the RLS algorithm performed a better estimation job than the LMS algorithm in the present work, both in speed and accuracy. Real-time parametric estimation techniques, such as RLS, are widely used in the automation and aerospace industries to support adaptive control systems and state estimators. In automation applications, the RLS and LMS estimators are in fact divided into two different technologies, respectively, self-tuning and autotuning [31], where the first refers to real-time online adaptive process and the second involves the following steps: on-demand user request, acquiring the data set with sufficient samples, run offline processing without time constraints, present the results to the user, and decide whether to deploy or not the parametric update. Autotuning is less costlier to implement and thus more common, whereas the self-tuning technology requires online real-time processing with time constraints and is more commonly applied in the aerospace industry [31, 57].

LSA-based estimation is becoming more popular in industrial settings, such as in single-input, single-output (SISO) systems and in MIMO systems; in linear and non-linear system identification; and in polynomial and state-space MIMO system realizations [58, 59]. Such an increase in the use of RLS for real-time applications is justified by the advances of microprocessors to cope with the time constraints of fast dynamical systems, where the sampling interval can go as short as a few nanoseconds. Modern engineering applications with great impact in our society are associated with discrete-time sampled systems and signals, thus requiring adequate and reliable digital estimation algorithms, such as LSA. At least from an engineering viewpoint, the COVID-19's pandemic is also a discrete-time system, since the publicly available data for infections, deaths, and recoveries are updated on a daily basis, thus generating discrete-time sequences with the sampling period of 24 h.

During the COVID-19 pandemic, the scientific community is struggling to study, design, and deploy engineering applications to better assist society with information that could somehow quantify the pandemic's degree of severity based on $R_0(t)$ estimates [60]. Different from traditional modeling approaches, we propose a discrete-time MIMO system realization instead of the continuous-time estimation using ordinary differential equations; the MIMO-based parametric estimation of $R_0(t)$ is obtained through data fusion in a closed-loop estimation arrangement (see **Figure 3**).

The innovation of our modeling approach is the use of a discrete-time MIMO setup based on sensor fusion, a well-established strategy employed in the design of aerospace navigation systems under the assumption that "one always gains by adding a new measurement in terms of navigation error, and this no matter how bad the additional measurement is" [39, Remark 4.8]. Thus, by fusing data from the number of susceptibles, infections, recoveries, deaths, and individual parameters of the three coupled dynamical equations instead of a single one based solely on infections, the chance of having a better estimate of $R_0(t)$ in the maximum-likelihood sense is higher. Of course, the drawback of our proposed approach is the impossibility to apply it to nations lacking public reports on death and recovery numbers.

Our comparative approach shows that the strict measures adopted by some countries, such as Germany, Italy, Spain, and New Zealand, managed to stabilize the epidemics. In others, such as the United States and Brazil, the delay in adopting such measures and lack of coordination proved decisive to keep the R_0 values high and with a high degree of uncertainty. The real-time estimation of model parameters such as R_0 provides important insight into the underlying epidemic process and provides robustness in the face of imperfect data. This strategy can be eventually implemented as an online tracking application providing information about the dynamics of the pandemic to health officials and the public at large.

DATA AVAILABILITY STATEMENT

The data sets analyzed for this study can be found in the COVID-19 data repository from The Center for Systems Science and Engineering of the Johns Hopkins University (JSU CCSE) [40, 41].

AUTHOR CONTRIBUTIONS

AS and AP conceived and wrote the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

The fund was Supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), under Grant No. 408559/2016-0 and Universidade Federal do Pará.

REFERENCES

1. *WHO Announces COVID-19 Outbreak a Pandemic*. Available online at: http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic (accessed May 18, 2020).
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China 2019. *N Engl J Med*. (2020) **382**:727–33. doi: 10.1056/NEJMoa2001017
3. *Global Situation Report No. 118*. Available online at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200517-covid-19-sitrep-118.pdf?sfvrsn=21c0dafe_6 (accessed May 18, 2020).
4. 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. doi: 10.7326/M20-0504
5. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N Engl J Med*. (2020) **382**:2081–90. doi: 10.1056/NEJMoa2008457
6. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission the Achilles' Heel of current strategies to control Covid-19. *N Engl J Med*. (2020) **382**:2158–60. doi: 10.1056/NEJMe2009758
7. El-Sadr WM, Justman J. Africa in the Path of Covid-19. *N Engl J Med*. (2020) **382**:e11. doi: 10.1056/NEJMp2008193
8. Kirby T. South America prepares for the impact of COVID-19. *Lancet Respir Med*. (2020) **8**:551–2. doi: 10.1016/S2213-2600(20)30218-6
9. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet*. (2020) **395**:931–4. doi: 10.1016/S0140-6736(20)30567-5
10. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nat Hum Behav*. (2020) **4**:451–9. doi: 10.1038/s41562-020-0887-9
11. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. *Int J Soc Psychiatry*. (2020) **66**:317–20. doi: 10.1177/0020764020915212
12. Xu S, Li Y. Beware of the second wave of COVID-19. *Lancet*. (2020) **395**:1321–2. doi: 10.1016/S0140-6736(20)30845-X
13. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. *Nature*. (2006) **442**:448–52. doi: 10.1038/nature04795
14. Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty identifiability, and forecasts. *Infect Dis Model*. (2017) **2**:379–98. doi: 10.1016/j.idm.2017.08.001
15. M'Kendrick AG. Applications of mathematics to medical problems. *Proc Edinb Math Soc*. (1925) **44**:98–130. doi: 10.1017/S0013091500034428
16. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A Math Phys Eng Sci*. (1927) **115**:700–21. doi: 10.1098/rspa.1927.0118
17. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. (2020) **368**:489–93. doi: 10.1126/science.abb3221
18. Weitz JS, Beckett SJ, Coenen AR, Demory D, Dominguez-Mirazo M, Dushoff J, et al. Modeling shield immunity to reduce COVID-19 epidemic spread. *Nat Med*. (2020) **26**:849–54. doi: 10.1038/s41591-020-0895-3
19. Viceconte G, Petrosillo N. COVID-19 R0: magic number or conundrum? *Infect Dis Rep*. (2020) **12**:8516. doi: 10.4081/idr.2020.8516
20. Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: considerations for public health applications. *Am J Publ Health*. (2014) **104**:e32–41. doi: 10.2105/AJPH.2013.301704
21. Riley S. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science*. (2003) **300**:1961–6. doi: 10.1126/science.1086478
22. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R0). *Emerg Infect Dis*. (2019) **25**:1–4. doi: 10.3201/eid2501.171901
23. Obadia T, Haneef R, Boëlle PY. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med Inform Decis Mak*. (2012) **12**:147. doi: 10.1186/1472-6947-12-147
24. Bettencourt LMA, Ribeiro RM. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE*. (2008) **3**:e2185. doi: 10.1371/journal.pone.0002185
25. Wallinga J. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol*. (2004) **160**:509–16. doi: 10.1093/aje/kwh255
26. Cauchemez S, Boëlle PY, Donnelly CA, Ferguson NM, Thomas G, Leung GM, et al. Real-time estimates in early detection of SARS. *Emerg Infect Dis*. (2012) **12**:110–3. doi: 10.3201/eid1201.050593
27. Cauchemez S, Boëlle P, Thomas G, Valleron A. Estimating in real time the efficacy of measures to control emerging communicable diseases. *Am J Epidemiol*. (2006) **164**:591–7. doi: 10.1093/aje/kwj274
28. 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
29. Binny RN, Hendy SC, James A, Lustig A, Plank MJ, Steyn N. Effect of Alert Level 4 on effective reproduction number: review of international COVID-19 cases. *medRxiv*. (2020). doi: 10.1101/2020.04.30.20086934
30. Nievergelt Y. A tutorial history of least squares with applications to astronomy and geodesy. *J Comput Appl Math*. (2000) **121**:37–72. doi: 10.1016/S0377-0427(00)00343-5
31. Astrom KJ, Wittenmark B. *Adaptive Control*. 2nd ed. Mineola, NY: Dover Publications (2008).
32. Brunton SL, Kutz JN. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge: Cambridge University Press (2019).
33. Smirnova A, Chowell G. A primer on stable parameter estimation and forecasting in epidemiology by a problem-oriented regularized least squares algorithm. *Infect Dis Model*. (2017) **2**:268–75. doi: 10.1016/j.idm.2017.05.004
34. Banks HT, Hu S, Thompson WC. *Modeling and Inverse Problems in the Presence of Uncertainty*. Boca Raton, FL: CRC Press (2014).
35. Shen J. A Recursive bifurcation model for predicting the peak of COVID-19 virus spread in United States and Germany. *medRxiv* [Preprint]. doi: 10.1101/2020.04.09.20059329
36. Lewis FL, Xie L, Popa D. *Optimal and Robust Estimation*. 2nd ed. Boca Raton, FL: CRC Press; Taylor & Francis Group (2008).
37. Park G, Choi SB. An integrated observer for real-time estimation of vehicle center of gravity height. *IEEE Trans Intell Transport Syst*. (2020) 1–12. doi: 10.1109/TITS.2020.2988508
38. Jaros R, Martinek R, Kahankova R, Koziorek J. Novel hybrid extraction systems for fetal heart rate variability monitoring based on non-invasive fetal electrocardiogram. *IEEE Access*. (2019) **7**:131758–84. doi: 10.1109/ACCESS.2019.2933717
39. Kabamba PT, Girard AR. *Fundamentals of Aerospace Navigation and Guidance. Cambridge Aerospace Series*. Cambridge: Cambridge University Press (2014).
40. *Novel Coronavirus (COVID-19) Cases Data–Humanitarian Data Exchange*. Available online at: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases (accessed May 8, 2020).
41. *CSSEGIS and Data/COVID-19*. Available online at: https://github.com/CSSEGISandData/COVID-19 (accessed May 8, 2020).
42. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press (1992).
43. Brauer F, Castillo-Chavez C. Epidemic models. In: *Texts in Applied Mathematics*. New York, NY: Springer New York (2011). p. 345–409. doi: 10.1007/978-1-4614-1686-9_9
44. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model*. (2020) **5**:271–81. doi: 10.1016/j.idm.2020.03.001
45. Anderson RM, May RM. Population biology of infectious diseases: part I. *Nature*. (1979) **280**:361–7.
46. Brauer F, Feng Z, Castillo-Chavez C. Discrete epidemic models. *Math Biosci Eng*. (2010) **7**:1–15. doi: 10.3934/mbe.2010.7.1

47. Gómez S, Arenas A, Borge-Holthoefer J, Meloni S, Moreno Y. Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *Europhys Lett*. (2010) **89**:38009. doi: 10.1209/0295-5075/89/38009
48. *RKI–Coronavirus SARS-CoV-2–Aktueller Lage-/Situationsbericht des RKI zu COVID-19*. Available online at: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Gesamt.html (accessed May 6, 2020).
49. *Kevin Systrom*. Available online at: http://systrom.com/blog/author/ksys1983/ (accessed May 4, 2020).
50. *Coronavirus Flares as States and Countries Ease Social Distancing Guidelines*. Available online at: https://www.washingtonpost.com/national-security/coronavirus-flares-as-states-and-countries-ease-social-distancing-guidelines/2020/05/09/cccb3c0c-9219-11ea-9e23-6914ee410a5f_story.html (accessed May 17, 2020).
51. Filho RL, Lichtenthaler DG. A dynamic model for Covid-19 in Brazil. *medRxiv*. (2020). doi: 10.1101/2020.05.10.20097550
52. *Brazil Registers 888 New Deaths From Coronavirus; Total Number of Deaths Exceeds 18 Thousand*. Available online at: https://www1.folha.uol.com.br/internacional/en/scienceandhealth/2020/05/brazil-registers-888-new-deaths-from-coronavirus-total-number-of-deaths-exceeds-18-thousand.shtml (accessed May 21, 2020).
53. *Brazil, Once a Leader, Struggles to Contain Virus Amid Political Turmoil*. Available online at: https://www.nytimes.com/2020/05/16/world/americas/virus-brazil-deaths.html (accessed May 20, 2020).
54. *Basic Reproduction Number of Novel Coronavirus in Belgium Falls to 0.6*. Available online at: https://www.vrt.be/vrtnws/en/2020/05/04/basic-reproduction-number-of-novel-coronavirus-in-belgium-falls/ (accessed May 22, 2020).
55. Graham BS. Rapid COVID-19 vaccine development. *Science*. (2020) **368**:945–6. doi: 10.1126/science.abb8923
56. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of Covid-19. *N Engl J Med*. (2020) **382**:2049–55. doi: 10.1056/NEJMsb2005114
57. Administrator N. *Simplify, Simplify–Streamlined Adaptive System Works–Capabilities Verified*. (2013). Available online at: http://www.nasa.gov/centers/dryden/news/X-Press/mrac.html
58. Silveira A, Silva A, Coelho A, Real J, Silva O. Design and real-time implementation of a wireless autopilot using multivariable predictive generalized minimum variance control in the state-space. *Aerosp Sci Technol*. (2020) **105**:106053. doi: 10.1016/j.ast.2020.106053
59. Vicario F. *OKID as a general approach to linear and bilinear system Identification* (Ph.D. thesis), New York, NY: Columbia University (2014).
60. Adam D. A guide to R-the pandemic's misunderstood metric. *Nature*. (2020) **5883**:346–8. doi: 10.1038/d41586-020-02009-w

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.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 12 November 2020
doi: 10.3389/fphy.2020.564061

Trend Analysis of COVID-19 Based on Network Topology Description

Jun Zhu[1], *Yangqianzi Jiang*[1], *Tianrui Li*[1], *Huining Li*[2] *and Qingshan Liu*[1]*

[1]School of Mathematics, Southeast University, Nanjing, China, [2]School of Information Science and Engineering, Southeast University, Nanjing, China

In this study, the trend of the epidemic situation of COVID-19 is analyzed based on the analysis method for network topology. Combining with the sliding window method, the dynamic networks with different topologies for each window are built to reflect the relationship of the data on different days. Then, the static statistical features on network topologies at different times are extracted during the dynamic evolution of complex networks. A new trend function defined on the average degree and clustering coefficient of the network is tailored to measure the characteristics of the trend. Through the value of the trend function, we can analyze the trend of the epidemic situation in real time. It is found that if the value of the trend function tends to decrease, it means that the epidemic will have to be effectively controlled. Finally, we put forward some suggestions for early control of the epidemic.

Keywords: COVID-19, sliding window, network topology, dynamic evolution, trend analysis

OPEN ACCESS

Edited by:
Zhen Wang,
Hong Kong Baptist University,
Hong Kong

Reviewed by:
Yuyao Wang,
Nanjing University of Science and Technology, China
Chengyi Xia,
Tianjin University of Technology, China

***Correspondence:**
Qingshan Liu
qsliu@seu.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *20 May 2020*
Accepted: *07 October 2020*
Published: *12 November 2020*

Citation:
Zhu J, Jiang Y, Li T, Li H and Liu Q (2020) Trend Analysis of COVID-19 Based on Network Topology Description. Front. Phys. 8:564061. doi: 10.3389/fphy.2020.564061

1. INTRODUCTION

Since December 2019, patients with pneumonia of unknown cause have appeared in some medical institutions. By now, the number of cases caused by coronavirus (COVID-19) has increased. The World Health Organization (WHO) declared the COVID-19 disease a pandemic on March 11, 2020. The cumulative confirmed cases have reached almost 3,220,000 as of May 1, 2020 worldwide. For new outbreaks, it is significant to understand the transmission dynamics of infection, which can help governments take effective measures to contain them and reduce the number of spread. In the survey of other two pandemics caused by coronavirus severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), scientists have put forward many effective measures to build the transmission models, such as the transmission analysis based on genome research (Qin et al., 2003) and the mathematical model of infection kinetics (Liang, 2020).

Since the outbreak of COVID-19, scholars have conducted relevant research through different models. Zhu and Chen give a statistical analysis of COVID-19 with early transmission model (Zhu and Chen, 2020). A data-based iterative prediction method is proposed to find growth rates under which the situation will be in control (Perc et al., 2020). Robust time series are used to complete statistical forecasts for the confirmed cases of COVID-19 by Fotios and Spyros (Petropoulos and Makridakis, 2020). In (Chen and Zhou, 2020), a Monte Carlo method is proposed to quantify the control efficacy, which is completed by calculating the mean number of secondary cases infected by a case with symptom onset every day. Moreover, a segmented Poisson model is adopted in Zhang et al. (2020) to analyze the new cases, which takes the governments' regulations into consideration. An extended SIR model is employed by Jia and Han to compare the epidemics trend in Italy and Hunan, China (Jia et al., 2020).

With the development of complex networks, the spread analysis of epidemics on complex networks has attracted wide attention in the literature. Based on the SIR model in complex networks, Xia et al. have investigated the effects of delaying the time to recovery and of nonuniform transmission on the propagation of diseases on structured populations (Xia et al., 2012; Xia et al., 2013). In (Small and Tse, 2005), Small and Tse propose a new four state model based on the transmission of SARS, where community is modeled as a small-world network of interconnected nodes. Wang et al. point out the spread of epidemics in small-world networks (Wang and Li, 2016). The prevalence of infectious diseases in the population, the spread of viruses on computer networks, and the spread of rumors in human society can all be regarded as the problem of information dissemination on the network, which belongs to the dynamic process of the network and can be dealt with machine learning (Silva and Zhao, 2016).

In the study of complex network diseases (Wang et al., 2019; Wu and Hadzibeganovic, 2020), individuals in the population are regarded as nodes in the network, and the connections between individuals are regarded as edges in the network, which establishes the topology of the network. Since the real network is usually small scale and scale-free, it is generally that the network under study is a Watts–Strogatz (WS) or Barabási-Albert (BA) scale-free network (Wu et al., 2019). After the establishment of network topology, a mathematical model that can reflect the dynamic characteristics of infectious diseases is able to be built according to the transmission characteristics and infectious diseases between individuals (Huang, 2008; Liu and Li, 2019; Lu and Liu, 2019; Zhou and Wu, 2019; Aadil et al., 2020). In this paper, we attempt to make use of empirical data and combine the characteristics of COVID-19 transmission to analyze the trend of COVID-19. We mainly use the knowledge of network topology to give the trend analysis of COVID-19, which networks are established based on the data from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (https://github.com/CSSEGISandData/COVID-19).

The article is organized as follows. **Section 1** introduces the process of relevant research. In **Section 2**, which is also the most significant part of the article, we present the construction of the networks and the topological features extracted from the networks. **Section 3** displays the networks we built and analyzes the epidemic situation in the four regions through the topological characteristics of the networks. We summarize the method we used and give some suggestions in **Section 4**.

2. METHODOLOGY

This section introduces the construction of the networks and the topological features extracted from the networks.

2.1. Networks Constructing

Here, we select four regions for the analysis, including Wuhan, South Korea, Russia, and Germany. The total number of confirmed cases is extracted for every day in each region. We get a time series $R_1 = \{x_1, x_2, \ldots, x_n\}$, where x_i is the number of diagnoses in the ith day, and n is the total days from the virus outbreak. Then, the growth rate of the number of diagnoses y_i in the ith day is described as follows:

$$y_i = \begin{cases} \dfrac{x_{i+1} - x_i}{x_i}, & x_i \neq 0 \\ 0, & x_i = 0 \end{cases} \tag{1}$$

From this treatment, the change in the daily diagnostic number can be seen more clearly. At the same time, the impact of the total local population on the growth rate of the number of confirmed cases can also be ignored. Then, we get a new time series on the growth rate of daily diagnoses $R_2 = \{y_1, y_2, \ldots, y_{n-1}\}$.

The dynamic evolution analysis method is an important way for data analysis based on the features of network topology. In dynamic evolution, the feature measurement of networks is a function of time. In the same evolution mode, two subnetworks obtained at different times have different features. Therefore, it is a very important and effective way to analyze and classify the network by using static statistical features at different times during the dynamic evolution of networks (Backes et al., 2009). Here, the sliding window method is used to extract the features of network topology for further observation. The key to selecting sliding windows is how to effectively maintain the quality and quantity of the original time series information, while minimizing the calculation complexity to the most extent (Li and Zhang, 2004; Li and Xiao, 2009). In this study, we apply the sliding window with the length of 9 days and the step size of 1 day. **Figure 1** shows the process of sliding windows for the time series data. Next, we will use the daily growth rate to build the networks.

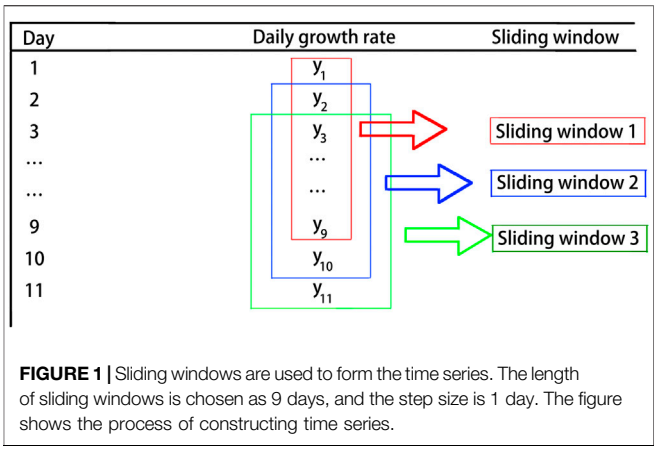


FIGURE 1 | Sliding windows are used to form the time series. The length of sliding windows is chosen as 9 days, and the step size is 1 day. The figure shows the process of constructing time series.

For one of the sliding window $R_i = \{y_i, y_{i+1}, \ldots, y_{i+8}\}$, using the nine-day growth rates as nodes, we calculate the Euclidean distance between any two nodes to get the 9×9 distance matrix D with

$$d_{ij} = \left| y_i - y_j \right|, \tag{2}$$

where d_{ij} is the element of D in row i and column j. The growth rate defined in **(1)** formulates the difference between the different nodes, so the connections in the network give a relevance description for

the data in the 9 days. Set the average value of the distance as the threshold denoted by δ. When $d_{ij} < \delta$, there is a connection between nodes i and j, that is, $a_{ij} = 1$ in the connection matrix A of the network. Otherwise, we stipulate a) if one of y_i and y_j is 0, $a_{ij} = 0$; b) if $d_{ij} < 0.0002$, $a_{ij} = 0$. Here, a) is to exclude the relationship between the current day and the remaining days if a certain growth rate is 0, and b) is an adjustment for the following consideration: at the time of the definition of the growth rate, the number of diagnoses per day on the previous day is increasing, and the growth rate is changing even if the number of growth is invariant. The construction of connection matrix is described as follows:

$$a_{ij} = \begin{cases} 1, & \text{if } 0.0002 < d_{ij} < \delta \text{ and } \min\{y_i, y_j\} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

2.2. Topological Features of Networks

The degree k_i of the node i is defined as the number of sides connected to the node. In an undirected and unauthorized network, the mathematical formula for degree can be expressed as follows:

$$k_i = \sum_{j=1}^{N} a_{ij}, \quad (4)$$

where a_{ij} is the element of connection matrix A and N is the total number of nodes. Then, we can get the average degree k of the whole network as follows:

$$k = \frac{1}{N} \sum_{i=1}^{N} k_i \quad (5)$$

The clustering coefficient is a coefficient that measures the degree of network aggregation, which can be formulated as follows:

$$C = \frac{1}{N} \sum_{i=1}^{N} \frac{2|e_i|}{k_i(k_i - 1)}, \quad (6)$$

where $|e_i|$ represents the number of connected edges between nodes in the neighborhood of node i (i.e., the number of triangles formed by node i and other two nodes in its neighborhoods) and k_i is the degree of node i. In general, the clustering coefficient of the network quantifies the connection between nodes in the network with $C \in [0, 1]$. If $C = 1$, all points in the network are connected. If the value of C is relatively small, the network connection is loose.

3. EXPERIMENTAL RESULTS

In this section, we combine the daily number of confirmed cases in Wuhan, South Korea, Russia, and Germany to build the networks and analyze the epidemic situation in the four regions through the topological characteristics of the networks.

3.1. Data Processing

We use the daily number of diagnoses from January 22, 2020 to May 16, 2020 as the data set. So, for each region, we can get the total number of diagnoses per day for 116 days. First, from **(1)**, the data are processed to calculate the 115-day daily diagnosis growth rate for each region. Then, using a sliding window of 9 days and a step size of 1 day, the time series data are divided into 107 periods, and 107 networks are constructed with nine nodes in each period. **Figure 2** shows the networks at the 43rd day of the four regions. The more connections the network has, the greater the change is of the 9-day growth rate. It should be emphasized that few connections cannot only indicate the control period of the disease but also the period of early outbreak.

3.2. Analysis of Network Topological Characteristics

We use the equations in **Eqs (5)** and **(6)** to calculate the average degree and clustering coefficient of each network, and the trend function is defined as $I = 0.5C + 0.5k$ combining the average degree and clustering coefficient. Taking Germany for example, **Figure 3A** shows the growth rate change chart of Germany in the first 80 days, and **Figure 3B** depicts the evolution of trend function. It can be seen from **Figure 3A** that the daily growth rate in the first 80 days has changed greatly, and it is difficult to find the regular pattern of epidemic development. However, it can be seen from **Figure 3B** that when the value of trend function is relatively large and stable, the epidemic situation has not been completely controlled. This situation can be seen from the data in the next few days. It shows that the network topology method proposed in this article is efficient for the analysis of epidemic situation.

The evolution of the trend function in the four regions is shown in **Figure 4**. In the figure, the value of the abscissa is the number of days passed from January 22, 2020, and the ordinate is the value of the trend function I. The larger the I value, the larger is the clustering coefficient and mean sum. The larger the clustering coefficient, the difference of growth rate of any 3 days is less than the threshold in 9 days, and the larger the average degree, the difference of the growth rate in 9 days is less than the threshold number of days. Therefore, when the daily diagnostic growth rate of 9 days becomes relatively small, the clustering coefficient and average degree will be relatively small. If the growth rate changes greatly in 1 day in the 9 days, the average threshold will become larger, and the number of connections will increase in the remaining 8 days, then the value of trend function will also increase.

From **Figure 4A** in Wuhan, there is a clear downward trend in the values of the trend function around March 2. The growth rate of the number of confirmed patients in the next 9 days has also dropped to 0.1% for the first time, and the growth rate is also declining in the next few days, gradually turning to 0. This indicates that the change trend is related to the change of the daily growth rate. Furthermore, if the values of trend function show a downward trend, it can be inferred that the growth rate of the region has dropped to a smaller value, and it can be considered that the epidemic situation in the region has been effectively controlled.

In **Figure 4D**, the values of trend function change in Germany can be seen that it has a small wave peak at first, and then

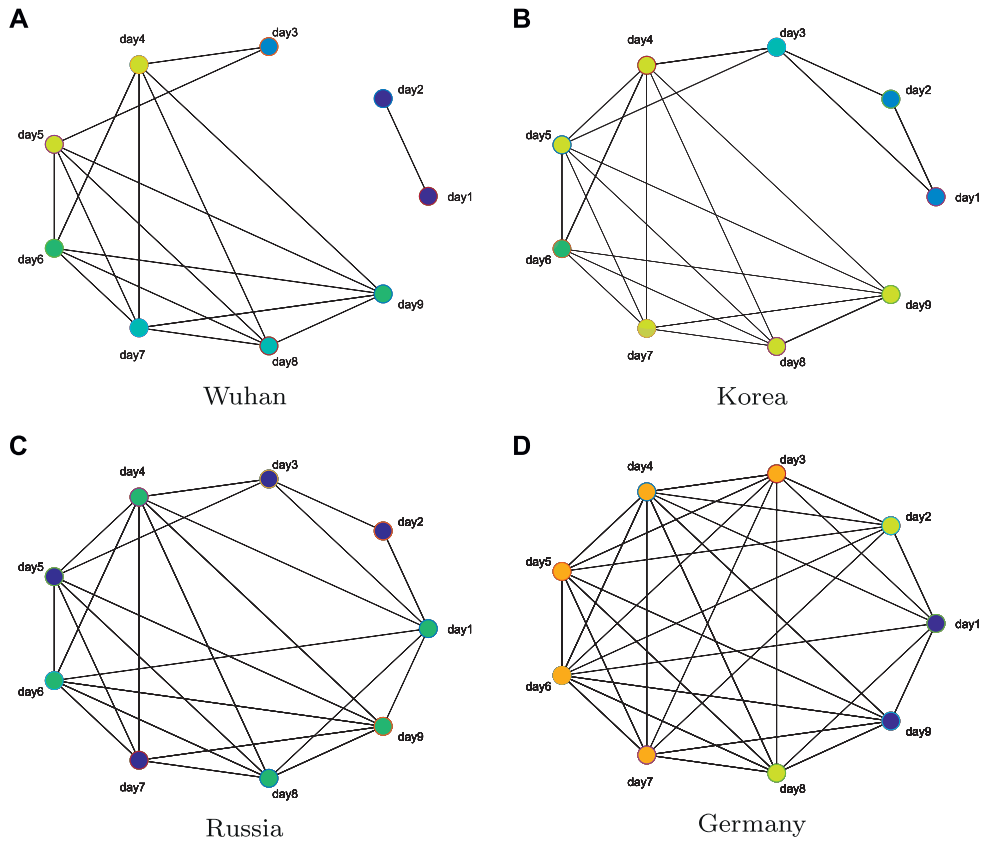


FIGURE 2 | Networks at the 43rd day of the four regions. The number of connections in the network reflects the change degree of nine-day growth rate.

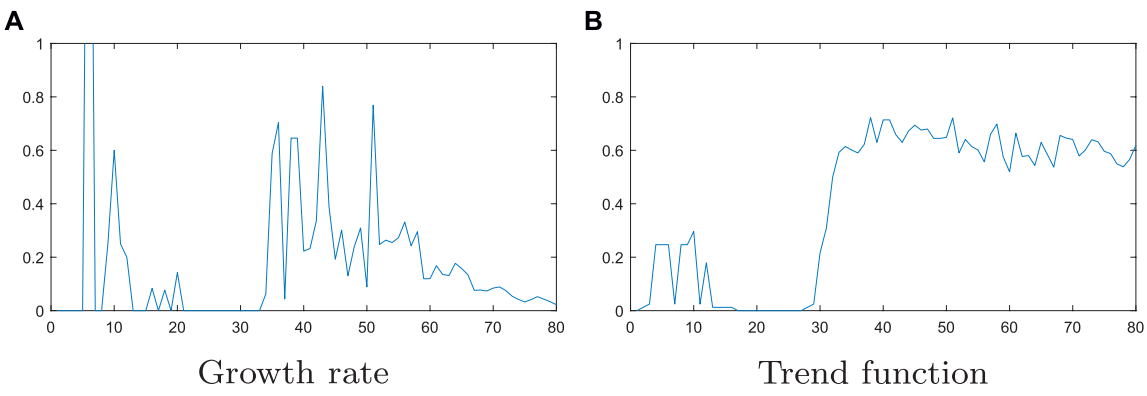


FIGURE 3 | Growth rate and the evolution of trend function of Germany.

suddenly increases until it stabilizes around 0.6. So, it can be speculated that there was a small outbreak in Germany between January 22 and February 2, and then it was effectively controlled, resulting in a growth rate of almost zero. But since February 22, there has been a second outbreak in Germany. The values of trend function have been fluctuating around 0.6, indicating that Germany's growth rate still remains high and the epidemic has not been effectively controlled. The above results are consistent with the local epidemic data in Germany, which proves that the method is effective.

From above analysis, we can analyze and predict the epidemic situation in South Korea and Russia. From **Figure 4**, it can be seen that the figure of South Korea has shown a clear downward trend since April 2, indicating that the epidemic situation in South

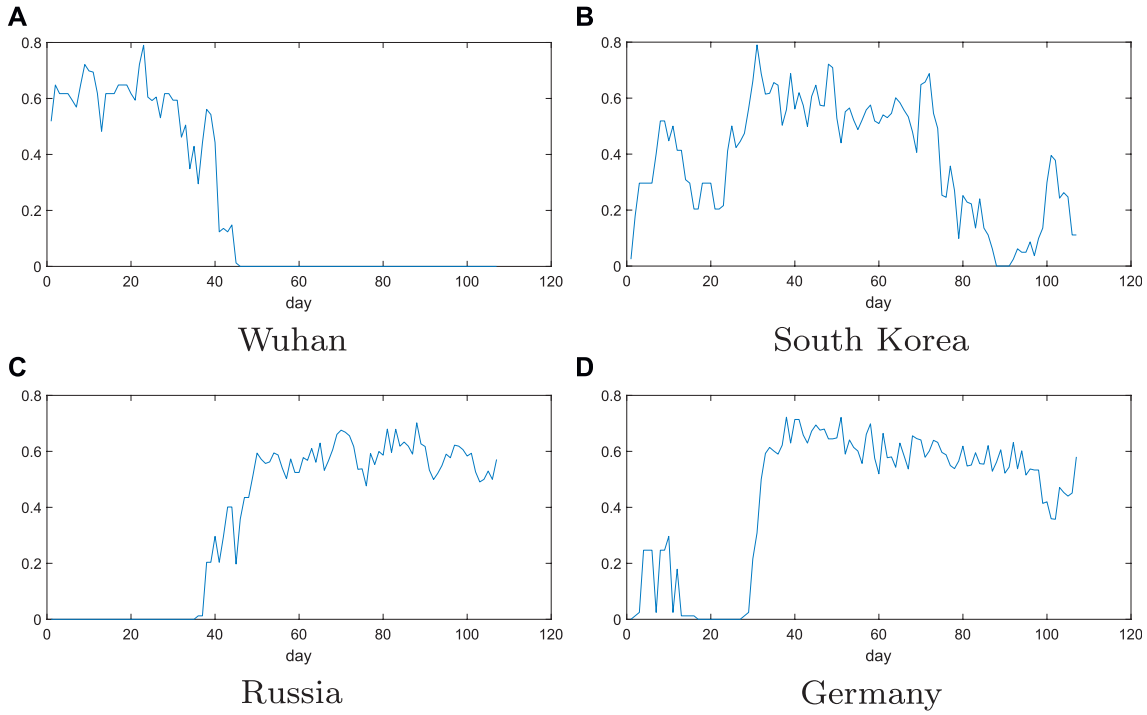


FIGURE 4 | Evolution of trend function for the four regions: Wuhan, South Korea, Russia, and Germany. The value of the abscissa is the number of days passed from January 22, 2020, and the ordinate is the value of the trend function I. The larger the I value, the larger is the clustering coefficient and mean sum.

Korea has been effectively controlled. However, there was a small upward trend at the 100th day. This indicates that the daily growth rate in South Korea has increased by a small margin recently. But, it can be controlled quickly. For Russia, where the values of trend function is still fluctuating around 0.6, which indicates a certain fluctuation in the growth rate of the daily confirmed population in Russia during April. We can also see that the growth rate is still relatively high, which shows that the Russian epidemic has not been effectively controlled, and the growth rate will not be significantly reduced in the near future, and more stringent measures are needed to control the development of the epidemic.

Note that the effective control of the epidemic in this article refers to the fact that the daily growth rate is almost zero, that is, there is almost no new infection, rather than the change in the daily growth rate of 0, or in other words, the next day is approximately equal to the daily growth rate of the previous day, as mentioned in some articles. For example, for the platform period mentioned in Perc et al. (2020), we understand it as the epidemic situation has been preliminarily controlled, and only when there is no new infection can it be considered to be effectively controlled.

4. DISCUSSIONS AND CONCLUSIONS

In this article, we proposed a trend analysis method based on network construction with sliding windows to extract the characteristics of network dynamic evolution over time and analyzed the epidemic trend in four typical regions. In the analysis, we found that some regions had better control of the epidemic, while others were still in the process of outbreak. So, we put forward some suggestions and hope that the epidemic situation in various countries can be effectively controlled as soon as possible.

The proposed method in this article is easy but efficient for the trend analysis of COVID-19. In general, since COVID-19 patients' mid-term course of disease develops rapidly, it is hard to accurately judge the cycle from mild to severe. Moreover, the issue of infectivity in the incubation period and the infectious power of those infected patients during the recovery period remains to be studied, which may be the cause of second outbreak in Germany. The intensity of different infection generation and the difference of infection are still unknown. The question of whether the virus will disappear or persist in the population remains to be resolved.

Many countries have taken effective measures to the epidemic, such as closing churches, bars, and gymnasiums. In severe cases, some countries such as China seal off the city from all outside contact to stop the spread of the plague. We can learn from the above analysis that Wuhan has got the epidemic under control in a relatively short time. In order to block the transmission chain of the virus, it is a very effective method to trace the confirmed patient's activity route and contacts. For countries like Russia

where the epidemic is still serious, which can be observed from the trend in **Figure 4**, they should consider to strengthen the isolation measures.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here https://github.com/CSSEGISandData/COVID-19.

AUTHOR CONTRIBUTIONS

JZ and QL designed and performed the research. JZ, YJ, TL, and HL wrote the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant 61876036.

REFERENCES

Aadil L., Adel S., Hamza E. M., Mustapha E. J., Mohamed E. F. (2020). Global dynamics of an epidemic model with incomplete recovery in a complex network. *J. Franklin Inst.* 357, 4414–4436. doi:10.1016/j.jfranklin.2020.03.010

Backes A. R., Casanova D., Bruno O. M. (2009). A complex network-based approach for boundary shape analysis. *Pattern Recogn.* 42, 54–67. doi:10.1016/j.patcog.2008.07.006

Chen D., Zhou T. (2020). Control efficacy on COVID-19. Populations and evolution. arXiv:2003.00305.

Huang S.-Z. (2008). A new SEIR epidemic model with applications to the theory of eradication and control of diseases, and to the calculation of. *Math. Biosci.* 215, 84–104. doi:10.1016/j.mbs.2008.06.005

Jia W., Han K., Song Y., Cao W., Wang S., Yang S., et al. (2020). Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Front. Med.* 7, 169. doi:10.3389/fmed.2020.00169

Li F., Xiao J. (2009). How to get effective slide-window size in time series similarity search. *J. Front. Comput. Sci. Tech.* 3, 105–112. [in Chinese]. doi:1673-9418/2009/03(01)-0105-08

Li J., Zhang D. (2004). Algorithms for dynamically adjusting the sizes of sliding windows. *J. Softw.* 15, 1800–1814. [in Chinese]. doi:1000-9825/2004/15(12)1800

Liang K. (2020). Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infect. Genet. Evol.* 82, 104306. doi:10.1016/j.meegid.2020.104306

Liu Q., Li H. (2019). Global dynamics analysis of an seir epidemic model with discrete delay on complex network. *Phys. Stat. Mech. Appl.* 524, 289–296. doi:10.1016/j.physa.2019.04.258

Lu Y., Liu J. (2019). The impact of information dissemination strategies to epidemic spreading on complex networks. *Phys. Stat. Mech. Appl.* 536, 120920. doi:10.1016/j.physa.2019.04.156

Perc M., Gorišek Miksić N., Slavinec M., Stožerr A. (2020). Forecasting COVID-19. *Front. Phys.* 8, 127. doi:10.3389/fphy.2020.00127

Petropoulos F., Makridakis S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS One* 15, e0231236. doi:10.1371/journal.pone.0231236

Qin E. d., He X., Tian W., Liu Y., Li W., Wen J, et al. (2003). A genome sequence of novel SARS-CoV isolates: the genotype, gd-ins29, leads to a hypothesis of viral transmission in south China. *Dev. Reprod. Biol.* 1, 101–107. doi:10.1016/s1672-0229(03)01014-3.

Silva T. C., Zhao L. (2016). "Case study of network-based semi-supervised learning: stochastic competitive-cooperative learning in networks." in *Machine learning in complex networks* (Cham, Switzerland: Springer International Publishing), 291–321.

Small M., Tse C. K. (2005). Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment. *Phys. Stat. Mech. Appl.* 351, 499–511. doi:10.1016/j.physa.2005.01.009

Wang B., Li P. (2016). Introduction of small world network. *Mod. Phys.* 28, 51–55. [in Chinese]. doi:10.13405/j.cnki.xdwz.2016.03.018

Wang Y., Yuan G., Fan C., Hu Y., Yang Y. (2019). Disease spreading model considering the activity of individuals on complex networks. *Phys. Stat. Mech. Appl.* 530, 121393. doi:10.1016/j.physa.2019.121393

Wu Q., Hadzibeganovic T. (2020). An individual-based modeling framework for infectious disease spreading in clustered complex networks. *Appl. Math. Model.* 83, 1–12. doi:10.1016/j.apm.2020.02.012

Wu Y., Gao L., Zhang Y., Xiong X. (2019). Structural balance and dynamics over signed BA scale-free network. *Phys. Stat. Mech. Appl.* 525, 866–877. doi:10.1016/j.physa.2019.04.038

Xia C., Wang L., Sun S., Wang J. (2012). An sir model with infection delay and propagation vector in complex networks. *Nonlinear. Dynam.* 69, 927–934. doi:10.1007/s11071-011-0313-y

Xia C.-Y., Wang Z., Sanz J., Meloni S., Moreno Y. (2013). Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks. *Phys. Stat. Mech. Appl.* 392, 1577–1585. doi:10.1016/j.physa.2012.11.043

Zhang X., Ma R., Wang L. (2020). Predicting turning point, duration and attack rate of COVID-19 outbreaks in major western countries. *Chaos Solit. Fractals* 135, 109829. doi:10.1016/j.chaos.2020.109829

Zhou R., Wu Q. (2019). Epidemic spreading dynamics on complex networks with adaptive social-support. *Phys. Stat. Mech. Appl.* 525, 778–787. doi:10.1016/j.physa.2019.03.107

Zhu Y., Chen Y. Q. (2020). On a statistical transmission model in analysis of the early phase of COVID-19 outbreak. *Stat Biosci.* 1–17. doi:10.1007/s12561-020-09277-0

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 12 November 2020
doi: 10.3389/fphy.2020.586899

Mathematical Modeling Based Study and Prediction of COVID-19 Epidemic Dissemination Under the Impact of Lockdown in India

Vipin Tiwari, Namrata Deyal and Nandan S. Bisht**

Department of Physics, Kumaun University, Uttarakhand, India

COVID-19 (SARS-CoV-2) is rapidly spreading in South Asian countries, especially in India. India is the fourth most COVID-19 affected country at present i.e., until July 10, 2020. With limited medical facilities and high transmission rate, the study of COVID-19 progression and its subsequent trajectory needs to be analyzed in India. Epidemiologic mathematical models have the potential to predict the epidemic peak of COVID-19 under different scenarios. Lockdown is one of the most effective mitigation policies adopted worldwide to control the transmission rate of COVID-19 cases. In this study, we use an improvised five compartment mathematical model, i.e., Susceptible (S)-Exposed (E)-Infected (I)-Recovered (R)-Death (D) (SEIRD) to investigate the progression of COVID-19 and predict the epidemic peak under the impact of lockdown in India. The aim of this study is to provide a more precise prediction of epidemic peak and to evaluate the impact of lockdown on epidemic peak shift in India. For this purpose, we examine the most recent data (from January 30, 2020 to July 10, 2020 i.e., 160 days) to enhance the accuracy of outcomes obtained from the proposed model. The model predicts that the total number of COVID-19 active cases would be around 5.8×10^5 on August 15, 2020 under current circumstances. In addition, our study indicates the existence of under-reported cases i.e., 10^5 during the post-lockdown period in India. Consequently, this study suggests that a nationwide public lockdown would lead to epidemic peak suppression in India. It is expected that the obtained results would be beneficial for determining further COVID-19 mitigation policies not only in India but globally as well.

Keywords: epidemic peak, SEIRD model, COVID-19, lockdown, India, under reporting

OPEN ACCESS

Edited by:
Aristides Moustakas,
University of Crete, Greece

Reviewed by:
Ankush Bhaskar,
National Aeronautics and Space Administration, United States
Subhas Khajanchi,
Presidency University, India

****Correspondence:***
Nandan S. Bisht
bisht.nandan@kunainital.ac.in
Vipin Tiwari
vipint92@gmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *24 July 2020*
Accepted: *07 September 2020*
Published: *12 November 2020*

Citation:
Tiwari V, Deyal N and Bisht NS (2020) Mathematical Modeling Based Study and Prediction of COVID-19 Epidemic Dissemination Under the Impact of Lockdown in India. Front. Phys. 8:586899. doi: 10.3389/fphy.2020.586899

INTRODUCTION

COVID-19 is a contagious disease of the Severe Acute Respiratory Syndrome Corona Virus (SARS-COV) family, and has emerged as the biggest health crisis of the twenty-first century across the globe. It has been declared as a global pandemic by World Health Organization (WHO) on March 11, 2020 just after a few months since its first case reported in Wuhan, China [1]. COVID-19 infected patients generally exhibit common symptoms like cough, fever and respiratory disorders. In the worst conditions, it might result in serious health issues like kidney failure, and pneumonia which might cause death of patients. Investigations on COVID-19 indicate that it transmits through respiratory droplets, as well as human to human transition [2].

The most serious issue with the COVID-19 pandemic is its rapidly growing transmission rate across the world. Until July 10, 2020, total confirmed cases in world are 1.2×10^7 and total number of deaths are 5.5×10^5 [3]. United States of America (USA), Brazil, Russia, India and United Kingdom (UK) are the top five most COVID-19 affected countries until July 10, 2020 [4]. Although researchers across the world are working on a vaccination for COVID-19 to suppress its rapid dissemination [5], an official vaccination of COVID-19 will not be available for some time. Therefore, the preliminary mitigation policies i.e., social distancing, public lockdown and improved testing efficiency are vital steps to control the transmission rate of COVID-19 worldwide [6].

Despite several mitigation policies, COVID-19 cases are in an alarming situation in South Asian countries, particularly in India. Being a developing nation and the second largest populous country in the world, it is very challenging to control the spread of COVID-19 in India. The first COVID-19 confirmed case was reported in Kerala, India on January 30, 2020 [4]. Under the lack of sufficient medical facilities in a country of 1.3 billion people, various preliminary containment strategies have been implemented by the government of India in the beginning phase of COVID-19 transmission [7–9]. Social distancing and several social awareness programs have been initiated as the first precautionary steps to prevent the possibility of human-to-human transmission. The primary goal of these strategies is to control the main cause of COVID-19 transmission i.e., the social interaction in public places i.e., schools, colleges, theaters, cultural and sports events excluding the essential public services like hospitals, daily grocery shops, banks, and police. On March 25, 2020 the Indian government announced the first official nationwide lockdown for 14 days [9]. Subsequently, the public lockdown has been implemented in five different phases in India. In the second and third phase, the lockdown has been extended up to April 15, 2020 and May 03, 2020 respectively to achieve better control on COVID-19 transmission in India. However, few relaxations (conditional inter-state travel, resuming domestic flights) in full lockdown have been provided by Indian government in the fourth (May 04- May 17, 2020) and fifth phases (May 18- May 31, 2020) of lockdown respectively. On June 01, 2020 onwards, the nationwide lockdown was partially lifted in India. Unfortunately, it resulted in a sudden growth in COVID-19 cases in the last 40 days. At present, the number of COVID-19 cases are dramatically increasing in India and tending toward its epidemic peak [10]. Therefore, it is necessary to study the dynamics and predict the future inflammation of COVID-19 in India.

Mathematical models have the potential to trace and predict the epidemic trajectory under different circumstances. Various mathematical, statistical models have been proposed to understand the dissemination trajectory for a pandemic [11–34]. Among these models, Susceptible (S)-Infected (I)-Recovered (R) model (SIR model) has been frequently used in past to predict the dynamics of various contagious diseases i.e., HIV virus [11], plague [12], SARS [13]. Recently, SIR model has also been applied for prediction of COVID-19 trajectory and its epidemic peak in all over the world [15–19]. Numerous relevant studies have been carried out in India as well [20–34]. A COVID-19 simulation models i.e., IndiaSIM have been proposed by Center for Disease Dynamics, Economics & Policy (CDDEP) to predict COVID-19 infected cases under different lockdown periods in India [25]. This model predicts the total infected cases as 10^6

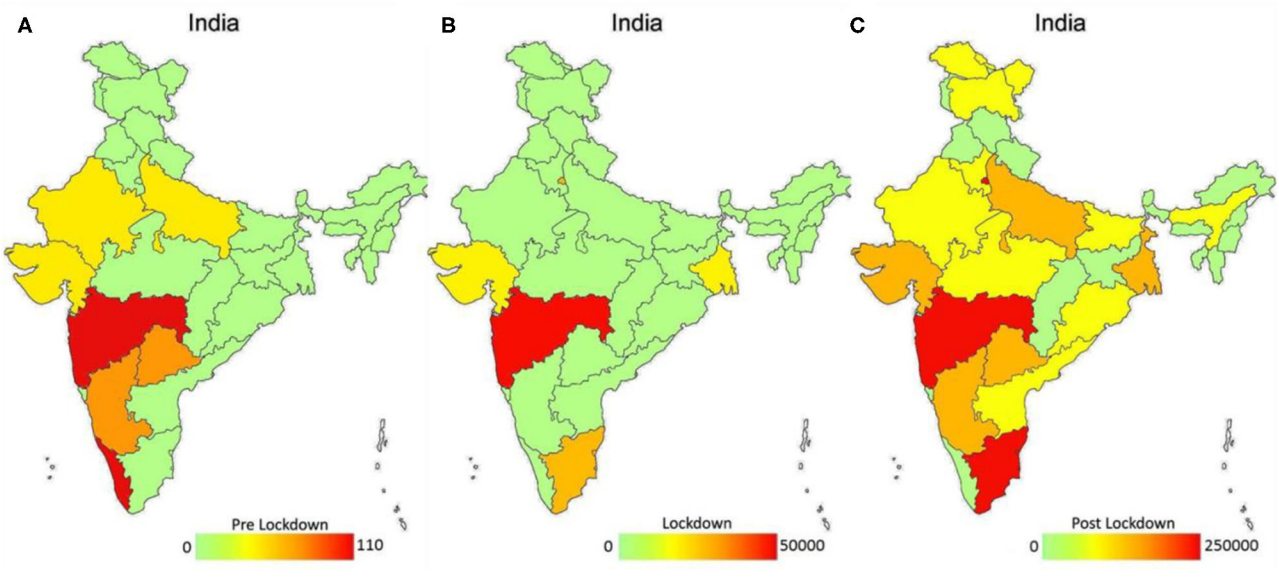


FIGURE 1 | Spatial distribution of COVID-19 reported confirmed cases under the effect of lockdown in India **(A–C)**. Color bar represents the spatial variation of confirmed cases from lowest (green color) to highest (red color) under three phases of lockdown i.e., pre-lockdown **(A)**, during lockdown **(B)** and post-lockdown **(C)** in India.

under pre-lockdown conditions and 1.5×10^6 under lockdown conditions, respectively. Pai et al. investigated the COVID-19 transmission dynamics under the effect of lockdown. This study predicts that the peak of active infected cases around 43,000 will occur in the middle of May, 2020 [29]. In another study, a compartmental mathematical model to predict COVID-19 dynamics has been proposed by analyzing data up to April 30, 2020. This study mainly focuses on the computation of basic reproduction number (R_0) for COVID-19 spread and its impact on epidemic transmission in India. Moreover, *Sarkar et al.* proposed a six compartment mathematical model i.e., susceptible (S), asymptomatic or pauci-symptomatic infected (A), symptomatic reported infected (I), unreported infected (U), quarantine (Q), and recovered (R) ["SAIUQR" model] to predict the COVID-19 dynamics in four major states of India i.e., Jharkhand, Gujarat, Andhra Pradesh, and Chandigarh [33]. However, such studies have been carried out at the very early stage of the pandemic. Further a major problem for such short-term predictions is the determination of crucial model parameters and corresponding impact of intervention strategies on the epidemic dynamics. Moreover, these studies mainly predict the number of active cases from COVID-19 at particular region whereas the mortality i.e., death rate is also a crucial factor to take into account while forecasting the influence of a fatal pandemic i.e., COVID-19. In addition, it is pointed out in earlier studies that under-reporting i.e., the difference between the reported cases and actual confirmed cases is another vital parameter in prediction of COVID-19 trajectory [16]. A data=based study on impact of abrupt changes in intervention policies (testing) and corresponding population synchrony with transmission dynamics of Bovine-tuberculosis has been discussed [14]. It depicts how testing interruptions can alter the epidemic

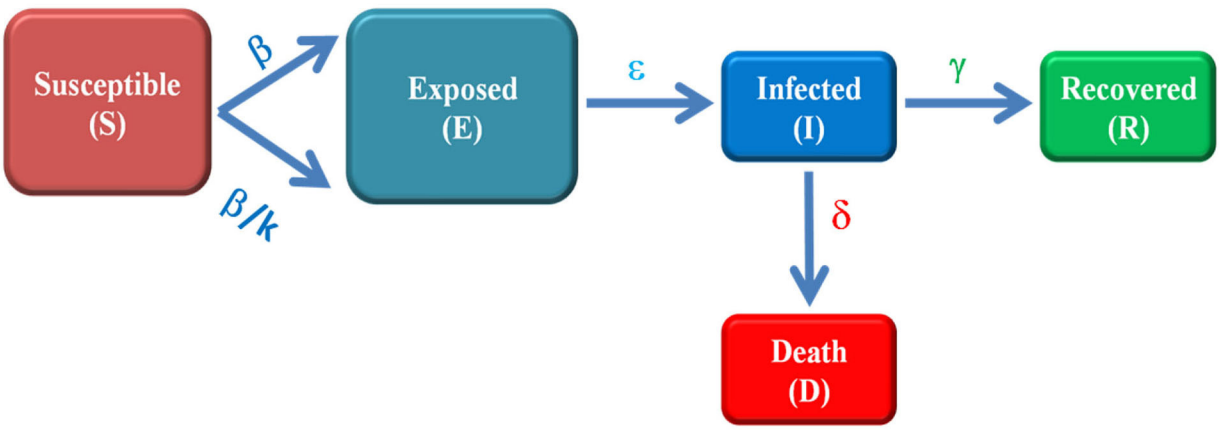


FIGURE 2 | Schematic of modified SEIRD model. Susceptible (S) is the number of un-infected persons, exposed (E) are the individuals which are infectious but not able to transmit the disease, infected (I) represent the active cases i.e., those who can transmit the disease, recovered (R) are the number of individuals who have been cured from disease and Death (D) represents the number of deaths from COVID-19. The parameters $\beta, \varepsilon, \gamma$ and δ are transmission rate, incubation rate, recovery rate and death rate, respectively. The transmission rate (β) is optimized to β/k, $(0.2 \leq k \leq 1)$ to account for the impact of lockdown.

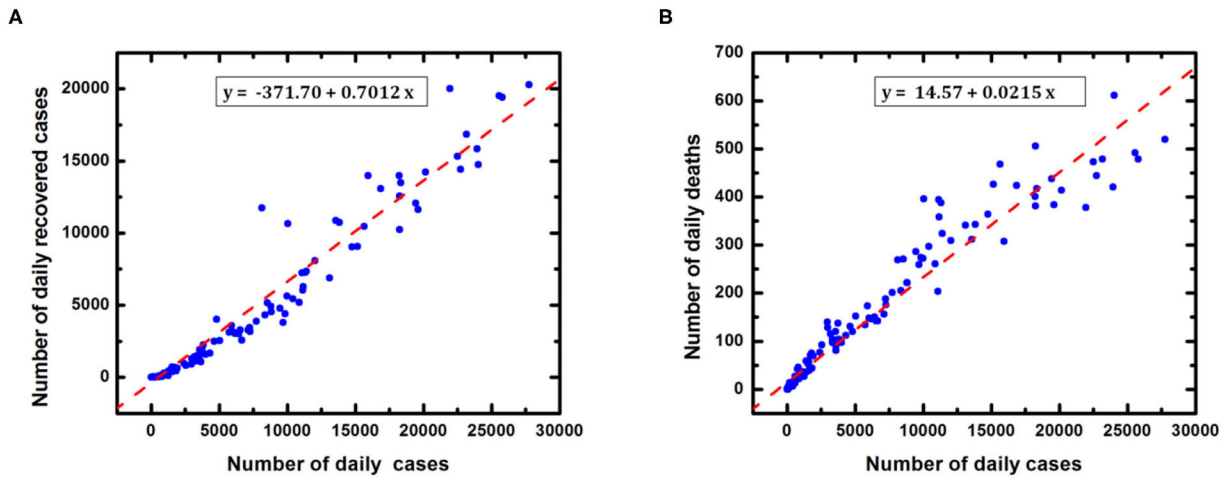


FIGURE 3 | Determination of recovery rate and death rate in India **(A)** Reported recovered cases **(B)** Reported deaths with respect to daily confirmed cases. The recovery rate (γ) and death rate (δ) have been estimated using linear regression method with $r = 0.9701$ and $p < 0.0001$.

parameters and shift the dynamics of disease, and suggests regular testing leads to more de-synchronized infected cases.

In this study, we propose an improvised five compartment mathematical model i.e., Susceptible (S)-Exposed (E)-Infected (I)-Recovered (R)-Death (D) (SEIRD model) to analyze the progression of COVID-19 and forecast the epidemic peak of COVID-19 pandemic under the influence of a nationwide public lockdown in India. We use robust statistical methods i.e., linear regression method and least square method to estimate the SEIRD model parameters by analyzing publicly available reported data of COVID-19 outbreak in India. In particular, a few crucial factors i.e., optimized transmission rate, effective reproduction number [R (t)] and under-reporting are incorporated in this study for accounting impact of public lockdown. Consolidation of these additional parameters is helpful to provide a broader picture of COVID-19 dissemination in India.

MATERIALS AND METHODS

Data Collection

The data has been collected from the open source COVID-19 dataset official websites [3, 4, 9]. Global data of COVID-19 has been retrieved from World Health Organization (WHO) official dashboard [3] whereas publicly available data sources [4, 9] have been used to perform data based study COVID-19 progression in India.

Figures 1A–C represents the spatial transmission of COVID-19 cases in India during three time periods i.e., pre-lockdown (January 30- March 24, 2020), lockdown (March 25-May 31, 2020) and post-lockdown periods (June 1-July 10, 2020). It clearly indicates that COVID-19 cases are least transmitted during the lockdown period as compared to pre-lockdown and post-lockdown across different states of India.

SEIRD Model

We used a five-compartment epidemic model i.e., Susceptible (S)-Exposed (E)-Infected (I)-Recovered (R)-Death (D) (SEIRD model) to study COVID-19 in India. The SEIRD model has been improvised to account for the effect of the containment policy (lockdown) by adjusting the contact rate (β) and reproduction number (R) accordingly. **Figure 2** represents a pictorial view of SEIRD model.

Theoretically, the SEIRD model can be represented by four differential equations [16]

$$\frac{dS}{dt} = -\beta I S \tag{1.1}$$

$$\frac{dE}{dt} = \beta I S - \varepsilon E \tag{1.2}$$

$$\frac{dI}{dt} = \varepsilon E - \gamma I \tag{1.3}$$

$$\frac{dR}{dt} = \gamma (1 - \delta) I \tag{1.4}$$

$$\frac{dD}{dt} = \delta \gamma I \tag{1.5}$$

Here $\beta, \varepsilon, \gamma$ and δ are contact rate/transmission rate, incubation rate, recovery rate and death rate, respectively and known as model parameters.

SEIRD Model Parameters

Equations (1.1–1.5) are solved in Matlab programming software (Mathworks, R2016a) using ordinary differential equations solver function *"ode45."* In this study, initial susceptible volume has been determined by multiplying the total population of India (N) with a factor 10^{-3} [20]. The total confirmed cases have been estimated by adding the number of infected (active) cases, recovered cases and number of deaths predicted from the model. The data is fitted with the SEIRD model to predict the possible total confirmed cases in India under the effect of lockdown. The fixed value of incubation period i.e., 5.2 days (incubation rate = 0.1923 day^{-1}) is used in our model [21]. However, the recovery rate (γ) and death rate (δ) has been calculated with the help of **Figures 3A,B**, respectively. To estimate the transmission rate (β), the least square method is used to optimize the value of β [23]. We have minimized the difference between reported cases taken from [4] and predicted the number of cases as

$$min|C^*(t) - C(t)|^2 \tag{2}$$

Here $C^*(t)$ represents the data based total number of confirmed cases and $C(t)$ represents the corresponding number of cases predicted by model. The numerical values of $C(t)$ can be computed as [28]

$$C(t) = I(t) + R(t) + D(t) \tag{3}$$

Further, we have assumed that β has been changed to β/k, $(0.2 \leq k \leq 1)$ under the impact of lockdown in India [29] and projected the number of total confirmed cases for the estimated β. Out of these estimated values of β, the most suitable value i.e., $\beta = 0.4809$ is used to obtain the best fit for predicted number of cases with reported number of cases. These SEIRD model parameters are listed in **Table 1**.

Effective Reproduction Number [R (t)]

The basic reproduction number (R_0) is the most crucial parameter in the SEIRD mathematical model. It determines how the disease is transmitting over the population during a particular time interval. The value of $R_0 > 1$ implies the disease is in

TABLE 1 | SEIRD model parameters.

S. No.	SEIRD model parameters	Value and source
1	β	0.4809 (estimated)
2	ε	0.1923 (adapted from [21])
3	Γ	0.7012 (calculated from **Figure 3A**)
4	δ	0.0215 (calculated from **Figure 3B**)

*Transmission rate (β) has been estimated from the model with 95% confidence interval (0.4521, 0.4922). The incubation period (ε) (5.2 days) is fixed and adapted from [21]. The recovery rate (γ) and death rate (δ) have been estimated using **Figures 3A,B**, respectively.*

outbreak state, $R_0 = 1$ indicates that the disease is still in system but in a stable state while $R_0 < 1$ represents the disease has diminished completely. In practice, the effective reproductive number R (t) measures the number of secondary cases generated by an infectious case once an epidemic is ongoing. In our study, we have considered R as R(t) i.e., as a function of time [17]. It is given as

$$R(t) = \frac{\beta(t)}{\gamma} \tag{4}$$

Case Fatality Risk/ratio (CFR) is another important parameter for epidemic study. It interprets the status of epidemic in terms of deaths. It is defined as [16]

$$CFR = \frac{Number\ of\ deaths\ daily}{Number\ of\ cases\ daily} \tag{5}$$

RESULTS AND DISCUSSION

At first, we performed a data-based study of COVID-19 progression in India during the time period of January 30-July 10, 2020 i.e., 160 days. We have observed the trend of COVID-19 under the impact of lockdown. **Figure 4A** represents the number of cases progression over time and indicates that the COVID-19 cases were negligible in pre-lockdown period. It implies that the lockdown has been implemented at the beginning phase of COVID-19 transmission in India. Further on comparing the cases in the lockdown and post lockdown periods, clearly our results show that the cases are much fewer during the lockdown period, and have increased abruptly just after lockdown lifted. Similarly, **Figure 4B** depicts the variation of testing samples under the impact of lockdown. It is observed that a negligible amount (20,864) of samples have been tested

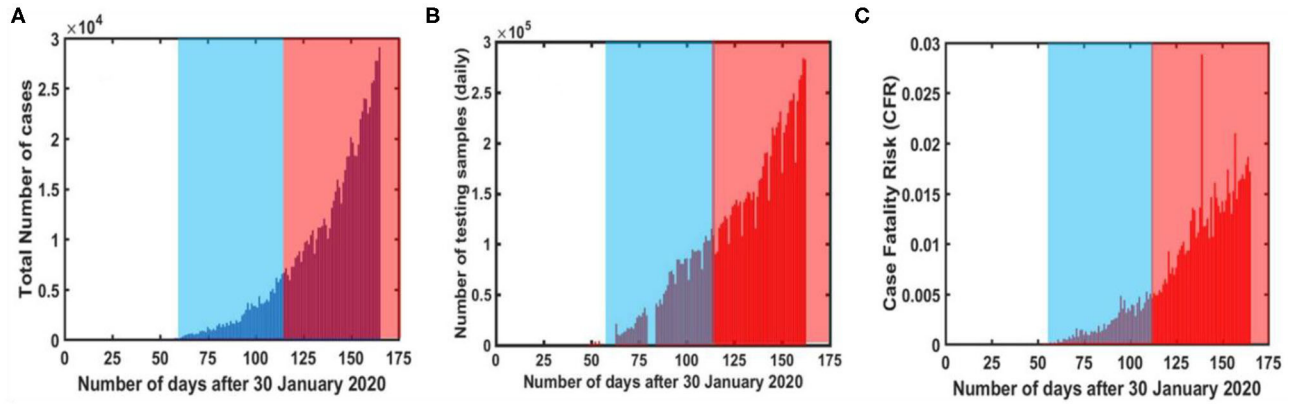


FIGURE 4 | COVID-19 reported data [5] based analysis during different time intervals in India **(A)** Reported number of daily cases **(B)** daily tested samples **(C)** Case Fatality Ratio (CFR). The time interval has been divided into three sections i.e., pre-lockdown, during lockdown (blue shaded area) and post-lockdown (red shaded area), respectively.

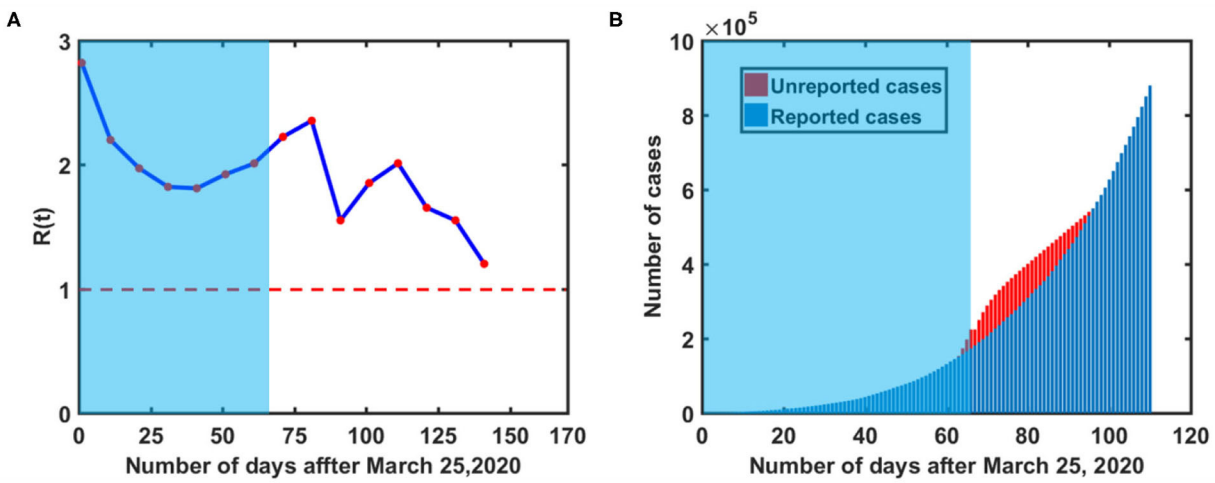


FIGURE 5 | COVID-19 reported data-based analysis during different time intervals in India, **(A)** Basic Reproduction number variation over time **(B)** estimation of underreported cases. Red colored dashed line in **(A)** represents the threshold value of reproduction number i.e., R(t) = 1. The red colored region in **(B)** represents possible under-reported cases in India. [Blue shaded region represents actual lockdown period].

during pre-lockdown period [3] which implies that a number of cases might have not been reported during the pre-lockdown period as well. In fact, first sample testing data has been reported on March 13, 2020 in India. Upon further analyzing the sample testing data, an increasing trend of testing samples was observed with respect to time. Apparently, a lesser number of samples had been tested during the lockdown period as compared to the post-lockdown period. Moreover, an identical ascending trend has been noticed for number of cases and number of testing samples. In addition, **Figure 4C** indicates that CFR is very low (1–3%) in India i.e., COVID-19 fatality is below average as compared to other countries in the world. However, CFR also increases rapidly in the post-lockdown period.

Figure 5A represents the variation in effective reproduction number at various phases of lockdown in India. A descending trend has been observed in R during the lockdown period, whereas it hikes again as the lockdown is lifted i.e., in post lockdown period. Further observations suggest that R (t) approaches unity around August 25, 2020. It indicates that the epidemic peak should appear in August 2020 in India. Apart from this, **Figure 5B** represents the possible under-reported cases in India. It can be defined as the difference in the total confirmed cases predicted from model and data based reported cases [17]. A significant number of underreported cases ($\sim 10^5$) has been observed in the beginning of post lockdown in India.

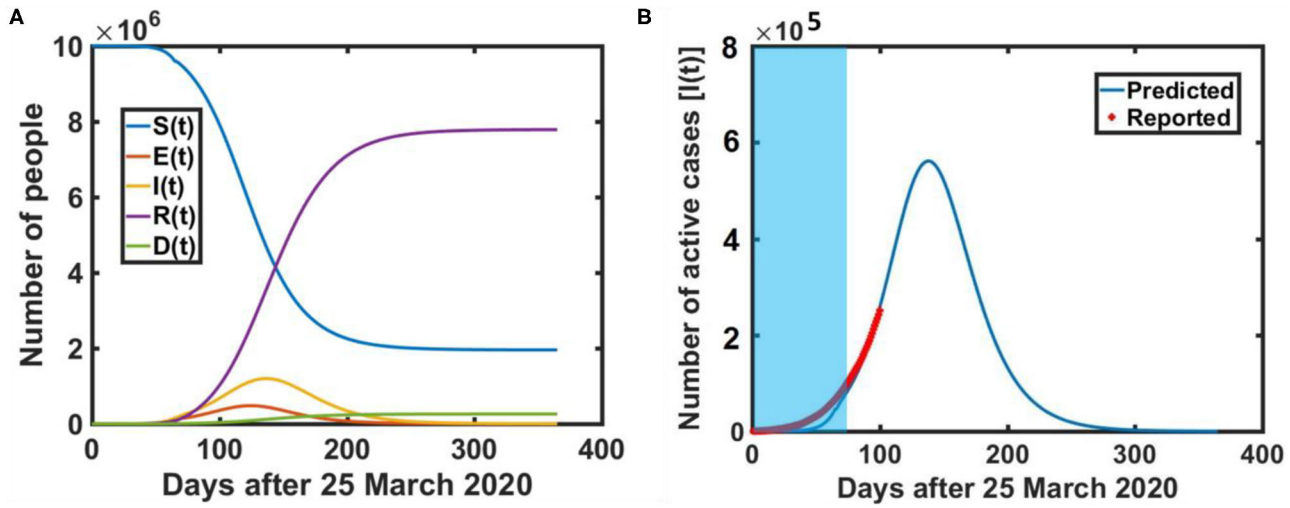


FIGURE 6 | Mathematical model prediction of COVID-19 dissemination in India, **(A)** Five compartment trajectory over time **(B)** COVID-19 epidemic peak i.e., maximum active cases prediction in India. [Blue shaded region represents the actual lockdown period]. The incubation period and recovery rate are taken as 5.2 days (incubation rate = 1/5.2 day^{-1}) is 0.7012 day^{-1} respectively.

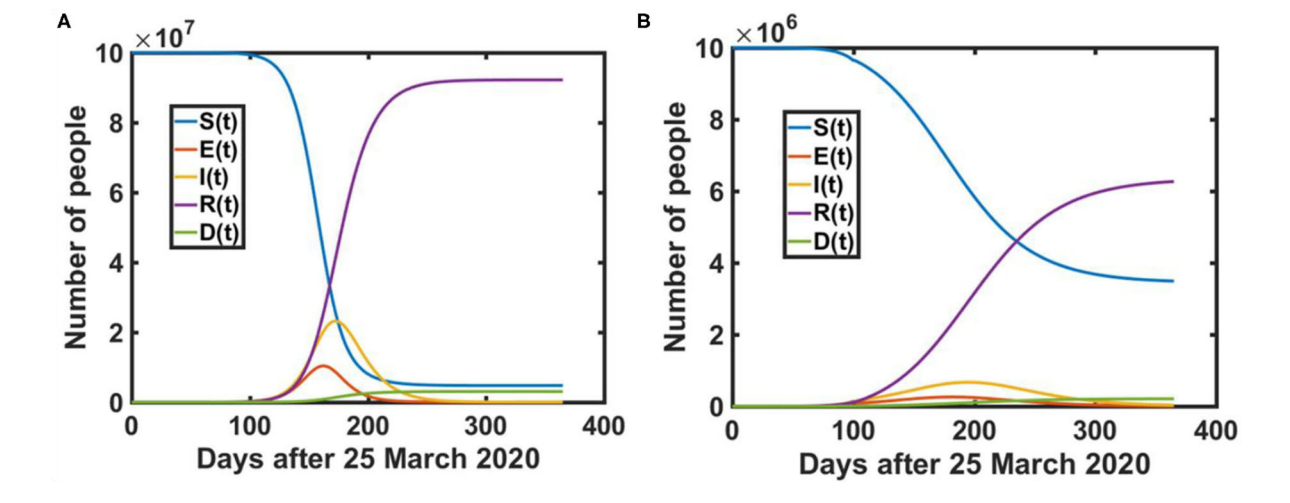


FIGURE 7 | Mathematical model prediction of COVID-19 dissemination in India under **(A)** without lockdown **(B)** Lockdown of 100 days. The incubation period and recovery rate are taken as 5.2 days (incubation rate = 1/5.2 day^{-1}) is 0.7012 day^{-1} respectively.

TABLE 2 | COVID-19 epidemic peak prediction using SEIRD model in India under lockdown effect.

S. no.	Lockdown period	Estimated COVID-19 epidemic peak	
		Date	Estimated maximum cases
1	0 days (hypothetical cases i.e., if no lockdown has been implemented in India)	September 25, 2020	2.1×10^7
2	65 days (March 25- May 31 2020) (actual case)	August 16, 2020	2.5×10^6
3	Lockdown of 100 days (extensive case)	September 15, 2020	1.2×10^6

Prediction of COVID-19 Epidemic Peak in India

To forecast the trajectory of COVID-19 in subsequent months, we have applied the SEIRD model to predict the trajectory of COVID-19 under the nationwide lockdown in India. **Figure 6A** represents the actual transmission dynamics of all five compartments (susceptible, exposed, infected, recovered and death) for COVID-19 in India. **Figure 6B** provides more clear insights of the epidemic peak in India based on actual model parameters. It is estimated that the maximum number of COVID-19 confirmed cases and active cases should be found around 2.5×10^6 and 5.8×10^5 respectively at mid-August 2020. The red colored markers represent the COVID−19 trajectory based on reported data [4]. On the other hand, **Figure 7A** shows the five-compartment (SEIRD) trajectory under the pre-lockdown condition in India. It indicates that the number of cases would be around 10^7 in that case. Moreover, **Figure 7B** depicts if a lockdown of 100 days could have been implemented instead of 65 days, the number of cases would be much lower i.e., 10^5. However, the peak would be shifted to mid-September 2020 in that case.

Table 2 represents the predicted numerical data under various lockdown possibilities in India. Observations show that the cases could be order of 10^7, if lockdown had not been implemented in India whereas a lockdown of 100 days instead of 65 days could suppress the cases to up to 50% of the actual cases prediction in India.

CONCLUSION

In this study, we have analyzed and predicted COVID-19 dissemination dynamics in India using a five-compartment mathematical model i.e., SEIRD model. The conventional SEIRD model has been improvised for accounting the impact of COVID-19 mitigation policy i.e., nationwide lockdown in India. The modified transmission rate [β/k, $(0.2 \leq k \leq 1)$] based on intervention strategies i.e., lockdown, time dependent reproduction number [R (t)] and determination of under reported cases are the key elements of our improvised SEIRD model. Moreover, the data has been examined up to July 10, 2020 in this study. The data based study suggests that COVID-19 inflammation has been controlled up to 50% by implementing public lockdown in India at the starting phase of disease transmission. Results indicated that the effective reproduction number has been varied within a range of 2 to 3 during the lockdown period. On the other hand, a relationship has been established between testing efficacy and disease transmission in an earlier study [34]. Our study strongly recommends improved testing efficiency (higher number of sample testing) as one of the robust strategies against COVID-19 spread worldwide. However, a significant number of un-reported cases ($\sim 10^5$) have been identified in early post-lockdown period in India. The possible reasons for this underreporting might be inconsistency in testing samples or abrupt release of lockdown.

In addition, the COVID-19 epidemic peak has also been estimated under different possible lockdown scenarios using the mathematical model. This study predicts the maximum number of total confirmed cases and maximum active cases around 2.5×10^6 and 5.8×10^5, respectively on August 15, 2020. Corresponding outcomes imply that COVID-19 cases could have been controlled up to a better extent with an extensive lockdown period. Although nationwide lockdown has been implemented at the emerging phase of COVID-19 in India, this was only for 65 days. Keeping in mind the population and medical facilities in India, such a short lockdown period might not be sufficient in order to suppress COVID-19 transmission. Our study suggests that the lockdown could have been extended up to 100 days to achieve better control over COVID-19 transmission. Apart from this, the researchers around the world are engaged in developing a vaccine for SARS-COV-2. Therefore, one can expect that the evolution of such pharmaceutical intervention strategies and other medical interventions against COVID-19 might modify the outcomes of this study.

In brief, this study recommends the nationwide public lockdown as an effective controlling policy for COVID-19 dissemination, not only in India but across the globe. It will definitely play a crucial role in epidemic peak suppression of COVID-19 dissemination. Although our improvised SEIRD model provides satisfactory results, one can include few additional parameters i.e., imported or exported cases, asymptotic cases, under-reported cases etc. to enhance the validity of mathematical model while predicting an epidemic dissemination.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://data.world/kim4597/world-covid19/workspace/intro.

AUTHOR CONTRIBUTIONS

VT, ND, and NB designed the study. VT and NB developed the mathematical model and simulation coding. ND contributed in manuscript writing and data interpretation. All authors contributed to the article and approved the submitted version.

REFERENCES

1. World Health Organization (WHO). *Corona Virus Disease (COVID-19) Outbreak Situation*. World Health Organization (2020). Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019
2. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-cov-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med.* (2020) 382:1177–9. doi: 10.1056/NEJMc2001737
3. WHO. *COVID-19-Global-Data.* WHO (2020). Available online at: https://covid19.who.int/info (accessed July 13, 2020).
4. Data. *World-COVID-19-Data.* (2020). Available online at: https://data.world/kim4597/world-covid19/workspace/intro (accessed July 14, 2020).
5. *Covid-19 Vaccine Trial on Humans Starts as UK Warns.* (2020). Available online at: https://edition.cnn.com/2020/04/23/health/coronavirus-vaccine-trial-uk-gbr-intl/index.html
6. Ferguson N M, Laydon D, Nedjati-Gilani Imai G, Ainslie K, Baguelin K, Bhatia S, et al. Impact of nonpharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Bull Math Biol.* (2020) 82:52. doi: 10.1007/s11538-020-00726-x
7. Stringent and Stingy. *Emerging-Market Lockdowns Match Rich-World Ones. The Handouts Do Not* (2020). Available online at: https://www.lastminutestuff.com/content/Emergingmarket-lockdowns-match-richworld-ones/6241717.html (accessed July 15, 2020).
8. *Indian Railways Ready to Modify 20000 Coaches Which…" -PIB.* (2020). Available online at: https://pib.gov.in/PressReleasePage.aspx?PRID=1609528 (accessed July 15, 2020).
9. Government of India (2020). Available online at: https://pib.gov.in/PressReleseDetail.aspx?PRID=1613216 (accessed July 16, 2020).
10. *COVID-19 Data Trend.* (2020). Available online at: https://www.covid19india.org (accessed July 16, 2020).
11. Anderson RM. The role of mathematical models in the study of HIV transmission and the epidemiology of AIDS. *J Acquir Immune Defic Syndr.* (1988) 1:241–56.
12. Didelot X, Whittles LK, Hall I. Model-based analysis of an outbreak of bubonicplague in Cairo in 180. *J R Soc Interface.* (2017) 14:20170160. doi: 10.1098/rsif.2017.0160
13. Cai QC, Jiang QW, Xu QF, Cheng X, Guo Q, Sun QW, et al. To develop a model on severe acute respiratory syndrome epidemic to quantitatively evaluate the effectiveness of intervention measures. *Chin J Epidemiol.* (2005) 26:153–8.
14. Moustakas A, Evans M, Daliakopoulos I, Yannis Markonis Y. Abrupt events and population synchrony in the dynamics of Bovine Tuberculosis. *Nat Commun.* (2018) 9:2821. doi: 10.1038/s41467-018-04915-0
15. Calvetti D, Hoover AP, Rose J Somersalo E. Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19. *Front Phys.* (2020) 8:261. doi: 10.3389/fphy.2020.00261
16. Mauger A, Barchitta M, Battiato S, Agodi A. Estimation of unreported novel coronavirus (SARS-CoV-2) infections from reported deaths: a susceptible–exposed–infectious–recovered–dead model. *JCM.* (2020) 9:1350. doi: 10.3390/jcm9051350
17. Saberi M, Hamedmoghadam H, Madani K, Dolk HM, Morgan AS, Morris JK, et al. Accounting for underreporting in mathematical modeling of transmission and control of COVID-19 in Iran. *Front Phys.* (2020) 8:289. doi: 10.3389/fphy.2020.00289
18. Chen Z, Yang J, Dai B. Forecast possible risk for COVID-19 epidemic dissemination under current control strategies in Japan. *Int J Environ Res Public Health.* (2020) 17:3872. doi: 10.3390/ijerph17113872
19. Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. *J Med Virol.* (2020) 92:645–59. doi: 10.1002/jmv.25750
20. Chatterjee S, Sarkar S, Chatterjee S, Karmakar K, Paul R. Studying the progress of COVID-19 outbreak in India using SIRD model. *Indian J Phys Soc Indian Assoc Cultiv Sci.* (2020) 1–17. doi: 10.1007/s12648-020-01766-8
21. Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and regression model based covid-19 outbreak predictions in India. *arXiv.* (2020). doi: 10.2196/preprints.19406
22. Dhanwant J, Ramanathan V. Forecasting Covid 19 growth in India using susceptible-infected-recovered (sir) model. *arXiv.* (2020). arXiv:2004.00696.
23. Kashyap V, Malhotra B. Progression of COVID-19 in Indian States - forecasting endpoints using sir and logistic growth models. *medRxiv.* (2020). doi: 10.1101/2020.05.15.20103028
24. Gupta S, Shankar R. Estimating the number of COVID-19 infections in Indian hot-spots using fatality data. *arXiv.* (2020). arXiv:2004.04025.
25. Sardar T, Nadim S, Chattopadhyay J. Assessment of 21 days lockdown effect in some states and overall India: a predictive mathematical study on covid-19 outbreak. *arXiv.* (2020) 139:1–10. doi: 10.1016/j.chaos.2020.110078
26. Singh R, Adhikari R. Age-structured impact of social distancing on the COVID-19 epidemic in India. *arXiv.* (2020). Available online at: https://arxiv.org/abs/2003.12055
27. Kaur T, Sarkar S, Chowdhury S, Sinha S, Jolly M, Dutta P. Anticipating the novel coronavirus disease (COVID-19) pandemic. *medRxiv.* (2020). doi: 10.1101/2020.04.08.20057430
28. Pai C, Bhaskar A, Rawoot V. Investigating the dynamics of COVID-19 pandemic in India under lockdown. *Chaos Solitons Fract.* (2020) 138:109988. doi: 10.1016/j.chaos.2020.109988
29. Bhandari S, Tak A, Gupta J, Patel B, Shukla J, Shaktawat AS, et al. Evolving trajectories of COVID-19 curves in India: prediction using autoregressive integrated moving average modelling. *Res Square.* (2020). doi: 10.21203/rs.3.rs-40385/v1
30. Samui P, Mondal J, Khajanchi S. A mathematical model for COVID-19 transmission dynamics with a case study of India. *Chaos Solitons Fract.* (2020) 140:110173. doi: 10.1016/j.chaos.2020.110173
31. Khajanchi S, Sarkar K. Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India. *Chaos.* (2020) 30:071101. doi: 10.1063/5.0016240
32. Sarkar K, Khajanchi S, Nieto J. Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fract.* (2020) 139:110049. doi: 10.1016/j.chaos.2020.110049
33. Khajanchi S, Sarkar K, Mondal J, Perc M. Dynamics of the COVID-19 pandemic in India. *arXiv.* (2020). doi: 10.21203/rs.3.rs-27112/v1
34. Moustakas A. Ranking the explanatory power of factors associated with worldwide new Covid-19 cases. *arXiv.* (2020). arXiv:2006.00971.

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 18 November 2020
doi: 10.3389/fphy.2020.578455

On the Time Shift Phenomena in Epidemic Models

Ayse Peker-Dobie[1,2], *Ali Demirci*[1]*, *Ayse Humeyra Bilge*[2] *and Semra Ahmetolan*[1]

[1]*Department of Mathematics, Faculty of Science and Letters, Istanbul Technical University, Istanbul, Turkey,* [2]*Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Kadir Has University, Istanbul, Turkey*

In the standard Susceptible-Infected-Removed (SIR) and Susceptible-Exposed-Infected-Removed (SEIR) models, the peak of infected individuals coincides with the inflection point of removed individuals. Nevertheless, a survey based on the data of the 2009 H1N1 epidemic in Istanbul, Turkey displayed a time shift between the hospital referrals and fatalities. An analysis of recent COVID-19 data and the records for Spanish flu (1918–1919) and SARS (2002–2004) epidemics confirm this observation. We use multistage SIR and SEIR models to provide an explanation for this time shift. Numerical solutions of these models present strong evidence that the delay between the peak of $R'(t)$ and the peak of $J(t) = \sum_i I_i(t)$ is approximately half of the infectious period of the epidemic disease. In addition, we use a quadratic approximation to show that the distance between successive peaks of I_i is $1/\gamma_i$, where $1/\gamma_i$ is the infectious period of the ith infectious stage, and we present numerical calculations that confirm this approximation.

Keywords: COVID-19, epidemic models, multistage Susceptible-Infected-Removed model, multistage Susceptible-Exposed-Infected-Removed model, time shift

OPEN ACCESS

Edited by:
Aristides Moustakas, University of Crete, Greece

Reviewed by:
Jose Roberto Castilho Piqueira, University of São Paulo, Brazil
Ibrahim Halil Aslan, Batman University, Turkey
Changjing Zhuge, Beijing University of Technology, China

***Correspondence:**
Ali Demirci
demircial@itu.edu.tr

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *30 June 2020*
Accepted: *12 October 2020*
Published: *18 November 2020*

Citation:
Peker-Dobie A, Demirci A, Bilge AH and Ahmetolan S (2020) On the Time Shift Phenomena in Epidemic Models. Front. Phys. 8:578455. doi: 10.3389/fphy.2020.578455

1. INTRODUCTION

From the early attempts [1–4] to recent studies, epidemic modeling which is applicable in a wide range of fields from informatics [5, 6] to chemistry [7–9] has drawn the attention of researchers in various disciplines. Since the basic compartmental model Susceptible-Infected-Removed (SIR) which is commonly used to model diseases for which the infection confers permanent immunity was introduced by Kermack and McKendrick in 1927 [4]; other compartmental models [10–12] have been developed to model diseases with different structures and dynamics. Especially in recent years, major outbreaks such as avian flu in 2005, swine influenza in 2006, and H1N1 influenza in 2009 have highlighted the need for more effective and reliable models to control the spread of disease and to provide a better knowledge for the prediction of future threats and for the development of stronger containment strategies. In Refs. [13–18], some results on the modeling of different types of epidemic diseases, the solution form of these models, the observation of global stability, and the determination of the final size of the epidemic are obtained. In Ref. [15], global stability criteria are derived for the SEIS model which can be regarded as a model with no immunity, and different SIR models are examined in Refs. [16, 17]. Moreover, the works [17, 18] provide some useful results on the final size of the epidemic for SIR models. With the same motivation of these works but rather a different contribution to literature, we use a multistage model [14] in this article to explain the time shift observed in several surveys such as Spanish flu (1918–1919) [19], SARS (2002–2004), the 2009 H1N1 in Istanbul, and recently, COVID-19 [20] (see **Figures 1–3**).

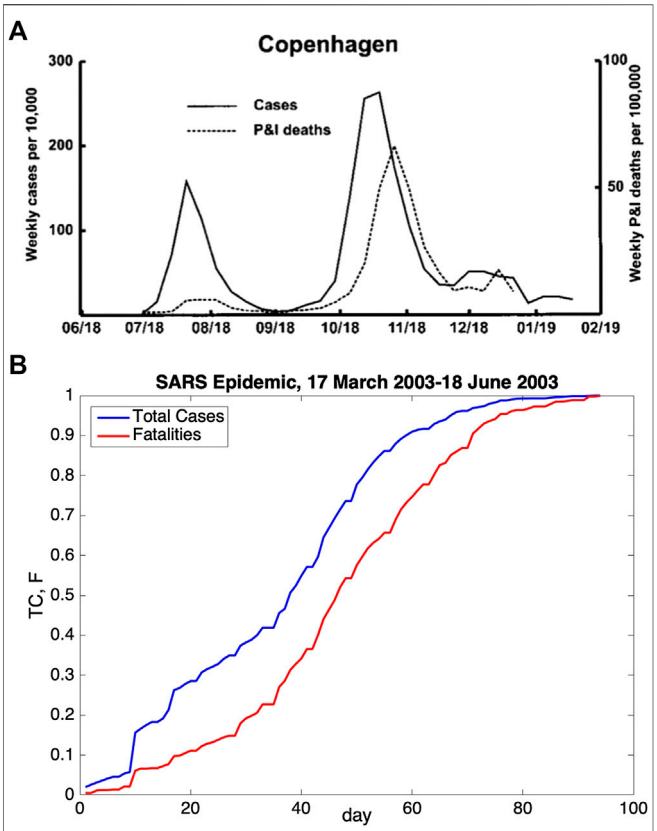


FIGURE 1 | (A) Weekly number of influenza cases and respiratory deaths (pneumonia and influenza) in Copenhagen, Denmark, during 1918–1919 are presented. This figure was taken from the study of Andreasen et al. [19] (copyright Andreasen et al. [19]). A shift occurred between the maxima of curves. **(B)** Daily number of total cases and cumulative number of the fatalities for the 2003 SARS epidemic in the world are shown. The data provided by WHO were used in the generation of this figure (https://www.who.int/csr/sars/country/en/) (last access: September 15, 2020). An explicit shift was also observed for the SARS epidemic.

In the literature, available observed data that are used for the ordinary differential equation system representing the classical SIR epidemic model is based on the curve of removed individuals. Usually, this curve is obtained by taking into consideration only the fatality data of the epidemic disease, whereas in some research studies, not only the fatality data but also the hospitalization data for the epidemic are taken into account in the modeling process. In Ref. [21], it is shown that there exists a delay between the peak of the hospitalization (infectious) curve and the inflection point of the fatality (removed) curve based on the data collected. The original contribution of this article to the literature is that we explain this time shift by the multidimensional form of SIR and SEIR models and also provide numerical evidence that the expected delay is approximately half of the infectious period of the epidemic disease for both of the multistage systems.

In the second section of this work, we give a brief summary of the classical epidemic SIR and SEIR models and define the multistage form of these models that will be used in further analysis. The graphs obtained by the numerical evaluations of the classical SIR and SEIR models and their multistage forms are also given. Analysis of these graphs confirms that the multistage SIR and SEIR models explain the time shift observed in several surveys. In the third section, the evaluation of delay for different epidemic parameters is presented by using the numerical evaluations of these multistage models. In the fourth section, the distance between the points where successive stages and hence any two stages assume their maximum is found approximately. The last section includes a summary of the results obtained in the previous sections as well as motivations for future analysis.

2. STANDARD EPIDEMIC MODELS AND EPIDEMIC MODELS WITH MULTIPLE INFECTIOUS STAGES

The SIR model is commonly used to model diseases for which the removed individuals are assumed to be immune to reinfection. In addition, the total population with constant size is divided into three distinct compartments the size of which change with time t. These compartments are called the susceptible class S, the infective class I, and the removed class R. Healthy individuals with no immunity are members of class S until they are infected with a pathogen and become capable of transmitting the disease to others. They move from the class S into the class I once they are infected and then from I to R once they recover or die. Childhood illnesses like measles or rubella are good examples for the SIR model.

The SIR epidemic model without vital dynamics, that is, the recruitment of new susceptible through birth or immigration as well as the loss through mortality or emigration are ignored, is defined by the following system of nonlinear ordinary differential equations:

$$\begin{aligned} SIR: \quad S' &= -\beta SI, \\ I' &= \beta SI - \gamma I, \\ R' &= \gamma I, \end{aligned} \tag{1}$$

where the coefficient β refers to the disease transmission rate and $1/\gamma$ represents the duration of infection period. Note that since $S' + I' + R' = 0$, we may assume $S + I + R = 1$ by the use of appropriate normalization.

The standard SIR model ignores a latent phase which is the delay between the time of the acquisition of infection and the onset of infectiousness. In order to define this latent phase, the introduction to the SIR model of an exposed class E whose members are individuals who have been infected with a pathogen but are not yet infectious due to the incubation period of pathogen yields the SEIR model. Chicken pox is suitable for the SEIR model, which is defined by the following system of ordinary differential equations

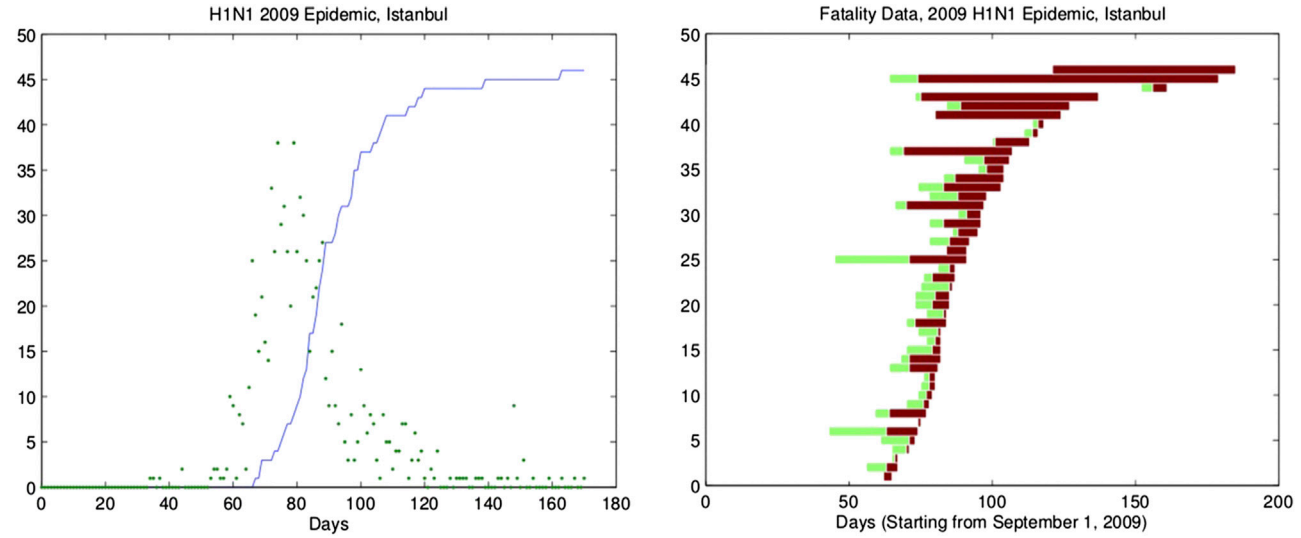


FIGURE 2 | Daily number of referrals to hospitals and cumulative number of fatalities for the 2009 H1N1 epidemic in Istanbul, Turkey. The scatter in the number of daily referral to hospitals indicates that the hospitalization rate varies during the epidemic. The asymmetry of the incidence curve is still observable. The fatalities shown here are adjusted to at most 15 days after referral to the hospital **(left panel)**. The duration of symptoms prior to hospitalization (light color) and the duration of hospitalization prior to death (dark color) for the fatalities in Istanbul, Turkey, due to 2009 H1N1 epidemic **(right panel)**.

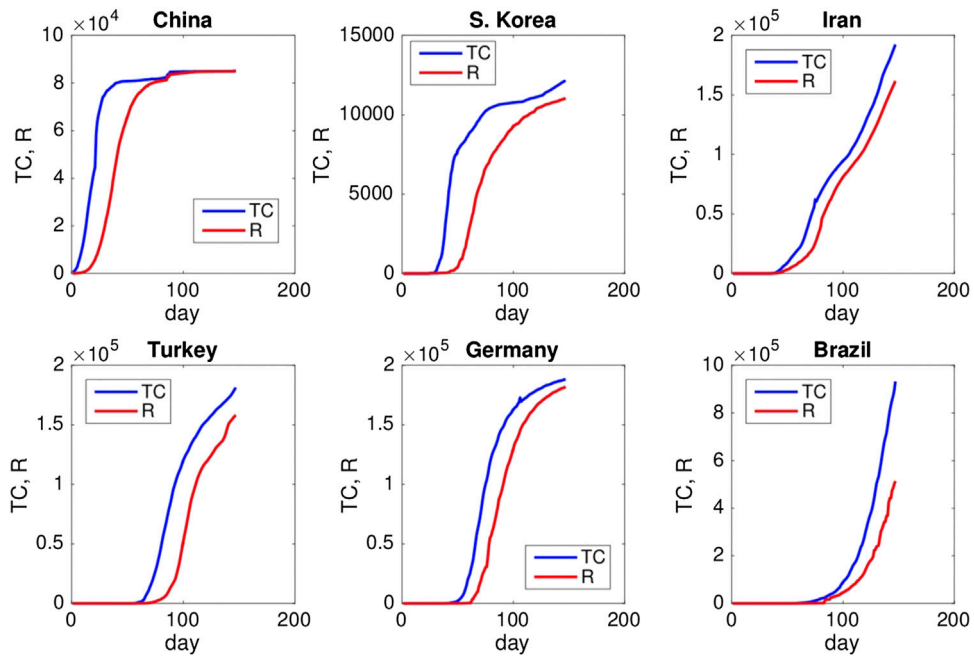


FIGURE 3 | Graphs of total cases (TC) and removed individuals (R) for selected countries. The dataset of each country is collected according to published official reports and available at the website http://www.worldometers.info/coronavirus/ (last access: June 28, 2020). Updated data are also available at the website http://epikhas.khas.edu.tr/. The last data in this work were collected on the 16th of June 2020. Data cover the period January 22–June 28, 2020, and in the following, "Day 1" corresponds to January 22, 2020.

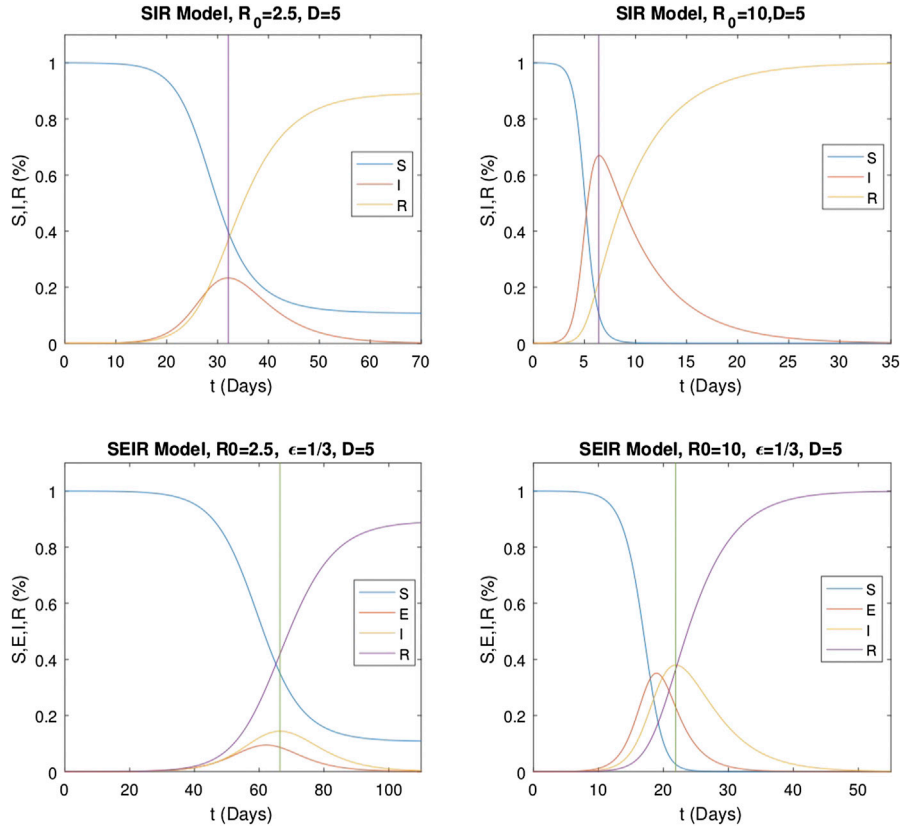


FIGURE 4 | Numerical solutions of classical SIR (1) and SEIR (2) models with R_0 = 2.5 and 10. Here, ϵ and γ are chosen as 1/3 and 1/5 (D = 5), respectively. Time shift does not occur in classical models.

$$SEIR: \quad \begin{aligned} S' &= -\beta SI, \\ E' &= \beta SI - \epsilon E, \\ I' &= \epsilon E - \gamma I, \\ R' &= \gamma I, \end{aligned} \tag{2}$$

where $1/\epsilon$ represents the mean exposed period. Note that we may assume $S + E + I + R = 1$ by the use of appropriate normalization.

These two models are suitable for mathematical modeling of seasonal diseases, but they fail to reproduce the time shift that was observed in the modeling of the 2009 H1N1 epidemic in Istanbul, Turkey [21], as shown in **Figure 2**. For COVID-19, a similar time shift is observed between the curves of total cases and removed individuals in China, South Korea, Iran, Turkey, Germany, and Brazil as shown in **Figure 3**. Publicly accessible data that have been released by the state offices of each country are used for this analysis.

In the literature, a similar time shift is also observed for Spanish flu [19] and SARS epidemics. The graphs for the data of these epidemics are shown in **Figure 1**. Weekly case and fatality reports for Copenhagen, Denmark, in 1918–1919 are displayed in **Figure 1A**. In this figure, it can clearly be seen that there is a time shift between the peak of the infectious cases and the peak of fatalities. Similarly, in **Figure 1A–B**, the time shift is also observable between cumulative cases and cumulative fatalities.

In this study, we use multiple infectivity periods [13, 14] to explain this delay that is unforeseen in the standard SIR model. The approach of multiple infectious stages consists of replacing the single infectious stage I with $N + 1$ substages denoted by I_i, which is the density of individuals in the ith infectious stage. Unlike the model in Ref. [14], each of these stages may have different infectivity β_i and a variable infectious period $1/\gamma_i$. In order to compare the solution curves with the ones for the standard SIR and SEIR models, we set

$$\frac{1}{\gamma_0} + \frac{1}{\gamma_1} + \cdots + \frac{1}{\gamma_N} = \frac{1}{\gamma}.$$

The multistage SIR and SEIR epidemic models are defined by the following systems:

$$Multistage\ SIR: \quad \begin{aligned} S' &= -S(\beta_0 I_0 + \beta_1 I_1 + \cdots + \beta_N I_n), \\ I'_0 &= S(\beta_0 I_0 + \beta_1 I_1 + \cdots + \beta_n I_N) - \gamma_0 I_0, \\ I'_1 &= \gamma_0 I_0 - \gamma_1 I_1, \\ &\cdots \quad \cdots\cdots \\ &\cdots \quad \cdots\cdots \\ I'_N &= \gamma_{N-1} I_{N-1} - \gamma_N I_N, \\ R' &= \gamma_N I_N, \end{aligned} \tag{3}$$

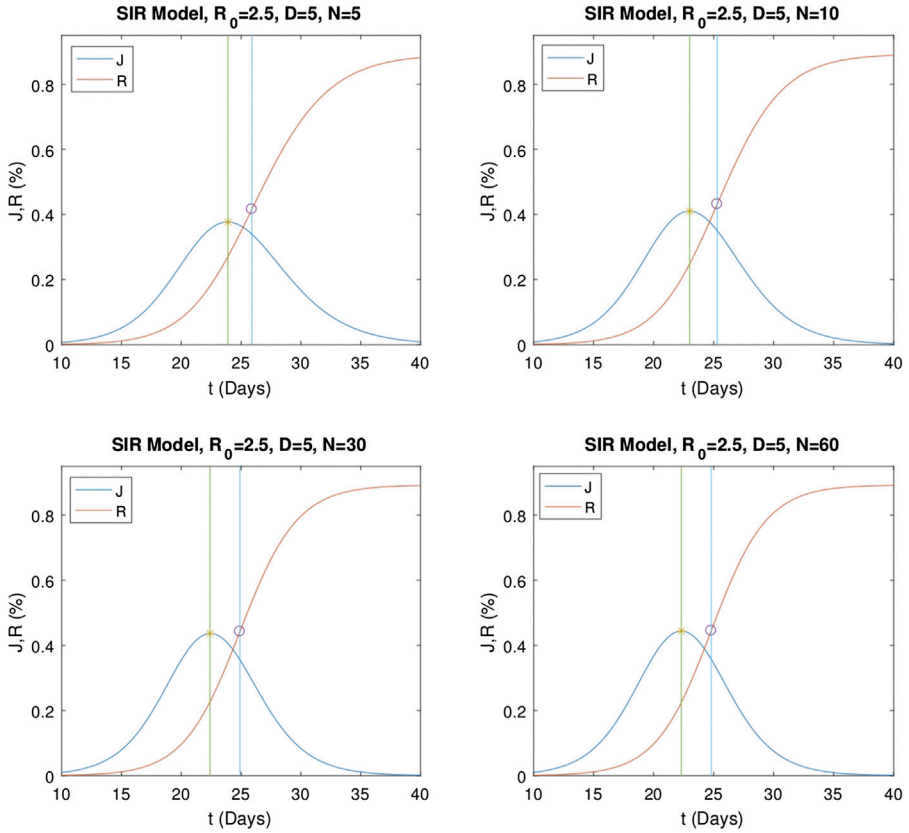


FIGURE 5 | Numerical solutions of the multistage SIR model for different stage numbers $N = 5, 10, 30$, and 60. Here, $R_0 = 2.5$ and $\gamma = 0.2 \times N$. In the graphs, * represents the location of the maximum value of J and o represents the location of the inflection point of R. Graphs show that there is a time shift between these points.

$$
\begin{aligned}
\textit{Multistage SEIR}: \quad & S' = -S(\beta_1 I_1 + \cdots + \beta_N I_N), \\
& E' = S(\beta_1 I_1 + \cdots + \beta_N I_N) - \epsilon E, \\
& I'_1 = \epsilon E - \gamma_1 I_1, \\
& I'_2 = \gamma_1 I_1 - \gamma_2 I_2, \\
& \cdots \quad \cdots\cdots \\
& \cdots \quad \cdots\cdots \\
& I'_N = \gamma_{N-1} I_{N-1} - \gamma_N I_N, \\
& R' = \gamma_N I_N.
\end{aligned} \tag{4}
$$

The multistage SIR and SEIR systems with $\beta_0 = \ldots = \beta_n$ and $\gamma_0 = \ldots = \gamma_n$ correspond to the choice of gamma-distributed "Infection Period Distribution" (IPD) in the integral equation formulation of the SIR model [14].

The linear parts of apparently different infectious stages for $i \geq 1$ in the multistage SIR model and $i \geq 2$ in the multistage SEIR model have a similar structure. We write the linear parts of each equation above as a system and then rearrange and rename as follows to keep the models as clear and simple as possible

$$
\begin{aligned}
& SJR: J = I_0 + \frac{\beta_1}{\beta_0} I_1 + \cdots + \frac{\beta_N}{\beta_0} I_N, \\
& S' = -\beta_0 SJ, \\
& I'_0 = \beta_0 SJ - \gamma_0 I_0, \\
& I'_i = \gamma_{i-1} I_{i-1} - \gamma_i I_i, \quad \textit{for } i = 1, \ldots N, \\
& R' = \gamma_N I_N, \\
& SEJR: J = I_1 + \frac{\beta_2}{\beta_1} I_2 \cdots + \frac{\beta_N}{\beta_1} I_N, \\
& S' = -\beta_1 SJ, \\
& E' = \beta_1 SJ - \epsilon E, \\
& I'_1 = \epsilon E - \gamma_1 I_1, \\
& I'_i = \gamma_{i-1} I_{i-1} - \gamma_i I_i, \quad \textit{for } i = 2, \ldots N, \\
& R' = \gamma_N I_N.
\end{aligned} \tag{5, 6}
$$

Subsequently, the numerical evaluations of these two systems, SJR and SEJR, defined above will be used for some of the structural

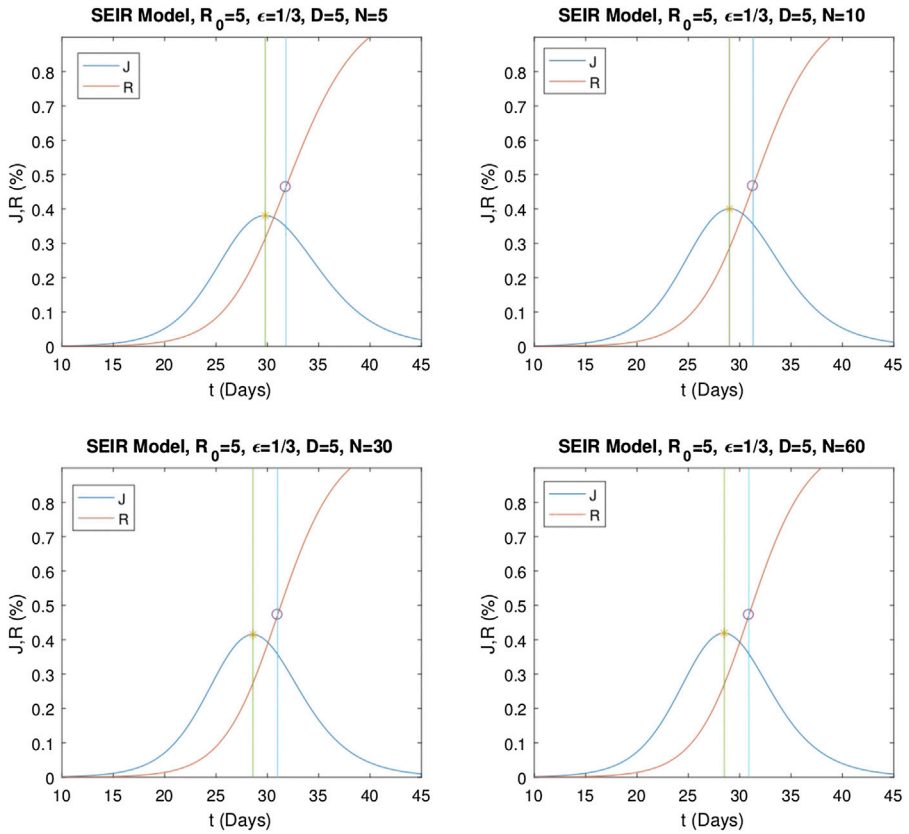


FIGURE 6 | Numerical solutions of the multistage SEIR model for different stage numbers $N = 5, 10, 30$, and 60. Here, $R_0 = 2.5$, $\varepsilon = 1/3$, and $\gamma = 0.2 \times N$. In the graphs, * represents the location of the maximum value of J and o represents the location of the inflection point of R. Graphs show that there is a time shift between these points.

comparisons of the classical models and multistage models. A MATLAB® ODE45 solver is used for all the numerical evaluations. In typical cases, initial conditions are chosen so that $S(0)$ is close to 1, whereas $E(0)$ and $I_i(0)$ are close to zero, and $R(0)$ is chosen so that the sum of all variables is 1, that is,

$$S(0) = 1 - 10^{-4},\ I_0(0) = 10^{-4},\ I_i(0) = 0,\ i = 1, \ldots, N,\ R(0) = 0; \tag{7}$$

and for the SEJR model,

$$S(0) = 1 - 10^{-4},\ E(0) = 5 \times 10^{-5},\ I_1(0) = 5 \times 10^{-5},\ I_i(0) = 0,\ i = 2, \ldots, N,\ R(0) = 0. \tag{8}$$

Numerical evaluations of the classical SIR and SEIR models are made for specific epidemic parameters, and the results are given in **Figure 4**. For the evaluations of the classical models, γ and ϵ are fixed as 1/5 and 1/3, respectively, whereas the value of R_0 is chosen to be 2.5 and 10, respectively. In all cases, the maximum of the infectious stage and the inflection point of the removed stage occur at the same point in time; hence, there is no time shift in these classical models.

Numerical evaluations of the multistage SIR model are repeated for various stage numbers. First, R_0 is set as 2.5, while the total number of stages, N, is given the values 5, 10, 30, and 60, and the corresponding graphs of the solutions are shown in **Figure 5**. In these evaluations, γ_i is chosen to be $0.2 \times N$, for $i = 0, 1, \ldots, N$. In **Figure 5**, the time shift between the maximum point of the curve, representing the sum of the infectious stages, $J(t)$ and the inflection point of $R(t)$ can clearly be seen. As predicted, the system given by **Eq. 5** confirms the delay and therefore seems adequate to explain the time shift between infectious and removed stages observed in Istanbul data [21]. Note that the time shifts for $N = 10, 30, 60$ seem to be equal but the one for $N = 5$ is smaller.

Similarly, numerical evaluations of the multistage *SEIR* model are obtained for various stage numbers. First, R_0 is set as five, while the stage number N is given the values 5, 10, 30, and 60, and the corresponding graphs of the solutions are shown in **Figure 6**. In these evaluations, ϵ and γ_i are chosen to be 1/3 and $0.2 \times N$, respectively, for $i = 1, \ldots, N$. **Figure 6** illustrates that just like it is seen in the SJR system, there exists a time shift between the maximum point of J and the inflection point of R in the SEJR model, with characteristics similar to the ones for the SIR model.

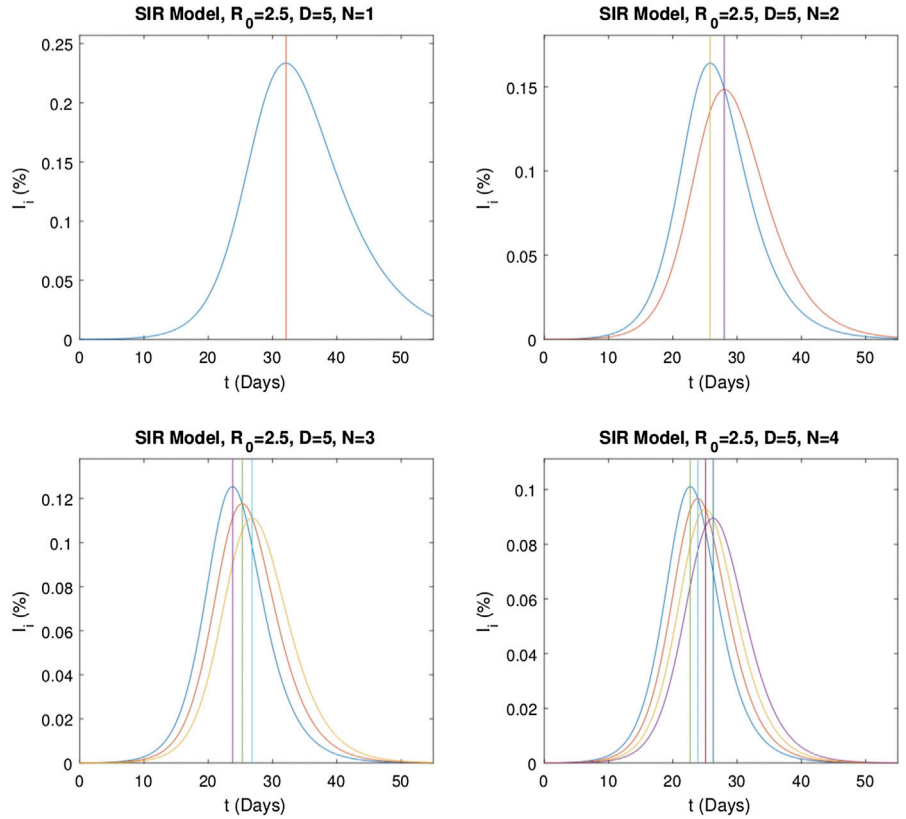


FIGURE 7 | Graphs of the independent infectious stages $I_0, \ldots I_N$ from numerical solutions of the multistage SIR model (5). Here, $R_0 = 2.5$ and $\gamma = 0.2 \times N$. In the graphs, vertical lines indicate the position of maximum points of the independent infectious stages. The distance between these vertical lines is approximately $1/\gamma$.

3. NUMERICAL RESULTS FOR MODELS WITH MULTIPLE INFECTIOUS STAGES

In this section, we investigate the dependency of the delay on the epidemic parameters by numerical evaluations of system **Eq. 5**. Graphs $I(_i(t)$ for $N = 1, 2, 3, 4$ for the SIR and SEIR models are given in **Figures 7** and **8**, respectively. To analyze the nature of the time shift for relatively large values of the stage number, the graphs of the solutions for the same epidemic parameters are obtained for N ranging from 2 to 16 in steps of two and the corresponding graphs are given in **Figure 9**. As it can be seen from **Figure 9**, the solution curves start to resemble as the stage number N increases. Finally, in **Figure 10**, we present the dependency of the time shift on the number of stages N, for $1 \leq N \leq 150$. Similar computations are repeated for the SEIR model and the results are presented in **Figures 11**, **12**.

For the SIR model, to investigate the effect of the basic reproduction number R_0 and the infectious period $1/\gamma$ on the infectious dynamics and the resulting delay, system (5) is solved with the initial conditions given by **Eq. 7** for some parameter values. To this end, the pair $(R_0, 1/\gamma)$ is chosen $(2.5, 5)$, $(5, 5)$, $(10, 5)$, $(2.5, 10)$, $(5, 10)$, $(10, 10)$, $(2.5, 20)$, $(5, 20)$, and $(10, 20)$, respectively, and the numerical evaluations for various infectious stages N are shown in **Figure 10**. It can be observed from this figure that the delay is almost half of the infectious period. This fact can also be seen in **Figure 9** where the time value of the maximum of J/N (normalized J) in time is located at the middle of time values of the maximum of the first infectious stage and the maximum of the last infectious stage. Comparison of panels of **Figure 10** shows that the change in the reproduction number R_0 for a fixed infection period has no effect on the delay. On the other hand, the delay depends on the infection period; in fact, it is approximately half of it for large N.

The same analysis is repeated for the multistage SEIR model. The system defined by **Eq. 6** is solved for the same epidemic parameters and initial conditions as above and with $\varepsilon = 1/3$, and the resulting graphs are given in **Figure 11**. As for the SIR model, the solution curves start to resemble for large N and the peak of $J(t)$ is located at the midpoint of the delay interval.

To illustrate the dependency of the delay on the system parameters of the *SEIR* model, the pairs (ε, γ) are chosen as $(1/3, 1/3)(1/3, (1/5), (1/5, 1/3), (1/5, 1/5)$, and variations of the delay for each of these cases are presented in **Figures 12A–D**, for $R0 = 5$ and $R0 = 7.5$. An analysis of these graphs yields that as N increases, the delay converges to a value. Moreover, it could easily be observed that the delay is independent of ϵ and R_0, but yet it is influenced by $1/\gamma$ (i.e., infectious period). Note that the delay for the multistage SEIR model is shorter than the delay observed in the multistage SIR model.

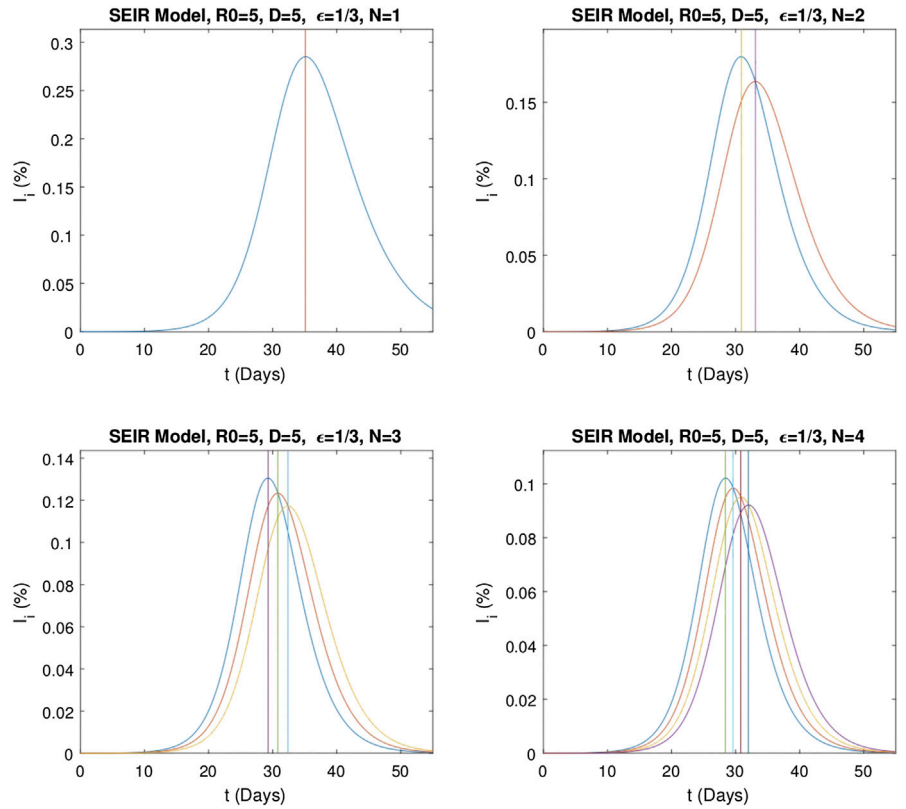


FIGURE 8 | Graphs of the independent infectious stages $I_1, \ldots I_N$ from numerical solutions of the multistage SEIR model (6). Here, $R_0 = 2.5$, $\varepsilon = 1/3$, and $\gamma = 0.2 \times N$. In the graphs, vertical lines indicate the position of maximum points of the independent infectious stages. The distance between these vertical lines is approximately $1/\gamma$.

4. ESTIMATION OF THE DELAY

In this section, we use a quadratic approximation to each $I_i(t)$ around the peak of $I_{i-1}(t)$ to show that within the validity of the quadratic approximation, the delays between successive peaks are $1/\gamma_i$.

Let t_i be the time where each I_i assumes its maximum and $1/\gamma_i$ be the corresponding infectious period for $i = 0, 1, \ldots N$. To determine the distance between the points t_i where each substage I_i assumes its maximum value, we use quadratic approximation of the Taylor series expansion of I_i at the point $t = t_{i-1}$ where,

$$I_i(t) = I_i(t_{i-1}) + I_i^{'}(t_{i-1})(t - t_{i-1}) + \frac{1}{2}I_i^{''}(t_{i-1})(t - t_{i-1})^2, \quad (9)$$

for $i \geq 1$. Differentiating

$$I_i^{'}(t) = I_i^{'}(t_{i-1}) + I_i^{''}(t_{i-1})(t - t_{i-1}) \quad (10)$$

and then substituting $t = t_i$ in **Eq. 10** and using the fact that $I_i^{'}(t_i) = 0$ since I_i reaches its maximum at t_i, one obtains

$$t_i - t_{i-1} = -\frac{I_i^{'}(t_{i-1})}{I_i^{''}(t_{i-1})}. \quad (11)$$

The multistage SIR model defined by the equations in **Eq. 5** suggests that for $i \geq 1$,

$$I_i^{'}(t) = \gamma_{i-1}I_{i-1}(t) - \gamma_i I_i(t). \quad (12)$$

Differentiating **Eq. 12** yields

$$I_i^{''}(t) = \gamma_{i-1}I_{i-1}^{'}(t) - \gamma_i I_i^{'}(t). \quad (13)$$

By considering the fact that $I_{i-1}^{'}(t_{i-1}) = 0$ since I_{i-1} assumes its maximum value at t_{i-1}, one gets the following by replacing $t = t_{i-1}$ in equation

$$I_{i-1}^{''}(t_{i-1}) = -\gamma_i I_i^{'}(t_{i-1}).$$

Substitution of the equation above in **Eq. 11** gives the approximate distance formula as follows:

$$t_i - t_{i-1} = \frac{1}{\gamma_i}. \quad (14)$$

Therefore, the distance between any t_i is $t_i - t_j = \sum_{k=j+1}^{i} \frac{1}{\gamma_k}$.

Results obtained by the numerical evaluations are compatible with **Eq. 14**. To observe the distance between the maximum points of the independent infectious stages, solutions of the multistage SIR model with respect to various infectious periods are chosen. In this

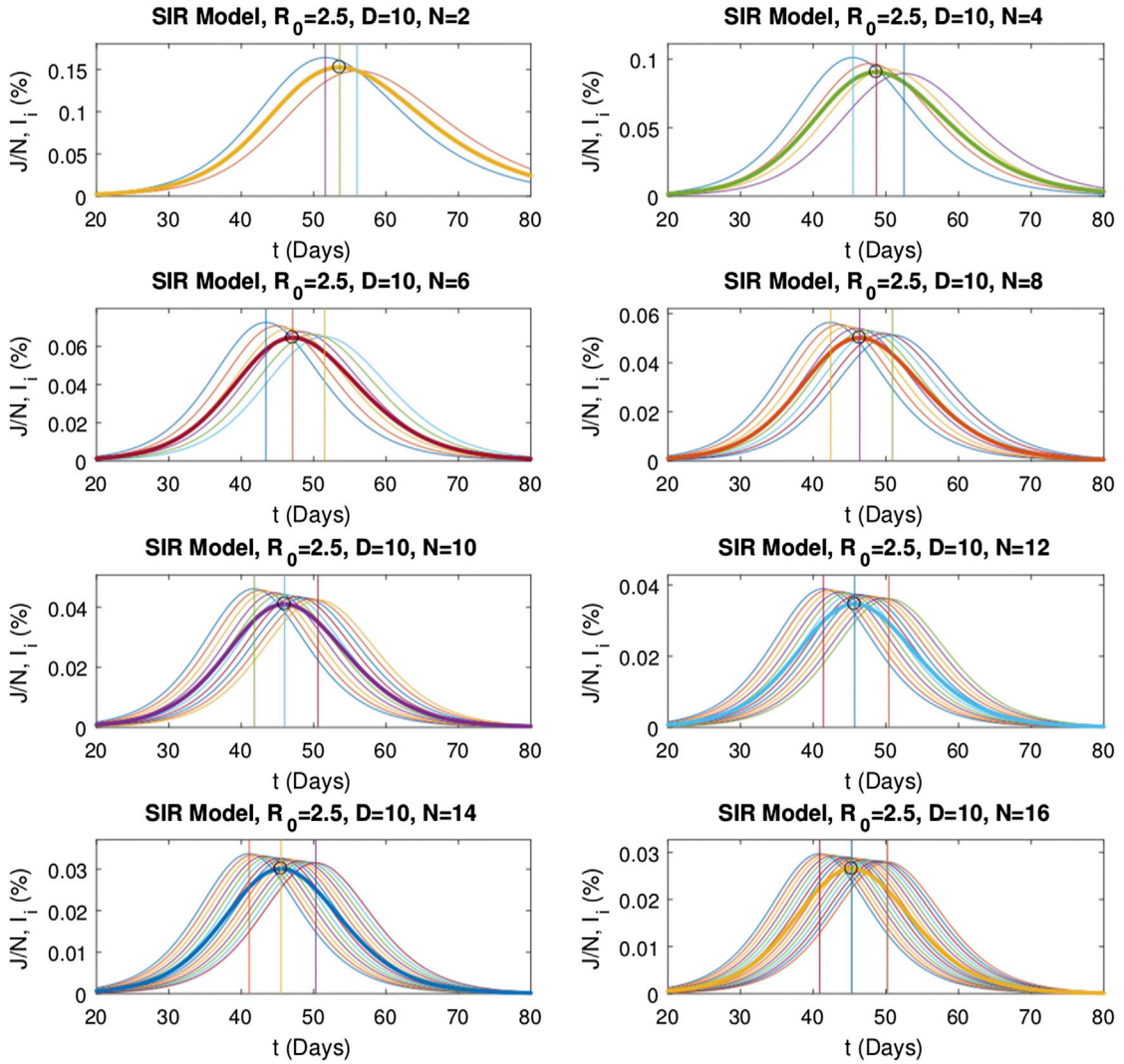


FIGURE 9 | Change of the position of maximum of *J*/*N* (normalized *J*) in time with respect to the number of stages *N* for the multistage SIR model. "o" indicates the position of the maxima of *J*/*N* (thick curves). Vertical lines represent the position of maximum points of the first stage, *J*/*N*, and the last stage, respectively.

respect, the basic reproduction numbers R_0 and γ_i (i.e., duration is 5) are set as 2.5 and $0.2 \times N$, respectively, and the related graphs of the solutions for various stage numbers ($N = 1, 2, 3, 4$) are given in **Figure 7**. Comparison of graphs in **Figure 7** reveals that the value of the difference of the points where successive stages reach their maximum is approximately $1/\gamma_i$.

It should be emphasized that **Eq. 14** is also valid for the multistage SEIR model. The distance between the maximum points of the independent infectious stages including the E stage is approximately $1/\gamma_i$. Since the proof is the same as in the SIR case, we do not repeat the derivation of **Eq. 14** again to avoid repetition. However, to observe the distance between the maximum points of the infectious stages numerically, the basic reproduction numbers R_0, ϵ, and γ_i (i.e., duration is 5) are set as 5, 1/3, and $0.2 \times N$. Then, the SEIR model is solved for $N = 1, 2, 3$, and 4, and the related graphs are given in **Figure 8**. As in the case of the SIR model, it is observed that the distance between the maxima is found to be approximately $1/\gamma_i$, too.

5. CONCLUSION

Epidemic data display a time shift between the peaks of infectious cases and fatalities. This time shift is not foreseen by the ordinary

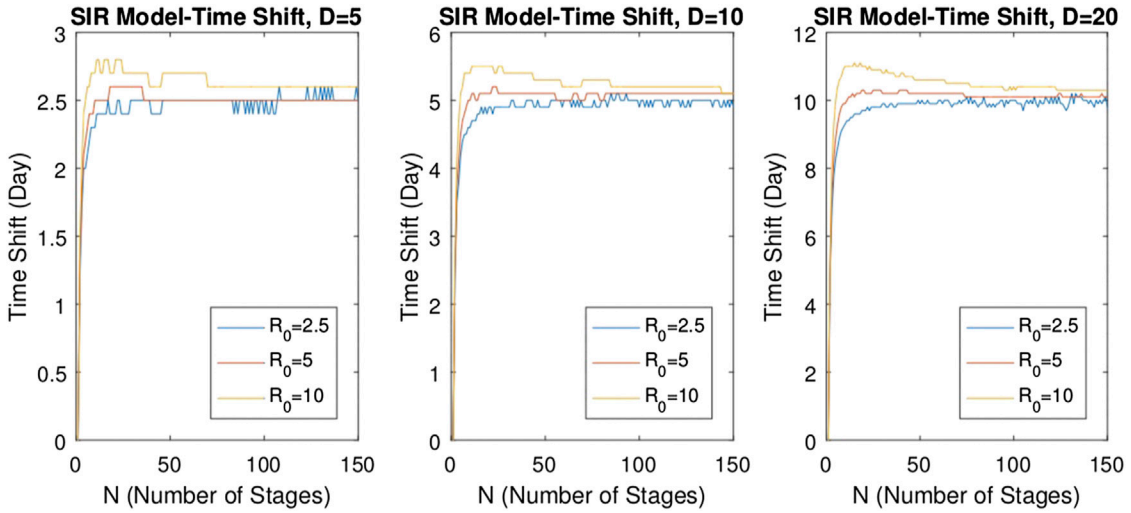


FIGURE 10 | Variations of time shift (delay) with respect to the number of stages N for R_0 = 2.5, 5, and 10; D = 5, 10, and 20 in the multistage SIR model. For all cases, the time shift becomes stable at a constant value after a critical stage number N.

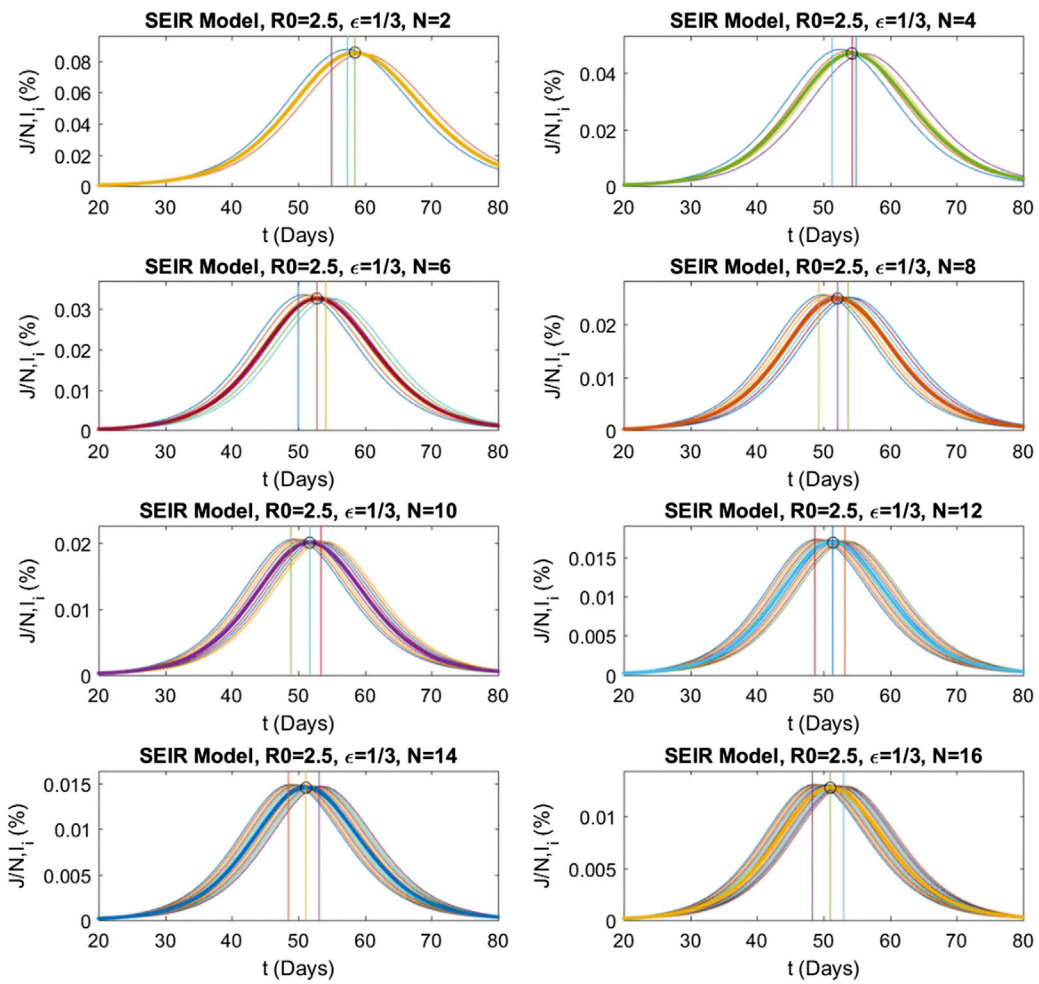


FIGURE 11 | Change of the position of maximum of J/N (normalized J) in time with respect to the number of stages N for the multistage SEIR model. "o" indicates the position of the maxima of J/N (thick curves). Vertical lines represent the position of maximum points of the first stage, J/N, and the last stage, respectively.

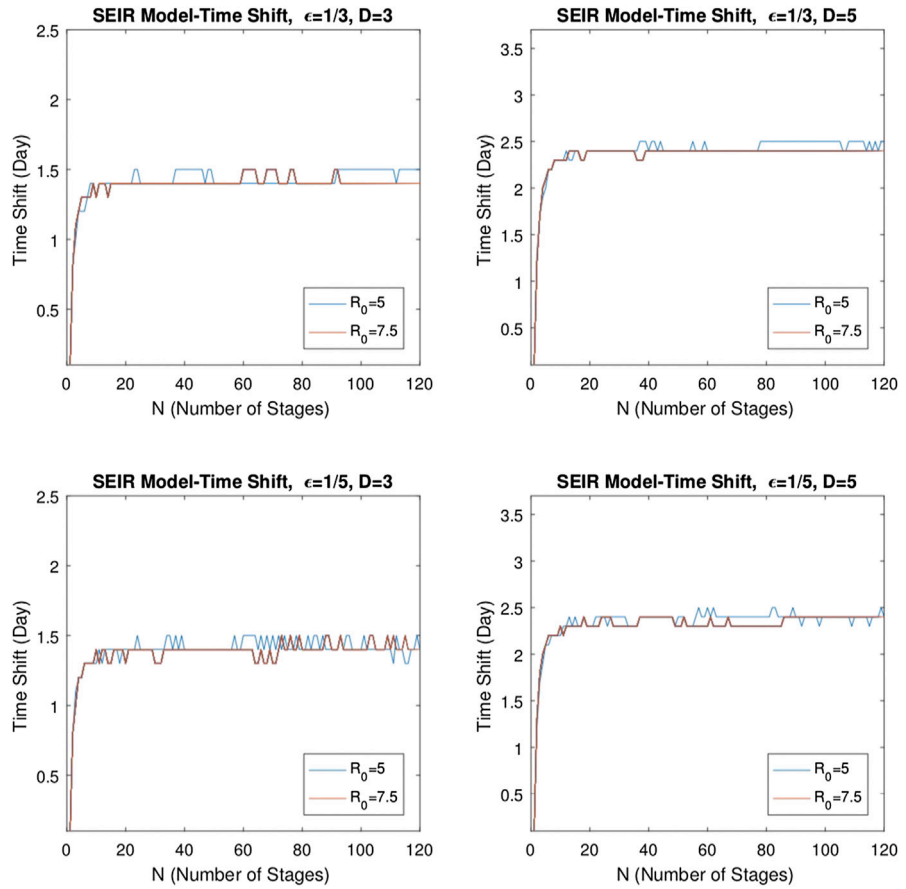


FIGURE 12 | Variation of time shift for R_0 = 5 and 7.5 in the multistage SEIR model. Here, ϵ = 1/3, 1/5, and D = 3, 5 are taken. For each case, delay becomes stable at a constant value after a critical stage number N.

differential equations for the SIR and SEIR models since in both of them, the derivative of $R(t)$ is proportional to $I(t)$. Nevertheless, this can be remediated by using gamma distributions instead of exponential distributions for the infectious period distribution (IPD) in the original SIR model of Refs. [4, 10] given in terms of integral equations [14], leading to a multistage model.

In this article, we propose a generalization of these multistage models by allowing the parameters to be unequal in different stages. We showed that within the validity of a quadratic approximation to $I_i(t)$, the distance between the points where each infectious stage reaches its maximum is approximately $1/\gamma_i$.

We solved the multistage models for a range of epidemic parameters, and we have seen that the solution curves reveal the time shift. While the delay varies for relatively small stage numbers, it is observed that the delay becomes nearly stable as the number of stages increases, and it is independent of the basic reproduction number. This fact supports the validity of the quadratic approximation and shows that the delay phenomenon observed in the infectious diseases defined by the epidemic models SIR and SEIR can be successfully explained by the multistage forms of these models.

In addition to a theoretical contribution, the existence and the estimation of the time shift between the progression of the infectious cases and the fatalities have a practical importance, in the sense that, in order to take timely actions, the severity of the epidemic should be measured in terms of the increase in the number of infectious cases.

Finally, we note that the importance of the effects of quarantine is realized during the COVID-19 pandemic. In the literature, there are two basic approaches, one of which adds a compartment to the model as in Ref. [22] and the other adds a function $\beta/(\gamma + Q(t))$, where $Q(t)$ is a time-dependent rate as in Ref. [23] and has the effect of shortening the duration of the infection period. This may explain the

relatively shorter delays in China (compared with South Korea) where strict quarantine was in effect.

DATA AVAILABILITY STATEMENT

The datasets analysed for this study can be found in the webpages: http://www.worldometers.info/coronavirus/ (last access: 05 November 2020), http://epikhas.khas.edu.tr/ (last access: 05 November 2020) and https://www.who.int/csr/sars/country/en/ (last access: 05 November 2020).

AUTHOR CONTRIBUTIONS

APD, AHB and SA performed theoretical results; AD, APD, AHB and SA performed computations; AD performed literature survey and APD, AD, AHB and SA wrote the paper.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at arxiv.org, arXiv:2004.13178.

REFERENCES

1. Bernoulli D (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. *Mem. Math. Phys. Acad. Roy. Sci., Paris*:1–45.
2. Hamer WH. *Epidemic disease in England: the evidence of variability and of persistency of type* London: Bedford Press (1906) 72 p.
3. Ross R. *The prevention of malaria* New York: J. Murray (1910) 669 p.
4. Kermack WO, McKendrick AG (1927). A contribution to the mathematical theory of epidemics. *Proc R Soc Lond–Ser A Contain Pap a Math Phys Character* 115:700–21.
5. Miller JC. A note on a paper by erik volz: sir dynamics in random networks. *J Math Biol* (2011) 62:349–58. doi:10.1007/s00285-010-0337-9
6. Miller JC, Volz EM. Incorporating disease and population structure into models of sir disease in contact networks. *PLoS One* (2013) 8:e69162. doi:10.1371/journal.pone.0069162
7. Bilge AH, Pekcan Ö, Gürol MV. Application of epidemic models to phase transitions *Phase Transitions* (2012) 85:1009–17. doi:10.1080/01411594.2012.672648
8. Bilge AH, Pekcan O, Kara S, Ogrenci AS. Epidemic models for phase transitions: application to a physical gel. *Phase Transitions* (2017) 90:905–13. doi:10.1080/01411594.2017.1286487
9. Bilge AH, Ogrenci AS, Pekcan O. Mathematical models for phase transitions in biogels. *Mod Phys Lett B* (2019) 33:1950111. doi:10.1142/s0217984919501112
10. Hethcote HW, van den Driessche P. An sis epidemic model with variable population size and a delay. *J Math Biol* (1995) 34:177–94. doi:10.1007/bf00178772
11. Hethcote HW, van den Driessche P. Two sis epidemiologic models with delays. *J Math Biol* (2000) 40:3–26. doi:10.1007/s002850050003
12. Cooke KL, Van Den Driessche P. Analysis of an seirs epidemic model with two delays. *J Math Biol* (1996) 35:240–60. doi:10.1007/s002850050051
13. Iggidr A, Mbang J, Sallet G, Tewa J-J. Multi-compartment models. *Discrete Contin. Dyn. Syst* (2007) 506–19.
14. Lloyd AL. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* (2001) 60:59–71. doi:10.1006/tpbi.2001.1525
15. Bame N, Bowong S, Mbang J, Sallet G, Tewa J-J. Global stability analysis for SEIS models with n latent classes. *Math Biosci Eng* (2008) 5:20–33. doi:10.3934/mbe.2008.5.20
16. Guo H, Li MY, Shuai Z. Global dynamics of a general class of multistage models for infectious diseases. *SIAM J Appl Math* (2012) 72:261–79. doi:10.1137/110827028
17. Sherborne N, Blyuss KB, Kiss IZ. Dynamics of multi-stage infections on networks. *Bull Math Biol* (2015) 77:1909–33. doi:10.1007/s11538-015-0109-1
18. Brauer F. A simple model for behaviour change in epidemics. *BMC Public Health* (2011) 11:1–5. doi:10.1186/1471-2458-11-s1-s3
19. Andreasen V, Viboud C, Simonsen L. Epidemiologic characterization of the 1918 influenza pandemic summer wave in copenhagen: implications for pandemic control strategies. *J Infect Dis* (2008) 197:270–8. doi:10.1086/524065
20. Ahmetolan S, Bilge AH, Demirci A, Peker-Dobie A, Ergonul O. What can we estimate from fatality and infectious case data using the susceptible-infected-removed (sir) model? a case study of covid-19 pandemic. *Front Med* (2020) 7:570. doi:10.3389/fmed.2020.556366
21. Bilge AH, Samanlioglu F, Ergonul O. On the uniqueness of epidemic models fitting a normalized curve of removed individuals. *J Math Biol* (2015) 71:767–94. doi:10.1007/s00285-014-0838-z
22. Piqueira JRC, Batistela CM. Considering quarantine in the sira malware propagation model. *Math Probl Eng* (2019) 2019:1–8. doi:10.1155/2019/6467104
23. Dandekar R, Barbastathis G (2020). Quantifying the effect of quarantine control in covid-19 infectious spread using machine learning. medRxiv preprint.

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 30 November 2020
doi: 10.3389/fphy.2020.586180

Modeling Growth, Containment and Decay of the COVID-19 Epidemic in Italy

Francesco Capuano *

Department of Industrial Engineering, Università degli Studi di Napoli "Federico II", Napoli, Italy

A careful inspection of the cumulative curve of confirmed COVID-19 infections in Italy and in other hard-hit countries reveals three distinct phases: i) an initial exponential growth (unconstrained phase), ii) an algebraic, power-law growth (containment phase), and iii) a relatively slow decay. We propose a parsimonious compartment model based on a time-dependent rate of depletion of the susceptible population that captures all such phases for a plausible range of model parameters. The results suggest an intimate interplay between the growth behavior, the timing and implementation of containment strategies, and the subsequent saturation of the outbreak.

Keywords: COVID-19, power-law growth, compartment model, containment strategies, unreported cases

OPEN ACCESS

Edited by:
Zhen Wang,
Hong Kong Baptist University, Hong Kong

Reviewed by:
Matúš Medo,
University of Electronic Science and Technology of China, China
Gui-Quan Sun,
North University of China, China

***Correspondence:**
Francesco Capuano
francesco.capuano@unina.it

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *22 July 2020*
Accepted: *30 October 2020*
Published: *30 November 2020*

Citation:
Capuano F (2020) Modeling Growth, Containment and Decay of the COVID-19 Epidemic in Italy. Front. Phys. 8:586180. doi: 10.3389/fphy.2020.586180

1. INTRODUCTION

The coronavirus disease 2019 (COVID-19) is an infectious respiratory illness caused by the newly discovered virus strain SARS-CoV-2 [1–3]. The on-going COVID-19 outbreak started in Wuhan (Hubei, China) in late December 2019 and spread quickly to all Chinese provinces and to several countries, prompting the World Health Organization (WHO) to declare a pandemic on March 11, 2020 [4]. The high human-to-human transmission rate and clinical severity of the infection have raised enormous concern on an international scale, with governments worldwide taking drastic and unprecedented measures to contain the disease spread. Despite the efforts, 34,804,348 people have been infected world-wide, and 1,030,738 have died as of October 4, 2020 [5].

The global emergency has issued a massive call-to-arms for researchers from several disciplines. Among the various fields of study, epidemic modeling is of utmost importance to inform policy makers about the required sanitary system capacity, to guide the design of containment strategies, and to assess their effectiveness in real time [6]. Humongous efforts are currently on-going to develop accurate mathematical models of the disease spread, ranging from *top-down* (i.e., static curve fitting) approaches [7, 8] to dynamic compartment models of various degree of complexity [9, 10]. The latter class is especially appealing, as it allows a rather straightforward incorporation and testing of physics-based hypotheses. Nonetheless, the modeling process is hindered by several factors, including incomplete knowledge of the disease, as well as the challenging incorporation of containment strategies and unreported cases [11]. Retrospective analysis of data from countries that have already overcome a turning point in the COVID-19 epidemic is highly valuable to inspire and calibrate novel mathematical models with predictive capabilities.

In this paper, we aimed to derive a physics-based dynamic compartment model able to adhere as much as possible to the qualitative and quantitative nature of the observed data. To this purpose we take Italy, one of the early hard-hit countries that has overcome a first epidemic wave, as a primary modeling source. We start by preliminarily analyzing the epidemic data in **Section 2**, drawing important qualitative observations about the scaling laws of the growth and decay of the outbreak.

Guided by these, in **Section 3** we propose a compartment model that incorporates key epidemiological features of the disease and is able to reproduce the observed trends. In **Section 4** we report and discuss the results of the proposed model, and compare them to official data from Italy. Concluding remarks are given in **Section 5**.

2. DATA ANALYSIS OF THE COVID-19 OUTBREAK IN ITALY

The outbreak in Italy started on February 21, 2020, when a local cluster of COVID-19 cases was identified in Codogno (Lodi). The Italian government immediately ordered lockdown for 11 hard-hit Northern-Italy municipalities; however, the rapid and seemingly uncontrolled growth of infections in the following days led to increasingly tight regulations on the entire national territory, including closure of schools (March 4) and, shortly after, quarantine (March 9). The measures proved to be effective and, following the decline of the number of confirmed cases, the containment measures were gradually released since May 4. As of early July, the epidemic was mostly under control, with a reproductive number < 1 in the majority of Italian regions [12].

It is instructive to preliminarily analyze the cumulative curve of laboratory-confirmed infections $C(t_i)$, as well as the curve of new daily cases, $\Delta C(t_i) = C(t_i) - C(t_{i-1})$. It is worth mentioning that confirmed infected individuals are immediately quarantined (either home-isolated or hospitalized). All data were downloaded from the publicly available database provided by the Johns Hopkins University Center for Systems Science and Engineering (JHU-CSSE) [13].

Figure 1A shows a logarithmic plot of confirmed COVID-19 cases in Italy in a restricted time interval (from February 22 to April 30), while **Figure 1B** reports new daily cases in a linear plot for the same time range. Three distinct phases can be inferred from static data analysis:

(1) an initial exponential growth, which is typical of unconstrained outbreaks, with best-fit exponent $\omega \approx 0.27$. The exponential growth phase lasts approximately until March 2, before undergoing a transition to a different scaling behavior;
(2) a phase of algebraic growth, starting approximately on March 6 and ending on March 18. The power-law scaling t^{α} is best revealed by examining the slope in the logarithmic plane

$$\alpha(t_i) = \frac{\Delta \log C(t_i)}{\Delta \log t_i}, \tag{1}$$

as shown in the inset of **Figure 1A**. The plot shows a distinct plateau within the mentioned time window with $\bar{\alpha} \approx 3.1$, which guided the definition of the starting/ending dates of the algebraic phase. The starting day of the power-law phase slightly precedes the national quarantine (ordered on March 9), presumably as a consequence of previous containment measures, such as lockdown of Northern areas and school closure;
(3) a relatively *slow* decay. The decay phase is deliberately assumed to start from the turning point of the epidemic, that occurred on March 23, after roughly 5 days of "transition" from the previous algebraic phase. The slow decay is evident from **Figure 1B**, which shows a markedly asymmetric bell-shaped curve with respect to the turning point.

Evidence of algebraic growth has been observed and described elsewhere for other countries, including China [14] as well as other European countries [15], with exponents generally ranging from 2 to 4. From a fundamental perspective, this behavior has been attributed to structural changes in the topology of the population network that supports the epidemic spreading. Several authors [15, 16] have resorted to the *small-world* hypothesis [17] to justify the algebraic growth. More in general, spatial networks where short links are favored over long-range connections have been shown to produce power-law exponents in close similarity with the observed COVID-19 dynamics [18]. Similarly, the *asymmetric* decay has been put into some (empirical) connection with the simultaneous presence of a persistent phase of algebraic growth, in contrast with other countries where the epidemic has been characterized by a rather symmetric rise-and-fall behavior [19]. In this regard, graphs with a power-law degree distribution are known to produce outbreaks characterized by a polynomial growth followed by an exponential decay [20]. A functional form of this type was successfully used to fit COVID-19 data from over 100 countries [21], supporting the evidence that the nature of the underlying network is key to the infectious dynamics, and that growth and decay of the epidemic are indeed intimately connected. Network effects and their relationship with COVID-19 epidemic trends are further discussed in Ref. 22.

The reported observations have important consequences on traditional modeling approaches. From the perspective of classical population growth models, standard logistic models appear to be inappropriate, as they provide symmetric S-shaped curves for the cumulative number of infections; in contrast, the generalized Richard's model (GRM) has been shown to provide a rather accurate description of the epidemic in Chinese provinces [8]. While the GRM can provide sub-exponential growth and asymmetric decline [23], it is known to lack clear epidemiological significance [24]. Compartment models are richer in terms of physics and allow addition of several degrees of freedom. The basic version of the celebrated Susceptible–Infectious–Removed (SIR) framework [25] can be easily transformed, under mild assumptions, into a logistic model [26], therefore suffering from the same above-mentioned limitations. Several refined SIR-like approaches have been proposed, particularly aiming at quantifying undetected cases [27] and at modeling the effect of containment policies [14, 28]. These have been modeled in most cases by means of a piecewise constant transmission rate (or, alternatively, by changing the local reproduction number) [29], or possibly by incorporating the effects of individual reaction [30].

Reconciling algebraic growth with mean-field models is not straightforward. Recently Ref. 21, observed that a general solution

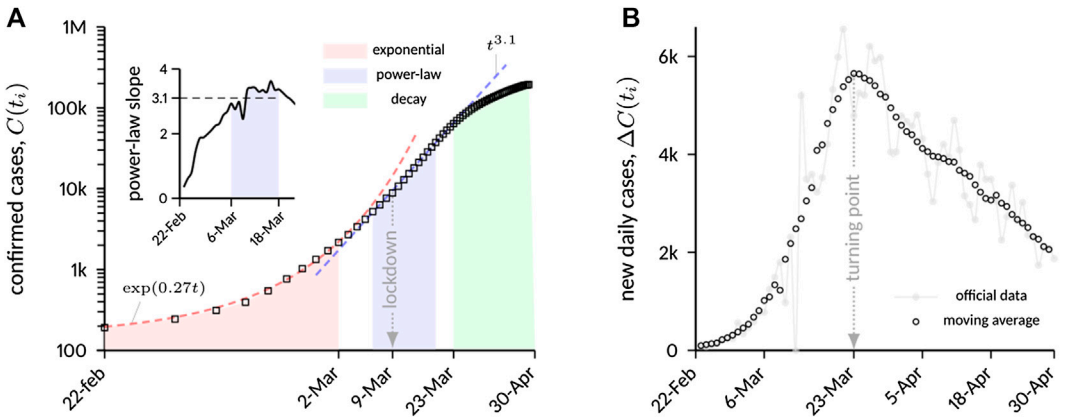


FIGURE 1 | (A) Cumulative number of laboratory-confirmed COVID-19 infections in Italy. A moving average with a time window of 7 days was applied to the data to filter weekly fluctuations in case reporting. After an initial exponential growth, on March 6 the increase starts to follow a scaling law t^{α} with average $\overline{\alpha} \approx 3.1$. **(B)** Number of new daily cases of laboratory-confirmed COVID-19 infections in Italy. The *turning point* is highlighted. Gray circles represent official data, highlighting fluctuations in data reporting, while black circles were obtained with a 7-days moving average.

of the type $I(t) \propto t^{\alpha}$, where I is the infectious compartment, is compatible with a SIR model provided that the corresponding basic reproduction number decays as t^{-1}. On the other hand, Maier and Brockmann [14] (M&B hereinafter) have found, and proved analytically, that a region of algebraic growth can be captured by a slightly modified SIR-like model that includes a (constant) depletion rate of susceptible individuals. The question arises whether such a modeling framework is able to capture a markedly asymmetric decay dynamics such as the one occurring in Italy.

We conjecture that the three phases reported above are inherently coupled by endogenous epidemic processes, and thus they can be accurately modeled into a unique framework that relies exclusively on physics-based assumptions and policy-driven changes of the model parameters. Inspired by M&B, we aim to propose a compartment model that is able to: 1) capture the power-law growth; 2) generate an asymmetric decay; 3) incorporate unreported cases.

3. PROPOSED COMPARTMENT MODEL

We propose a compact compartment model of SEIQR-type (Susceptible–Exposed–Infectious–Quarantined–Removed), whose schematic is shown in **Figure 2**. Although many more compartments can be added to increase the level of detail and supposedly the realism of the model, we rather opted for a parsimonious framework, while focusing primarily on the reproduction of the observed scaling laws. We note explicitly that the inclusion of the Q-compartment is highly warranted for countries characterized by strong government interventions, since identified infectious individuals are typically quarantined and therefore they are no longer able to spread the disease. On the contrary, the necessity of taking the incubation period into account via the E-class is currently debated, mainly due to incomplete understanding of the relationship between incubation and latency periods for SARS-CoV-2 infection.

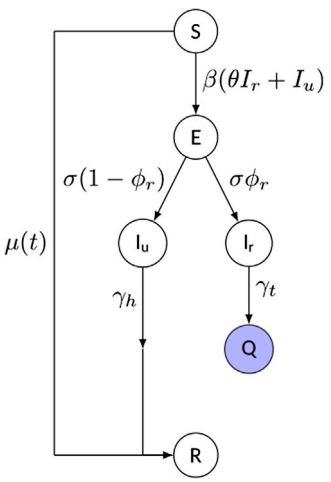


FIGURE 2 | Schematic of the model and of its parameters (see text for details). The "quarantined" variable $Q(t)$ coincides, by construction, with the time-series of official laboratory-confirmed infections $C(t_i)$, when evaluated at the same sample points; $Q(t)$ is thus the only *known* variable in the context of a data-driven calibration of the model.

However, a certain time lag between infection and infectiousness (in either clinical or subclinical cases) has often been observed [2, 31]; therefore, we chose to include the exposed class with an average incubation period γ^{-1}.

Additional key features of the model include:

- a time-dependent rate of depletion of the susceptible population, $\mu(t)$. Following M&B, we conjecture that starting (roughly) from the implementation of lockdown measures, and due to increased perception of the disease, a fraction of the S-class becomes unsusceptible due to self-protection, isolation, and adoption of preventive measures (e.g., wearing masks, hand-washing, etc.). However, and

importantly, preliminary tests showed that a constant rate of depletion leads to strong under-prediction of the number of infections during the decay phase. Here we propose a time-dependent, decaying rate of depletion of susceptible individuals,

$$\mu(t) = \begin{cases} 0, & t < t^*, \\ \mu_0 e^{-(t-t^*)\tau}, & t \geq t^*, \end{cases} \quad (2)$$

where μ_0 is the initial rate of the depletion, t^* is the start of the containment measures and τ is a timescale parameter. Such a modeling paradigm allows for a non-zero asymptotic residual of susceptible individuals, and is an essential component for capturing the strongly asymmetric behavior observed for the curve of new daily cases;

- an explicit distinction between reported and unreported cases. The infectious class has been split into reported (I_r) and unreported (I_u) individuals, modulated by a reporting rate ϕ_r. Since the policy in Italy consists into testing only symptomatic individuals, it is reasonable to assume that reported and unreported cases roughly coincide with asymptomatic (i.e., subclinical) and symptomatic (i.e., seeking medical attention) cases. Therefore, individuals belonging to the I_r class are infectious as long as they are tested and quarantined at a rate γ_t, and thus transferred to the Q-class. By construction, the variable $Q(t)$ coincides with the time-series of laboratory-confirmed infections $C(t_i)$. On the other hand, for individuals belonging to the Iu class, the disease goes unrecognized and they continue mixing with the susceptible population, before spontaneously recovering at a rate γ_h. We assume a bilinear transmission rate for reported and unreported cases: $\theta\beta$ for symptomatic individuals, and β for subclinical ones, with $\theta \in [0, 1]$. Although counter-intuitive, this choice was driven by the fact that people with symptoms that are severe enough to require medical attention are likely to reduce social contacts, therefore diminishing their transmission rate with respect to asymptomatic individuals (despite having, supposedly, a higher viral load and thus contagious potential).

Note that, for the sake of simplicity, we do not explicitly model the subsequent recovery/death of quarantined individuals, nor we distinguish between hospitalized and home-isolated patients; however, information in this regard can be inferred a-posterior once the temporal dynamics of $Q(t)$ and $I(t)$ is known. It is also worth to underline that the removed (R) class contains both recovered (from the unreported compartment) and protected individuals. Possible re-infection is not modeled as it is currently considered to be unlikely [32].

In summary, the temporal dynamics of the compartments is governed by the following system of ordinary differential equations:

$$\dot{S}(t) = -\beta s(\theta I_r + I_u) - \mu(t)S, \quad (3a)$$

$$\dot{E}(t) = \beta s(\theta I_r + I_u) - \sigma E, \quad (3b)$$

$$\dot{I}_u(t) = \sigma(1 - \phi_r)E - \gamma_h I_u, \quad (3c)$$

$$\dot{I}_r(t) = \sigma\phi_r E - \gamma_t I_r, \quad (3d)$$

$$\dot{Q}(t) = \gamma_t I_r, \quad (3e)$$

$$\dot{R}(t) = \gamma_h I_u + \mu(t)S. \quad (3f)$$

where $s = S/N$ and N is the total (constant) population. The time dependency of the main variables has been omitted at the right-hand sides for the sake of clarity. All the parameters appearing in the model are constant and positive real numbers, with $\theta, \phi_r \in [0, 1]$, except for $\mu(t)$ that has explicit dependence on time. The basic reproduction number R_0 of this model can be inferred by a next-generation matrix approach; following [33]; the matrices F and V can be defined as

$$F = \begin{bmatrix} 0 & \beta & \beta\theta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} \sigma & 0 & 0 \\ \sigma(\phi_r - 1) & \gamma_h & 0 \\ -\sigma\phi_r & 0 & \gamma_t \end{bmatrix},$$

and R_0 is the spectral radius of FV^{-1}, yielding

$$R_0 = \frac{\beta\theta\phi_r}{\gamma_t} + \frac{\beta(1 - \phi_r)}{\gamma_h}. \quad (4)$$

The model defined by **Eq. 3** has eight free parameters; some of them can be set *a-priori* based on epidemiological, clinical or policy-related evidence. The incubation time σ^{-1} has been found by different clinical groups to be around 5 days [2, 31], thereby yielding $\sigma = 1/5$, a value also used in similar modeling studies [34, 35]. The rate of testing γ_t is country-specific; in Italy, the "Istituto Superiore di Sanità" (ISS) has reported an average time from symptoms onset to diagnosis (via pharyngeal swab) of 5 days, within the time range considered in this work. The infection duration in subclinical cases without medical treatment has been estimated to be in the range of 5–10 days [36]; after preliminary sensitivity tests, we set $\gamma_h = 1/5$. The starting time of the susceptible population removal t^* is expected to coincide roughly with the starting time of national lockdown; we left, however, t^* as a free parameter for manual calibration.

4. RESULTS AND DISCUSSION

Numerical simulations of the outbreak in Italy were obtained upon integration of **Eq. 3** for the time range from February 24 to June 30, 2020. **Table 1** summarizes the values of the entire set of model parameters, including those assigned *a-priori* and the ones inferred from the data. After several preliminary tests aimed to circumvent overfitting issues, we chose to optimize the parameter space containing $\{\beta, \theta, \mu_0, \tau, E(0)\}$, while ϕ_r and t^* were left free to be manually varied in a parametric way until an optimal fit was found. The optimization procedure was carried out using an interior-point method implemented in MATLAB (Mathworks,

Natick, United States); the optimization target was based on a blend of the cumulative number of confirmed cases and the daily number of new cases, that were weighted 10 and 90%, respectively. The infectious compartments were initialized as follows: $I_r(0) = \phi_r E(0)$, $I_u(0) = (1-\phi_r)E(0)$, while the other initial conditions are either known from data or can be found by subtraction, with $N = 60 \times 10^6$.

Main results are shown in **Figure 3**. **Figure 3A** reports the temporal evolution of $Q(t)$, as compared with the time-series of laboratory-confirmed infections $C(t_i)$. The agreement is very good for all the phases of the outbreak. Interestingly, the best-fit value for t^* was March 9, i.e., the start of the national quarantine, which indirectly proves the robustness of the proposed framework in reproducing the underlying epidemic mechanisms at play. The inset of **Figure 3A** reports the power-law slope provided by the model, which is in very good agreement with the observed one, proving that the model is capable of reproducing the peculiar algebraic growth phase of the COVID-19 pandemic reported in several countries.

Of note, the model was able to capture the decay phase of the epidemic correctly without any further change of the parameters. This circumstance suggests a strong link between the decay dynamics and the timing and implementation of the containment strategies. It is conjectured that an incomplete depletion of the susceptible population during the lockdown (provided by the exponential decay of $\mu(t)$) naturally sets the observed *slow* decay. It can be further hypothesized in light of the results that the behavioral changes established during the containment phase persist throughout the subsequent phases of the epidemic. Since lower values of the timescale τ over which the susceptible removal occurs yield a faster decay, this parameter be re-interpreted as an "effectiveness index" of the implementation of containment policies.

We also tested a variant of the model that more closely resembles the original M&B approach, with a constant depletion rate of the susceptible population $\mu = \mu_0$. Upon re-running the optimization procedure with $\tau = 0$, we obtained similar values as those reported in **Table 1**, but it was not possible to match the decay phase correctly. This variant of the model provides a reasonable agreement only up to the turning point of the epidemic (with the peak value of new daily cases highly over-estimated), while strongly under-estimating the number of infections during the decay phase, as clearly seen in **Figure 3B**. An additional steep change of the model parameters would be necessary, although this would be hardly justifiable in terms of physical assumptions or containment policies.

In light of the above mentioned results and observations, it is instructive to directly quantify the effects of $\mu(t)$ on the epidemic dynamics. To this aim, it is worth to preliminarily observe that, under a disease-free linearization (i.e., assuming that containment dominates over infection) the normalized number of susceptible for $t \gg t^*$ is not null (as in the case $\tau = 0$) but converges to

$$s_{\text{limit}} = \frac{S_0}{N} e^{-\mu_0/\tau}, \tag{5}$$

TABLE 1 | Optimal values of the estimated model parameters for the simulation of the outbreak in Italy.

Parameter	Description	Value
β	Transmission rate	0.95
θ	Transmission factor for reported (symptomatic) cases	0.58
t^*	Starting time of depletion of susceptible	9-Mar
μ_0	Initial rate of depletion of susceptible	0.16
τ	Timescale factor for the rate of depletion of susceptible	0.1
ϕ_r	Fraction of reported (symptomatic) individuals	0.6
σ	Rate at which exposed individuals become infectious	0.2
γ_t	Rate of testing for symptomatic individuals	0.2
γ_h	Rate of recovery for unreported (subclinical) individuals	0.2
$E(0)$	Initial number of exposed individuals	2,000

showing that τ is explicitly responsible for the incomplete removal. Although it is possible, in principle, to derive a closed-form expression for $Q(t)$ under some mild simplifying hypotheses, the analytic treatment becomes very involved and it is cumbersome to extract immediate information. The reader is referred to Ref. 14 for an analytic demonstration of how the depletion of susceptible is responsible for the power-law growth, while the effects of μ_0 and τ on the decay are best revealed through parametric analysis. **Figure 4** shows the effects of the containment-related parameters on ΔQ, with the other modeling parameters being held equal to the optimal solution reported in **Table 1**. **Figure 4A** shows that lower values of μ_0 (i.e., weaker depletion) tend to shift the curve toward higher values of ΔQ and delay the turning point, while the overall shape is roughly preserved. On the other hand, **Figure 4B** demonstrates that values of $\tau > 0$ contribute to asymmetrize the curve of new daily cases, effectively slowing the decay while preserving the growth dynamics. Since τ is directly related to the residual number of susceptible individuals during the course of the epidemic, see **Eq. 6**, these findings reinforce our claim that the observed asymmetry is directly related to a (spatially-averaged) incomplete depletion of the susceptible population during the containment phase.

With regards to the functional form of the time-varying depletion rate $\mu(t)$, the exponential decay was chosen as the most *natural* one with a clear and intuitive biological significance. However, we also explored whether alternative expressions for $\mu(t)$ are able to reproduce the essential characteristics of the COVID-19 epidemic. To this end, we consider a simplified version of the proposed compartment model, wherein the E-class is removed and $\phi_r = \theta = 1$, i.e., all infected individuals are identified and quarantined:

$$\begin{aligned} \dot{S}(t) &= -\beta s I - \mu(t) S, \\ \dot{I}(t) &= \beta s I - \gamma I, \\ \dot{Q}(t) &= \gamma I. \end{aligned} \tag{7}$$

In such case, assuming that in a certain time interval $Q(t) \approx t^{\alpha}$, and that containment dominates over infection, simple analytic manipulations yield

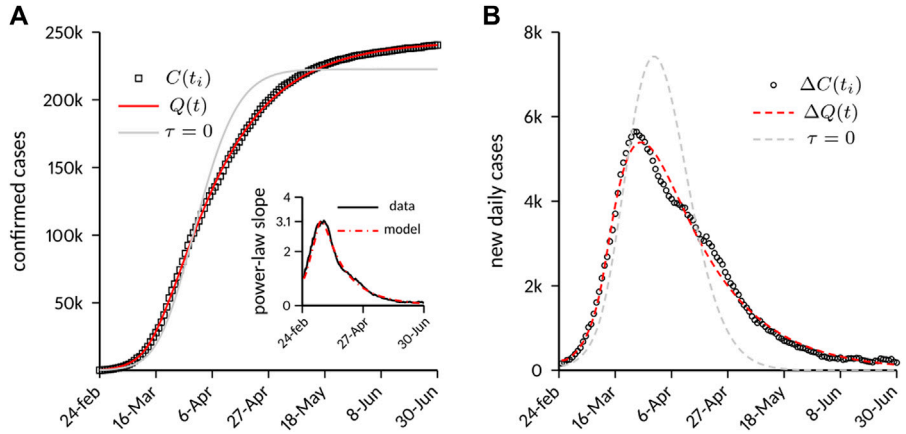


FIGURE 3 | Case numbers in Italy compared to model predictions. **(A)** Cumulative number of laboratory-confirmed infections. The model predictions are shown both with the full model and with a variant obtained by setting $\tau = 0$ (constant depletion rate of the susceptible population). The inset reports the local power-law slope for the moving-averaged official data and the full-model prediction. **(B)** Number of new daily cases.

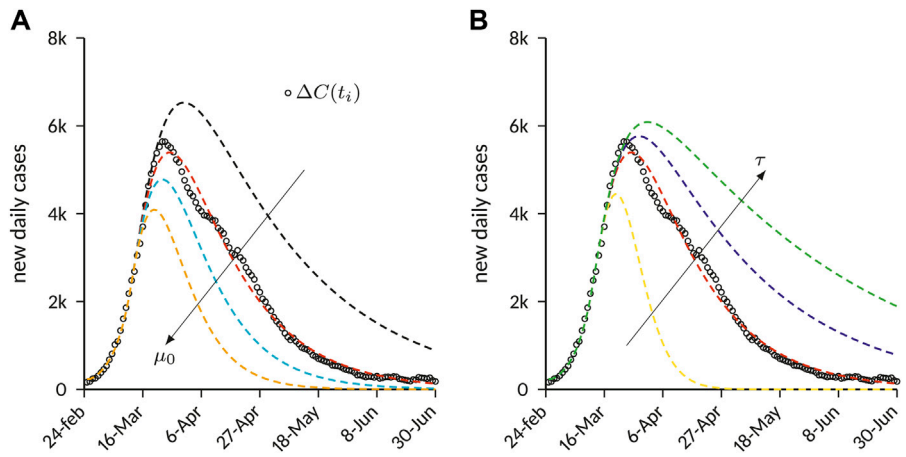


FIGURE 4 | Effects of containment on the number of new daily cases predicted by the SEIQR compartment model. The results are obtained with the parameters provided in **Table 1**, and by varying the parameters μ_0 and τ. The red curve corresponds to the optimal solution. Moving-averaged official data are also reported. **(A)** Effect of μ_0 for values in the set $[0.14, 0.16, 0.18, 0.22]$. **(B)** Effect of τ for values in the set $[0.0, 0.1, 0.12, 0.13]$.

$$\mu(t) \approx \frac{c_1}{t + c_2 t^2}, \tag{8}$$

where c_1 and c_2 are constants. We tested the behavior of the complete model by plugging **Eq. 8** into **Eq. 3**; we were able to generate outbreaks (not shown here) in good agreement with the official data and for similar values of the model parameters as those reported in **Table 1**. While this circumstance promotes the correctness of the underlying modeling hypotheses, additional tests and analyses are required to determine which functional form is most suitable for $\mu(t)$, and whether other decay laws (e.g., Gaussian) can be instead ruled out to describe the containment process.

The role of subclinical infected with regards to the spread of the disease is one of the most debated epidemiological and biomedical issues in the context of the COVID-19 outbreak [37]. Accurate studies in small-scale, "laboratory-like" contexts with blanket testing showed that a substantial fraction of SARS-CoV-2-positive individuals can display very mild symptoms or even remain completely asymptomatic throughout the course of the infection, while having a viral load comparable to that of symptomatic patients [38–40]. Here we found a very good fit for values of the reporting rate in the range $\phi_r \in [0.55, 0.65]$, yielding a fraction of unreported (subclinical) cases $(1 - \phi_r)$ in line with previous modeling and experimental estimates [27, 38].

It is worth to remark, in general, that we were able to generate outbreaks in good agreement with official data for a certain range and combination of the model parameters. While we did not quantify sensitivity, nor we carried out specifical statistical analyses of the confidence interval of such parameters, we chose the set of values that yielded the best accordance with

the observed scaling laws (exponential growth rate, peak power-law slope, decay dynamics) and a reasonable value for the basic reproduction number. For the best-fit reported in **Table 1**, we get $R_0 = 3.53$, very much in line with previous literature results [41].

5. CONCLUDING REMARKS

In this study, we started from the preliminary observation that in countries such as Italy, the cumulative number of infections displays an initial exponential amplification, followed by a power-law growth and a markedly asymmetric saturation behavior. We conjecture that these three phases are intimately connected and can be described by the interplay between the contagion process and the behavioral changes/containment policies acting on the population over comparable time scales.

Based on this hypothesis, and inspired by previous work by Ref. 14, we proposed a SEIQR-type compartment model with the following salient features: i) a compartment for unreported (asymptomatic) infected, that are supposed to play a major role in the spread of the disease; ii) a time-dependent rate of depletion of the susceptible population, which allows for a non-zero asymptotic residual of susceptible individuals.

The model was able to accurately reproduce the entire epidemic course in Italy, including quantitative agreement with exponential growth and power-law slope, for a plausible range of the model free parameters, with only *one* steep change of the parameters driven by the introduction of strict containment strategies. This circumstance suggests that the timing and implementation of containment policies (as quantified by model parameters t^*, μ_0 and τ) may effectively establish the decay characteristics of the epidemic. In particular we conjecture that the slow decay of the Italian epidemic can be attributed to an incomplete depletion of susceptible individuals during the containment phase.

The proposed modeling framework could be profitably used to gain insights and predictions on the so-called "second wave" of the epidemic, which (as of early October 2020) appears to be hitting Italy and other countries. To this purpose, the model could be modified by 1) introducing a *release* rate of protected individuals (from compartment R to S), that reflects the relaxation of containment measures as well as the gradually diminished perception of the pandemic; 2) accounting for a different testing policy: while only symptomatic individuals were tested during the first wave of the epidemic, the introduction of contact tracing and screening procedures has certainly contributed to increase the reporting rate ϕ_r, whereas the simultaneous improvements in testing efficiency and COVID-related infrastructures has led to a decrease in the testing time ϕ_t.

This study presents a number of simplifications and limitations that, however, do not affect the main conclusions. Among others, the number of confirmed infections is influenced by the number of tests performed each day, an aspect that was only weakly incorporated into the model via the (constant) fraction of reported cases ϕ_r. Even though many countries experienced initial difficulties with the testing capacity, the robustness and universality of the observed epidemic trends [19] strongly suggest that the scaling laws at play are primarily driven by the described containment-related mechanisms, with the scale of testing playing a minor role in the process [22]. Furthermore, the COVID-19 pandemic has been shown to have a strong geographical character with localized outbreaks, due to so-called *superspreaders* or family clusters [42]. In the context of a well-mixed framework, such as the one proposed in this paper, this phenomenon could be modeled by considering a specific infectious compartment with a much higher transmission rate, similarly as in Ref. 43. This aspect was not accounted for here and is left for future work.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

FC designed the study, conducted the data analysis, developed the mathematical model, interpreted the data and wrote the article.

ACKNOWLEDGMENTS

The author is indebted to Valerio Carandente, Marco Cozzo, Vincenzo Di Savino and Andrea Piccolo for insightful discussions.

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* (2020). 382:727–33. doi:10.1056/NEJMoa2001017.
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* (2020). 382:1199–207. doi:10.1056/NEJMoa2001316.
3. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature* (2020). 579:265–9. doi:10.1038/s41586-020-2008-3.
4. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report–51 (2020). p. 9.
5. World Health Organization. Coronavirus disease 2019 (COVID-19) weekly epidemiological update—5 october 2020 (2020). https://www.who.int/publications/m/item/weekly-epidemiological-update—5-october-2020.
6. Adam D. Special report: the simulations driving the world's response to COVID-19. *Nature* (2020). 580:316–8. doi:10.1038/d41586-020-01003-6.
7. Shen CY. Logistic growth modelling of COVID-19 proliferation in China and its international implications. *Int J Infect Dis* (2020). 96:582–9. doi:10.1016/j.ijid.2020.04.085..
8. Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world (2020). arXiv preprint arXiv:2003.05681.

9. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* (2020). 5: e261–70. doi:10.1016/S2468-2667(20)30073-6.
10. Jit A, Russell T, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* (2020). 20:P553–8. doi:10.1016/S1473-3099(20)30144-4.
11. Jewell NP, Lewnard JA, Jewell BL. Predictive mathematical models of the COVID-19 pandemic. *J Am Med Assoc* (2020). 323:1893–4. doi:10.1001/jama.2020.6585.
12. Ministero della Salute (2020). COVID-19 weekly monitoring reportAvailable from: http://www.salute.gov.it/nuovocoronavirus (Accessed July 6, 2020).
13. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* (2020). 20:533–4. doi:10.1016/S1473-3099(20)30120-1.
14. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* (2020). 368: 742–6. doi:10.1126/science.abb4557.
15. Manchein C, Brugnago EL, da Silva RM, Mendes CFO, Beims MW. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies. *Chaos* (2020). 30:041102. doi:10.1063/5.0009454.
16. Ziff AL, Ziff RM. Fractal kinetics of COVID-19 pandemic (2020). Net York, NY: medRxiv.
17. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* (1998). 393:440–2. doi:10.1038/30918.
18. Medo M. Contact network models matching the dynamics of the COVID-19 spreading (2020). arXiv preprint ArXiv:2003.13160.
19. Singer HM. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Phys Biol* (2020). 17(5):055001. doi:10.1088/1478-3975/ab9bf5.
20. Vazquez A. Polynomial growth in branching processes with diverging reproductive number. *Phys Rev Lett* (2006). 96:038702. doi:10.1103/physrevlett.96.038702.
21. Bodova K, Kollar R. Emerging algebraic growth trends in SARS-CoV-2 pandemic data. *Phys Biol* (2020). 20:42. doi:10.1088/1478-3975/abb6db.
22. Thurner S, Klimek P, Hanel R. A network-based explanation of why most COVID-19 infection curves are linear. *Proc Natl Acad Sci USA* (2020). 117: 22684–9. doi:10.1073/pnas.2010398117.
23. Chu KH. Fitting the Gompertz equation to asymmetric breakthrough curves. *J Environ Chem Eng* (2020). 8:103713. doi:10.1016/j.jece.2020.103713.
24. Li J, Lou Y. Characteristics of an epidemic outbreak with a large initial infection size. *J Biol Dynam* (2016). 10:366–78. doi:10.1080/17513758.2016.1205223.
25. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc Roy Soc Lond* (1927). 115:700–21. doi:10.1098/rspa.1927.0118.
26. Postnikov EB. Estimation of COVID-19 dynamics "on a back-of-envelope": does the simplest SIR model provide quantitative parameters and predictions? *Chaos Solit Fractals* (2020). 135:109841. doi:10.1016/j.chaos.2020.109841.
27. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* (2020). 368:489–93. doi:10.1126/science.abb3221.
28. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* (2020). 8:e488–96. doi:10.1016/S2214-109X(20)30074-7.
29. Munday G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* (2020). 26:855–60. doi:10.1038/s41591-020-0883-7.
30. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infect Dis* (2020). 93:211–6. doi:10.1016/j.ijid.2020.02.058.
31. 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.
32. Deng W, Bao L, Liu J, Xiao C, Liu J, Xue J, et al. Primary exposure to SARS-CoV-2 protects against reinfection in rhesus macaques. *Science* (2020). 369: 818–23. doi:10.1126/science.abc5343.
33. Lv P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci* (2002). 180:29–48. doi:10.1016/S0025-5564(02)00108-6.
34. Sun G-Q, Wang S-F, Li M-T, Li L, Zhang J, Zhang W, et al. Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources. *Nonlinear Dynam* (2020). 16:1–13. doi:10.1007/s11071-020-05770-9.
35. Li M-T, Sun G-Q, Zhang J, Zhao Y, Pei X, Li L, et al. Analysis of COVID-19 transmission in Shanxi Province with discrete time imported cases. *Math Biosci Eng* (2020). 17:3710–20. doi:10.3934/mbe.2020208.
36. Li Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *J Am Med Assoc* (2020). 323:1406–7. doi:10.1001/jama.2020.2565.
37. Wang M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19. *N Engl J Med* (2020). 382: 2158–60. doi:10.1056/NEJMe2009758.
38. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo′. *Nature* (2020). 584, 425–9. doi:10.1038/s41586-020-2488-1.
39. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship. *Euro Surveill* (2020). 25:2000180. doi:10.2807/1560-7917.ES.2020.25.10.2000180.
40. Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill* (2020). 25:200256. doi:10.2807/1560-7917.ES.2020.25.12.2000256.
41. Flasche M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc Natl Acad Sci USA* (2020). 117:10484–91. doi:10.1073/pnas.2004978117.
42. Rinaldo Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* (2020). 395:e47. doi:10.1016/s0140-6736(20)30462-1.
43. Ndaïrou F, Area I, Nieto JJ, Torres DFM. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solit Fractals* (2020). 135:109846. doi:10.1016/j.chaos.2020.109846.

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.

frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 30 November 2020
doi: 10.3389/fams.2020.571544

Application of a Susceptible, Infectious, and/or Recovered (SIR) Model to the COVID-19 Pandemic in Ecuador

Pablo Espinosa[1], *Paulina Quirola-Amores*[2] *and Enrique Teran*[1,2]*

[1]Colegio de Ciencias de la Salud, Universidad San Francisco de Quito, Quito, Ecuador, [2]Instituto de Microbiología, Colegio de Ciencias Biológicas y Ambientales-COCIBA, Universidad San Francisco de Quito, Quito, Ecuador

The coronavirus disease 2019 (COVID-19) pandemic is wreaking havoc in healthcare systems worldwide. COVID-19 was reported for the first time in Wuhan (China) and the first case in Ecuador was confirmed on February 27, 2020. Several determinants are taken into consideration for the establishment of asymptomatic or critical illness, and are necessary to predict the dynamics and behavior of a pandemic. We generated a Susceptible, Infectious, and/or Recovered model and reflected upon the COVID-19 pandemic in Ecuador. For the entire Ecuadorian population, we estimated that the reproduction number (*R*0) was 2.2, with 88% susceptible/infected individuals. To stop a national epidemic, a quarantine for 3–4 months is required, and when 55% of the population has been immunized (equivalent to 110 days since the first report of a COVID-19 case), a real decrease of new cases will be observed. The effectiveness of quarantine should be analyzed retrospectively, and not as a result of contemporary control of the COVID-19 epidemic.

Keywords: R0, COVID-19, herd immunity, susceptible infected-recovered model, attack rate

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Wen-Xuan Wang,
Beijing University of Posts and Telecommunications (BUPT), China
Liangli Yang,
University of Science and Technology Beijing, China

***Correspondence:**
Enrique Teran
eteran@usfq.edu.ec

Received: *11 June 2020*
Accepted: *24 September 2020*
Published: *30 November 2020*

Citation:
Espinosa P, Quirola-Amores P and Teran E (2020) Application of a Susceptible, Infectious, and/or Recovered (SIR) Model to the COVID-19 Pandemic in Ecuador. Front. Appl. Math. Stat. 6:571544. doi: 10.3389/fams.2020.571544

1. INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic that is wreaking havoc in healthcare systems worldwide is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Several phylogenetic studies have determined the origin of this zoonotic virus and, simultaneously, the potential reservoirs and amplifying species (especially mammals) [1–3].

The first report of SARS-CoV-2 infection was in December 2019 in Wuhan (Hubei Province, China). Worldwide spread of SARS-CoV-2 was inevitable due to its easy dissemination (as droplets), apparently low mortality, and high incidence of asymptomatic cases (≥15.8% of infected people) [1–5].

In <3 months, 60% of countries reported COVID-19 cases, particularly in Europe (e.g., Italy, Spain). Then, SARS-CoV-2 infection was documented in North America and, one week later, in Brazil, Ecuador, and Chile.

For COVID-19, lethality has been estimated to be 4–6%, mortality to be 1–1.5%, and severe illness to occur in ≥4.7% of diagnosed cases. The recovery time has been estimated to range from 8 to 11.5 days, but it is related to multiple factors [6–8].

The Ministry of Public Health (MSP) of Ecuador reported the first case of COVID-19 on February 27, 2020. It was an Ecuadorian woman who returned from Spain to Guayaquil on a commercial flight and had a welcoming party. Then, other cases at the northern (Tulcan) and southern (Machala) borders appeared (arising potentially from Colombia and Peru, respectively) [8]. By the beginning of

June, according to official reports, there were ≥43,120 confirmed cases, with a higher prevalence in people aged 21–60 years, lethality of 4–7%, and a mortality rate (<1%) [8, 9].

Mathematical models such as the Susceptible, Infectious, and/or Recovered (SIR) model are used to predict different scenarios related to epidemiologic factors and possible outcomes to assess epidemic spread. The reproduction number (*R*0) is used to estimate the capacity of viral transmission from one person to another. The effective reproduction number (*Re* or *Rt*) refers to the number of infected people in a specific time interval based on the *R*0. The susceptible fraction (*Fso*) takes into account the percentage of infected people that might exist in an epidemic. These indices show how a population and a virus are related to make predictions and consider the consequences at a given moment [8, 10, 11].

We aimed to generate SIR models based on data provided by the Ecuadorian MSP to estimate *R*0, *Rt*, *Fso*, time in quarantine, and the required percentage of immunized people to consider a reduction in the total incidence of COVID-19 in Ecuador. This strategy would allow the creation of a new SIR model readily without the need to have a large volume of regional data or national data. The possibility of developing retrospective and predictive models in this or other conditions based on an *R*0 calculation (observed here as a dynamic value) and the generation of a dynamic *Rt* value that will serve as a reference to make decisions in real-time, along with another type of model for quarantine and immunization, would be very helpful.

2. METHODOLOGY

2.1. Data Source

The data used for these models were from daily reports produced by the MSP. The data comprised of cumulative cases, new cases, excluded cases, deceased cases, cases who recovered from COVID-19, number of testing samples, and samples already collected but waiting to be processed from day-1 to day 30 of the epidemic (https://www.gestionderiesgos.gob.ec/informes-de-situacion-covid-19-desde-el-13-de-marzo-del-2020/) [9]. Values were taken from day 1 to day 30 of the epidemic considering the range with the highest reliability of data.

The SIR model, predictions/estimations, and tabulation data were processed using Excel™ within Office™ v16.37 (Microsoft, Redmond, WA, United States). Model parameters were verified using Vensim™ v8.0.9 (Salisbury, United Kingdom) and RStudio™ v4.0 (R Foundation for Statistical Computing, Vienna, Austria). Graphs were generated using Prism™ v8.4.2 (GraphPad, San Diego, CA, United States).

2.2. Estimated Infection Rate for the COVID-19 Epidemic in Ecuador

We used the parameters related to the number of individuals infected per day, the trend of the epidemic, mean number of samples processed, and samples waiting to be processed to determine the percentage of "true" work carried out by the MSP. Then, we compared the number of daily infected cases and estimated the rate of increase (cases/day). This was done using **Eq. 1**, in which the numerator is the number of reported cases in a final period of time and the denominator is the number of reported cases in an initial time [12, 13].

$$rate\ of\ increase\left[\frac{cases}{day}\right] = \frac{Infected\ t2}{Infected\ t1} \quad (1)$$

$$contagion\ rate = (1 - rate\ of\ increase)*100\% \quad (2)$$

The result using **Eq. 1** revealed whether the number of cases increased or not from 1 day to another, and also predicted the number of new cases for the next day. Thus, it also estimated the daily rate of contagion (**Eq. 2**), and the percentage of contagion in a certain period, based on the rate of increase. Both rates enabled the observation of dynamic behavior with *R*0 and *Rt*.

2.3. Calculation of the Minimum Number of Cases Required to Generate the COVID-19 Epidemic in Ecuador

A mathematical model based on sigmoid curves was applied to estimate the minimum number of cases required to generate an epidemic ("community infection"). To calculate the constants for the equation, two methods were employed. First, **Eq. 3** (*K*1) was obtained from the curve for the number of positive case reported by the MSP using logarithmic regression, where *t* is the time of infection, and *I* is the number of people infected in time *t*. Second, **Eq. 4** (*K*2) was considered as a differential logarithm regarding time, where *Ti* is the initial infection time, *Tf* is the final infection, *If* denotes finally-infected individuals and *Ii* denotes initially-infected people in an estimated period of time [14].

$$K1 = \frac{\sum Ln(t) - Ln(I) - \frac{\sum Ln(t)*\sum Lb(I)}{n}}{\sum (Ln(t))^2 - \frac{\sum (Ln(t))^2}{n}} \quad (3)$$

$$K2 = \frac{Ln(If) - Ln(Ii)}{Tf - Ti} \quad (4)$$

For K1 all the study data was used and in *K*2 data were from shorter ranges. Then, *K*2 provides daily evolution while *K*1 provides a general view of the epidemic. To calculate *It* (**Eq. 5**) (the number of infections over time in an epidemic), *Ic* was the number of infections present when the epidemic started. Based on the results, K1 (**Eq. 3**) was 0.1972, and *K*2 (**Eq. 4**) was 0.2386.

$$It = \left[\frac{1}{1 + Ice^{-kt}}\right]Ic^2 \quad (5)$$

2.4. Duration of Quarantine

To determine the duration of quarantine and its subsequent removal, we needed to calculate the recovery time of the population. Hence, we took the minimum number of cases after the first recovery of COVID-19, and a hypothetical daily peak of infections (for our purpose this was considered to be 100 daily cases [15]) and a threshold of cases after this peak for considering lifting the quarantine. Hence, when this number of infected cases was reached, with no increase in the last few days,

quarantine could be lifted. We assumed that all the daily reports from the MSP detailed the maximum number of cases, so we would know how long it would take for the recovery of these infected individuals [14].

$$It = \frac{If}{1 + (If - Reo)e^{-kIft}} \tag{6}$$

$$k = Ln\left[\frac{\frac{If}{It} + 1}{If}\right] \tag{7}$$

$$td = \frac{Ln\left[\frac{\frac{If}{Ir} - 1}{If}\right]}{k} \tag{8}$$

In **Eq. 6**, *It* is the number of recoveries in a period, *If* is the maximum number of infections in a particular period, *Reo* is the number of people who recovered on day-0 after reaching the maximum number of infected individuals (usually 1). In **Eq. 7**, *k* is the constant of proportionality. In **Eq. 8**, *td* is the duration of quarantine, *Ir* is the maximum number of infected people minus the threshold of 100 infected individuals [15] needed for lifting the quarantine ($Ir = If - 100$) and *If* is the maximum of cases for each day of quarantine.

2.5. SIR Models for COVID-19

We wished to model and predict the behavior of the COVID-19 epidemic in Ecuador. Hence, we took into account three mathematical models based on SIR, differed in the use of different parameters for the equations and the calculation of *R*0, where we did not consider the rate of asymptomatic cases (not a well-established value at the time of analyses).

The first model (**Eqs 9–16**) was based on the number of infections generating new cases. The parameters for construction were the recovery rate (95%), mortality rate (5%), time needed for recovery (17 days), *R*0 (taken from international databases [5], and assuming that 40% of the population was under total quarantine (that is equivalent to 4% less probability of becoming infected). *N* was the total population of Ecuador and the *R*0 was the ratio between the number of infected people and the number of people who recovered and those who died (**Eqs 13**, **16**) [13, 16–18].

$$\%P = \frac{NCoM}{IiTDSi} \tag{9}$$

$$CoM = \frac{IiTDSi}{N\%P} \tag{10}$$

$$Re = Ii*\left(\frac{TR}{DI}\right) \tag{11}$$

$$M = Ii*\left(\frac{TM}{DI}\right) \tag{12}$$

$$R0 = \frac{CoM}{(Re + M)} \tag{13}$$

$$Rt = \frac{CoM(t)}{(Re(t) + M(t))} \tag{14}$$

$$dRt = R0*\left(\frac{1}{1 + e^{\left(\frac{So}{N}\right)}}\right)*\left(\frac{So}{N}\right)*\left(e^{\left(\frac{So}{N}\right)}\right) \tag{15}$$

$$R0m = \sqrt[n]{R01*R02*R0n} \tag{16}$$

Sf (**Eq. 17**) represents the number of vulnerable individuals, *If* (**Eq. 23**) is the number of infected people, *Rf* (**Eq. 21**) is the number of people who recovered from COVID-19, *Mf* (**Eq. 22**) is the number of people who died; all of these parameters were taken after a certain period of time. *DI* represents the duration of illness in days, *TD* (**Eq. 10**) is the percentage of infections (*R*0), *%P* (**Eq. 9**) is the probability of becoming infected, *TR* (**Eq. 11**) is the percentage of people who recovered, *TM* (**Eq. 12**) is the percentage of individuals who died, and *Rt* (**Eqs 12**, **15**) is the *R*0 that varies through time and the function of infected people.

The second model (**Eqs 17–26**) takes into consideration differential equations from a standard model that were integrated subsequently with national data where the percentage of people who recovered (γ) (**Eq. 17**) was 95%, mortality (μ) (**Eq. 22**) was 5%, and the time for recovery (*TR*) (**Eqs 9**, **11**, **20**) was 17 days. To calculate *R*0 for Ecuador, we considered standard calculations (**Eqs 21**, **26**). *Rt* (**Eqs 14**, **15**, **24**) was based on the variation in the percentage of susceptible people during a particular time, and *p* (**Eq. 17**) was the relationship between the number of people who recovered from COVID-19 and the infection rate [13, 19].

$$p = \frac{\gamma}{\beta} \tag{17}$$

$$p = \frac{N - Sf}{Ln\left(\frac{\gamma N - \gamma R}{\gamma So}\right)} \tag{18}$$

$$R0 = \frac{\beta}{\gamma}So \tag{19}$$

$$Sf = Soe^{-\left[\frac{N - So}{p}\right]} \tag{20}$$

$$Rf = \left(\frac{\gamma}{TR}\right)Ro \tag{21}$$

$$Mf = \mu Ii \tag{22}$$

$$If = Sf = Ii - Ri - Mi \tag{23}$$

$$Rt = Ro*e^{\left(\frac{So}{N}\right)} \tag{24}$$

$$dRt = R0*\left(\frac{1}{1 + e^{\left(\frac{So}{N}\right)}}\right)*\left(\frac{So}{N}\right)*e^{\left(\frac{So}{N}\right)} \tag{25}$$

$$R0m = \sqrt[n]{R01*R02*R0n} \tag{26}$$

The third model (**Eqs 27–36**) was based on the infection rate calculated with the equation for the rate of increase, so the β value (**Eq. 27**) was assigned to it. For the recovery rate (γ), it was considered that 17 days was the minimum time required to recover, and a 3% mortality percentage (μ) in accordance to the MSP report [10, 13].

$$\beta = 1 - rate \tag{27}$$

$$\gamma = \frac{1}{recuperation\ time} \tag{28}$$

$$R0 = \frac{\beta}{(\gamma + \mu)} \tag{29}$$

$$Infection(Con) = \beta*Ii*R0*\left(\frac{So - Ii}{N}\right) \tag{30}$$

$$If = So - Con - Re - Mu \tag{31}$$

$$Recuperation\,(Re) = If * \gamma \tag{32}$$

$$Dead\,(Mu) = If * \mu \tag{33}$$

$$Rt = Ro * e^{\left(\frac{So}{N}\right)} \tag{34}$$

$$dRt = R0 * \left(\frac{1}{1 + e^{\left(\frac{So}{N}\right)}}\right) * \left(\frac{So}{N}\right) * \left(e^{\left(\frac{So}{N}\right)}\right) \tag{35}$$

$$R0m = \sqrt[n]{R01 * R02 * R0n} \tag{36}$$

2.6 Variation of the SIR Model With Different Values of *R*0

Taking into consideration the first proposed SIR model (**Eqs 9–16**), we generated different *R*0 values (**Eqs 13**, **16**) and observed their behavior during a particular time. This strategy allowed for the ascertainment of how *Rt* and the epidemic cycle were related so that we could observe it and identify possible "bottlenecks", when the number of susceptible people decreased (for whatever reason). Simultaneously, we compared the derivative of *Rt* (**Eq. 15**), which allowed us to observe the maximum number of infections when *Rt* varied, and worked to adjust a national *R*0 [14]. For model validation, we applied a difference between the curve for infected people (from the MSP report) and the number of infections generated with the three SIR models. Then, the curve (of these three models) that best fits to the real data reported by the MSP, was selected for further analysis [14].

2.7. Susceptible Factors and the Minimum Percentage Required for Immunization Using *Rt*

For the *Rt* of each model, we calculated the susceptible factor (*Fso*) (**Eq. 37**). This strategy allowed us to observe the highest percentage of infections (based on the variation in *Rt*) to generate a threshold that estimated the direction of the epidemic and where *R*0 was located in Ecuador [10, 19]. With *Fso* and *R*0 generated in each model, we could predict the minimum percentage (*%Pi,v*) (**Eq. 38**) required for immunization (by vaccination or herd immunity) to stop COVID-19 spread [10, 19].

$$fSo\,(t) = 1 - e^{-Rt} \tag{37}$$

$$\%Pi,\ v = 1 - \frac{1}{R0} \tag{38}$$

2.8. Vaccination and Immunization

After the derivation of *%Pi,v* from **Eq. 38**, we considered the incorporation of the *Rt* toward the population (*N*). Considering the variation in infection, we could recognize how long it would take for a vaccination strategy to obtain results and stop infection. We could also observe the number of immunized individuals and those who would not be infected during a particular time [10, 18, 20].

$$Immunizate\ population\,(Piz) = N * \left(1 + \left(\frac{1}{2Rt^2}\right)\right) \tag{39}$$

$$Infected\ population\,(Pif) = N - N * \left(1 + \left(\frac{1}{2Rt^2}\right)\right) \tag{40}$$

3. RESULTS

Upon analyses of all the reported data by the MSP, a marked delay in sample processing was noticed (40% CI = ±0.0035) and an increase in the number of new daily cases of 28.5% (*CI* = ±0.00146) was reported. Hence, the infection curve was delayed in 28.5% of cases (~40% of undetermined cases). This observation was confirmed when the rate of increase per day remained the average between the number of delayed samples and number of reported cases. This resulted in 1.20 reported cases per day (CI = ±0.0103). When we analyzed the number of cases with COVID-19 reported by the MSP, there was a 30% delay in confirmatory diagnose. This problem was most evident at day 33 of the epidemic when there was 60% delay in confirmatory diagnoses (**Figure 1**).

We used *K* values (**Eqs 3**, **4**) to determine the minimum number of cases required to generate an epidemic in Ecuador. In our simulation using *K1*, five cases were needed for an epidemic in Ecuador. Also, the model using *K2* showed that with six cases, an exponential communitarian infection was possible (**Figure 2**), and that recovery from COVID-19 could take 18 days (CI = ± 0.0945).

While observing the trend in the increase in infected cases, one key question was how long the quarantine should last until there were 100 infected cases. We calculated that quarantine should be lifted after approximately 80–110 days (CI = ±0.2001) because the maximum number of cases had been reached (**Figure 3**).

Simultaneously with obtaining results, we generated three SIR models with variation of some parameters to obtain *R*0 and observed which model was the best fit.

When the number of infected people was compared for each model, model 1 (SIR 1) obtained the closest values to the true situation. Hence, we chose model 1 to observe the dynamic in cases once quarantine had been implemented (**Figure 4**). Using this model, *R*0 = 2.2 (CI = −1.644, +2.75). Importantly, with each model, an *R*0 value was generated so a different *Rt* analysis was undertaken to detect the appearance of a bottleneck in a susceptible population. We found that, for each model, the *R*0 was between 1 and 3 (CI = −1.865, +1.974) (**Table 1**).

Using these values, we calculated the maximum number of people susceptible to SARS-CoV-2 infection for each *R*0 value (95.02%, CI = ±1.27). The same calculation showed that a maximum of 66.67% (CI = ±2.70) of the same population would need to be immunized to stop the epidemic (**Figure 5**).

When quarantine parameters were generated with SIR 1, the peak of the maximum number of infected individuals would be reached on day 154 of the epidemic (65 more days than if quarantine was not applied). The number of new cases would decrease (64.6%, CI = ±2.35) but not disappear, demonstrating that when the quarantine time increased, the number of infected cases was redistributed. This phenomenon was observed when the different *Rt* values (generated by a variation in parameters)

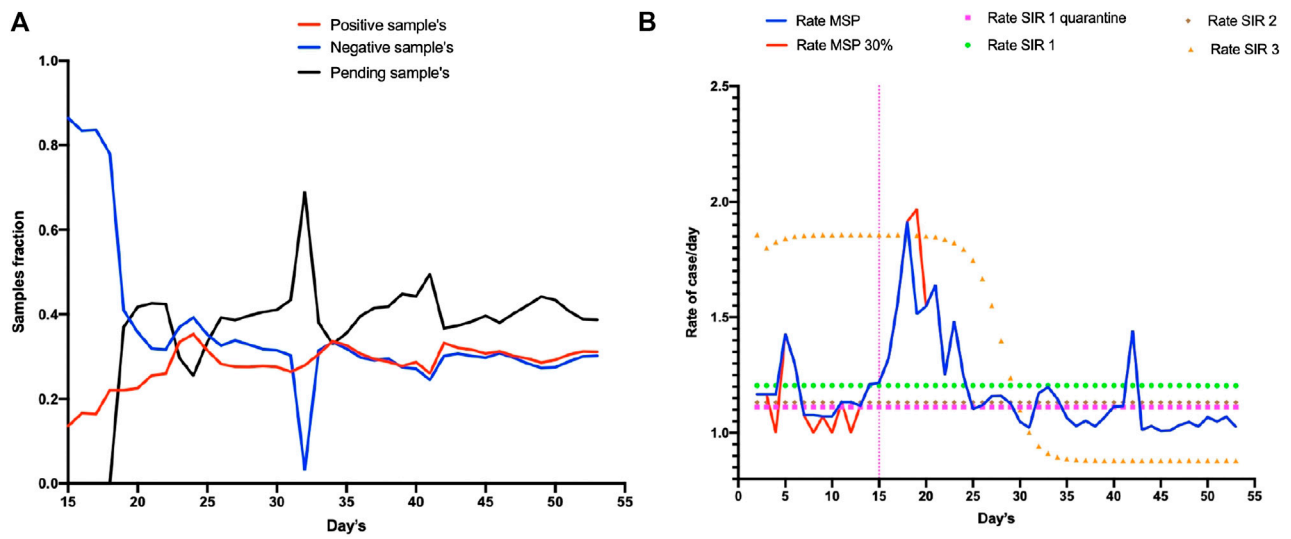


FIGURE 1 | Report of the COVID-19 epidemic in Ecuador in January 2020. Section **(A)** shows the diagnoses of people in Ecuador from 15 February to 23 April 2020 (15 at 55 days of the epidemic) whereby red denotes the number of positive cases, blue denotes the number of negative cases, and black the number of samples. **(B)** shows in blue line the increase rate of cases in MSP report, red line represents the possible real number of reported cases (including 30% delay in confirm diagnose). The model SIR 1 showed the more accurate tendency (orange line).

were analyzed (**Figure 6**) and demonstrated that, for each unit increase in *Rt*, the epidemic would increase by 10 days (CI = ±0.295).

To confirm the obtained *R*0 with SIR 1, we analyzed the variation in *Rt* with its derivative, in addition to the variation in *Fso* with regard to *Rt*. As the bottleneck appeared with *Rt*, without considering quarantine, the value of *Rt* remained in this bottleneck demonstrating that, in reality, *R*0 was between 1.5 and 2.5 CI = −1.51, +2.48. When comparing these data with *Fso*, we found that the maximum number of cases using this *R*0 value

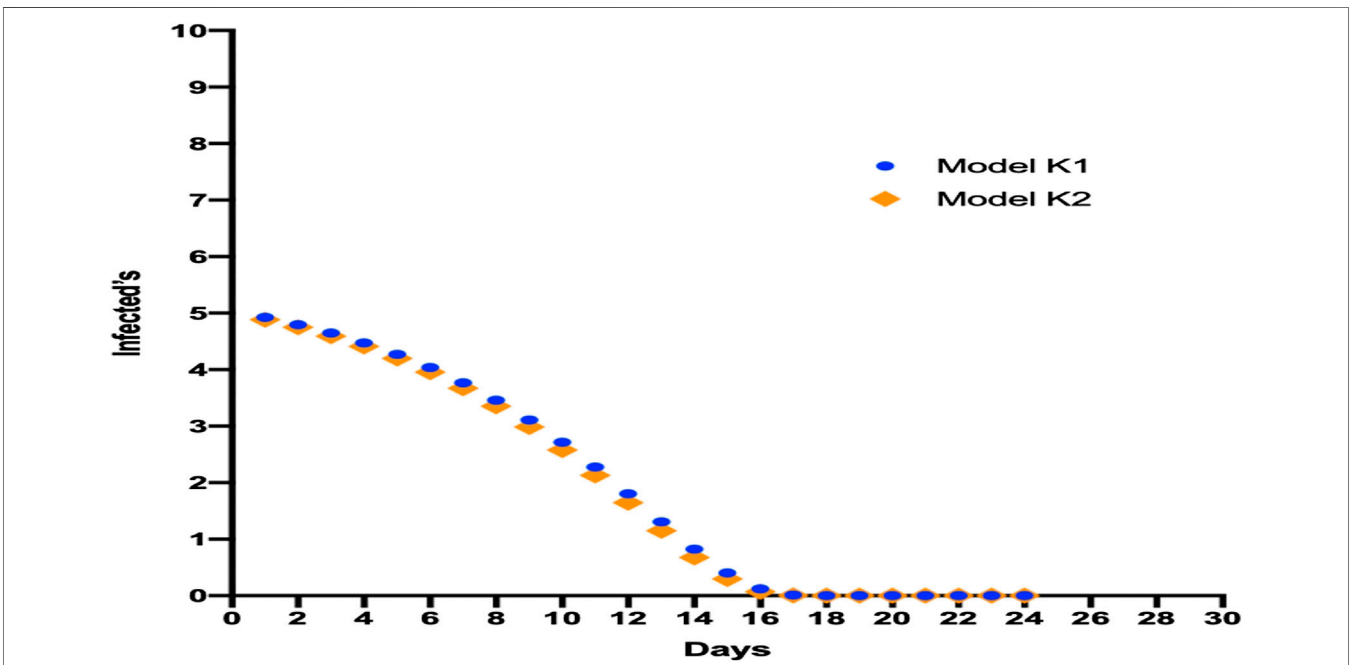


FIGURE 2 | Model for the minimum number of cases required for an epidemic in Ecuador showing different simulations with various K constants for observing this number, communitarian infection, and the average time needed for recovery.

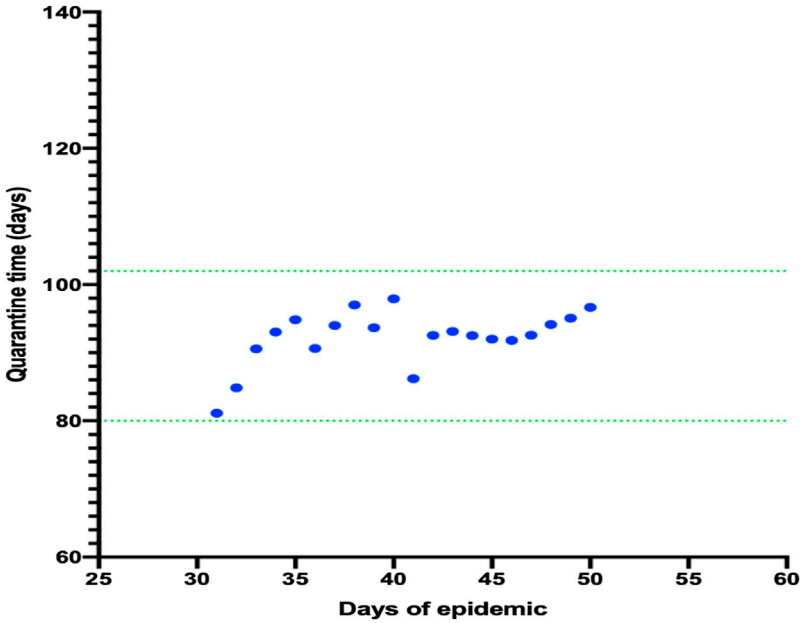


FIGURE 3 | Model for the number of days required for quarantine and the maximum number of daily cases. It was observed that the number of days necessary to finish the quarantine remained the same despite the increasing number of cases daily.

was 80%, with a maximum time of 90 days (when the epidemic would begin to decrease). However, if quarantine was applied, this value reached 86% of infected people at a time of 140 days (**Figure 7**).

Upon application of all the different data of *R*0, an issue was observed when an immunization appeared (even for vaccination or herd immunity). The most important question is how long will it take to observe if this strategy is successful. Our model

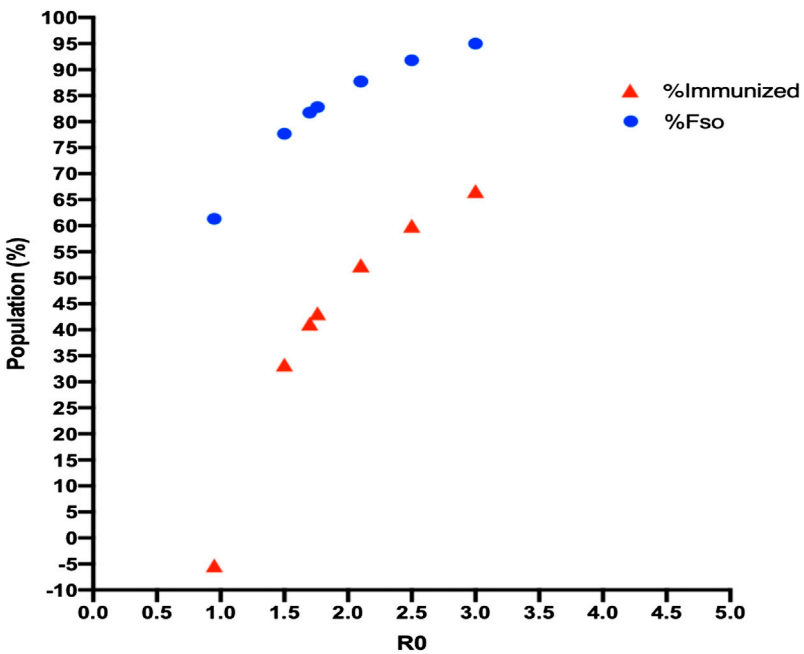


FIGURE 4 | *Fso* and herd-immunity values with *R*0 for each model. An increase in *R*0 for each SIR model led to an increase in the infected population and herd immunization, but neither of these values reached 100% of the population.

TABLE 1 | *R*0 values generated with three SIR models showing the maximum number of people who may be infected (Fso), and herd immunity for each model.

	R0	*R0m*	*Fso* (%)	Herd immunity (%)
*R*0 Ref. 28	2.1	2.1	87.75	52.3
SIR 1	1.73	2.20 (CI= −1.644, +2.75)	81.73	41.18
	3		95.02	66.67
	2.1		87.75	52.38
SIR 2	1.76	1.92	82.80	43.18
	2.1	(CI= −1.865, +1.974)	87.75	52.38
SIR 3	0.95	1.19	61.33	−5.26
	1.5	(CI= −1.042, +1.337)	77.69	33.33

R0m *is the average of the* R0 *value calculated in each model.*

showed that, without quarantine from day-80 to day 100, the introduction of vaccination, the rebound of immunized individuals, and a decrease in the number of infected cases applying quarantine between day 100 to day 120 would result in the same outcome (**Figure 8**).

4. DISCUSSION

The model used in our study is not the only model to predict the behavior of COVID-19. One must consider the dynamic of SARS-CoV-2 transmission, availability of resources, and delay in diagnosis when choosing the best model. For Ecuador, an important issue was the large number of people with pending SARS-CoV-2 tests. This factor, together with asymptomatic individuals and underreporting of "true" infected people (by consensus, it is 55%), demonstrated why our model was not well-fitted with the data reported by the MSP [9, 23].

Although the estimated time needed for a patient to recover from COVID-19 was 18–20 days, some scholars have stated it is 8–10 days using SIR models. According to the MSP, it takes 20 days, which poses problems for healthcare units in terms of space and resources [5, 23, 24]. Determining this value is a key component in the prediction based on SIR models because it is based on local data. Subsequently, extrapolations can be made in the hope that they are not too far from reality.

In our study, the time for recovery from COVID-19 was based on a *K* model and observing how the recovery curve was adjusted to the true situation. Mistakes can be generated because of the issues mentioned above as well as from the reports of patients who have recovered from COVID-19. Even though this value cannot be observed directly, it can be estimated through the report of deaths by the MSP and comparing them with the registry of the National Institute of Statistics of Ecuador. This comparison resulted in a difference of ≥20% and has been observed in other studies too [23–26].

Through an assessment of the first 100 cases of COVID-19, a marked difference in acceleration and control of the epidemic was made in Taiwan. The Taiwan government isolated and quarantined all individuals related to each COVID-19 case, and tracked all asymptomatic individuals. This strategy showed that, according to the proposed model, quarantining the first six cases rapidly and aggressively would have created a different scenario from that seen presently but, as the author mentions, the switch from theory to practice is not easy [25]. Another important observation is the rate of increase of cases. At the national level, the rate of increase was constant (the number of cases increased in 5–7 days), as has been observed on a global scale [9, 27]. Hence, locally obtained data were a reflection of

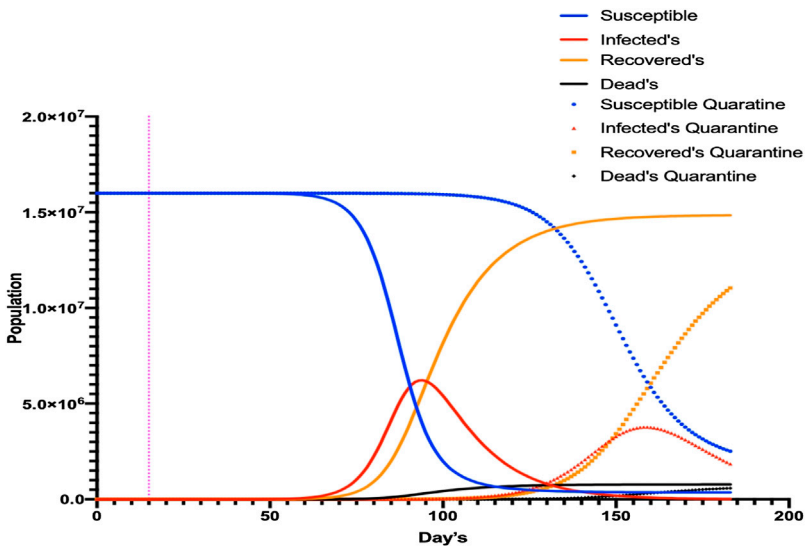


FIGURE 5 | SIR1 model with modified parameters for quarantine. Different simulations were undertaken for COVID-19, with and without quarantine. The number of infected individuals decreased but the duration of infectivity was prolonged by >200 days.

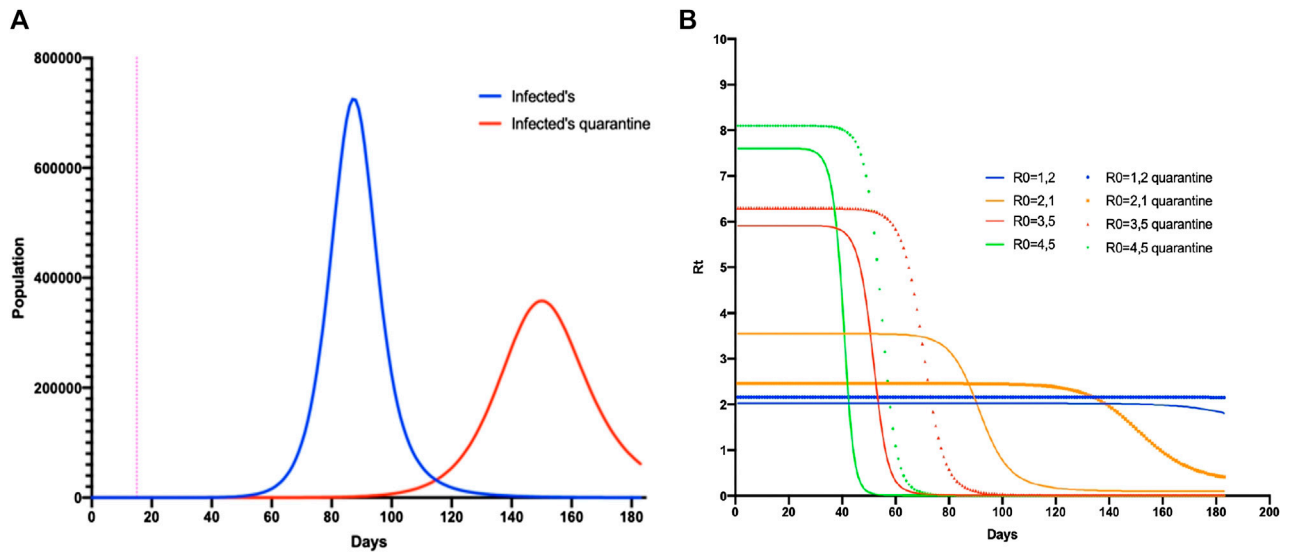


FIGURE 6 | SIR1 model comparing the parameters regulated with quarantine. **(A)** The number of infected people without quarantine (blue) and with quarantine (red). **(B)** Simulation of *Rt* in the SIR1 model varying the parameters with and without quarantine over the duration of the epidemic.

what occurred in the past when SARS-CoV-2 spread in Asia and Europe [9, 27].

Our model estimated that lifting the quarantine should occur approximately 2–3 months after the peak of infection has been reached. In China, after the peak of infection (after 17 days), the quarantine was lifted during day 13 to day 14. Such lifting is dependent upon time and the capacity of the health system of a particular country. In addition, a strict quarantine works not only by holding back the appearance of new cases, but also allows the health system to recover and prepare for new cases [5, 27].

The mortality rate for COVID-19 has been estimated to be about 3–7%, and results in another problem for SIR models. In our model, we took a mortality rate of 5% (in other models, it varies between 3 and 4.5%). However, one must consider other causes of death unrelated to COVID-19 that can affect a model indirectly [5, 23, 24, 29]. The increase in *R*0 determines how effective the strategies taken by local authorities are. In our study,

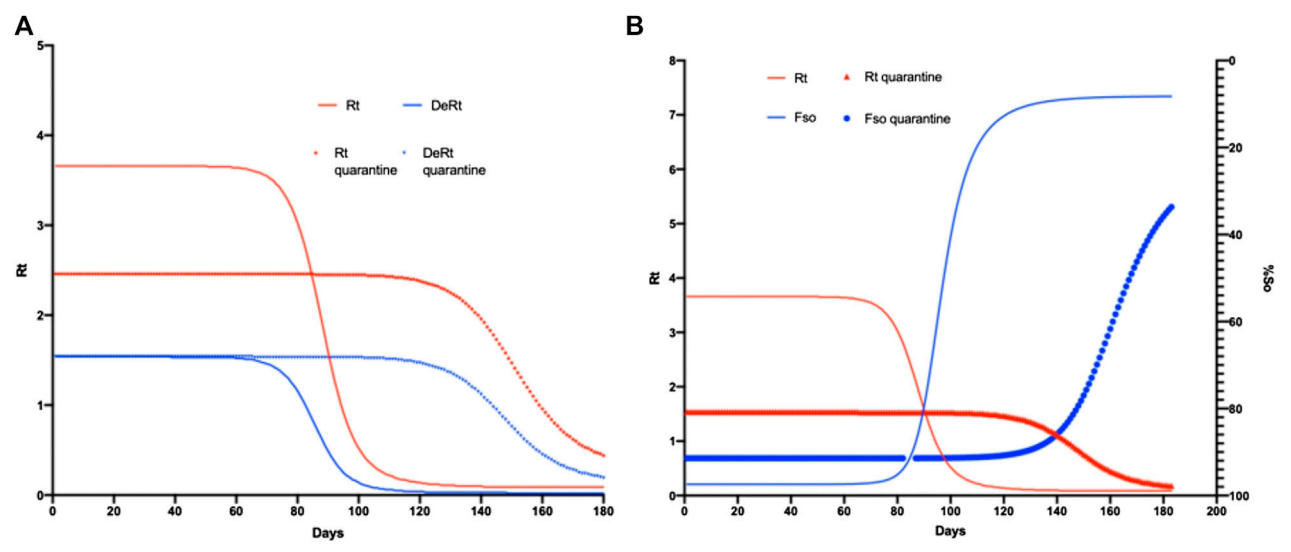


FIGURE 7 | SIR1 model comparing regular parameters with quarantine. In **(A)** the simulation of *Rt* is observed in the SIR1 model varying the parameters with and without quarantine throughout the epidemic duration. **(B)** shows the number of infections without quarantine (blue) and with quarantine (red).

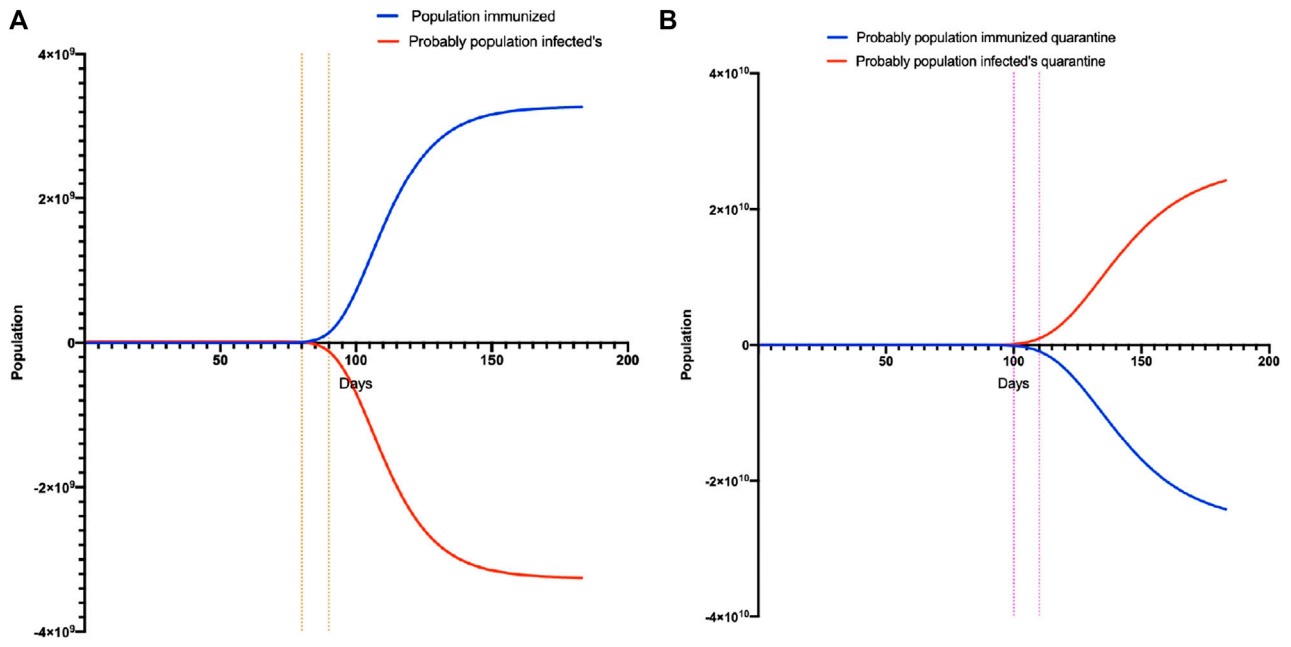


FIGURE 8 | Calculation model to reach herd immunity. **(A)** is the model without quarantine and shows the time that must pass for herd immunity to occur. **(B)** is the model with quarantine and shows the time that must pass for herd immunity to occur.

$R0$ = 2.2, which is not too far from the international trend (Ye et al. estimated the same value of $R0$, and the World Health Organization value is between 1.4 and 2.5).

The time of infection is generated based on $R0$ (which depends on whether the value increases or decreases). If we compare the $R0$ for other diseases caused by viruses of the same family (Middle East respiratory syndrome-related coronavirus, <1; SARS, 2–4), this value does not reflect the lethality for each one [5, 23, 27, 30, 31]. Hence, defining $R0$ is not easy and is mired in controversy. Each population could "fix" their value and reflect the variation in parameters of the equation used (β and α) in each SIR model. The latter could be constructed and calculated in different ways depending on data reliability. However, the $R0$ is very important because from it the minimum number of immunized people required to stop the epidemic, and the maximum number of individuals susceptible to infection, can be calculated [10, 11, 32, 33].

Some problems were identified when the $R0$ was calculated on the basis of the total population. However, in reality, interventions are applied only in specific or priority groups. We analyzed different $R0$ values for the COVID-19 epidemic in Ecuador and, despite an increase in $R0$, 100% of a population will not be infected. With the prediction models used for smallpox prevention, the $R0$ was 10, but the maximum percentage of infected individuals in a vulnerable population was 70% and, upon introduction of a vaccine, any epidemic can be prevented with a coverage of 60% of the population [6, 16].

Complex and dynamic clinical phenomena occur during disease development. The immune response, nutritional state, chronic diseases, or co-infections can interfere (accelerate or decelerate) epidemic expansion, and are not taken into account when the $R0$ is calculated [11]. Hence, Li and colleagues [11] postulated the interpretation and analyses of $R0$ in four categories. Their model explains the rapid expansion of a disease but with a relatively slow recovery process.

Rt provides a clearer picture of the time of increase of an epidemic according to the percentage of susceptible people over time. Initially, the Rt shows an acceleration in the number of infected individuals, then becomes constant at the end of the epidemic, and reaches a value of 0, thereby allowing calculation of an adjusted $R0$ adapted to constant changes or population variation over time [33, 34]. A bottleneck has been observed in the Rt as the number of days of an epidemic increase, and the $R0$ is closer to reality because, with fewer people to infect, the true infective capacity of a virus is observed. However, this scenario was not entirely accurate in our model ($R0$ = 1.97). Even though our $R0$ is lower than that reported worldwide, some scholars have calculated an $R0$ value between 0 and 1. This does not mean that an epidemic is improving or ending; in some cases, this value shows a latency period [11, 32, 34, 35]. For a more accurate adjustment of the $R0$, we applied Fso based on our obtained Rt value. As seen in **Figure 7**, the point where both curves intersect corresponds to the $R0$ and simultaneously coincides with the bottleneck observed in Rt (**Figure 5**), and we showed a similar value of $R0$ to that obtained in previous studies [10, 33].

The minimum number of infected people required to stop an epidemic has been mentioned [9, 23, 33, 36]. However, this value must be considered very carefully because these infected individuals will probably die. Importantly, 60% of the population should be immunized (based on $R0$) to stop the

COVID-19 epidemic in Ecuador [9, 23, 33, 36]. This concept is seen more clearly with "collective immunity" or "collective protection", which refers to the fact that a group of infected people who recover will prevent the disease from spreading to the entire population [21, 22, 37]. By applying this concept to our SIR models, this collective protection will be reached with 60% of the infected population but, as stated by Rashid and colleagues, there are three main problems of assuming this value as real.

The first problem is the heterogeneity of the population (i.e., the different immune responses that each person presents to SARS-CoV-2). Hence, even though 60% of the population can be infected, this does not ensure that people can generate an immune state that protects the remaining susceptible members of the population [35, 37].

The second problem refers to the *R*0 and the criterion of the equation for *%Pi,v*. Rashid et al. explained that, because the population is composed mostly of people at a high risk of contagion, the *R*0 is overestimated. Hence, collective protection does not correspond to the one that has been calculated. In these cases, one must include in the equation the efficiency of the country's vaccination campaigns to understand how the epidemic can be controlled [20, 35, 37]. This was not applied in the present study because a vaccine for COVID-19 has not yet been found.

The third problem identified by Rashid et al. is the possible impact of co-circulation of other pathogens (depends on the characteristics of each population) during an epidemic. This could modified SIR models through the calculated *R*0 with less or more impact in predictions, and not are fitted to reality.

We evaluated the impact of the quarantine implemented in Ecuador using two variables: the percentage of people quarantined (used in the SIR model) and *Rt* variation (by the effect of this measure). Quarantine reduces the number of people on the streets, and decreases the number of SARS-CoV-2 infections in a particular period [23, 24]. We took as a reference the percentage of people quarantined as 40%. Other models have considered 10%, and observed a decrease of 50% in the number of expected cases that would be observed if no health strategy was implemented. This 50% reduction does not mean that the cases will not exist, only that they will be redistributed over the duration of the epidemic. Hence, the epidemic would last 300 days due to the quarantine instead of 180–200 days, thereby ensuring that the health system does not collapse [29, 30]. Although quarantine has been a good strategy for control of infectious diseases, it is not considered a good strategy from dynamic and economic viewpoints because external factors will affect *Fso* (the maximum number of susceptible people that can be infected). In our model, *Fso* increased by 10% (from 80 to 90%) if quarantine was applied in 40% of the population. This increase occurred because the longer the quarantine lasts, the greater the probability that the population will not comply with it. Since quarantine was applied in Ecuador, the unemployment rate increased from 3.9 (2019) to 4.4% (2020). This increase forced people to take to the streets in search of work and increased the number of people exposed to SARS-CoV-2, and, consequently, *Fso* increased [24, 34, 38].

We also estimated the number of people with acquired immunity to SARS-CoV-2 based on the *R*0. We showed that a decrease in the number of newly infected cases would be observed by day 80 to day 100 of the epidemic. Other scholars say that this value depends on the responsiveness of the local government to report and identify cases, diagnostic capacity, and the delay in sample processing [9, 23, 33, 36].

We identified a delay in sample processing in ≥40% of cases, this parameter has not been reported before. Another value that can be estimated through the *R*0 and *Fso* is the duration of infection. If 40% of the population were in quarantine, we estimated that the duration of infection would increase to 15 days, which was also observed by Li et al. and Massed et al. Both research teams agreed that the duration of infection and number of immunized individuals should not be taken into account to estimate the reduction of an epidemic. A latency period denotes the time when reinfections occur or the epidemic disappears. Hence, if we know the *R*0, we could estimate the percentage of infected people required to stop the epidemic, which is equivalent to the one during this latent period. Our model determined it to be 65% but, in reality, this value will not be observed until it reaches 90% of infected people and only then will we observe a decrease in the infection rate [11, 23, 24, 33].

Although SIR models provide accurate data for the number of patients who have recovered from COVID-19, infected cases, and exposed patients, they require a large volume of reliable data. SIR models could be used in countries with better contingencies and strategies than those in most Latin American countries. Traditional models are dimensionless in time (results are not affected whether days or months are used) so the infection rate will not fit well with dynamic populations. For our models, we used a specific unit of time that allowed for the daily adjustment and discreet visualization of an increase or decrease of the epidemic. In the context of public health, the use of our model could enable prompt redistribution of resources according to infection trends.

A weakness of our models was not providing the percentage of asymptomatic cases and exposed cases. This led to a reduction of the power of prediction for infected people (≥15%). Although this percentage could affect the fit between the real curve and model curve, this range of variation is acceptable because the exact proportion of asymptomatic COVID-19 individuals in Ecuador is not known. A lack of reliable and updated data was the main limitation of our models. This resulted from a deficiency in the structure of surveillance measurements taken, identification of cases, sample-processing capacity, as well as weakness in recollection, updating, and data analyses from the MSP during the COVID-19 epidemic. The MSP is the only entity that possesses full access to the data, and those data are not open for external analyses or validation. In addition, the number of reported cases and tests performed changed dramatically over time without robust justification from the MSP. Another important limitation was the assumption of some of the parameters for our models. They were obtained from countries from the same region with similar

behavior of SARS-CoV-2 infection/COVID19 so did not affect our results.

5. CONCLUSIONS

For the entire Ecuadorian population, we estimated that $R0 = 2.2$, with 88% susceptible/infected individuals. To stop a national epidemic, a quarantine for 3–4 months is required, and when 55% of the population has been immunized (equivalent to 110 days since the first report of a COVID-19 case), a real decrease of new cases will be observed. The effectiveness of quarantine should be analyzed retrospectively, and not as a result of contemporary control of the COVID-19 epidemic.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

PE and ET conceptualize the study, PE and PQ obtained data and analyzed it, PE set up the mathematical models, PE and PQ run the models, ET reviewed results and provided feedback. PE and PQ drafted the manuscript. ET drafted the final version. All authors revised and approved final version.

REFERENCES

1. Anthony SJ, Johnson CK, Greig DJ, Kramer S, Che X, Wells H, et al. Global patterns in coronavirus diversity. *Virus Evolut* (2017) 3(1):1–15. 10.1093/ve/vex012.
2. Lal SK. Molecular biology of the SARS-coronavirus. In: Lal SK, editor *Climate change 2013—the physical science basis*. Vol. 53. Cambridge, United Kingdom: Cambridge University Press (2010) p. 1–30). doi:10.1017/CBO9781107415324.004
3. Miller SL, Gill J, Webb GR. The proximal origin of the hamstrings and surrounding anatomy encountered during repair. *J Bone Joint Surg* (2007) 89(1):44–8. doi:10.2106/jbjs.f.00094.
4. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, et al. Virology, epidemiology, pathogenesis, and control of covid-19. *Viruses* (2020) 12(4):1–17. doi:10.3390/v12040372.
5. Ye Q, Wang B, Mao J, Fu J, Shang S, Shu Q. Epidemiological analysis of COVID-19 and practical experience from China. *J Med Virol* (2020) 12:23–7. doi:10.1002/jmv.25813
6. Kaul D. An overview of coronaviruses including the SARS-2 coronavirus—molecular biology, epidemiology and clinical implications. *Curr Med Res Pract* (2020) 10(2):54–64. doi:10.1016/j.cmrp.2020.04.001.
7. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* (2020) 109, 102433. doi:10.1016/j.jaut.2020.102433.
8. SNGRE, CIES. *Covid-19 en el ecuador tendencia temporo espacial de la pandemia del 27 de febrero al 19 de abril del 2020 INFORME*. New York City, NY: MINTEL (2020)
9. MSP. *Informes de Situación e Infografias—COVID 19—desde el 29 de Febrero del 2020—Servicio Nacional de Gestión de Riesgos y Emergencias*. Quito, Ecuador: Ministerio de Salud Pública (2020)
10. Bjornstad ON. *Epidemics models and data using R*. Berlin, Germany: Springer (2010) doi:10.1007/978-3-319-97487-3
11. Li J, Blakeley D, Smith RJ. The failure of R0. *Comput Math Methods Med* (2011) 2011:1–17. doi:10.1155/2011/527610
12. Vynnycky E, White R. *An introduction to infectious disease modelling*. Oxford, UK: Oxford University Press (2010)
13. Keeling MJ, Rohani P. *Modeling infectious diseases through Contact networks*. Princeton, NJ: Princeton University Press (2008)
14. Li MY. An introduction to mathematical modeling of infectious diseases. In: L Ken M Golden editors. *An introduction to mathematical modeling of infectious diseases*. 2nd ed. Berlin, Germany: Springer International Publishing (2018) doi:10.1007/978-3-319-72122-4
15. NHCPRC. *Latest developments in epidemic control. National Health Commission of the People's Republic of China* (2020)
16. Jones JH. Notes on R 0. *Building* (2011) 1–19:17–33. doi:10.1186/isrctn10820810
17. Kretzschmar M. Measurement and modeling: infectious disease modeling. In: *International encyclopedia of public health*. 2nd ed. Vol. 4. Amsterdam, Netherlands: Elsevier. doi:10.1016/B978-0-12-803678-5.00229-0
18. Tarigan PB. Veterinary epidemiology. *J Chem Inf Model* (2013) 53(9):58–67. doi:10.1017/CBO9781107415324.004
19. Balkew TM. *The SIR model when S(t) is A multi-exponential function*. East Tennessee State University (2010) 12 p.
20. Smith DR. Herd immunity. *Vet Clin North Am Food Anim Pract* (2019) 35(3):593–604. doi:10.1016/j.cvfa.2019.07.001.
21. Chen J, Ye L, Zhou MY, Cheng YR, Wang M-W, Feng ZH. Herd immunity and COVID-19. *Eur Rev Med Pharmacol Sci* (2020) 24(8):4064–5. doi:10.26355/eurrev_202004_20978.
22. Kwok KO, Lai F, Wei WI, Wong SYS, Tang JWT. Herd immunity—estimating the level required to halt the COVID-19 epidemics in affected countries. *J Infect* (2020) 80(6):e32–3. doi:10.1016/j.jinf.2020.03.027.
23. Castro P, De los R, Juan C, González S, Merino P, Ponce J. Modelización y simulación de la propagación del virus sars-cov-2 en Ecuador. *MODEMAT* (2020) Available from: https://observatoriocovid19.sv/doc/biblioteca/internac/Informe-Covid19-Modemat.pdf.
24. Aguirre H, Orellana D, Armas R. *Escenarios de Propagación de COVID-19 al 21 de Abril y Relajamiento de Medidas de Mitigación*. Cuenca, Ecuador: Universidad de Cuenca (2020) p. 1–14.
25. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Lin H-H. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Intern Med* (2020) 180:1156–63. doi:10.1001/jamainternmed.2020.2020.
26. INEC. Estadísticas Vitales registro estadístico de Nacidos vivos y Defunciones 2016. *INEC* (2016) 1, 315. doi:10.1016/j.jhazmat.2015.06.018
27. Bulut C, Kato Y. Epidemiology of COVID-19. *Turk J Med Sci* (2020) 50(SI-1), 563–70. doi:10.3906/sag-2004-172.
28. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. COVID-19, SARS and MERS: are they closely related? *Clin Microbiol Infect* (2020) 22:74–8. doi:10.1016/j.cmi.2020.03.026
29. Stratton SJ. COVID-19: not a simple public health emergency. *Prehosp Disaster Med* (2020) 35(2):119. doi:10.1017/s1049023x2000031x.
30. Panovska-Griffiths J. Can mathematical modelling solve the current Covid-19 crisis? *BMC Publ Health* (2020) 20(1):1–3. doi:10.1186/s12889-020-08671-z.
31. Uras U. *Coronavirus: comparing COVID-19, SARS and MERS|news|Al Jazeera* (2020) Available from: https://www.aljazeera.com/news/2020/04/coronavirus-comparing-covid-19-sars-mers-200406165555715.html
32. Breban R, Vardavas R, Blower S. Theory versus data: how to calculate R0? *PLoS One* (2007) 2(3):e282. doi:10.1371/journal.pone.0000282.
33. Massad E. Ethical and Transborder issues. *Global Health Inform* (2017) 12:232–63. doi:10.1016/B978-0-12-804591-6.00012-4
34. Akpan N. *How to measure your nation's response to coronavirus*. London, UK: National Geographic (2020)
35. Metcalf CJE, Ferrari M, Graham AL, Grenfell BT. Understanding herd immunity. *Trends Immunol* (2015) 36(12):753–55. doi:10.1016/j.it.2015.10.004.

36. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Dig Health* (2020) 2:e201–8. doi:10.1016/S2589-7500(20)30026-1
37. Rashid H, Khandaker G, Booy R. Vaccination and herd immunity. *Curr Opin Infect Dis* (2012) 25(3):243–9. doi:10.1097/qco.0b013e328352f727
38. Lombeida E, Moreno L, Fabara C. *Pobreza y desigualdad Ecuador* (2010) Available from: http://www.ecuadorencifras.gob.ec/documentos/web-inec/POBREZA/2018/Junio-2018/Informe_pobreza_y_desigualdad-junio_2018.pdf

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.

frontiers in Physics

BRIEF RESEARCH REPORT
published: 04 December 2020
doi: 10.3389/fphy.2020.578024

Reduced Treatment Sensitivity of SARS-CoV-2 After Multigenerational Human-to-Human Transmission

Qiang Wang[†], *Tong Yang*[†] *and Yang Wang*[*]

Navy 971 Hospital of PLA, Qingdao, China

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has become a rapidly spreading worldwide pandemic, seriously threatening global public health security. At present, there is still no vaccine or specific drug available for the virus. Therefore, epidemiological prevention and control is crucial to block further spread of the disease. In this present study, based on the public information of COVID-19 in Qingdao, Shandong Province, China, we analyzed the epidemiological characteristics and treatment effects of 60 confirmed cases. It was revealed that the hospital stay of patients were imported from Wuhan or infected by Wuhan patients was obviously shorter than that of those who were imported from other regions or infected by them. In addition, in Hubei-related patients, the earlier transmission generation, which refers to the imported patients (G0) and the first-generation infections (G1), was associated with reduced cure time. Therefore, we speculated that the treatment sensitivity of SARS-CoV-2 decreased within interpersonal transmission in the early stage of this epidemic.

Keywords: COVID-19, human-to-human transmission, treatment sensitivity, generation, region

OPEN ACCESS

Edited by:
Chengyi Xia,
Tianjin University of Technology, China

Reviewed by:
Xiao Han,
University of California, Davis, United States
Zhishuang Wang,
Tianjin University of Technology, China

***Correspondence:**
Yang Wang
wy971@yandex.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *30 June 2020*
Accepted: *02 November 2020*
Published: *04 December 2020*

Citation:
Wang Q, Yang T and Wang Y (2020) Reduced Treatment Sensitivity of SARS-CoV-2 After Multigenerational Human-to-Human Transmission. Front. Phys. 8:578024. doi: 10.3389/fphy.2020.578024

INTRODUCTION

As of October 20, 2020, a total of 40652577 COVID-19 cases and 1119664 deaths have been confirmed globally. To date, the outbreak of COVID-19 in China has been effectively controlled [1]. On the contrary, the worldwide pandemic is still spreading rapidly, with massive increases in the number of cases daily. Therefore, it is necessary to analyze the epidemiological characteristics of COVID-19 for further insights into this disease. By stratified analysis of patients in a city, subgroup analyses can clarify relevant factors that affect therapeutic efficiency, thus providing important evidence for the improvement of the treatment and prevention strategy for COVID-19.

In this study, publicly available data were collected from 60 confirmed COVID-19 patients in Qingdao, Shandong Province (information from the official website of Qingdao Municipal Health Commission). The correlation analysis was performed between treatment effect (hospital stay) and related factors. After multigenerational human-to-human transmission, our study indicated reduced treatment sensitivity of SARS-CoV-2 to the current therapeutic strategy, which could be inferred from the fact that the cure time of Wuhan-imported patients and their associated infections is significantly short than that in other patients. Meanwhile, the earlier transmission generation (G0 and G1) of SARS-CoV-2 was correlated with shorter cure time in patients who can be directly traced to Hubei Province.

TABLE 1 | Characteristics of COVID-19 patients in Qingdao (N = 60).

Variable	Statistics (N = 60)
Age (years)	
Mean ± SD	45.63 ± 18.54
Median (range)	46 (1–90)
Gender	
Male	27
Female	33
Infection type	
Imported	27
Infected	33
Infection region	
Wuhan	20
Hubei (except Wuhan)	24
Others	16
Generation	
Imported (G0)	27
G1	16
G2 or G3	17
Diagnostic time	
Mean (range), days	2.05 (0–12)
≤1	39
2–4	13
≥5	8

METHODS

From January 21 to February 23, 2020, information of 60 cases of COVID-19 confirmed in Qingdao was collected from the official website of Qingdao Municipal Health Commission. The public information includes patients' age, gender, infection type (imported or locally infected), the source of imported cases, contact history of infected cases, and the date of onset of symptoms, admission, diagnosis, and discharge. According to the information, diagnostic time, hospital stay, and cure time of patients could be calculated. Cure time was defined as the period from the diagnosis of COVID-19 to the discharge of the patient. We regarded the cure time as an indicator of the sensitivity of treatment. Pearson's chi-squared test was carried out to evaluate the relationship of potential influence factors with cure time. Subgroup analyses were applied to evaluate the therapeutic efficiency affected by age, gender, infection type, generation, imported region, and diagnostic time. SPSS (version 16.0) and Review Manager (version 5.3) were adopted for statistical analyses. All statistical tests were 2-tailed, and p-value < 0.05 was regarded significant.

RESULTS

Baseline Characteristics

By February 23, 2020, a total of 60 COVID-19 cases were diagnosed in Qingdao with real-time fluorescence polymerase chain reaction detection of SARS-CoV-2 RNA [2] (**Table 1**). All patients were treated based on Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment (published by the National Health Commission of the People's Republic of China) after the diagnosis. The standard clinical cure was defined as temperature recovery for more than 3 days, with relieved respiratory symptoms and two consecutive negative results for respiratory pathogenic nucleic acid (sampling interval at least 1 day). The patients included 27 males and 33 females. For the 60 patients, the median age was 46 years, ranging from 1 to 90 years. According to the infection type, 27 cases were imported from other regions and 33 cases were locally infected. There were 13 imported patients from Wuhan, 8 from Hubei Province (excluding Wuhan), and 1 from Thailand, thus causing 7, 16, and 2 local infections, respectively. Another 2 cases were imported from Henan Province and 3 cases from Shandong Province, without interpersonal transmission. Another 8 cases were infected by local suspected cases. The average incubation period for the traceable 54 cases was 10.57 days, ranging between 1 and 31 days. Up to March 2nd, 59 cases were cured. The cure rate was 98.33%, and the average hospital stay for the cured cases was 17.69 ± 6.09 days. An 85-year-old female patient died from COVID-19.

Virus Transmission Scenarios

Totally, 27 confirmed imported cases were identified up to March 2. Among them, 10 cases caused local transmission and induced 33 infected patients (**Figure 1**). Other 22 imported cases displayed no human-to-human transmission. The maximum number of transmissions was 11 from a single imported case. The imported case (Case X2) was identified as a confirmed patient from Tianmen, Hubei Province. There were 16 first-generation infections, 12 second-generation infections, and 5 third-generation infections. Of note, four cases (Cases 50, 52, 56, and 58) were detected as asymptomatic infections, that is, without typical symptoms including fever.

Analysis of Clinical Cure

Up to March 2, a total of 59 patients cured after systematic treatment in Qingdao, with an average treatment time of 16.31 (range 6–43) days. Taking the median treatment time (16 days) as the cutoff point, 59 patients were divided into the shorter cure time (≤16 days) and longer cure time (>16 days) groups. Correlation analysis was conducted for relevant factors and cure time (**Table 2**). The results illustrated no significant difference of age between patients with shorter cure time and those with longer cure time. Moreover, the infection type (imported or infected by interpersonal transmission of virus) also demonstrated no significant correlation with therapeutic effects. It is worth noting that the imported region of SARS-CoV-2 exhibited significant correlation with therapeutic efficiency (p = 0.002). The patients from Wuhan or infected by Wuhan cases displayed shorter cure time than other patients. Given that COVID-19 first broke out from Wuhan and spread to other regions, the results suggested that SARS-CoV-2 became more difficult to remove after multiple generations of human-to-human transmission. Further analysis of the virus transmission generation also supported significantly prolonged cure time in the second and third generation of infected patients in Hubei-associated transmission chain (p = 0.040). However, in terms

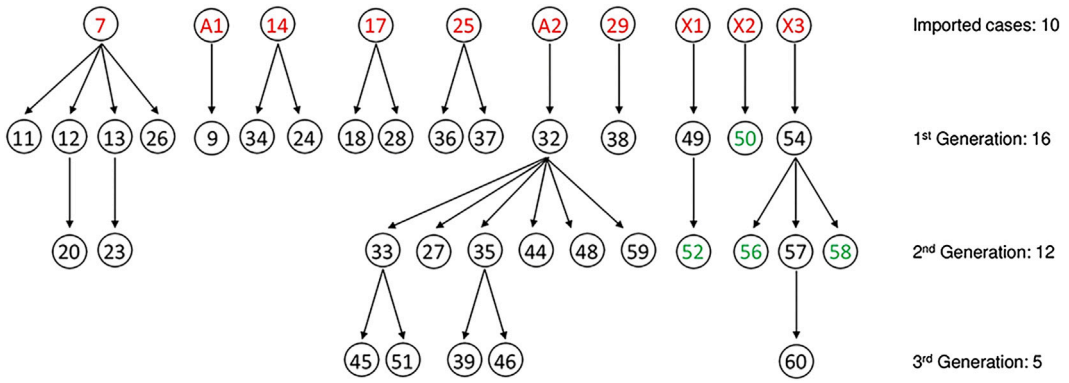


FIGURE 1 | SARS-CoV-2 transmission scenarios in Qingdao. Totally, 10 imported COVID cases (red number) cased 33 cases of human-to-human transmission (black or green number), among which 4 cases did not (green number) display typical symptoms of infection, such as fever. A1 was a suspected infected case from Wuhan. A2 was a confirmed infected case from Tianmen, Hubei. X1-3 was a suspected case in Qingdao.

TABLE 2 | Correlation analysis of therapeutic efficiency in cured COVID-19 patients in Qingdao (N = 59).

Variables	Cases with shorter cure time (≤16 d)	Cases with longer cure time (>16 d)	*p* value
	Age (years)		
≤46	19	13	0.253
>46	12	15	
	Gender		
Male	18	9	0.046
Female	13	19	
	Infection type		
Imported	16	11	0.343
Infected	15	17	
	Imported region		
Wuhan	16	4	0.002
Others	15	24	
	Generation (all 59 cured cases)		
G0 and G1	24	18	0.266
G2 and G3	7	10	
	Hubei-associated (43 cases)		
G0 and G1	21	10	0.040
G2 and G3	4	8	
	Diagnostic time		
≤1 d	21	18	0.779
>1 d	10	10	

of 59 cured patients, there was no significant statistical significance between transmission generation of SARS-CoV-2 and cure time of patients. Furthermore, we analyzed the influence of duration time from symptom onset to diagnosis, and no significant difference was observed (p = 0.779). Interestingly, it was clear that gender has a significant impact on the cure time of patients. Compared with female patients, male patients exhibited a trend of reduced hospital stay (p = 0.046).

Further pooled analysis (**Figure 2**) supported that the infection from other regions, rather than Wuhan (HR = 2.08, 95% CI: 1.32–3.27), was a detrimental factor for therapeutic efficiency and was probably related to later generation of interpersonal transmission of SARS-CoV-2. Female gender was also indicated as a promising risk factor for therapeutic efficiency (HR = 1.64, 95% CI: 1.00–2.70). By contrast, increased transmission generation was not identified as a significant risk factor for therapeutic efficiency (HR = 1.39, 95% CI: 0.74–2.59). What is more, no significant favor was observed in age, infection type, or diagnostic time for the cure efficiency.

DISCUSSION

The limitation of this study was the relatively modest number of identified COVID-19 cases in Qingdao because the results of the regional analysis may be different from those of overall research to some extent. However, because of government's efforts in epidemic management, up to now, only 65 local cases were

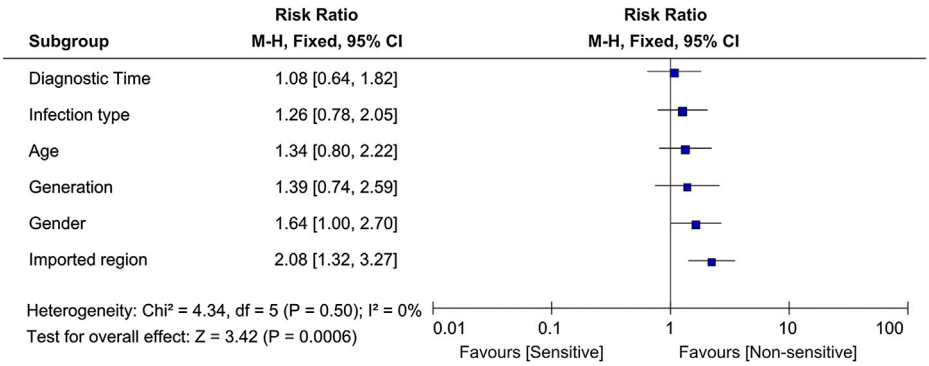


FIGURE 2 | Subgroup analyses of therapeutic efficiency through age, gender, infection type, generation, imported region, and diagnostic time in 59 cured COVID-19 patients. Dashed vertical line corresponds to zero heterogeneity of the therapeutic effect. The variables were compared as follows: age, >46 years vs ≤46 years; gender, female vs male; infection type, local transmission vs imported; generation, G2-3 vs G0-1; imported region, other regions vs Wuhan; and diagnosis time, >1 vs. ≤1 day.

confirmed in Qingdao and the last 5 cases confirmed after March 11 were all infected by patients imported from abroad. Considering the possible impact of different sources of virus, these 5 cases were not enrolled in our cohort. With a population of 9.2 million, Qingdao is a large regional city with broad medical and health coverage in the Jiaodong Peninsula. Its detailed and accessible epidemiology information of COVID-19 provided valuable research resources for the treatment and management of this disease. Up to March 2, the mortality rate of COVID-19 in Qingdao was 1.67%, lower than the national level (5.18%, 4,746/91,574, up to October 20, information from the official website of the National Health Commission of the People's Republic of China). The average time from symptom onset to diagnosis for the infected patients is 2.05 days, which is a reliable guarantee for high cure rate [3].

The systemic treatment of patients was based on Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment that was published by the National Health and Health Commission of China. The adjustment of this guidance may have some impact on the treatment effect because its content was constantly updated, and the treatment strategies were also changed according to clinical treatment effects. However, since all patients in this cohort were admitted to the hospital in a relatively short period (33 days) and the improvement of the clinical guidance was gradual, we believe that the impact of treatment adjustments on the analysis of results is limited and acceptable.

Laboratory diagnosis identified 4 infected patients from close-contact cases (Cases 50, 52, 56, and 58), without typical symptoms including fever [4]. Some infected patients did not display a significant inflammatory response but were positive for the virus detection, which were potential sources for further transmission. Moreover, the cured patients revealed favorable prognosis, and pulmonary fibrosis was rarely observed [5]. In this study, the only died patient was an elderly patient with various underlying diseases, including diabetes, hypertension, coronary heart disease, and heart failure. It was worth noting that the time from symptom onset to the diagnosis of this patient was 12 days. Long diagnosis time may also be a detrimental factor for patients. The correlation of related factors, including old age, underlying disease, and diagnostic time with treatment effect still needs further analysis with larger series of samples [5].

As it infects patients and causes pandemic, SARS-CoV-2 evolves at the same time. In March 2020, Tang et al. [6] proposed that SARS-CoV-2 developed into two subtypes, namely, L type and S type, among which L type was more aggressive and infectious, accounting for about 70% of infections. Forster's team [7] put forward three major variants in 160 complete SARS-CoV-2 genomes and named them A, B, and C. Types A and C have a high proportion outside East Asia (i.e., among Europeans and Americans). In contrast, type B is the most common type in East Asia. In early March, Korber et al. [8] identified seven D614G mutations on spike protein from 187 viral sequences, and its frequency increased rapidly when this mutant strain entered the population. In many cases, G614 became the main local form within a few weeks. The team obtained no evidence that the virus carrying the mutation caused more severe clinical symptoms. However, the levels of viral RNA in the nose and mouth of patients infected with G614 appeared to be slightly higher than those in patients with D614. Subsequent studies proved that G614 virus may have enhanced ability to enter cells [9, 10]. Moreover, a medical team from Hubei Province, China, compared the clinical characteristics of patients admitted before and after January 23, and they discovered that the initial symptoms of recently infected patients had become more insidious, and even after being cured and discharged from hospital, virus was rediscovered in three patients [11]. The authors speculate that the new coronavirus may gradually evolve into an influenza-like virus and even hide in asymptomatic patients for a long time [11].

As a reasonable inference, we speculate that in the process of long-term coexistence of the host, with improved adaptation to human cells, the virus may become easier to be transmitted or more difficult to be removed by the host. In cases of this cohort, this adaptation may be manifested in the prolongation of the cure time. This prolongation does not contradict with the decline of COVID-19 mortality worldwide. First, due to the increasing attention paid by countries around the world, the enhancement of admission and treatment capacity, and the standardization of diagnosis and treatment methods, the reduction in death rate of COVID-19 were predictable. Moreover, the evolution of SARS-CoV-2 does not necessarily lead to increased virulence. On the contrary, in the process of adapting to humans, certain functions of highly pathogenic CoV in suppressing host immunity may be weakened or lost [12]. And, this may be conducive to the survival of the virus in the host without increasing the severity of the disease. At present, there exist no specific treatment and vaccine available for the disease [13]. However, the application of treatment, even symptomatic or supportive, is beneficial to the cure of patients. In fact, the treatment sensitivity reflects the virus' ability to survive in the human body and the difficulty of being cleared in the presence of human intervention. To some extent, it is an indicator of the adaptability of the virus to the host.

This study revealed a reduced cure effect in infectious patients from other regions than those from Wuhan, which was correlated with decreased treatment sensitivity with the interpersonal transmission of SARS-CoV-2. Although neutralizing IgG extracted from the serum of cured COVID-19 patients is equally effective in blocking cell entry of both the D614 and 614G variants of SARS-CoV-2 at this stage [14], we cannot guarantee that new mutations that may threaten the development of antivirus drugs or vaccines will not appear. Before the advent of specific treatments for COVID-19, universal epidemiological control is of great significance to combat the pandemic. Population movement between regions is an important factor influencing the spread of epidemics [15], while virus testing along with case isolation and community quarantine has been proved to be able to prevent rapid local outbreak [16]. Physical distancing of 1 m or more and wearing masks are also believed to significantly reduce the risk of virus transmission [17]. In addition, the government's promotion of positive preventive measures can contribute to the inhibition of the outbreak [18]. All in all, epidemiology is a powerful weapon against COVID-19, and controlling the source of infection and cutting off the route of transmission should not be ignored at any time.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: http://wsjsw.qingdao.gov.cn/n28356065/index.html.

AUTHOR CONTRIBUTIONS

Guarantor of integrity of the entire study: YW; Study concepts and study design: QW and YW; Literature research, data collection and analysis: QW and TY; All the authors revised and approved the paper.

FUNDING

This research was supported by grants from the National Natural Science Foundation of China (81972793, 81502283), Shandong province Medical Health Science and Technology Project (2018WS447), Qingdao Science and Technology Program for Benefiting People Special Project 2019 (19-6-1-25-nsh, 2019-WJZD170), and Qingdao Outstanding Health Professional Development Fund.

REFERENCES

1. Burki T. China's successful control of COVID-19. *Lancet Infect Dis* (2020) 48(2):151–3. doi:10.1007/s15010-020-01409-4
2. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* (2020) 25:45. doi:10.2807/1560-7917.es.2020.25.3.2000045
3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* (2020) 395: 497–506. doi:10.1016/s0140-6736(20)30183-5
4. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* (2020) 395:514–23. doi:10.1016/s0140-6736(20)30154-9
5. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* (2020) 395:507–13. doi:10.1016/s0140-6736(20)30211-7
6. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. *Nat Sci Rev* (2020) 7:1012–23. doi:10.1093/nsr/nwaa036
7. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci USA* (2020) 117:9241–3. doi:10.1073/pnas.2007433117
8. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *Nat Sci Rev* (2020) 12:122–8. doi:10.1101/2020.04.29.069054
9. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* (2020) 182:812–27. doi:10.1101/2020.04.29.069054
10. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. Preprint repository name [Preprint] (2020). Available from: https://www.biorxiv.org/content/10.1101/2020.09.01.278689v1.
11. Chen Z, Jijia h., Zhang Z, Jiang S, Wang T, Shi Z, et al. Caution: clinical characteristics of COVID-19 patients are changing at admission. *SSRN Electron J* (2020) 4:32–9. doi:10.1101/2020.03.03.20030833
12. Fung SY, Yuen KS, Ye ZW, Chan CP, Jin DY. A tug-of-war between severe acute respiratory syndrome coronavirus 2 and host antiviral defence: lessons from other pathogenic viruses. *Emerg Microb Infect* (2020) 9:558–70. doi:10.1080/22221751.2020.1736644

13. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *J Am Med Assoc* (2020) 323:1824–36. doi:10.1001/jama.2020.6019
14. Tan Y, Liu F, Xu X, Ling Y, Huang W, Zhu Z, et al. Durability of neutralizing antibodies and T-cell response post SARS-CoV-2 infection. *Front Med* (2020) 12:23–7. doi:10.1007/s11684-020-0822-5
15. Yin Q, Wang Z, Xia C, Dehmer M, Emmert-Streib F, Jin Z. A novel epidemic model considering demographics and intercity commuting on complex dynamical networks. *Appl Math Comput* (2020) 386:125517. doi:10.1016/j.amc.2020.125517
16. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* (2020) 584:425–9. doi:10.1038/s41586-020-2488-1
17. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schunemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* (2020) 395:1973–87. doi:10.1016/j.jvs.2020.07.040
18. Wang Z, Xia C, Chen Z, Chen G. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern* (2020) 15:1–9. doi:10.1109/tcyb.2019.2960605

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 05 December 2020
doi: 10.3389/fphy.2020.613168

Epidemiological Model With Anomalous Kinetics: Early Stages of the COVID-19 Pandemic

Ugur Tirnakli[1]* *and Constantino Tsallis*[2,3,4]

[1]Department of Physics, Faculty of Science, Ege University, Izmir, Turkey, [2]Centro Brasileiro de Pesquisas Fisicas and National Institute of Science and Technology for Complex Systems, Rio de Janeiro, Brazil, [3]Santa Fe Institute, Santa Fe, NM, United States, [4]Complexity Science Hub Vienna, Vienna, Austria

We generalize the phenomenological, law of mass action-like, SIR and SEIR epidemiological models to situations with anomalous kinetics. Specifically, the contagion and removal terms, normally linear in the fraction I of infected people, are taken to depend on $I^{q_{up}}$ and $I^{q_{down}}$, respectively. These dependencies can be understood as highly reduced effective descriptions of contagion via anomalous diffusion of susceptible and infected people in fractal geometries and removal (i.e., recovery or death) via complex mechanisms leading to slowly decaying removal-time distributions. We obtain rather convincing fits to time series for both active cases and mortality with the same values of (q_{up}, q_{down}) for a given country, suggesting that such aspects may in fact be present in the early evolution of the COVID-19 pandemic. We also obtain approximate values for the effective population N_{eff}, which turns out to be a small percentage of the entire population N for each country.

Keywords: COVID-19, pandemics, complex systems, nonextensive statistical mechanics, epidemiological models

OPEN ACCESS

Edited by:
Matjaž Perc,
University of Maribor, Slovenia

Reviewed by:
Tassos Bountis,
University of Patras, Greece
Oscar Sotolongo,
University of Havana, Cuba
Airton Deppman,
University of São Paulo, Brazil

***Correspondence:**
Ugur Tirnakli
ugur.tirnakli@ege.edu.tr

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *01 October 2020*
Accepted: *02 November 2020*
Published: *05 December 2020*

Citation:
Tirnakli U and Tsallis C (2020) Epidemiological Model With Anomalous Kinetics: Early Stages of the COVID-19 Pandemic. Front. Phys. 8:613168. doi: 10.3389/fphy.2020.613168

1 INTRODUCTION

The classic and still widely used SIR and SEIR epidemiological models [1] represent contagion and removal in analogy with the law of mass action in chemistry, corresponding to a mean-field approach based on the assumption of homogeneous mixing. The latter hypothesis constitutes an oversimplification, particularly for the COVID-19 pandemic, due to strong government intervention (social distancing; lockdown) and underreporting as the number of cases grows beyond testing capacity. Diverse aspects are discussed, assuming homogeneous or nonhomogeneous mixing, in epidemiological models in general [2–4], as well as in the current pandemic [5–15].

Epidemic models can be formulated on varying levels of detail, from individual agents in geographically realistic settings to models of large populations without spatial structure. Each level has its own benefits and costs; the study of an ensemble of models is expected to yield a more reliable description than any single approach in isolation. In chemical kinetics of processes involving anomalous diffusion and/or complex conformational pathways, effective descriptions typically employ noninteger power-law terms where the mean-field or mass-action analysis involves integer powers of concentrations, as in the analysis of reassociation of folded proteins [16, 17]. With this motivation, we consider SIR- and SEIR-like models in which the contagion and removal terms depend on $I^{q_{up}}$ and $I^{q_{down}}$, respectively, instead of depending linearly on I, as they do in mean-field/homogeneous descriptions. Such generalization is

consistent with anomalous human mobility and spatial disease dynamics [18, 19] and emerges naturally within statistical mechanics based on nonadditive entropies [20] as we show in what comes next.

Let us now follow along lines close to [16], which provided a satisfactory description of reassociation in folded proteins [17]. Consider the equation

$$\frac{dy}{dt} = ay^q \left(q \in \Re;\ t \geq 0;\ y(0) > 0\right). \tag{1}$$

Its solution is given by

$$y(t) = y(0)\, e_q^{a\,[y(0)]^{q-1}\,t}, \tag{2}$$

with $e_q^z \equiv [1 + (1-q)z]^{1/(1-q)}$ ($e_1^z = e^z$), a function that emerges naturally in the nonadditive-entropy-based statistical mechanics [20]. We have a monotonically increasing function $y(t)$ for $a > 0$ (with infinite support if $q \leq 1$ and finite support if $q > 1$) and a monotonically decreasing function for $a < 0$ (with infinite support if $q \geq 1$ and finite support if $q < 1$). Notice an important point that will permeate through this entire paper: if $q = 1$, *only then* the coefficient a (which characterizes the scale of the evolution of $y(t)/y(0)$ with time) is *not* renormalized by the initial condition $y(0)$. If $q \neq 1$, the effective constant $\{a\,[y(0)]^{q-1}\}$ will differ from a; the difference can be very important depending on the values of q and $y(0)$.

2 GENERALIZED MODELS

2.1 *q*-SIR Model

The SIR set of equations is (see [1] for instance) as follows:

$$\begin{aligned} \frac{dS}{dt} &= -\beta\, S\, \frac{I}{N} \\ \frac{dI}{dt} &= \beta\, S\, \frac{I}{N} - \gamma\, I \\ \frac{dR}{dt} &= \gamma\, I, \end{aligned} \tag{3}$$

with $\beta > 0$, $\gamma > 0$, and $S + I + R = N = constant$, N is the total population, $S \equiv susceptible$, $I \equiv infected$, and $R \equiv removed$ (*removed* means either recovered or dead). Now let us q-generalize this model as follows:

$$\begin{aligned} \frac{d\left(S/N_{eff}\right)}{dt} &= -\beta\, \frac{S}{N_{eff}} \left(\frac{I}{N_{eff}}\right)^{q_{up}} \\ \frac{d\left(I/N_{eff}\right)}{dt} &= \beta\, \frac{S}{N_{eff}} \left(\frac{I}{N_{eff}}\right)^{q_{up}} - \gamma \left(\frac{I}{N_{eff}}\right)^{q_{down}} \\ \frac{d\left(R/N_{eff}\right)}{dt} &= \gamma \left(\frac{I}{N_{eff}}\right)^{q_{down}} \end{aligned} \tag{4}$$

with $q_{up} \leq 1$ and $q_{down} \geq 1$, where the bilinear term is generalized into a nonbilinear one and the *effective* population $N_{eff} = \rho N$ with $\rho \leq 1$. These equations generically have a single peak for $I(t)$. In all cases, we have $S(t) + I(t) + R(t) = N_{eff}$; moreover, $0 \leq S(t)/N_{eff}$, $I(t)/N_{eff}$, and $R(t)/N_{eff} \leq 1$. Consistently, in the set of **Eq. 4**, it is enough to retain the first two. Let us qualitatively compare the SIR and q-SIR models given by **Eqs 3** and **4**, respectively, by focusing on the β term; i.e., let us compare $\beta_{SIR}\,[S(t)/N]\,[I(t)/N]$ with $\beta_{qSIR}\,[S(t)/N_{eff}]\,[I(t)/N_{eff}]^{q_{up}} = \frac{\beta_{qSIR}}{[I(t)/N_{eff}]^{1-q_{up}}}\,[S(t)/N_{eff}]\,[I(t)/N_{eff}]$; we remind that $\rho = 1$ yields $N_{eff} = N$. It follows that roughly $\beta_{SIR} \simeq \frac{\beta_{qSIR}}{[I(t)/N_{eff}]}$. Since, before the peak, $I(t)$ steadily *increases* with time, a fixed value for $(1 - q_{up}) > 0$ acts qualitatively as a phenomenological time-dependent $\beta_{SIR}(t)$ which *decreases* with time. These tendencies are similarly realistic since they both reflect, each in its own manner, the generic action of pandemic authorities to isolate people in order to decrease the contagion represented by the β term in both models.

The particular limit $R(t) \equiv 0$ (hence $S(t) + I(t) = N_{eff}$) in **Eq. (4)** yields $I(t) = I(0)\, e_{q_{down}}^{-\gamma\,[I(0)/N_{eff}]^{q_{down}-1}\,t}$ if $\beta = 0$, and $I(t) = \frac{I(0)\, e^{\beta t}}{(I(0)/N_{eff})\, e^{\beta t} + [1 - I(0)/N_{eff}]}$ if $(\gamma, q_{up}) = (0, 1)$. For generic $q_{up} < 1$ and $R(t) \equiv 0$, **Eq. (4)** yields

$$\int_{S(0)/N_{eff}}^{S/N_{eff}} \frac{dx}{x(1-x)^{q_{up}}} = -\beta t, \tag{5}$$

and hence,

$$\begin{aligned} -\beta t = &\frac{\left(1 - S/N_{eff}\right)^{1-q_{up}} \Gamma\left(q_{up}\right)\, {}_2\tilde{F}_1\left(1, 1; q_{up}+1; N_{eff}/S\right)}{S/N_{eff}} \\ &- \frac{\left(1 - S(0)/N_{eff}\right)^{1-q_{up}} \Gamma\left(q_{up}\right)\, {}_2\tilde{F}_1\left(1, 1; q_{up}+1; N_{eff}/S(0)\right)}{S(0)/N_{eff}} \end{aligned}$$

where Γ is the Gamma function and $\tilde{F}$ is the regularized hypergeometric function. As an illustration, let us consider $q_{up} = 1/2$. It follows $\ln\left[\frac{1-\sqrt{1-S/N_{eff}}}{1+\sqrt{1-S/N_{eff}}} \times \frac{1+\sqrt{1-S(0)/N_{eff}}}{1-\sqrt{1-S(0)/N_{eff}}}\right] = -\beta t$; hence,

$$\frac{I(t)}{N_{eff}} = 1 - \frac{S(t)}{N_{eff}} = \left\{\frac{1 + \sqrt{I(0)/N_{eff}} - \left[1 - \sqrt{I(0)/N_{eff}}\right] e^{-\beta t}}{1 + \sqrt{I(0)/N_{eff}} + \left[1 - \sqrt{I(0)/N_{eff}}\right] e^{-\beta t}}\right\}^2. \tag{6}$$

Before the peak, $I(t)$ increases nearly exponentially if $q_{up} = 1$ and is roughly characterized by $I(t) \simeq t^{\frac{1}{1-q_{up}}}$ if $q_{up} < 1$. After the peak, $I(t)$ decreases exponentially if $q_{down} = 1$ and is roughly characterized by $I(t) \simeq t^{-\frac{1}{q_{down}-1}}$ if $q_{down} > 1$. These various aspects are illustrated in **Figure 1**. An important remark is necessary at this point. The possibility for nonbilinear coupling $\beta\,(S/N_{eff})\,(I/N_{eff})^{q_{up}}$ between subpopulations seems quite natural since nonhomogeneous mixing involves complex dynamics and networks for the susceptible and infected people, as well as for the infecting agent of the disease. But why would it be necessary to also allow, at the present phenomenological level, a nonlinear behavior for the one-subpopulation term $\gamma\,(I/N_{eff})^{q_{down}}$ itself? The answer might be found in nontrivial (multi)fractal-path-like relaxation mechanisms such as the one that is known to happen in reassociations in folded proteins [16, 17].

We have checked that the q-SIR model provides functions $I(t)$ that are numerically close but different from the quite performing ansatz in [5], namely, $I(t) \propto t^{\bar{\alpha}} e_{\bar{q}}^{-\bar{\beta} t^{\bar{\gamma}}}$, whose behavior before and after the peak is, respectively, $t^{\bar{\alpha}}$ and $t^{\bar{\alpha}-\bar{\gamma}/(\bar{q}-1)}$. Still, both approaches have power-law behaviors before and after the peak. We also checked $I(t) \propto e_{q_{up}}^{\beta_{up} t} \times e_{q_{down}}^{-\beta_{down} t}$, and the results are once again numerically close but nevertheless different from the ansatz in [5].

2.2 q-SEIR Model

The q-SIR model is not capable (for any choice of its parameters) of correctly fitting the epidemiologically crucial function $I(t)$ for the COVID-19 available data for various countries. Since this generalization of the simplest model does not provide a useful tool for COVID-19 data, we addressed a more sophisticated one, namely, a four-compartment model known as SEIR. Therefore, we next q-generalize the SEIR model with no vital dynamics (no births; no deaths), which is given by

$$\begin{aligned} \frac{dS}{dt} &= -\beta S \frac{I}{N} \\ \frac{dE}{dt} &= \beta S \frac{I}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{7}$$

with $S + E + I + R = N$, where E stands for *exposed*. We can generalize it as follows:

$$\begin{aligned} \frac{d\left(S/N_{eff}\right)}{dt} &= -\beta \frac{S}{N_{eff}} \left(\frac{I}{N_{eff}}\right)^{q_{up}} \\ \frac{d\left(E/N_{eff}\right)}{dt} &= \beta \frac{S}{N_{eff}} \left(\frac{I}{N_{eff}}\right)^{q_{up}} - \sigma \frac{E}{N_{eff}}, \\ \frac{d\left(I/N_{eff}\right)}{dt} &= \sigma \frac{E}{N_{eff}} - \gamma \left(\frac{I}{N_{eff}}\right)^{q_{down}} \end{aligned} \tag{8}$$

where once again we have generalized the bilinear couplings between subpopulations into nonbilinear ones and the linear γI term into a nonlinear one. Its particular instance $E(t) \equiv 0$ (hence, $S(t) + I(t) + R(t) = N_{eff}$) precisely recovers the q-SIR model, as defined here above. Notice that the *cumulative function* $C(t)$ of $I(t)$ is given by $C(t) \equiv \int_0^t dt' I(t') = \int_0^t dt' \left[\frac{1}{\gamma} \frac{dR(t')}{dt'}\right]^{1/q_{down}}$, which differs from the expression $C(t) = R(t)$ currently used in the SEIR model. It is of course possible to further generalize the above q-SEIR set of four equations by allowing in the right hand S^{q_s} (with $q_s \neq 1$) instead of S and E^{q_e} (with $q_e \neq 1$) instead of E, but no need has emerged to increase the number of free parameters of the model, since the allowance for $q_{up} < 1$ and for $q_{down} > 1$ appears to be enough for satisfactorily reproducing all the relevant features of the COVID-19 available data. Indeed, the variable which is epidemiologically crucial for avoiding a medical-hospital collapse in a given region is $I(t)$, and this time dependence generically appears to be very satisfactorily described by just allowing the possibility for $q_{up} \neq 1$ and/or $q_{down} \neq 1$. Notice that $\beta\,(I/N_{eff})^{q_{up}}$ is a *convex* function of (I/N_{eff}) for $0 < q_{up} < 1$ and $\gamma\,(I/N_{eff})^{q_{down}}$ is a *concave* function of (I/N_{eff}) for $q_{down} > 1$. These tendencies, illustrated in **Figure 2**, as well as the numerical values for the various coefficients of the model, are in agreement with the available medical/epidemiological evidence [21–25]. Notice also that, through $\tau \equiv \gamma t$ and $(\bar{I}/N_{eff}) \equiv (I/N_{eff})^{q_{down}}$, we can eliminate, without loss of generality, two fitting parameters (e.g., γ and q_{down}) within the set of **equations (8)**. Finally notice that if we consider $\beta = \sigma = 0$, we precisely recover **Eq. (1)** and its analytical solution in **Eq. (2)**. Therefore, even if the general analytical solution of the set of **equations (8)** is not available (due to mathematical intractability), the initial conditions naturally renormalize (in a nontrivial manner) the coefficients (β, σ, γ) of the model, and the effective population N_{eff} becomes a fitting parameter of the model. These renormalizations disappear of course if $q_{up} = q_{down} = 1$, i.e., for the standard SEIR model; concomitantly N_{eff} $(= \rho N$, with $\rho \leq 1)$ ceases being a fitting parameter and can be directly taken from the actual population N of the particular region under focus. For other mathematical aspects of nonlinear models such as the present one, the reader may refer to [26].

3 APPLICATION OF Q-SEIR MODEL TO COVID-19 PANDEMIC

In **Figures 3** and **4**, we have illustrations of this model for realistic COVID-19 cases. We identify the present variable I with the number of *active cases*,[1] as regularly updated online [27]. We verify that the description provided by the q-SEIR model for nonhomogeneous epidemiological mixing is indeed quite satisfactory for the early stages of the pandemic (before an unpredictable but possible second wave). Let us also mention that we have not followed here a road looking for the minimal number of free parameters, but rather a road where various realistic elements are taken into account, even if at the fitting-parameter level some of them might be redundant. Any further model yielding a deeper, or even first-principle, expression of exponents such as q_{up} and q_{down} in terms of microscopic/mesoscopic mechanisms is very welcome. This is by no means a trivial enterprise but, if successfully implemented, this would probably follow a road analogous to anomalous diffusion, namely, from Fourier's heat

[1]At this point, it should be noted that the variable I in the model represents both detected plus undetected active cases although, of course, the database includes only detected ones as active cases. However, since these cases are roughly proportional to each other (with a proportionality coefficient which might change from country to country), we use the model variable I to fit the active cases reflecting the incorporation of this proportionality into N_{eff} and hence into ρ.

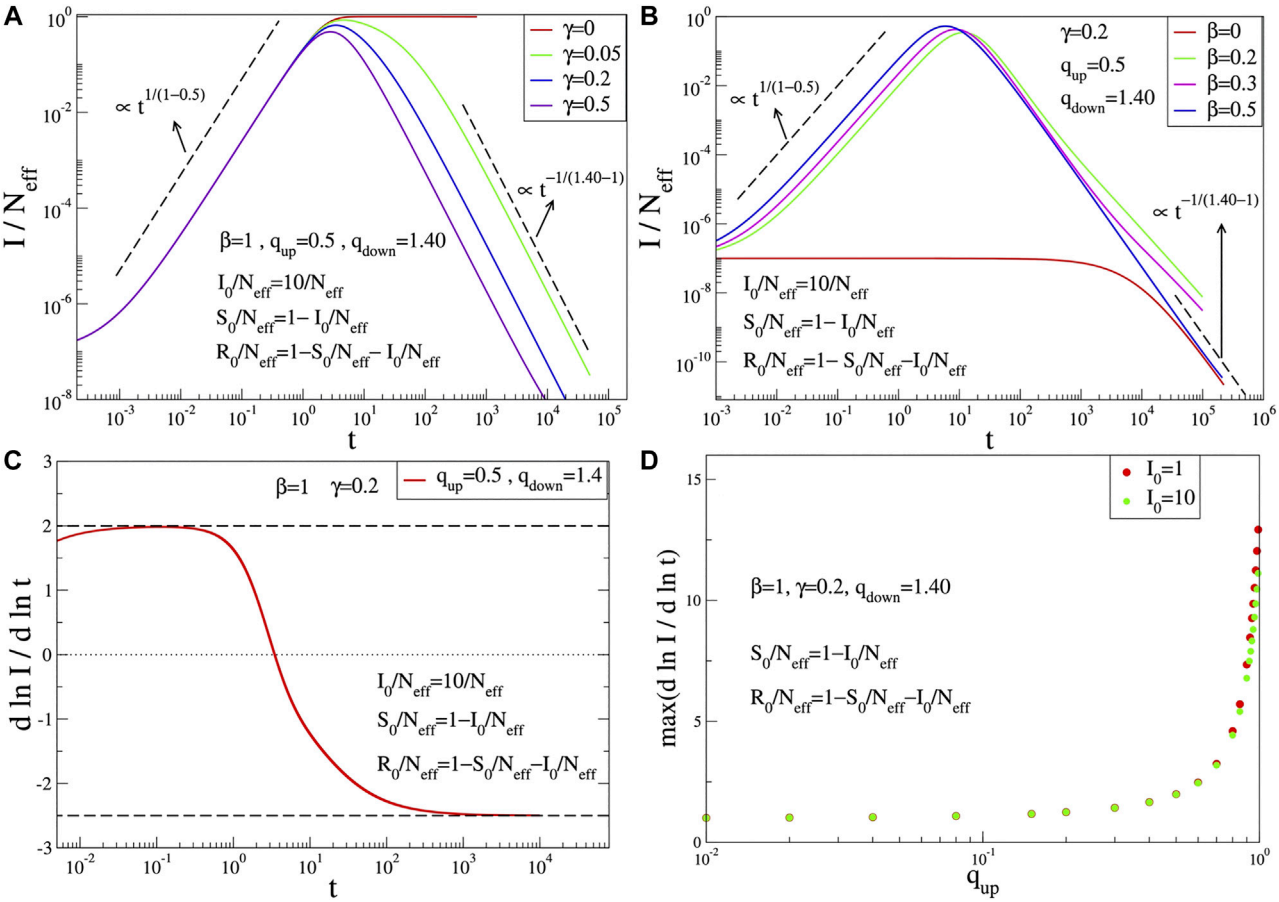


FIGURE 1 | Time evolution of $I(t)$ for the q-SIR equations with $N_{eff} = 10^8$; we consider t to run virtually from zero to infinity (not necessarily during only the typical range of real epidemics, say 1,000 days). **(A)** Fixed β and various values of γ. For $\gamma = 0$ and $\forall\beta$, $I(t)/N_{eff}$ precisely recovers the expression given in **Eq. (6)**. The slope at the first inflection point for increasing time is given by $max[d\ln I/d\ln t] = 1/(1-q_{up})$, $\forall\,\gamma$. **(B)** Fixed γ and various values of β. For $\beta = 0$, the q_{down}-exponential function is precisely recovered. **(C)** Time dependence of the slope of $[d\ln(I/N_{eff})/d\ln t]$, which appears to be bounded between $1/(1-q_{up})$ (= 2 in this example) and $-1/(q_{down}-1)$ (= −2.5 in this example) (dashed lines). The dotted line corresponds to the position of the peak of $I(t)$, and the maximum (minimum) corresponds to the left (right) inflection point. **(D)** The maximal slope of $I(t)/N_{eff}$ as a function of q_{up}. The high value at $q_{up} = 1$ reflects the divergence expected in the limit $q_{up} = q_{down} = 1$. Indeed, in this limit, the present q-SIR model recovers the standard SIR model, which increases exponentially (and not as a power-law) toward the corresponding peak. Notice also that this slope decreases when I_0/N_{eff} increases.

equation through Muskat's *Porous Medium Equation* [28] to Plastino and Plastino nonlinear Fokker-Planck equation [29], which in turn implied the scaling law $\bar{\alpha} = 2/(3-q)$ [30] ($\bar{\alpha}$ being defined through the scaling between x^2 and $t^{\bar{\alpha}}$ and q being the index value of the q-Gaussian solution for the nonlinear Fokker-Planck equation). This scaling law recovers, for $q = 1$, the Brownian motion scaling $\langle x^2\rangle \propto t$ and was impressively validated within 2% error in granular matter [31]. This phenomenological line was later legitimated on the basis of microscopic overdamped mechanisms in at least a wide class of systems, namely, providing $q = 0$ for the motion of vortices in type-II superconductors [32], and later extended to D-dimensional $1/r^{\lambda}$ short-range repulsive interactions ($\lambda/D \geq 1$), leading to $q = 1 - \lambda/D$ [33]. These approaches were shown to satisfy the zeroth law of thermodynamics, an H-theorem, and Carnot's cycle efficiency, with microscopically established analytical equations of state [34–37]. An attempt to follow along similar lines for the present q-SEIR model would surely be a very interesting challenge. In summary, we have q-generalized, through **Eq. (8)**, the SEIR epidemiological model. By solving this set of deterministic equations given the initial conditions and its parameters, we obtain $[S(t), E(t), I(t), R(t)]$, as well as the cumulative function $C(t)$. We have focused on $I(t)$ because the hardest quantities for satisfactorily fitting are the number of active cases and that of deaths and also because those are the most crucial quantities for making correct sanitary and epidemiological decisions.

From a general perspective, let us stress that the law of mass action, the Arrhenius relaxation law, and the Kramers

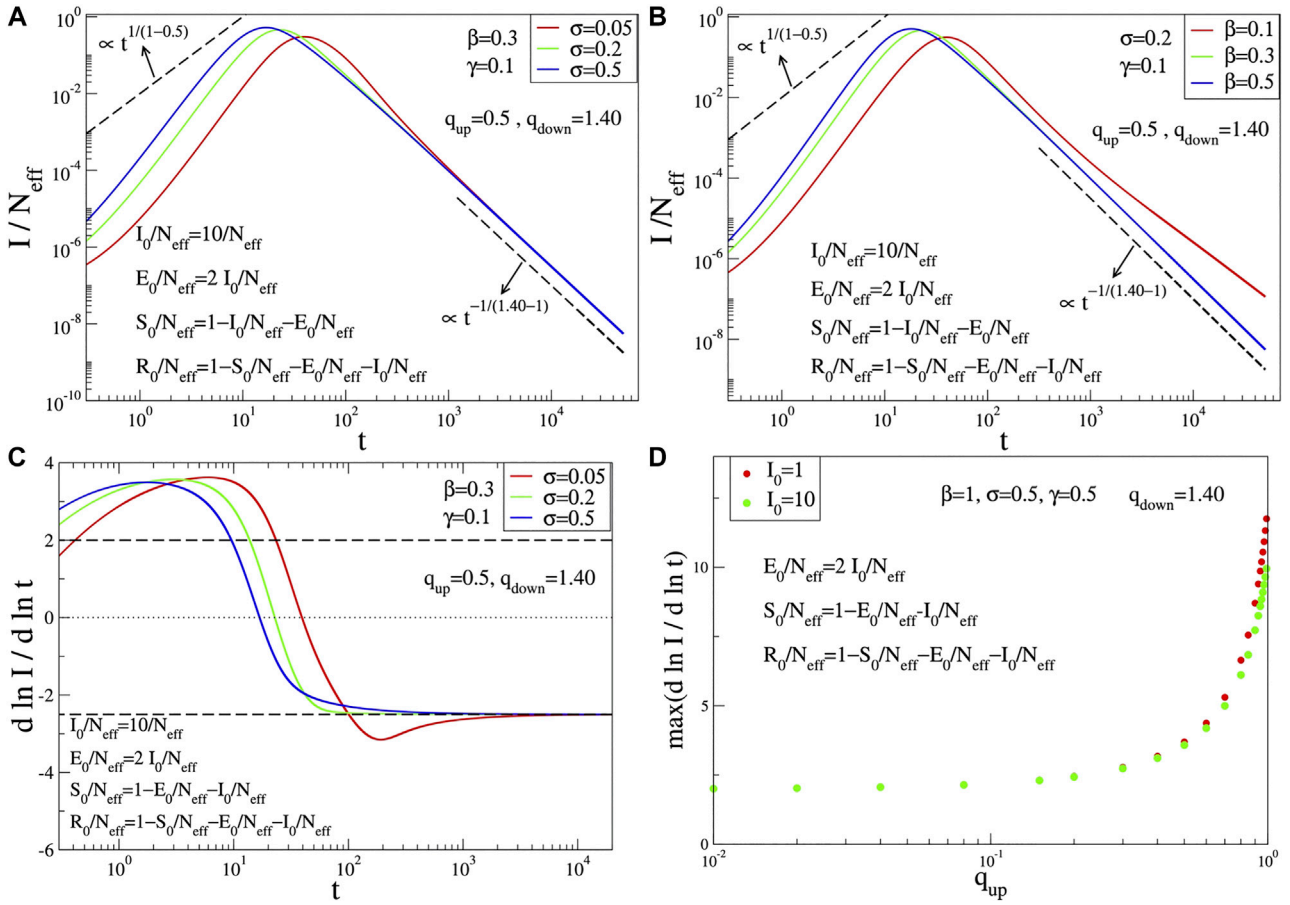


FIGURE 2 | Time evolution of $I(t)$ for the q-SEIR equations with $N_{eff} = 10^8$; we consider t to run virtually from zero to infinity (not necessarily during only the typical range of real epidemics, say 1,000 days). **(A)** Fixed (β, γ) and various values of σ. **(B)** Fixed (σ, γ) and various values of β. **(C)** Time dependence of the slope of $[d \ln I/d \ln t]$, which, in contrast with the q-SIR equations, is *not* bounded between $1/(1 - q_{up}) > 0$ and $-1/(q_{down} - 1) < 0$ (dashed lines). The dotted line corresponds to the position of the peak of $I(t)$, and the maximum (minimum) corresponds to left (right) inflection points. **(D)** The maximal slope of $I(t)/N_{eff}$ as a function of q_{up}. The high value at $q_{up} = 1$ reflects the divergence expected in the limit $q_{up} = q_{down} = 1$. Indeed, in this limit, the present q-SEIR model recovers the standard SEIR model, which increases exponentially (and not as a power-law) toward the corresponding peak. Notice also that this slope decreases when I_0/N_{eff} increases.

mechanism [38] of escape over a barrier through normal diffusion constitute pillars of contemporary chemistry. They are consistent with Boltzmann-Gibbs (BG) statistical mechanics and constitute some of its important successes. However, they need to be modified when the system exhibits complexity due to hierarchical space and/or time structures. It is along this line that a generalization has been proposed based on nonadditive entropies [20], characterized by the index q ($q = 1$ recovers the BG frame): see, for instance, [16, 39–43]. It is along these same lines that lies the present q-generalization of the standard SEIR model.

At the level of the numerical performance of the present q-SEIR model for the COVID-19 pandemic, it advantageously compares with models including time-dependent coefficients [44–47]. For instance, the SEIQRDP model [44, 45, 47] includes seven equations with several coefficients, two of them phenomenologically being time-dependent. It does fit rather well the COVID-19 reported data until a given date. However, the q-SEIR, which includes four (instead of seven) equations with several coefficients, all of them being fixed in time, fits definitively better the same data for all the countries that we have checked: see illustrations in **Figure 5**.

At this stage, let us emphasize a rather interesting fact. Neither the SIR nor the SEIR models distinguish the *dead* from the *recovered*, within the *removed* (R) subpopulation. However, the same values for (q_{up}, q_{down}) fit satisfactorily *both* the numbers of active cases and of deaths per day for a given country, as shown in **Figures 3** and **4**. At this point, it would be worth noting that although the present model, unlike, for example, the SEIRD model [48, 49], does not distinguish deaths from healings, the deceased cases will still be roughly proportional to infected people. It is this proportionality, we believe, which makes us obtain reasonable fits using the variable I of the model, again incorporating the proportionality into N_{eff}

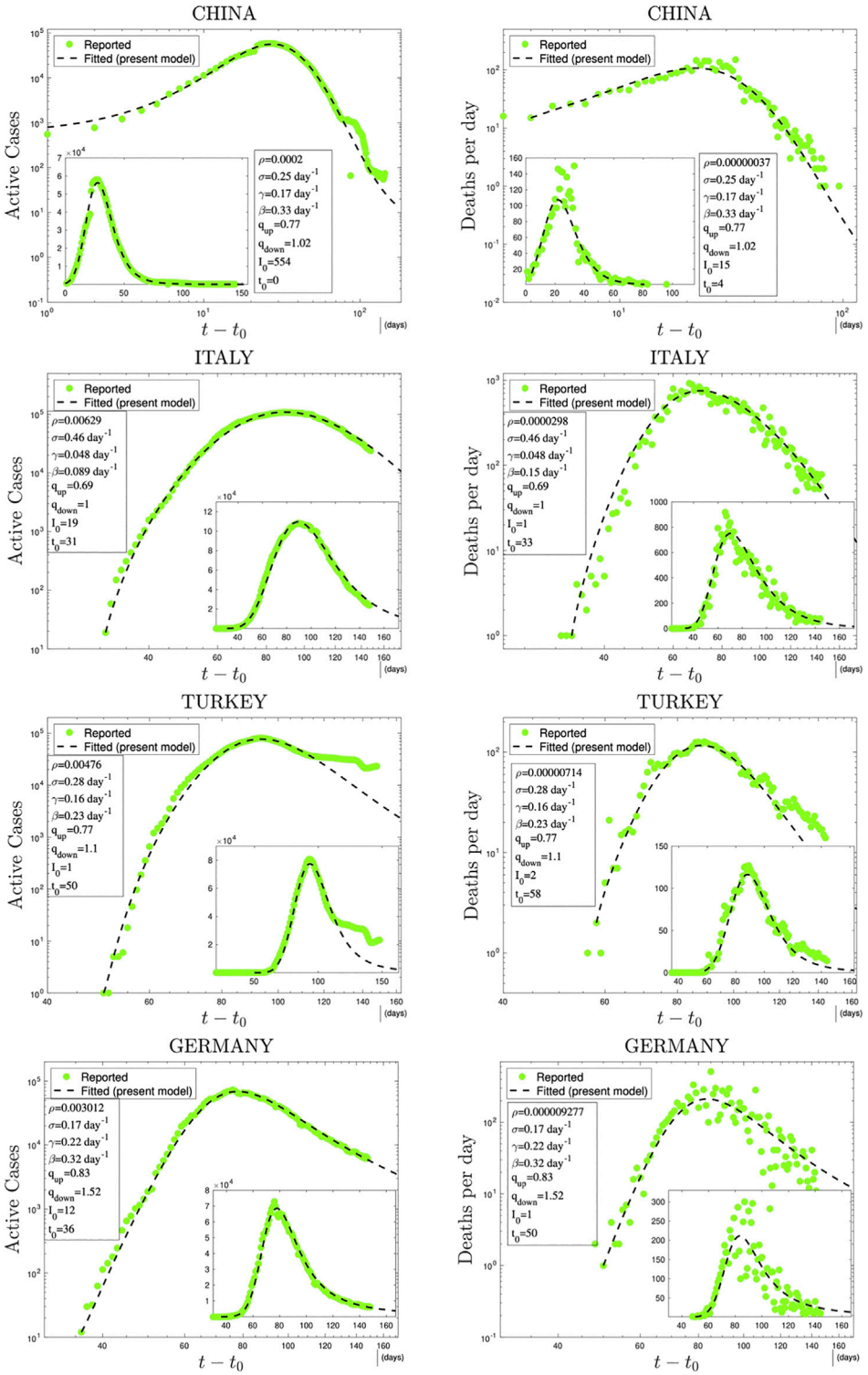


FIGURE 3 | Time evolution of available data for COVID-19 numbers of active cases (probably under-reported in most cases) and deaths per day [27] and their (linear scale) Least Squares Method fittings with $N_{eff}\,I(t)$ from the q-SEIR model; $\rho = N_{eff}/N$, where N is the population of the country. Notice that, (i) by convention, $t_0 = 0$ for China; (ii) parameters such as ρN are particularly relevant for sanitary-epidemiological decisions, and, as it is natural, $\rho\,(deaths) \ll \rho\,(active)$ for any given country; (iii) for any given country, the values of (q_{up}, q_{down}) are the same for both curves of active cases and of deaths; (iv) the dates of the peaks of the active cases and deaths per day do not necessarily coincide; (v) the values that emerge for β/γ (*reproduction number* or *growth rate*), $1/\beta$ (*exposition time*), $1/\gamma$ (*recovering time*), and $1/\sigma$ (*incubation time*) are consistent with those currently indicated in the literature [21–25]; (vi) we considered all the data reported until June 13, excluding some very initial transients or sudden anomalous discrepancies (e.g., in China, Turkey, and Brazil).

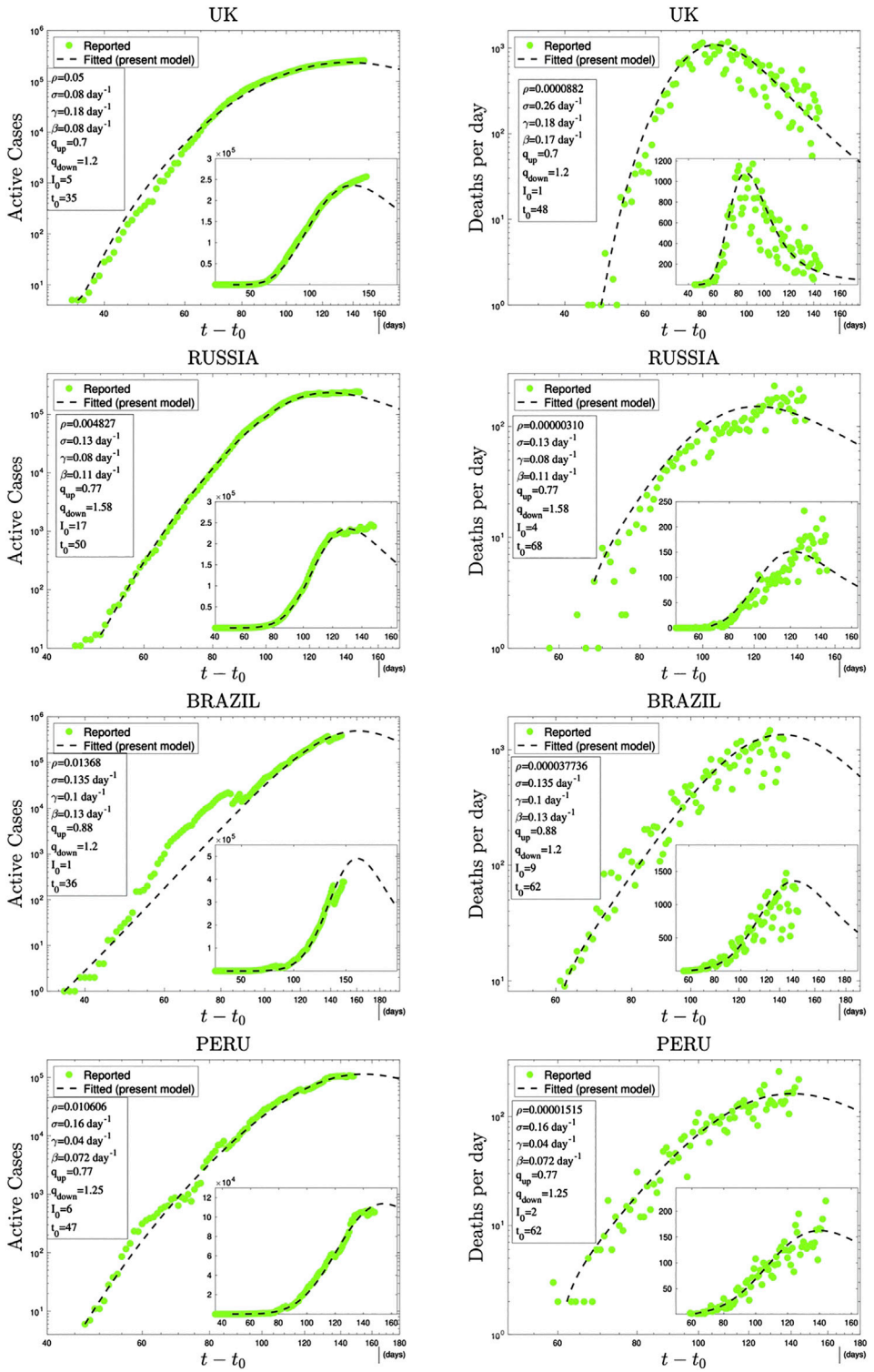


FIGURE 4 | Continuation of **Figure 3**.

and hence into ρ. It is also evident from these figures that the data for deaths per day are more scattered than those for active cases. This is due to the fact that the real data for the former are much less than for the latter. This is also the reason for the significantly small ρ values of deaths per day compared to those of active cases.

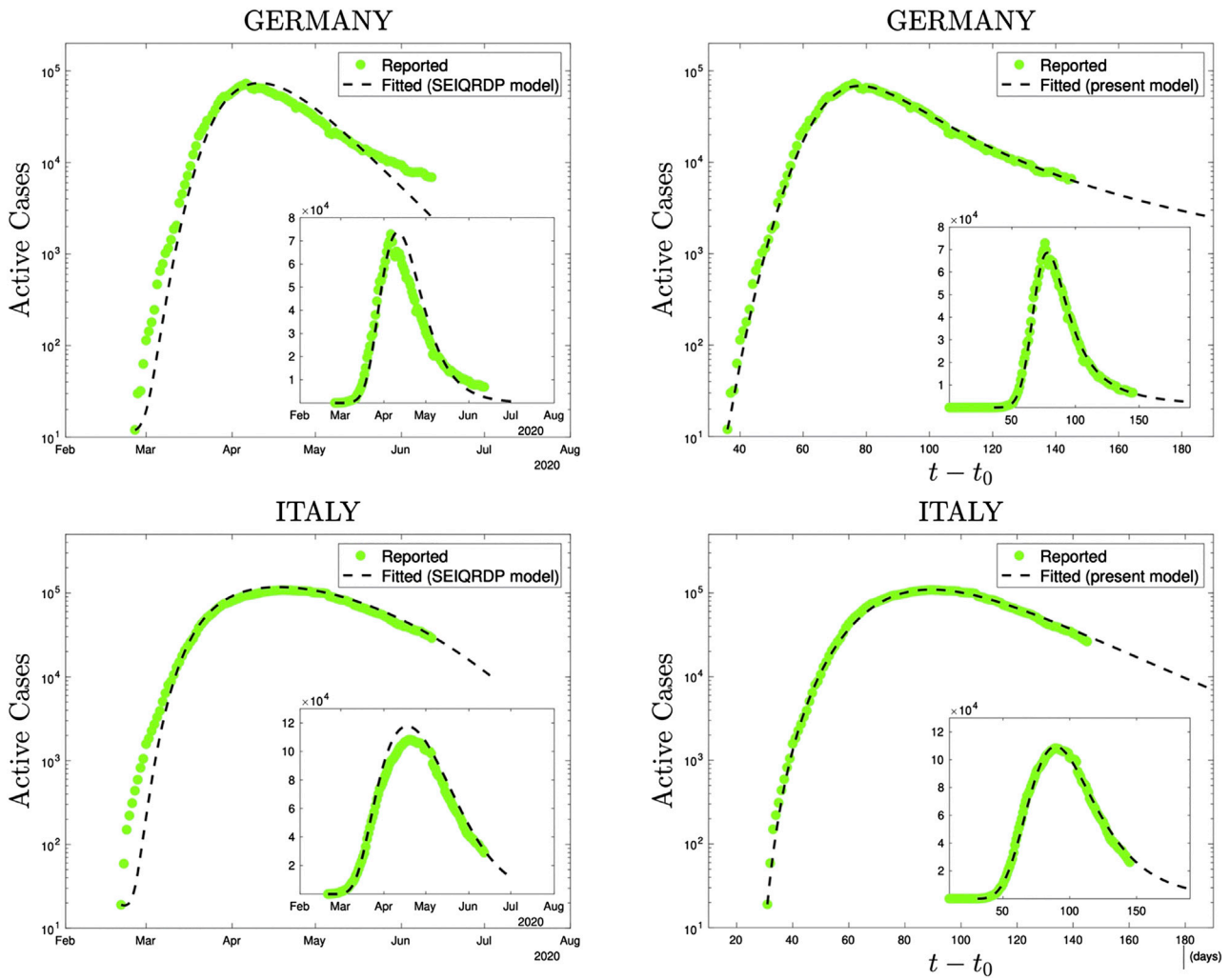


FIGURE 5 | Comparison, using precisely the same reported data (green dots), of the SEIQRDP model (left plots) and the q-SEIR model (right plots) for the time series of Germany (from February 26th to June 13th 2020) and of Italy (from February 21st to June 13th 2020). To obtain the SEIQRDP fittings (seven linear/bilinear equations satisfying $S + E + I + Q + R + D + P = N$ and including several fixed as well as two time-dependent coefficients), we have used the online program [47]. To obtain the q-SEIR fittings (four not necessarily linear/bilinear equations satisfying $S + E + I + R = N_{eff}$ and including only fixed parameters, two of them being the nonlinear exponents q_{up} and q_{down}), we have used a Standard Least Squares Method (linear scales).

4 CONCLUSION

To conclude, let us remind that the q-SEIR model recovers, as particular instances, the q-SIR model introduced here, as well as the traditional SEIR and SIR ones. It has, however, an important mathematical difference with the usual epidemiological models. Virtually all these models (SIR, SEIR, SAIR, SEAIR, SIRASD, SEAUCR, and SEIQRDP) are defined through equations that are multilinear in their variables; i.e., that are linear in each one of its variables. This multilinearity disappears in models such as the present q-SIR and q-SEIR ones if either q_{up} or q_{down} differs from unity. Consequently N_{eff} definitively plays a different role since it sensibly enters within the set of fitting parameters of the q-generalized models; its precise interpretation remains to be elucidated, perhaps in terms of the sociogeographical circumstances of that particular region. Last but not least, let us stress that the aim of the present q-SEIR model is to *mesoscopically describe a single epidemiological peak*, including its realistic *power-law* growth and relaxation in the time evolution of the number of active cases, and by no means to qualitatively address possibilities such as the emergence of two or more peaks, a task which is (sort of naturally, but possibly less justified on fundamental grounds) attainable within approaches using traditional (multilinear) models where one or more coefficients are allowed to phenomenologically depend on time by realistically adjusting their evolution along the actual epidemics. Alternatively, it is always possible to approach the two-peak case by proposing a linear combination of two q-SEIR curves

starting each of them at two different values of the departing time (t_0).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

AUTHOR CONTRIBUTIONS

All authors contributed equally to the article.

ACKNOWLEDGMENTS

We have greatly benefitted from very fruitful discussions with R. Dickman, T. Pereira and D. Eroglu, as well as from partial financial support by CNPq and Faperj (Brazilian agencies).

REFERENCES

1. Daley DJ Gani JM. *Epidemic modelling: an introduction*. Cambridge: Cambridge University Press (1999).
2. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface* (2007) 4:879. doi:10.1098/rsif.2007.1100
3. Burr TL Chowell G. Signatures of non-homogeneous mixing in disease outbreaks, *Math Comput Model* (2008) 48:122. doi:10.1016/j.mcm.2007.09.009
4. Milwid RM, O'Sullivan TL, Poljak Z, Laskowski M, Greer AL. Comparing the effects of non- homogenous mixing patterns on epidemiological outcomes in equine populations: a mathematical modelling study, *Sci Rep* (2019) 9:3227. doi:10.1038/s41598-019-40151-2
5. Tsallis C Tirnakli U. Predicting COVID-19 peaks around the world. *Front Phys* (2020) 8, 217. doi:10.3389/fphy.2020.00217
6. Sebastiani G, Massa M, Riboli E. COVID-19 epidemic in Italy: evolution, projections and impact of government measures. *Eur J Epidemiol* (2020) 35: 341. doi:10.1007/s10654-020-00631-6
7. Chowdhury R, Heng K, Shawon MSR, Goh G, Okonofua D, Ochoa-Rosales C, et al. Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries, *Eur J Epidemiol* (2020) 35:389. doi:10.1007/s10654-020-00649-w
8. Wang L, Chen H, Qiu S, Song H, Evaluation of control measures for COVID-19 in Wuhan, China. *J Infect* (2020) 81:318. doi:10.1016/j.jinf.2020.03.043
9. Manchein C, Brugnago EL, da Silva RM, Mendes CFO, Beims MW. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies, *Chaos* (2020) 30:041102. doi:10.1063/5.0009454
10. Ashurov R Umarov S. *Determination of the order of fractional derivative for subdiffusion equation*. arxiv 13468 (2020).
11. Ziff AL Ziff RM, et al. *Fractal kinetics of COVID-19 pandemic (with update 3/1/20)* medRxiv preprint (2020). doi:10.1101/2020.02.16.20023820
12. Vasconcelos GL, Macedo AMS, Ospina R, Almeida FAG, Duarte-Filho GC, Brum AA, et al. Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *PeerJ* (2020) 8:e9421. doi:10.7717/peerj.9421
13. Curado EMF Curado MR. A discrete-time-evolution model to forecast progress of COVID-19 outbreak. *PLoS ONE* (2020) 15:e0241472. doi:10.1371/journal.pone.0241472
14. Costa GS, Cota W, Ferreira SC, *Metapopulation modeling of COVID-19 advancing into the countryside: an analysis of mitigation strategies for Brazil*. medRxiv preprint (2020). doi:10.1101/2020.05.06.20093492
15. Pires MA, Crokidakis N, Cajueiro DO, de Menezes MA, Queiros SMD. What is the potential for a second peak in the evolution of SARS-CoV-2 in Brazil? Insights from a SIRASD model considering the informal economy. *Preprint* (2020) 2005:09019. doi:10.1016/j.meegid.2020.104502
16. Tsallis C, Bemski G, Mendes RS. Is re-association in folded proteins a case of nonextensivity?, *Phys Lett* (1999) 257:93. doi:10.1016/s0375-9601(99)00270-4
17. Austin RH, Beeson K, Eisenstein L, Frauenfelder H, Gunsalus IC, Marshall VP. Activation energy spectrum of a biomolecule: photodissociation of carbonmonoxy myoglobin at low temperatures. *Phys Rev Lett* (1974) 32: 403. doi:10.1103/physrevlett.32.403
18. Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel. *Nature* (2006) 439:26. doi:10.1038/nature04292
19. Brockmann D, David V, Morales Gallardo A. Human mobility and spatial disease dynamics. In: C Chmelik, N Kanellopoulos, J Kärger, D Theodorou, editors *Diffusion fundamentals III*. Leipzig: Leipziger Universitatsverlag (2009).
20. Tsallis C. Possible generalization of Boltzmann–Gibbs statistics. *J Stat Phys* (1988) 52:479. doi:10.1007/bf01016429
21. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* (2020) 5: e261–270. doi:10.1016/s2468-2667(20)30073-6
22. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel Coronavirus-infected pneumonia. *N Engl J Med* (2020) 382:1199. doi:10.1056/NEJMoa2001316
23. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha FA, Röst G. Risk assessment of novel Coronavirus COVID-19 outbreaks outside China. *JCM* (2020) 9:571. doi:10.3390/jcm9020571
24. Yuan G, Li M, Lv G, Lu ZK. Monitoring transmissibility and mortality of COVID-19 in Europe. *Int J Infect Dis* (2020) 95:311–5. doi:10.1016/j.ijid.2020.03.050
25. Wang G, Wang Z, Dong Y, Chang R, Xu C, Yu X, et al. Phase-adjusted estimation of the number of Coronavirus disease 2019 cases in Wuhan, China. *Cell Discov* (2020) 6:10. doi:10.1038/s41421-020-0148-0
26. Liu W-m, Levin SA, Iwasa Y. Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. *J Math Biol* (1986) 23:187. doi:10.1007/bf00276956
27. https://www.worldometers.info/coronavirus/#countries; https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases
28. Muskat M. *The flow of homogeneous fluids through porous media*. New York: Springer (1937).
29. Plastino AR Plastino A. Non-extensive statistical mechanics and generalized Fokker-Planck equation, *Phys Stat Mech Appl* (1995) 222:347. doi:10.1016/0378-4371(95)00211-1
30. Tsallis C Bukman DJ. Anomalous diffusion in the presence of external forces: exact time-dependent solutions and their thermostatistical basis. *Phys Rev E* (1996) 54:R2197. doi:10.1103/physreve.54.r2197
31. Combe G, Richefeu V, Stasiak M, Atman APF. Experimental validation of nonextensive scaling law in confined granular media, *Phys Rev Lett* (2015) 115: 238301. doi:10.1103/physrevlett.115.238301
32. Andrade JS, Jr., da Silva JFT, Moreira AA, Nobre FD, Curado EMF. Thermostatistics of overdamped motion of interacting particles, *Phys Rev Lett* (2010) 105:260601. doi:10.1103/physrevlett.105.260601
33. Moreira AA, Vieira CM, Carmona HA, Andrade JS, Jr., Tsallis C. Overdamped dynamics of particles with repulsive power-law interactions. *Phys Rev E* (2018) 98:032138. doi:10.1103/physreve.98.032138
34. Curado EMF, Souza AMC, Nobre FD, Andrade RFS. Carnot cycle for interacting particles in the absence of thermal noise. *Phys Rev E* (2014) 89: 022117. doi:10.1103/physreve.89.022117
35. Andrade RFS, Souza AMC, Curado EMF, Nobre FD. A thermodynamical formalism describing mechanical interactions, *Europhys Lett* (2014) 108: 20001. doi:10.1209/0295-5075/108/20001
36. Nobre FD, Curado EMF, Souza AMC, Andrade RFS. Consistent thermodynamic framework for interacting particles by neglecting thermal noise, *Phys Rev E* (2015) 91:022135. doi:10.1103/physreve.91.022135
37. Souza AMC, Andrade RFS, Nobre FD, Curado EMF. Thermodynamic framework for compact q-Gaussian distributions. *Phys Stat Mech Appl* (2018) 491:153. doi:10.1016/j.physa.2017.09.013

38. Kramers HA. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* (1940) 7(4):284. doi:10.1016/s0031-8914(40)90098-2
39. Lenzi EK, Anteneodo C, Borland L. Escape time in anomalous diffusive media. *Phys Rev E* (2001) 63:051109. doi:10.1103/physreve.63.051109
40. Aquilanti V, Borges EP, Coutinho ND, Mundim KC, Carvalho-Silva VH. *From statistical thermodynamics to molecular kinetics: the change, the chance and the choice*. Rendiconti Lincei: Scienze Fisiche e Naturali (2018).
41. Mundim KC, Baraldi S, Machado HG, Vieira FMC. Temperature coefficient (Q10) and its applications in biological systems: beyond the Arrhenius theory. *Ecol Model* (2020) 431:109127. doi:10.1016/j.ecolmodel.2020.109127
42. Tsallis C, Gell-Mann M, Sato Y. Asymptotically scale-invariant occupancy of phase space makes the entropy Sq extensive. *Proc Natl Acad Sci Unit States Am* (2005) 102:15377–82. doi:10.1073/pnas.0503807102
43. Gazeau JP Tsallis C. Möbius transforms, cycles and q-triplets in statistical mechanics. *Entropy* (2019) 21:1155. doi:10.3390/e21121155
44. Peng L, Yang W, Zhang D, Zhuge C, Hong L. *Epidemic analysis of COVID-19 in China by dynamical modeling*. medRxiv (2020). doi:10.1101/2020.02.16.20023465
45. Godio A, Pace F, Vergnano A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *IJERPH* (2020) 17:3535. doi:10.3390/ijerph17103535
46. Lopez L Rodo X. *A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics*. medRxiv (2020). doi:10.1101/2020.03.27.20045005
47. Cheynet E. *Generalized SEIR epidemic model (fitting and computation)*. Berlin: Springer (2020).
48. Piccolomini EL Zama F. Monitoring Italian COVID-19 spread by a forced SEIRD model. *PLoS ONE* (2020) 15:e0237417. doi:10.1371/journal.pone.0237417
49. Casas PF, Carrasco VG, Subirana JG. SEIRD COVID-19 formal characterization and model comparison validation. *Appl Sci* (2020) 10:5162. doi:10.3390/app10155162

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 21 December 2020
doi: 10.3389/fphy.2020.573322

Epidemiological Characteristics of COVID-19 in Mexico and the Potential Impact of Lifting Confinement Across Regions

Cristy Leonor Azanza Ricardo[1] *and Esteban A. Hernandez-Vargas*[2]*

[1]Centro de Física Aplicada y Tecnologia Avanzada, Universidad Nacional Autónoma de México, Juriquilla, Mexico, [2]Instituto de Matemáticas, Universidad Nacional Autonoma de Mexico, Juriquilla, Mexico

The novel coronavirus SARS-CoV-2 has paralyzed our societies, leading to self-isolation and quarantine for several days. As the 10th most populated country in the world, Mexico is on a major threat by COVID-19 due to the limitations of intensive care capacities, about 1.5 hospital beds for every 1,000 citizens. In this paper, we characterize the COVID-19 pandemic in Mexico and projected different scenarios to evaluate sharp or gradual quarantine lifting strategies. Mexican government relaxed strict social distancing regulations on June 1, 2020, deriving to pandemic data with large fluctuations and uncertainties of the tendency of the pandemic in Mexico. Our results suggest that lifting social confinement must be gradually sparse while maintaining a decentralized region strategy among the Mexican states. To substantially lower the number of infections, simulations highlight that a fraction of the population that represents the elderly should remain in social confinement (approximately 11.3% of the population); a fraction of the population that represents the confined working class (roughly 27% of the population) must gradually return in at least four parts in consecutive months; and to the last a fraction of the population that assumes the return of students to schools (about 21.7%). As the epidemic progresses, deconfinement strategies need to be continuously re-adjusting with the new pandemic data. All mathematical models, including ours, are only a possibility of many of the future, however, the different scenarios that were developed here highlight that a gradual decentralized region deconfinement with a significant increase in healthcare capacities is paramount to avoid a high death toll in Mexico.

Keywords: COVID-19, confinement, control strategies, epidemiology, risk, Mexico

OPEN ACCESS

Edited by:
Lin Wang,
University of Cambridge,
United Kingdom

Reviewed by:
Sen Pei,
Columbia University, United States
Lin Yang,
Hong Kong Polytechnic University,
Hong Kong
Gaoxi Xiao,
Nanyang Technological University,
Singapore

***Correspondence:**
Esteban A. Hernandez-Vargas
esteban@im.unam.mx

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *16 June 2020*
Accepted: *15 September 2020*
Published: *21 December 2020*

Citation:
Azanza Ricardo CL and Hernandez-Vargas EA (2020) Epidemiological Characteristics of COVID-19 in Mexico and the Potential Impact of Lifting Confinement Across Regions. Front. Phys. 8:573322. doi: 10.3389/fphy.2020.573322

1. INTRODUCTION

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the virus behind the 2019 coronavirus disease (COVID-19) with alarming levels of spread and death tolls worldwide. With more than four million confirmed cases and 292,046 deaths [1], COVID-19 pandemic has spread 212 countries moving the epicentre from China to Europe and consequently to America [1]. While potential vaccines and antiviral drugs are under fast development [2], epidemiological models have underlined the relevance of social distancing interventions as the main weapon so far to mitigate COVID-19 pandemic [3].

Epidemiological models have played a central role to advance our understanding in SARS-CoV-2 transmission [3–6]. In its early stages, the epidemic can double in size every 7.4 days [7]. The case fatality rate for COVID-19 ranges from 0.3 to 1% [1] up to 20% [8]. The basic reproductive number has been computed roughly 2.2 (95% CI, 1.4–3.9) [4, 7]. The lesson learned from China, Italy, and the United States pointed out that COVID-19 can quickly result in high demands of healthcare capacities deriving in the collapsing of hospitals of well-resourced nations [9, 10].

Developing countries with limitations of intensive care beds are highly jeopardized by COVID-19. Mexico ranks as the 10th most populated country in the world with about 127 million people [11]. About 60.4% of the population is economically active whereof 56.2% depends on the informal workforce [11]. Albeit the economic impact and levels of moderate poverty, Mexico adopted on March 15, 2020 social confinement, which was relatively early compared to the confirmed number of cases and deaths [12]. By the cut-off date of May 25, 2020, the public health strategy in Mexico has resulted in 71,105 infected cases and 7,633 deaths by COVID-19 [12].

The public health strategy followed by Mexico for the COVID-19 pandemic is the sentinel model, which was implemented by Mexico in 2009 against the pandemic of influenza H1N1 [13]. The sentinel model consists on three phases. In stage 1 of the outbreak, the model applies case-by-case monitoring. In phase two, the model focuses on places where the epidemic is growing, at this stage a community-based surveillance is applied. In phase three, which is the peak infection phase, the central problem of pandemic surveillance is not to monitor the growth of the pandemic but to ensure that hospital capacities are not exceeded. Mexico has an extremely limited healthcare capacity, approximately 1.5 beds for every 1,000 citizens. Based on early data of the pandemic and assuming social confinement restrictions, the Sub-secretary of Prevention and Health Promotion of Mexico presented mathematical model predictions with a peak of the pandemic Mexico between 8 and 10 May with an end of the pandemic on June 25, 2020 [12]. Mexico is planning to lift the strict social distancing regulations on June 1, 2020. However, an abruptly lifting of social confinement would likely result in new waves of new COVID-19 cases and high death tolls.

While the SIR model has a simple structure, this has been central to understand different epidemiological aspects of COVID-19 such as the reproductive number, number of infected cases in short time scales, and the effects of lifting confinement strategies in different countries [3–5, 14]. Simple models can help us to think clearly and to interpret pandemic data sets on quantitative grounds. Consequently, simulations can infer predictions of different possible scenarios. If the predictions are confirmed, the model has proven to be a justified simplification of the complex reality and may be further used. If the predictions are wrong, the model assumptions can be reconsidered or extended, which is very important during pandemics. In this paper, we fit a mathematical model for COVID-19 [15] using data of the COVID-19 epidemic in Mexico [12] as well as its public health capacities and demographic conditions [11]. Different lifting confinement scenarios are evaluated for the main regions of Mexico in order to inform public makers to tailor decentralized region strategies through the Mexican territory with the ultimate goal to minimize deaths.

2. THE SENTINEL MODEL FOR TRACKING COVID-19 IN MEXICO

Epidemiological models have been proposed to represent COVID-19 pandemic in different countries [15, 16]. These contain many variables resulting in a significant number of parameters difficult to estimate with epidemiological data at the moment of the pandemic. Instead of considering the exposed compartment and then divided into presymptomatic, symptomatic, and asymptomatic, we assume that the exposed hosts to the virus go from susceptible to presymptomatic. Consequently, presymptomatic cases would progress to symptomatic or to recover. In this fashion, we do not increase the complexity of the model. Furthermore, several assumptions would need to be integrated in order to match the sentinel model used by Mexico against COVID-19 [13]. The model writes as follow:

$$\frac{dS}{dt} = -\beta\frac{SA}{N} + \tau C - qS, \tag{1}$$

$$\frac{dC}{dt} = qS - \tau C, \tag{2}$$

$$\frac{dA}{dt} = \beta\frac{SA}{N} - \eta A, \tag{3}$$

$$\frac{dI}{dt} = \epsilon\eta A - \delta I, \tag{4}$$

$$\frac{dR}{dt} = (1-\epsilon)\eta A + \delta I. \tag{5}$$

The susceptible population is represented by S, from which the total country population (N) is about 127 million [11]. The susceptible population would pass to the presymptomatic compartment (A). Note that in the sentinel model applied in Mexico is that testing is about 10% of the suspected cases while 100% of hospitalized cases and deaths. Once the infected cases (I) are listed in the epidemiological data sets by the Mexican Government, the infected case should be either hospitalized or at home under confinement. Based on the fact that presymptomatic shedding is very common for COVID-19 [17], it is a reasonable assumption to consider that new infections are mainly driven by the presymptomatic population with the term $\beta\frac{SA}{N}$. The presymptomatic phase has been reported on average 5–6 days, but it can be up to 14 days [1, 3]. The presymptomatic population will leave this compartment with a rate η. A fraction ϵ of the presymptomatic population that will present severe symptoms or complications, and consequently counted in the reports of infected cases (I). This is approximately 20% [1, 3, 18], that is $\epsilon = 0.2$. The other fraction $1-\epsilon$ will move to the compartment of remove (R), those cases that recover or die. The average infection time is about 3–7 days [19], while the clinical recovery is 2–6 weeks after symptoms [20, 21]. Thus, we assume the recovery rate of infected cases (δ) in 10 days.

TABLE 1 | Parameters of the sentinel model (1)–5) for tracking COVID-19 in Mexico.

Parameters	Value (range)	References
Cases with severe symptoms, ϵ	0.2 (0.15–0.3)	[1, 3, 18]
Recovery time, δ	2 (1.5–3) weeks	[20, 21]
Factor for estimate presymptomatic cases, f	10 (8–12)	[22]
Reduction in the mobility C_S	60 (50–70)%	[23]
Confinement rate, q	1 (0.9–1.1)	Fixed
Infection rate, β	1.48	Estimated
Incubation period, $1/\eta$	0.55	Estimated
Reproductive number, R	2.7 (2.4–3.6)	Estimated

The first confirmed case of COVID-19 in Mexico was reported on February 28th, 2020 [12]. On the day of social confinement, March 15, 2020, the number of infected cases (I_0) for the whole country was 12 [12]. Due to the change in social movement as well as the very low number of infected cases between the first confirmed case till the day of social confinement, we perform our parameter fitting at the first day of social confinement. The initial number of presymptomatic population (A_0) is approximately a factor f of 8–12 respect to the confirmed cases [22], which approximates $A_0 = fI_0$.

The compartment C is the confined population which entered with a rate q from the susceptible population. τ the effect of population deconfinement, which can be attributed to government policies as well as resistant/forgetting factor by the population to keep adequately the confinement. Estimations by the COVID-19 community mobility report of Google [23] suggest a reduction in the mobility of about 50–70%. Thus, the ratio between $\tau/q = 2/3$ would represent the 60% confinement percent in the steady state. If we consider a fast confinement rate, that is $q = 1$, then $\tau = 2/3$.

2.1. Re-sampling Strategy for Parameter Distributions

A re-sampling strategy was employed to fit the parameters β and η. The range of parameter values in **Table 1** served to generate 3,000 sets of random parameters with specific statistical distributions as presented in **Supplementary Figure S1**. This set of parameters consequently was used to fix parameters and perform 3,000 fitting repetitions by minimizing the root mean square (RMS) difference between the model predictive output ($\bar{y}_i$) and the reported data of infected cases (y_i). Among different several stochastic optimization algorithms, the minimization of RMS is performed using the Differential Evolution (DE) algorithm [24] with our Python library *PDEparams* [25]. Using the python library *st.t.interval* and parameter fitting distributions based on 3,000 samples, the 95% confidence interval (CI) is computed.

3. SIMULATION RESULTS

Social confinement started in Mexico on March 15, 2020, which was early compared to the confirmed number of cases and deaths in comparison to other countries [12]. Albeit of this early confinement, Mexico had an exponential growth of infected cases and many deaths, thus confinement needed to be extended. Pandemic data in Mexico is very heterogeneous among regions as well as in time scales. This may be attributed to different social behavior, working class, and public health policies.

3.1. Scenario 1: COVID-19 Pandemic Under Confinement

The first scenario to investigate, Scenario 1, is the reports of infected cases from 15 March till May 25, 2020 [12]. This time period is when a major part of the population remained in social confinement, about 50–70% [23]. The embedded panel in **Figure 1** presents the best fit and sampling procedure for the whole country, Mexico City (CMX) and the State of Mexico (MEX), while the other regions with major number of infected cases is shown in **Supplementary Figure S2**. Simulations show that while the early social confinement would delay the initial expected peak of the pandemic on middle may [12] till July-August, the number of infected per day would reach very high numbers, an order of 10^4–10^5.

In fact, reproductive numbers for the whole country as well as major regions affected with COVID-19 were between 2.3 and 3.6 (**Supplementary Table S1**). This underlines that even social restrictions SARS-CoV-2 was still spreading among the country and major populated cities. The heterogeneity of the infections and their respective public health necessities varied significantly among different geographic regions of Mexico. Note in **Figure 2** that some states had R_0 larger than 4, which is not reasonable. This is attributed to the data of the respective states are having a small number of infected cases, deriving in bias predictions.

Considering the reported number of infected cases, **Figure 3** presents the percent of the COVID-19 infected cases that require Intensive Care Units beds (ICU), intubation facilities (INT) and hospital beds (HOS) from 12 April till August 15, 2020 [26]. Large box plots imply that the requirements of the state were changing a lot in time e.g., Zacatecas (Zac) and Colima (Col). On the other hand, small box plots highlight that public health requirements for the state were very consistent during the reported period e.g., CMX.

Assuming the percents of required hospital beds with respect to the reported infected cases are consistent in the future (about 38% for the whole country, see **Figure 3** and **Supplementary Table S2**), the most optimistic peak of the pandemic presented in **Figure 1** could be about 254,174 infected cases, which would result in a saturation of hospitals. That is, the whole country would require to have about 97,180 hospital beds, while the number available of beds for COVID-19 pandemic reported for the whole country is about 49,083 [27]. This is a major public concern as Mexico would only be able to provide approximately 50% for the peak of COVID-19 pandemic, otherwise hospitals capacities for non-COVID-19 patients would be reduced.

On March 24, an official communication [27] reported that 40% of the available hospital beds in the country were dedicated to the COVID-19 epidemic. That is, a total of 49,083 hospital beds, 256 intensive care units, 5,523 ventilators and 2,446 intensive care beds. Furthermore, a re-conversion program was in process, to increase the health capacity [27]. In spite of social confinement, simulations pointed out a high threat to Mexico (**Figure 1**), that is for the most optimistic scenario, the pandemic peak would be about 338

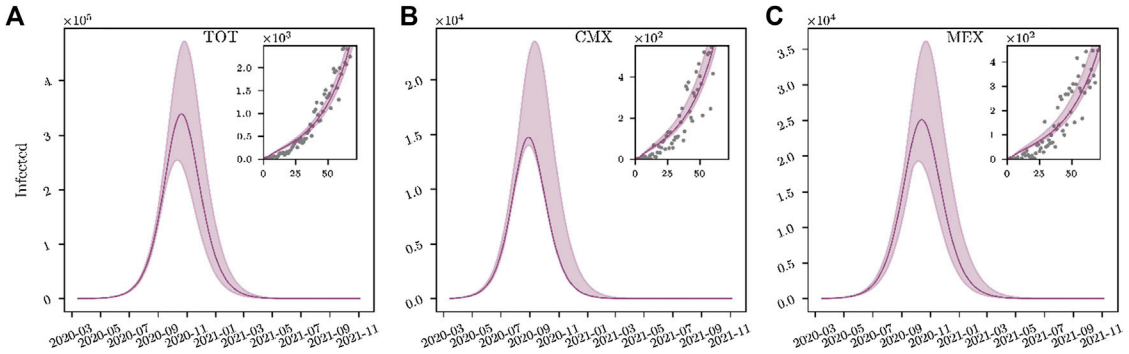


FIGURE 1 | Scenario 1: Maintaining COVID-19 Lockdown. Mathematical model (1)–(5) is fitted with pandemic data from 15 March to May 25, 2020. Shadows indicate the range of 3,000 fits of infected cases considering the range of parameters in **Table 1**. Simulations are performed till November 2021.

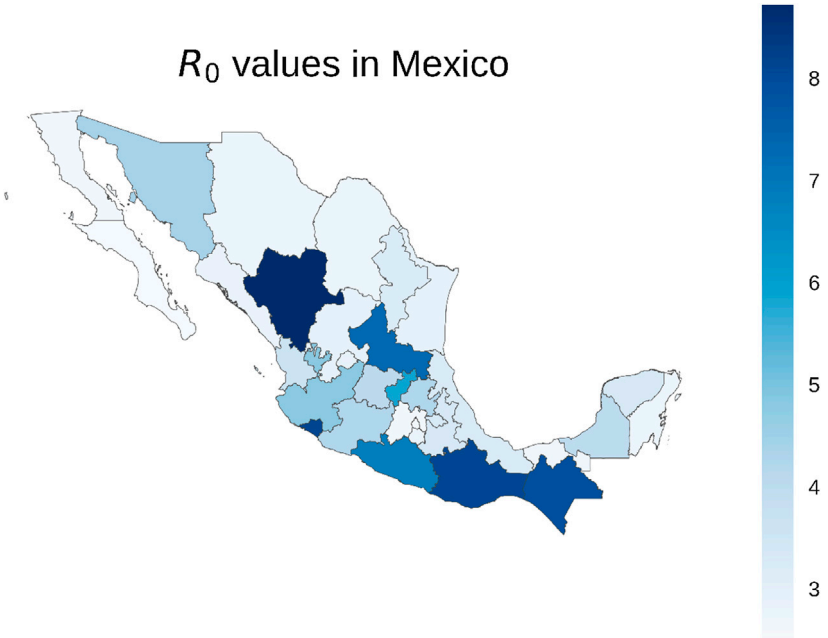


FIGURE 2 | Reproductive number (R_0) for all Mexico states obtained from the model fits up to June 25, 2020.

thousand infected people. If the ratios of hospitalization observed so far are consistent in the next months (about 38% of infected cases, see **Figure 3**), Mexico would be almost at the maximum capacity, approximately 98% hospital bed occupancy, although beds of non-COVID-19 patients are considered for COVID-19 patients. Therefore, social confinement lifting would very likely result in a hospital system breakdown.

3.2. Scenario 2: Complete Lifting Confinement on June 1, 2020

Mexico was planning to lift social distancing restrictions on June 1, 2020, however, due to the continuous increase in infected cases, confinement has been irregular [12]. Based on the nominal values presented in **Table 1**, we develop different possible scenarios to gradually lift social confinement and explain the on-going pandemic. To this end, we will study first the hypothetical case of a complete lifting confinement on June 1, 2020 (Scenario 2). This would help to understand what has been the implication of social confinement on the initial estimations. Panels (a), 2) and 3) in **Figure 4** present a simulation scenario with a sharp and massive return of activities on June 1, 2020. This would result in a major increase of infected cases, about two orders of magnitude higher than the scenario with social confinement in **Figure 1**.

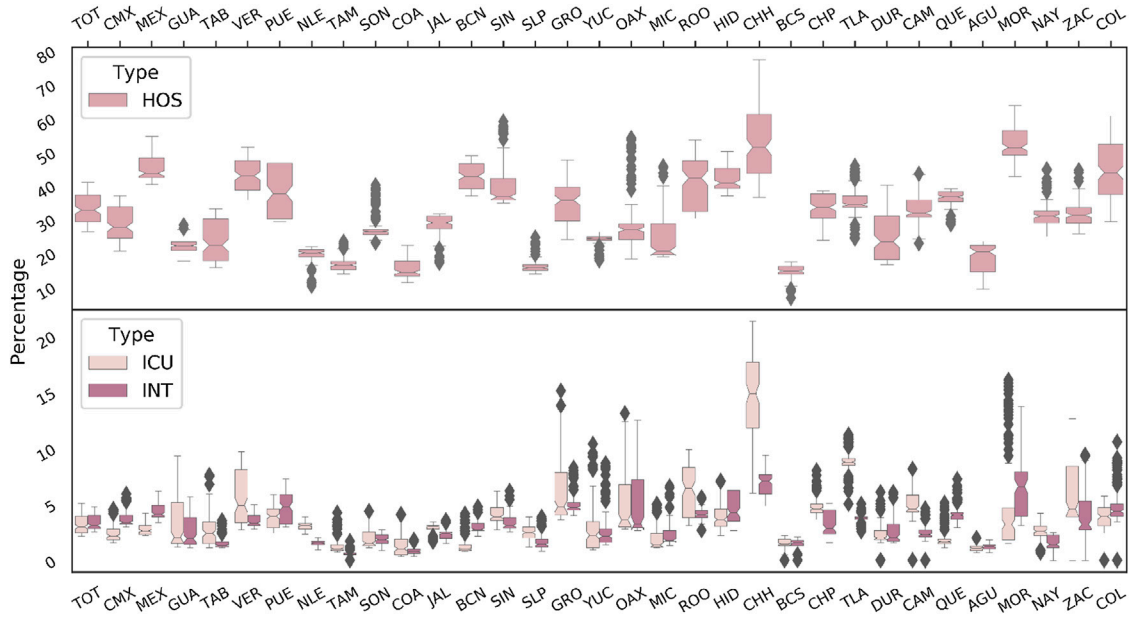


FIGURE 3 | National requirements of Intensive Care Units beds (ICU), intubation facilities (INT) and hospital beds (HOS). These percents are based on the reported data of infected cases that require ICU, INT or HOS. Box plots are composed by data sets from 12 april till 15 August. Therefore, those states with small box plots reveal a very stable needs in public health capacities in time (e.g., CMX). Data was extracted from Ref. [26].

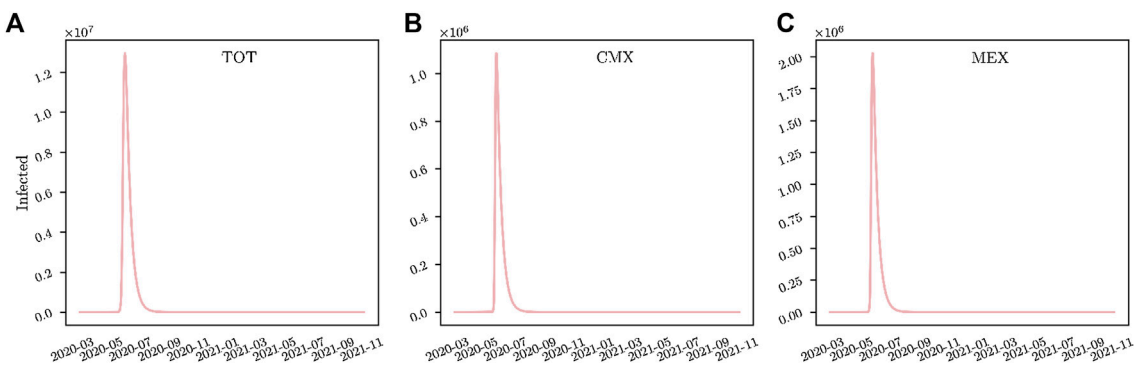


FIGURE 4 | Scenario 2: Complete lifting Confinement on June 1, 2020. Panels **(A)** Present a sharp and massive return of people for social deconfinement on June 1, 2020 for the whole country; **(B)** CMX, and; **(C)** MEX, respectively.

3.3. Scenario 3: Orderly Gradual Lifting Confinement on June 1, 2020

Considering that the confinement in Mexico was about 60% of the population [23], we evaluate a step-wise social confinement lifting based on a gradual deconfinement of different age groups. Based on Mexican reports by the INEGI [11], Mexican population can be divided into three major age groups: G_1 are those between 0 and 24 years old (approximately 42.94% of the population), G_2 are those between 25 and 59 years old (approximately 45.75% of the population), and G_3 are those between 60 years old (approximately 11.32% of the population).

Note that our model does not include any differentiation in infection risk and contact patterns among elderly, working class and students. As a result, it is not able to distinguish those sub-populations. Therefore, any conclusion about the order for lifting confinement of any sub-populations is purely based on assumptions. What can be represented in our model is the fraction of population lifted in each phase.

Thus, we consider fractions of the population that would represent the elderly, group G_3, should stay in social confinement. The remaining population in confinement, about 48.7%, is divided into two groups: the first are students from 3 to 17 years old composed approximately 21.7% [11], the second is

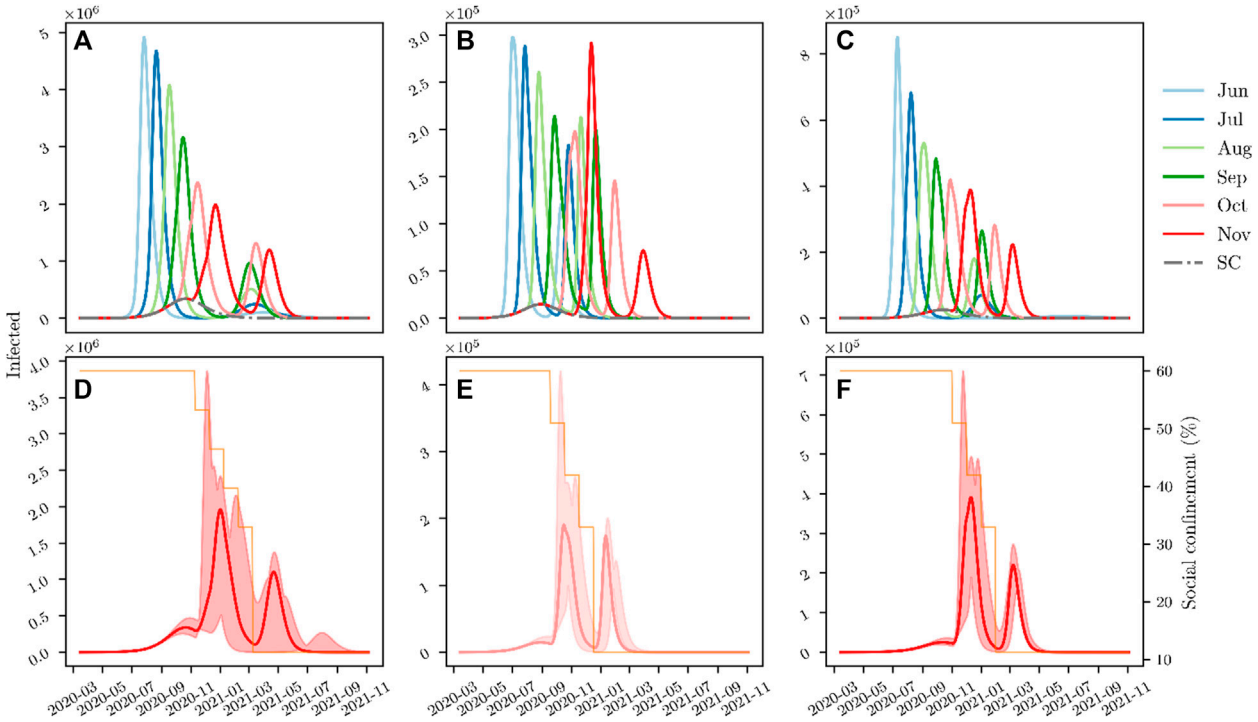


FIGURE 5 | Scenario 3: Gradual lifting Confinement on June 1, 2020. Panels **(A)** shows a step-wize social confinement lifting at different time points for the whole country; **(B)** CMX, and; **(C)** MEX, respectively. Panels **(D)** underlined the best scenario for confinement lifting in the whole country; **(E)** CMX, and; **(F)** MEX, respectively. This consists on keeping in confinement a fraction of the population equivalent to the size of the elderly, while a fraction representing the confined working force of Mexico (about 27%) returns to activity in four parts (three steps for CMX and MEX), each one in consecutive months, while a fraction assuming the students would come back to school one month after the working group.

the group of the working force of Mexico that stayed in social confinement, around 27% [11].

The initiation of activities of a fraction of the population proportional to the working class and students is tailored to reduce the peak of the pandemic. **Figure 5A–C** shows a step-wise deconfinement strategy starting at different months (from June till November), returning a fraction proportional to the working class in four consecutive months (or three for CMX and MEX) and at the last the returning of students to schools. Independently of the initiation time of deconfinement, simulations in **Figure 5** highlight that confinement lifting must be avoided before passing the peak of the pandemic. If the initiation of activities is before the first peak, this would result in a much higher number of infected cases, and consequently death tolls.

Similar outcomes are obtained for CMX and MEX (**Figure 5B, C**) as well as in major regions of Mexico (**Supplementary Figure S3**). **Figure 5D–F** underlines that the best time to initiate deconfinement in CMX and MEX would be between October and November. While this step-wise confinement lifting strategy is the best alternative to mitigate the peak of the pandemic, it would still derive a second wave of the pandemic equivalent to the first one under social restrictions. Estimations of hospital beds requirements for deconfinement strategies for different Mexican regions are presented in **Supplementary Table S2**.

3.4. Scenario 4: Compliance of the Population Under Social Restrictions and Gradual Lifting Confinement on June 1, 2020

Using the model (1)–(5) to recapitulate the pandemic data from June 1 till August 15, 2020 was not possible to achieve changing only the social confinement (%C). Therefore, the model (1)–(5) was extended with the **Eq. 6**, which would mimic the deconfinement and social behavior dynamics in Mexico. Deconfinement initiates at day d_S^C and ends at day d_E^C. **Eq. 6** contains a damped cosine function to represent the social behavior that promotes individuals to enter and leave confinement. The decay constant ($\lambda = 2/(d_E^C - d_S^C)$) is referred to the half of deconfinement time period to assure a strong amplitude damping at d_E^C.

Note that **Eq. 6** is a possible scenario among many others to explain the epidemiological data in Mexico. Of course this is a hypothetical scenario, but in principle any damped oscillating function would represent the pandemic data.

$$C\%(d) = \begin{cases} C_S & d < d_S^C \\ C_S + \dfrac{d - d_S^C}{d_E^C - d_S^C}(C_E - C_S) + A_C e^{-2\left(\frac{d-d_S^C}{d_E^C-d_S^C}\right)} \cos\left(\dfrac{d - d_S^C}{p_C}\right) & d_S^C < d < d_E^C \\ C_E & d > d_E^C \end{cases} \tag{6}$$

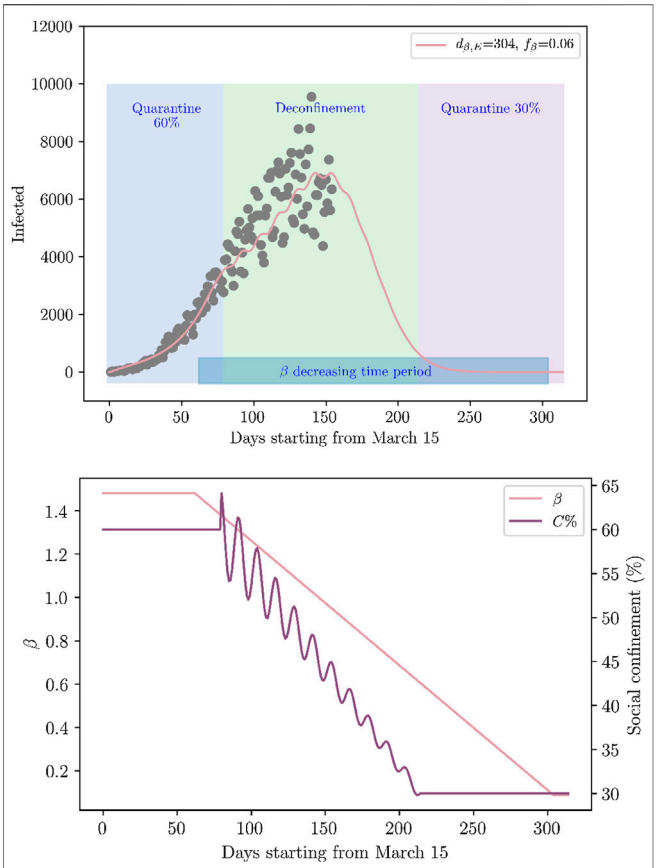


FIGURE 6 | Possible scenario to represent COVID-19 pandemic data. Pandemic data and model simulation (upper panel). Confinement lifting scheme and β modification (lower panel). The percent of the population in confinement (%C) with **Eq. 6** and social distance measures affecting the value of β with **Eq. 7**. These equations with parameter values at **Table 2** represent a possible scenario to reconcile the pandemic data from June 1 till August 15.

To further adapt to the pandemic data, we incorporate social distance (f_β) and the use of mouth cover strategies that could have decreased the new infections. This starts at day d_S^β, and ends at day d_E^β. This is represented by the following equation:

$$\beta(d) = \begin{cases} \beta_0 & d < d_S^\beta \\ \beta_0\left[1 + \left(f_\beta - 1\right)\dfrac{d - d_S^\beta}{d_E^\beta - d_S^\beta}\right] & d_S^\beta < d < d_E^\beta \\ f_\beta\beta_0 & d > d_E^\beta \end{cases} \quad (7)$$

The dynamics of social distance and deconfinement is illustrated in the lower panel in **Figure 6** with parameter values in **Table 2**. **Figure 6** shows the pandemic data and model simulations. We would like to remark this is just a possible scenario among possibly many others that recapitulate the pandemic data of Mexico.

TABLE 2 | Parameter values for confinement lifting scheme and social distance measures in **Eqs 6**, **7**. β_0 corresponds to the fitted value up to May 25, 2020.

d_S^C	d_E^C	C_S	C_E	A_C	p_C
June 01, 2020 (79)	October 14, 2020 (214)	60%	30%	5%	2
d_S^β	d_E^β	β_S	β_E	β_0	f_β
May 15, 2020 (62)	January 12, 2021 (304)	β_0	$f_\beta\beta_0$	1.48	0.06

4. DISCUSSION

The year 2020 has revealed one of the biggest pandemics reported in history, the novel coronavirus SARS-CoV-2 causes a severe and potentially fatal acute respiratory syndrome (COVID-19). Epidemiological reports by countries with strong public health capacities have uncovered the potential of COVID-19 to saturate hospitals in short time [1]. Therefore, COVID-19 is a major threat to developing countries because of the limitations of intensive care beds. While no vaccine or antiviral drug is likely to be available soon, the only remaining tool against COVID-19 is social confinement. However, a prolonged lock-down will hugely affect societies, education, and economy.

In this paper, we investigate the dynamics of the COVID-19 pandemic in Mexico, which adopted on March 15 social confinement to avoid the fast spread of the virus and the eventual collapse of the public health services [12]. Mexico is a highly populated country with major levels of moderate poverty. A strict social confinement has been extremely difficult to apply in a country with 56.2% of the population working informally [11]. Mexico originally plan to lift the social confinement on June 1, 2020, however, the high death tolls in a limited public health system forced to extend the confinement.

While keeping the elderly in social confinement until a vaccine is available, the best strategy to contain the epidemic in Mexico would be to return social activities after the first peak of the pandemic, returning a fraction of the population equivalent to the confined working class (about 27% of the population) in four equal parts, and at the last the returning of students to schools (see **Figure 5**). Considering the best deconfinement strategy, Mexico would need at least a three fold increase in current hospital capacities. Note that this is an upper view of the pandemic at country level, however, the fate decision of pandemic evolution has shown great differences not only among countries but also among states inside of countries. Therefore, social confinement lifting should be tailored for each city in Mexico. Mexican Government may need to consider in deconfinement strategies not only on the age factor in COVID-19 patients **Supplementary Figure S4** but also people with chronic diseases such as diabetes, COPD and obesity as shown in **Supplementary Figure S5**. This is also a main concern, as in Mexico 65% of the population is overweight, 30% of the population is obese, and about 10% of the population had diabetes [28].

The mathematical model (1)–(5) was fitted with data from 15 March to May 25, 2020 (**Figure 3**). For validation the model was used to predict an independent data set from May 25 till June 1st, **Supplementary Figure S6** shows how the model can predict the tendency of the pandemic. However, on June 1, 2020,

confinement restrictions were relaxed in a growing phase of new infected cases and deaths. However, data sets from June 1 till 15 shows a deceleration of the infected cases. Simulations in **Figure 6** shows that the dispersion of data sets of infected cases could be attributed to changes in social behavior to confinement as well as social distance measures. A key aspect to represent pandemic data in **Figure 6** was the combination of deconfinement from 60% to 30% and a decrease in β starting two weeks before deconfinement initiation.

In the same spirit of the epic phrase by Goerge E.P. Box that "All models are wrong, but some are useful", our model as well as others for COVID-19 has several limitations. In particular, mathematical models that are fitted during the emerging phase, exponential growth, of an outbreak that have potential sources of bias [29]. Furthermore, the data of infected cases in Mexico is very limited because of the implementation of the sentinel model. While sampling is based on 10% of the suspected cases while 100% of hospitalized cases and deaths, mass testing for stages 2 and 3 is considered irrelevant. In fact, reports by the OECD (COVID-19) [30] underlines that Mexico was only testing 0.6–0.8 per every thousand, occupying in this way Mexico the last place among countries with high number of infected cases. With 5,623 new tests on May 4, 2020, Mexico reported the highest number of samples per day [30]. Therefore, with this sample rate Mexico will be unable to register the dimension of the epidemic, actually, it is expected that Mexico will have a flat region as soon as the number of infected cases per day passes over 7,560. At the peak of the pandemic Mexico would need to test at least 18 per every thousand to provide a better vision of the problem. The summary of activities to reduce the COVID-19 Pandemic in Mexico would be:

- Return social activities after the first peak of the pandemic.
- Keep the elderly and citizens with chronic diseases such as diabetes, COPD and obesity in social confinement until a vaccine is available.
- Lift confinement to a fraction of the population equivalent to the confined working class (about 27% of the population) in four equal parts after the first peak during confinement.
- Returning of students to schools by February 2021.
- Testing for COVID-19 should be at least 18 per every thousand.
- Increase at least three fold hospital capacities dedicated for COVID-19.
- Continuous social distance and mouth cover use to decrease the rate of new infections.

Traveling is having a key aspect in COVID-19 pandemic. Previous studies explored the consequences for travel [6], particularly focusing on airplane transportation but attempting to give a gross approximation to terrestrial movement since this is the main form of population movement in Mexico. Given the large population size of migrants in Mexico, the transmission risk of this population is high. Thus, airports should test passengers arriving from certain countries when they enter Mexican territory, potentially allowing them to spend less time in quarantine if they test negative.

Note that this study does not consider the economic effects but it was only based to have enough hospital capacities. With the progressing of information of COVID-19 pandemic as well as new data of infected cases in Mexico, our simulations would be more accurate and hopefully less drastic scenarios. Nevertheless, our model predictions highlight critical scenarios and suggest tailored public health strategies for social confinement lifting in combination with a significant increase in the health care system capacities.

AUTHOR'S NOTE

This manuscript has been released as a preprint at medRxiv (31).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://serendipia.digital/2020/03/datos-abiertos-sobre-casos-de-coronavirus-covid-19-en-mexico.

AUTHOR CONTRIBUTIONS

CR performed the simulations. EH-V envisaged the project. All the authors discussed and wrote the paper.

FUNDING

This research was funded by the Universidad Nacional Autonoma de Mexico (UNAM), CONACYT, and the Alfons und Gertrud Kassel-Stiftung. This work received support from Luis Aguilar of the Laboratorio Nacional de Visualización Cientifica Avanzada (LAVIS) at UNAM-Juriquilla.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.573322/full#supplementary-material.

REFERENCES

1. CDC. Coronavirus diseases (COVID-2019). (2020). Situation Reports.
2. Lowe D A close look at the frontrunning coronavirus vaccines as of may 1 (updated)—in the pipeline. (2020). Tech. Rep.
3. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* (2020) 395:931–934. doi:10.1016/S0140-6736(20)30567-5
4. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* (2020) 3099:1–7. doi:10.1016/s1473-3099(20)30144-4

5. Weitz JS, Beckett SJ, Coenen AR, Demory D, Dominguez-Mirazo M, Dushoff J, et al. Modeling shield immunity to reduce COVID-19 epidemic spread. *Nat Med* (2020) 26:849–854. doi:10.1038/s41591-020-0895-3
6. Acuña-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
7. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med* (2020) 382:1199–1207. doi:10.1056/NEJMoa2001316
8. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis* (2020) 3099:30195. doi:10.1016/s1473-3099(20)30195-x
9. Li R, Rivers C, Tan Q, Murray MB, Toner E, Lipsitch M. The demand for inpatient and ICU beds for COVID-19 in the US: lessons from Chinese cities. *medRxiv* (2020) 16:20033241. doi:03.0910.1101/2020.03.09.20033241
10. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* (2020) 368(6493):860–8. doi:10.1126/SCIENCE.ABB5793
11. INEGI. Encuesta Nacional de Ocupación y Empleo (ENOE), población de 15 años y más de edad. (2019). Tech Rep.
12. Secretaria de Salud MEXICO. Coronavirus (COVID-19)-Comunicado técnico diario—secretaría de Salud—gobierno—gob.mx. (2020). Tech Rep.
13. de Salud S. Manual para la vigilancia epidemiológica de Influenza. *Dirección General de Epidemiologia* (2014) 1–84.
14. López L, Rodó X. The end of social confinement and COVID-19 re-emergence risk. *Nat Human Behav* (2020) 4:746–55. doi:10.1038/s41562-020-0908-8
15. Peng L, Yang W, Zhang D, Zhuge C, Hong L Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv* (2020) 1–12. doi:10.1101/2020.02.16.2002.06563
16. Zhao H, Feng Z. Staggered release policies for covid-19 control: costs and benefits of relaxing restrictions by age and risk. *Math Biosci* (2020) 326:108405. doi:10.1016/j.mbs.2020.108405
17. 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
18. LeungLau R, Okel LC, Dorigatt I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* (2020) 20(6):669–77. doi:10.1016/S1473-3099(20)30243-7
19. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* (2020) 581(7809):465–9. doi:10.1038/s41586-020-2196-x
20. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19–24 February 2020 (2020).
21. Phua J, Weng L, Ling L, Egi M, Lim C-M, Divatia JV, et al. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *Lancet Respir Med* (2020) 8:506–17. doi:10.1016/s2213-2600(20)30161-2
22. Reporte de la Secretaria de Salud MEXICO (8 de Abril 2020). *Versión estenográfica. Conferencia de prensa. Informe diario sobre coronavirus COVID-19 en México* (2020).
23. Google. *COVID-19 community mobility reports*.
24. DuWeng R, Price K. Differential Evolution–a simple and efficient adaptive scheme for global optimization over continuous spaces. *J Global Optim* (1997) 11:341–59. doi:10.1023/A:1008202821328
25. Parra-Rojas C, Hernandez-Vargas EAPDE. Parameter fitting toolbox for partial differential equations in python. *Bioinformatics* (2019) 36(8):2618–9. doi:10.1093/bioinformatics/btz938
26. General de Epidemiología D. *Datos Abiertos para COVID-19–dirección General de Epidemiología* | Secretaría de Salud | Gobierno | gob.mx.
27. Velázquez Ramírez MDC. *Capacidad instalada en México para enfrentar al coronavirus COVID-19* (2020).
28. Levaillant M, Lièvre G, Baert G. Ending diabetes in Mexico. *Lancet* (2019) 394:467–8. doi:10.1016/S0140-6736(19)31662-9
29. Britton T, Tomba GS. Estimation in emerging epidemics: biases and remedies. *J R Soc Interface* (2019) 16:20180670. doi:10.1098/RSIF.2018.0670
30. OECD. Testing for COVID-19. A way to lift confinement restrictions. *OECD* (2020). Tech. Rep. April.
31. Azanza Ricardo CL, Hernandez Vargas EA. The risk of lifting covid-19 confinement in Mexico. *medRxiv* (2020) 1–18. doi:10.1101/2020.05.28.20115063

Disclaimer: The results expressed in this report should not be construed to represent the views of any agencies or the Mexican government.

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fphy.2020.587483

The Impact of Population Migration on the Spread of COVID-19: A Case Study of Guangdong Province and Hunan Province in China

Guo-Rong Xing[1], *Ming-Tao Li*[2], *Li Li*[3] *and Gui-Quan Sun*[1,4]*

[1]Complex Systems Research Center, Shanxi University, Taiyuan, China, [2]School of Mathematics, Taiyuan University of Technology, Taiyuan, China, [3]School of Computer and Information Technology, Shanxi University, Taiyuan, China, [4]Department of Mathematics, North University of China, Taiyuan, China

On the eve of the Spring Festival in 2020, the coronavirus disease 2019 (COVID-19) was reported. Subsequently, the Chinese government at all levels took emergency measures to control the spread of COVID-19 among people. Guangdong and Hunan are large population floating provinces. The spread of COVID-19 is affected by population migration. Before the Spring Festival, Guangdong and Hunan Provinces dominated population export and import, respectively; after the Spring Festival, the trend of population flow was reversed by the resumption of work. Taking Guangdong and Hunan as examples, we establish a three-stage dynamical model to study the impact of population migration on the spread of COVID-19. The result reveals that Guangdong Province mainly emigrated the population and the scale of infection was reduced before the Spring Festival. However, the situation in Hunan Province was just the opposite. After the Spring Festival, work resumption was taking place across China and the migration of Guangdong may cause a second outbreak of the epidemic. While people in Hunan leave the province to work, the migration of population will have little effect on the spread of COVID-19.

Keywords: COVID-19, dynamical modeling, migration, basic reproduction number, work resumption

OPEN ACCESS

Edited by:
Chengyi Xia,
Tianjin University of Technology, China

Reviewed by:
Chen Chu,
Yunnan University of Finance and Economics, China
Qing Cheng,
National University of Defense Technology, China
Jianquan Li,
Shaanxi University of Science and Technology, China

***Correspondence:**
Gui-Quan Sun
Sungquansun@126.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *26 July 2020*
Accepted: *28 September 2020*
Published: *23 December 2020*

Citation:
Xing G-R, Li M-T, Li L and Sun G-Q (2020) The Impact of Population Migration on the Spread of COVID-19: A Case Study of Guangdong Province and Hunan Province in China. Front. Phys. 8:587483. doi: 10.3389/fphy.2020.587483

1. INTRODUCTION

At the end of 2019, unknown pneumonia was reported. The World Health Organization (WHO) officially named the new coronavirus as COVID-19 on February 11, 2020 [1, 2]. The main transmission routes of COVID-19 are direct transmission, aerosol transmission, and contact transmission [3]. Early, Li et al. analyzed the data of 425 confirmed cases in Wuhan to determine the epidemiological characteristics of NCIP [4]. As the time of the outbreak is approaching the Spring Festival, passenger traffic from Wuhan to all parts of the country was large, which caused the rapid spread of COVID-19. In order to control the epidemic effectively, the Chinese government took the lockdown strategy in Wuhan on January 23 [5]. Everyone who came from Hubei Province has to be registered and then quarantined at home for 14 days. The flights, railways, and highways taken by people from Wuhan were closely followed up and announced in detail after registration. In order to control the spread of COVID-19, the government took a series of measures, such as prohibiting all transport in and out [6, 7] and disinfecting and taking body temperature in public places. For individuals, people must wear their own masks when going out.

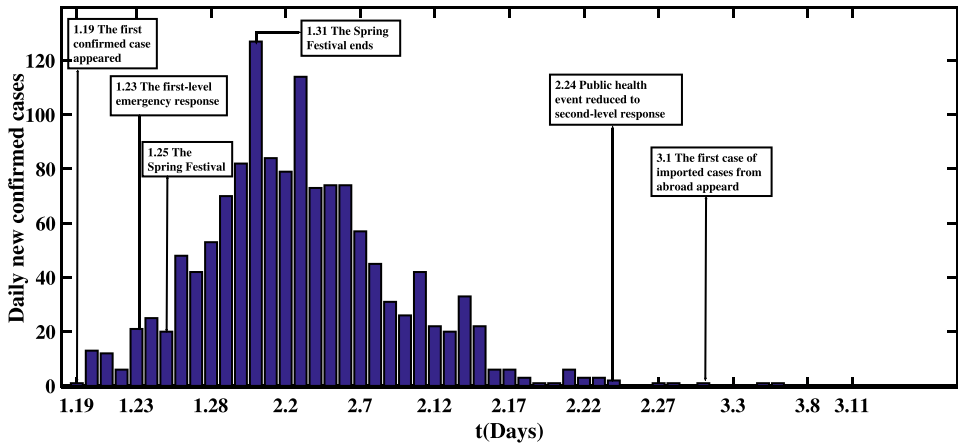


FIGURE 1 | Transmission scheme of COVID-19 in Guangdong Province. Guangdong Province launched the first-level response to major health incidents on January 23 and the Spring Festival ends on January 31.

At present, the epidemic is still spreading globally. Many research teams have been studying the impact of COVID-19 transmission in real time. The study of dynamical modeling to predict the epidemic rule plays an important role in disease control [8–11]. Typically, Li et al. analyzed the impact of different lockdown time in Wuhan on the transmission of COVID-19 in Shanxi Province [12]. Tang et al. focused on the time to resume work and assessed the risk of the epidemic [13–15]. Furthermore, Huang et al. predicted the trend of COVID-19 and calculated the parameters of this epidemic [16]. Moreover, lots of scholars established dynamical modeling to analyze the influence of different quarantine measures on the spread of COVID-19 [17–19]. For example, Gao et al. established dynamical modeling showing that the epidemic was controlled by reducing the contact rate [20]. Some works were dedicated to assessing the effectiveness of epidemic prevention and control measures [21–23]. Meanwhile, great academics have discussed the issue of media raising awareness of epidemic prevention and control [24, 25].

For the provinces with a large flow of people, Guangdong and Hunan naturally attract great attention from society. Compared with other provinces in China, Guangdong Province has the largest number of migrant people. We collected data on reported confirmed cases of COVID-19 in Guangdong from January 19 to March 11 released by the Health Commission of Guangdong Province [26]. **Figure 1** illustrates the time series of confirmed COVID-19 cases in Guangdong Province. On January 19, the National Health Commission verified the first confirmed case of pneumonia with imported COVID-19 infection in Guangdong Province. Guangdong Province initiated the first-level emergency response on January 23. **Figure 1** shows that the peak of daily new confirmed cases is 127 cases on January 31, and the epidemic will disappear at the end of February. Based on Baidu migration data [27], **Figure 2** indicates the number of migrant people in Guangdong and Hunan around the Spring Festival proportion. Before the Spring Festival, most of the people of Hunan Province immigrated from Guangdong, Hubei, and Jiangxi Provinces on a large scale, among which Guangdong Province had the largest immigration population accounting for more than half of the total. However, the main emigration population of Guangdong Province flowed to Hunan Province, Guangxi Zhuang Autonomous Region, and Jiangxi Province. After the Spring Festival, work resumption was taking place across China. Most of the emigrants from Hunan Province went to work in Guangdong Province. At this time, the trend of population flow in the two provinces was opposite to that before the Spring Festival.

In order to reflect the COVID-19 variations and describe the impact of population migration on the spread of COVID-19, we establish the corresponding dynamical modeling in three different time periods around the Spring Festival and calculate the basic reproduction number for each stage. Using the actual data published by the Health Commission, the parameters were estimated and 95% confidence interval was obtained to verify the rationality of the dynamical modeling [26, 28]. Finally, we draw practical conclusions based on dynamic modeling, which provides some help for resuming work.

We conduct a detailed study of immigration and emigration of population, which is helpful to the prevention and control of areas with large population flows all over the world. At present, there are few teams to study the spread of COVID-19 in provinces. We not only consider the control characteristics in different provinces but also combine the features of the Spring Festival to establish multiple-stage dynamical modeling on the time scale, which can more accurately predict the development trend of the epidemic.

2. METHODS

2.1. Data Sources

In our study, we use Baidu migration data to estimate the immigration and emigration rates of Guangdong and Hunan Provinces around the Spring Festival [27]. In addition, we used

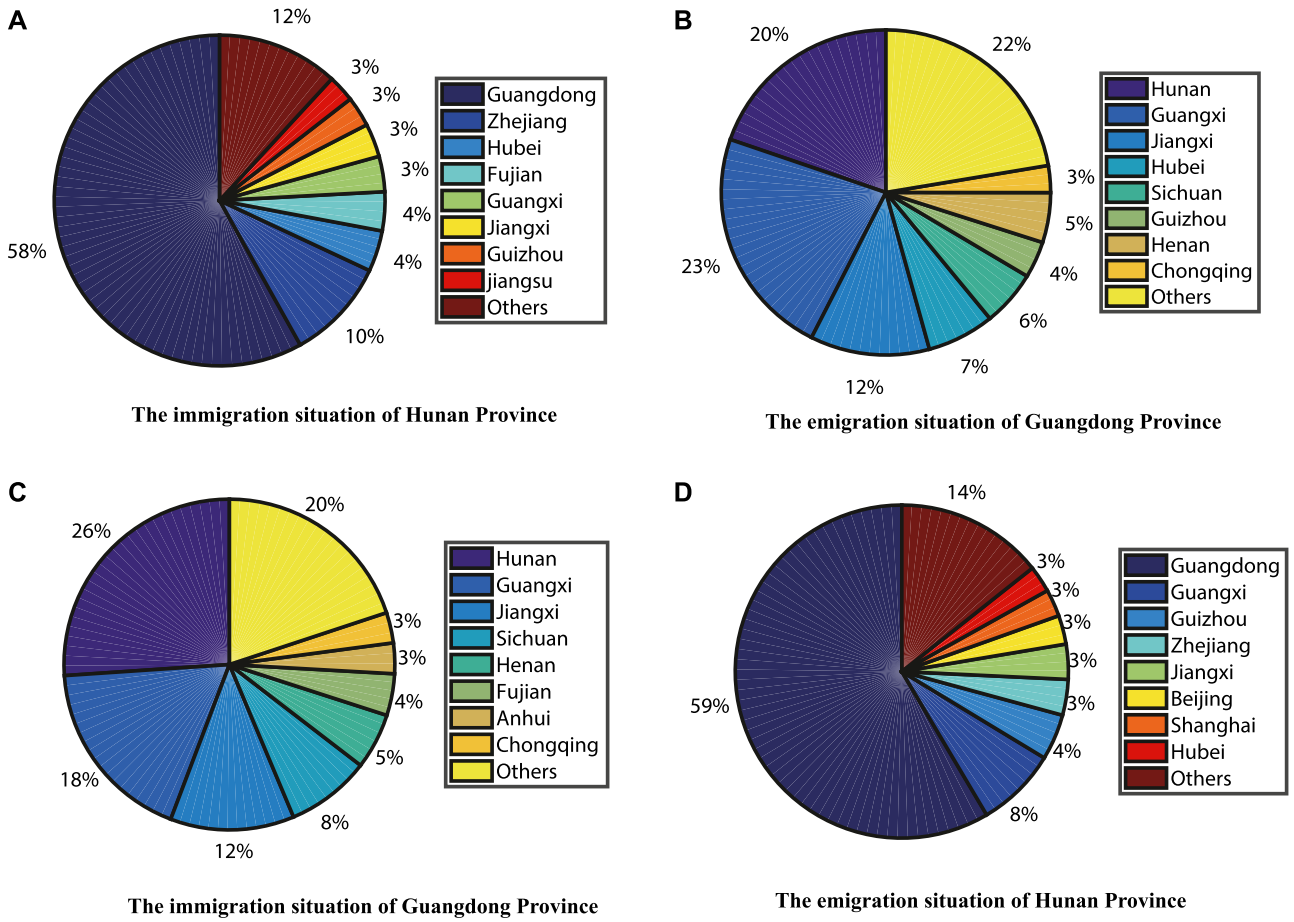


FIGURE 2 | Migration of population in Hunan Province and Guangdong Province [27]. **(A)** The immigration situation of Hunan Province before the Spring Festival. **(B)** The emigration situation of Guangdong Province before the Spring Festival. **(C)** The immigration situation of Guangdong Province after the Spring Festival. **(D)** The emigration situation of Hunan Province after the Spring Festival.

the information released by Guangdong and Hunan Provincial Health Commissions to collect the number of cumulative and daily new confirmed COVID-19 cases [26, 28]. According to the information released by the National Bureau of Statistics, we obtained the number of permanent residents in Guangdong and Hunan Provinces at the end of this year [29].

2.2. Dynamical Modeling Construction and Analysis

Based on the knowledge of COVID-19, it is necessary to establish dynamical modeling to further understand the transmission mechanism [30–33]. Most infectious diseases are studied in three parts: susceptible population, transmission route, and infected population [34–37]. According to the characteristics of COVID-19, the total population $N(t)$ is divided into susceptible individuals $S(t)$, exposed individuals $E(t)$, infected individuals $I(t)$, confirmed individuals $Q(t)$, and removed individuals $R(t)$. We make the following assumptions:

(1) The exposed individuals $E(t)$ have the ability to infect susceptible individuals S(t).
(2) The removed individuals $R(t)$ have immunity during the epidemic.
(3) Individuals' birth and death rates are ignored.

In order to describe the spread of COVID-19, we establish a corresponding dynamical model based on the policies of epidemic control in Guangdong and Hunan Provinces. The following takes Guangdong Province as an example of mathematical analysis. We study the dynamic modeling in three stages. The first stage: the first confirmed case appeared until the Spring Festival (1.19–1.24). Each province has hospitals that treat and isolate cases. In order to prevent large-scale infection, the confirmed cases $Q(t)$ were treated in the province, indicating that they did not migrate. **Figure 3** demonstrates the transmission of COVID-19 and population migration in detail. Based on the above assumptions, we establish the COVID-19 transmission dynamical modeling **Eq. (1)**:

$$\begin{cases} \dfrac{dS(t)}{dt} = -\dfrac{\beta_1 SE + \beta_2 SI}{N} + (a-b)S, \\ \dfrac{dE(t)}{dt} = \dfrac{\beta_1 SE + \beta_2 SI}{N} - \delta E + (a-b)E, \\ \dfrac{dI(t)}{dt} = \delta E - mI + (a-b)I, \\ \dfrac{dQ(t)}{dt} = mI - \gamma Q, \\ \dfrac{dR(t)}{dt} = \gamma Q. \end{cases} \tag{1}$$

Here, δ denotes the transfer rate of exposed individuals to infected individuals, so the incubation period is $1/\delta$. The recovery rate of confirmed individuals is γ. The parameter m is the confirmation rate of infected individuals. Susceptible individuals are infected with COVID-19 through contact with exposed individuals and infected individuals. In particular, the novel coronavirus incubation period is infectious, so the parameters β_1 and β_2 are the transmission incidence rates from exposed individuals and infected individuals to susceptible individuals, respectively; a and b denote the immigration rate and emigration rate, respectively.

The second stage: during the Spring Festival (1.24–1.30), the Guangdong Provincial Government has increased control efforts to investigate suspected cases. In addition, people's awareness of the epidemic has gradually increased. We assume that the population has not moved across provinces during this period. The dynamical modeling (2) is as follows:

$$\begin{cases} \dfrac{dS(t)}{dt} = -\dfrac{\beta_1 SE + \beta_2 SI}{N}, \\ \dfrac{dE(t)}{dt} = \dfrac{\beta_1 SE + \beta_2 SI}{N} - \delta E, \\ \dfrac{dI(t)}{dt} = \delta E - mI, \\ \dfrac{dQ(t)}{dt} = mI - \gamma Q, \\ \dfrac{dR(t)}{dt} = \gamma Q. \end{cases} \tag{2}$$

The third stage: due to the nationwide resumption of work, the population began to move across the provinces. At this stage, the prevention and control measures were quite perfect, and the relevant departments immediately quarantine infected individuals and monitor the people who are in close contact with them. The dynamical modeling **Eq. (3)** is as follows:

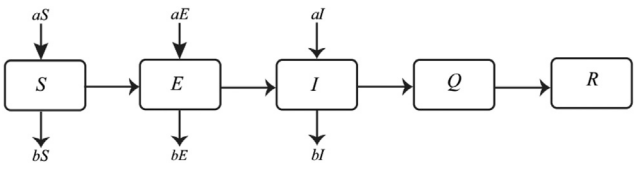


FIGURE 3 | Before the Spring Festival, the flowcharts of COVID-19 infectious disease model. $S(t)$, $E(t)$, $I(t)$, $Q(t)$, and $R(t)$ represent susceptible, exposed, infected, confirmed, and removed populations.

$$\begin{cases} \dfrac{dS(t)}{dt} = -\dfrac{\beta_1 SE}{N} + (a-b)S, \\ \dfrac{dE(t)}{dt} = \dfrac{\beta_1 SE}{N} - \delta E, \\ \dfrac{dI(t)}{dt} = \delta E - mI, \\ \dfrac{dQ(t)}{dt} = mI - \gamma Q, \\ \dfrac{dR(t)}{dt} = \gamma Q. \end{cases} \tag{3}$$

Before and after the Spring Festival, Guangdong Province was opposite to the overall population mobility of Hunan Province. In the modeling, parameters a and b are used to describe the migration situation of the two provinces (shown in **Table 1**).

Since the modeling in the three stages is roughly the same, we only conduct mathematical analysis on the first stage. Below, we calculate the basic reproduction number R_0 by using the method of the next-generation matrix [38]. Denote

$$\mathcal{F} = \begin{pmatrix} \dfrac{\beta_1 SE + \beta_2 SI}{N} \\ 0 \\ 0 \end{pmatrix}, \quad \nu = \begin{pmatrix} \delta E - (a-b)E \\ -\delta E + mI - (a-b)I \\ -mI + \gamma Q \end{pmatrix}, \tag{4}$$

By calculating the Jacobian Matrix at the disease-free state E_0, we have

$$F = \begin{pmatrix} \beta_1 & \beta_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \delta - a + b & 0 & 0 \\ -\delta & m - a + b & 0 \\ 0 & -m & \gamma \end{pmatrix}. \tag{5}$$

Then,

$$FV^{-1} = \begin{pmatrix} \dfrac{1}{\delta + b - a} & 0 & 0 \\ \dfrac{\delta}{(\delta + b - a)(m + b - a)} & \dfrac{1}{m + b - a} & 0 \\ \dfrac{m\delta}{\gamma(\delta + b - a)(m + b - a)} & \dfrac{m}{\gamma(m + b - a)} & \dfrac{1}{\gamma} \end{pmatrix}. \tag{6}$$

So

$$R_0 = \rho\left(FV^{-1}\right) = \frac{\beta_1}{\delta + b - a} + \frac{\beta_2 \delta}{(m + b - a)(\delta + b - a)}. \tag{7}$$

If $R_0 > 1$, we can get the endemic equilibrium $E^* = (S^*, E^*, I^*, Q^*, R^*)$.

TABLE 1 | The migration relationship between Guangdong Province and Hunan Province.

Time	The relationship between *a* and *b*	
	Guangdong Province	**Hunan Province**
Before the Spring Festival	$a < b$	$a > b$
During the Spring Festival	$a = b$	$a = b$
After the Spring Festival	$a > b$	$a < b$

$$\begin{cases} S^{\star} = \dfrac{(\delta + \mathrm{b} - \mathrm{a})(\mathrm{m} + \mathrm{b} - \mathrm{a})\mathrm{N}}{(\beta_1(\mathrm{m} + \mathrm{b} - \mathrm{a}) + \beta_2\delta)\mathrm{I}^{\star}}, \\ E^{\star} = \dfrac{\mathrm{m} + \mathrm{b} - \mathrm{a}}{\delta}\mathrm{I}^{\star}, \\ I^{\star} = \dfrac{(\mathrm{a} - \mathrm{b})\delta\mathrm{N}}{\beta_1(\mathrm{m} + \mathrm{b} - \mathrm{a}) + \beta_2\delta}, \\ Q^{\star} = \dfrac{\mathrm{m}}{\gamma}\mathrm{I}^{\star}, \\ R^{\star} = \mathrm{m}\mathrm{I}^{\star}. \end{cases} \tag{8}$$

3. RESULTS

3.1. Parameters Estimation

In our study, the first stage is the early phase of the epidemic when infected individuals and exposed individuals are freely contacted with susceptible people in public places, so $\beta_1 = \beta_2$. In the second stage, during the Spring Festival, the communities investigated the activities of residents. Anyone who came back from Wuhan and had a history of contact with Wuhan personnel should be quarantined at home for 14 days. Therefore, we assume $\beta_2 = \frac{1}{14}\beta_1$. In the third stage, the medical conditions and control measures were perfect; it is assumed that $\beta_2 = 0$. The parameter γ, which depends on the actual infection cases, does not affect the parameters' estimation of the model.

Since the epidemic occurred before the Spring Festival, the initial value of the susceptible population $S(t)$ equals 1.1521×10^8 of the permanent population in Guangdong Province [29]. By the time the hospital detected the first case, many people had already been infected. Based on the data of confirmed cases reported by Guangdong health and Health Commission on the second day, we assume that the initial value of infected individuals $I(t)$ is 10. Since the COVID-19 has an average incubation period of 5 days, we assume that the initial value of exposed individuals $E(t)$ is 110 based on the data from the Health Commission of Guangdong Province about 5 days after the release of the report. According to the characteristics of the COVID-19 incubation period [39], we chose $\delta_1 = \delta_2 = \frac{1}{3.4}$, $\delta_3 = \frac{1}{6.8}$. Before the Spring Festival, the emigration rate is much greater than the immigration rate. According to Baidu migration data [27], we assume that $a = 0.01, b = 0.24$.

Similarly, we assume that the initial value of the susceptible population $S(t)$ equals 6.91838×10^7 of the permanent population in Hunan Province [29]. The initial value of exposed individuals $E(t)$ is 110, and the number of infected individuals $I(t)$ is 21. Moreover, we chose $\delta_1 = \delta_2 = \frac{1}{3.1}$, $\delta_3 = \frac{1}{6}$. Before the Spring Festival, the immigration rate in Hunan Province is much bigger than the emigration rate according to Baidu migration data [27], so $a = 0.12, b = 0.001$.

In this work, we estimate the parameters m and β_1 by the least-squares method. **Figure 4** unveils the time evolution of the cumulative and the daily new confirmed cases in three stages. The parameter values of our simulation are shown in **Table 2**. **Figure 4** indicates that the cumulative value of confirmed cases in Guangdong Province reaches 1356 and COVID-19 will disappear after 40 days. The peak value time of daily new confirmed cases is January 31, which is the same date as reported.

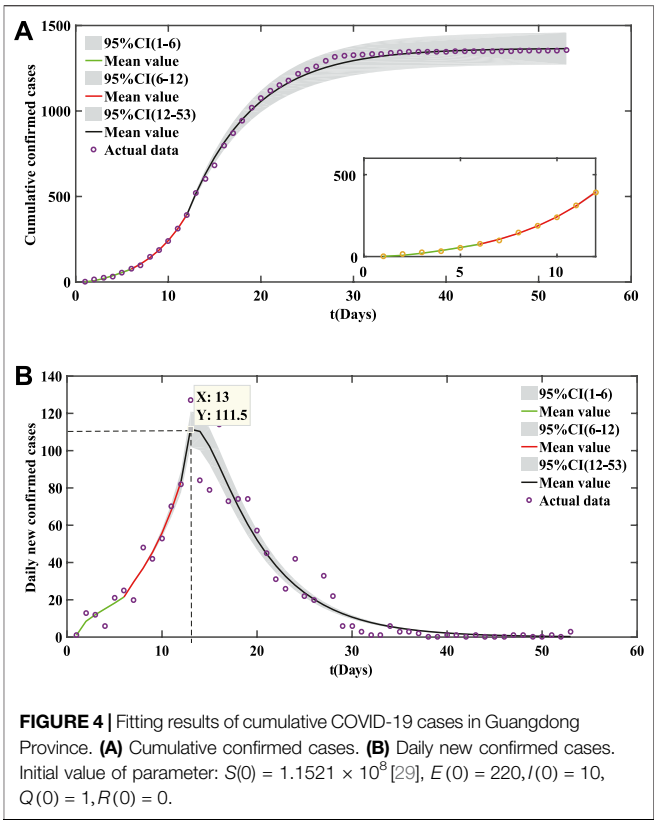


FIGURE 4 | Fitting results of cumulative COVID-19 cases in Guangdong Province. **(A)** Cumulative confirmed cases. **(B)** Daily new confirmed cases. Initial value of parameter: $S(0) = 1.1521 \times 10^8$ [29], $E(0) = 220, I(0) = 10$, $Q(0) = 1, R(0) = 0$.

Figure 5 shows the simulation results of the cumulative and the daily new confirmed cases in Hunan Province. The parameter values of the fitting result are shown in **Table 3**. From January 21 to February 29, the cumulative number of confirmed cases in Hunan Province reached 1018. It can be seen that the daily new confirmed cases in Hunan Province were slower than that in Guangdong Province from the simulation results. Our dynamical model indicates that the epidemic peak in Hunan Province occurred on February 2 and will disappear at the end of February.

3.2. Impact of Population Migration

In order to discuss the impact of population migration around the Spring Festival on the spread of COVID-19, we analyze the parameters a (immigration rate) and b (emigration rate) of Guangdong and Hunan Provinces which have been shown in **Figures 6**, **7**, respectively.

Due to the large number of external population in Guangdong Province, the emigration rate of the population needs to be discussed before the Spring Festival; we find that the greater the emigration rate of the population, the smaller the final scale of the cumulative confirmed cases (shown in **Figure 6A**) and the lower the peak value of daily new confirmed cases (shown in **Figure 6C**).

Through **Figures 6B,D**, we can see the impact of the immigration rate on the confirmed cases of COVID-19 after the Spring Festival when many industries began to resume work. As we can see, if the

TABLE 2 | Parameter estimation of COVID-19 in Guangdong Province.

Parameter	Estimated value			Source
	The first stage	The second stage	The third stage	
β_1	0.3346	0.4648	0.0010	Estimated
β_2	0.3346	0.0332	0	Estimated and calculated
a	0.01	0	0.1	[27]
b	0.24	0	0.005	[27]
δ	$\frac{1}{3.4}$	$\frac{1}{3.4}$	$\frac{1}{6.8}$	[37]
m	0.1836	0.2291	0.3432	Estimated
γ		$\frac{1}{30}$		Estimated
R_0	1.0923	1.6229	0.0068	calculated

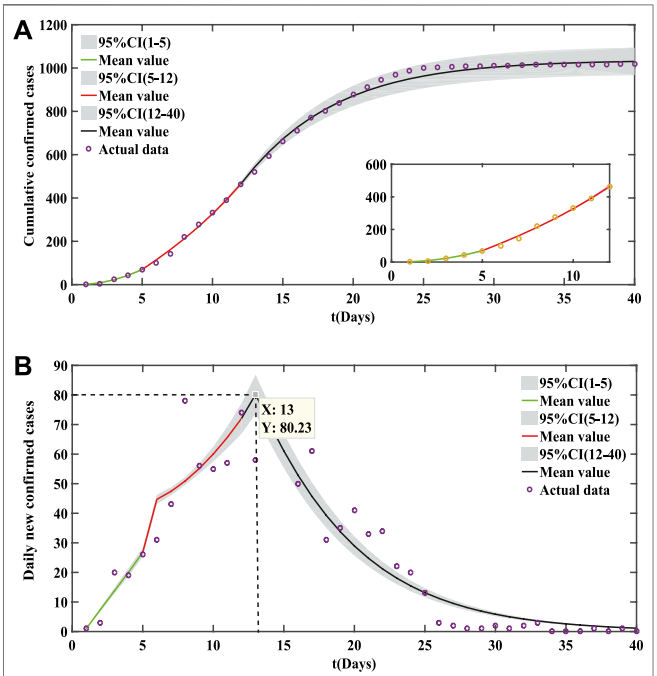


FIGURE 5 | Fitting results of cumulative COVID-19 cases in Hunan Province. **(A)** Cumulative confirmed cases. **(B)** Daily new confirmed cases. Initial value of parameter: $S(0) = 6.91838 \times 10^7$ [29], $E(0) = 110, I(0) = 21$, $Q(0) = 1, R(0) = 0$.

immigration rate is small, the final scale of confirmed cases will not change, whereas if the immigration rate is too large, the number of confirmed cases will increase with the increase of immigration rate, and the epidemic has a second outbreak. Relevant departments of Guangdong Province should strengthen the restrictions on the number of external population when resuming work.

For the population flow in Guangdong Province, Hu et al. studied the evaluation and prediction of the COVID-19 variations at different input population [40]. Both our research and that by Hu et al. draw the conclusion that there will be a risk of secondary outbreak when the input population flow is so severe that the population movement changes. Taking January 27 to February 20, 2020, as the research period, they established the *SEIRQ* dynamical modeling to simulate and predict the epidemic in Guangdong Province. In their work, imported population $p(t)A(t)$ and exported population $B(t)$ were both constants. Through the four time points of February 6, February 16, February 24, and March 5, 2020, Hu et al. discussed the impact of imported population and quarantine strategies on the epidemic in Guangdong Province, including cumulative and daily new confirmed cases. In this study, our dynamical modeling uses index input and output to represent the actual data immigration rate and emigration rate. In addition, we study the dynamical modeling in three phases around the Spring Festival, which reflects COVID-19 variations.

On the contrary, Hunan Province has a large number of people who work in other provinces, so we discuss the

TABLE 3 | Parameter estimation of COVID-19 in Hunan Province.

Parameter	Estimated value			Source
	The first stage	The second stage	The third stage	
β_1	0.1798	0.3961	0.0022	Estimated
β_2	0.1798	0.0283	0	Estimated and calculated
a	0.12	0	0.01	[27]
b	0.01	0	0.11	[27]
δ	$\frac{1}{3.1}$	$\frac{1}{3.1}$	$\frac{1}{6}$	[37]
m	0.1998	0.2291	0.3308	Estimated
γ		$\frac{1}{30}$		Estimated
R_0	3.8828	1.3265	0.0132	Calculated

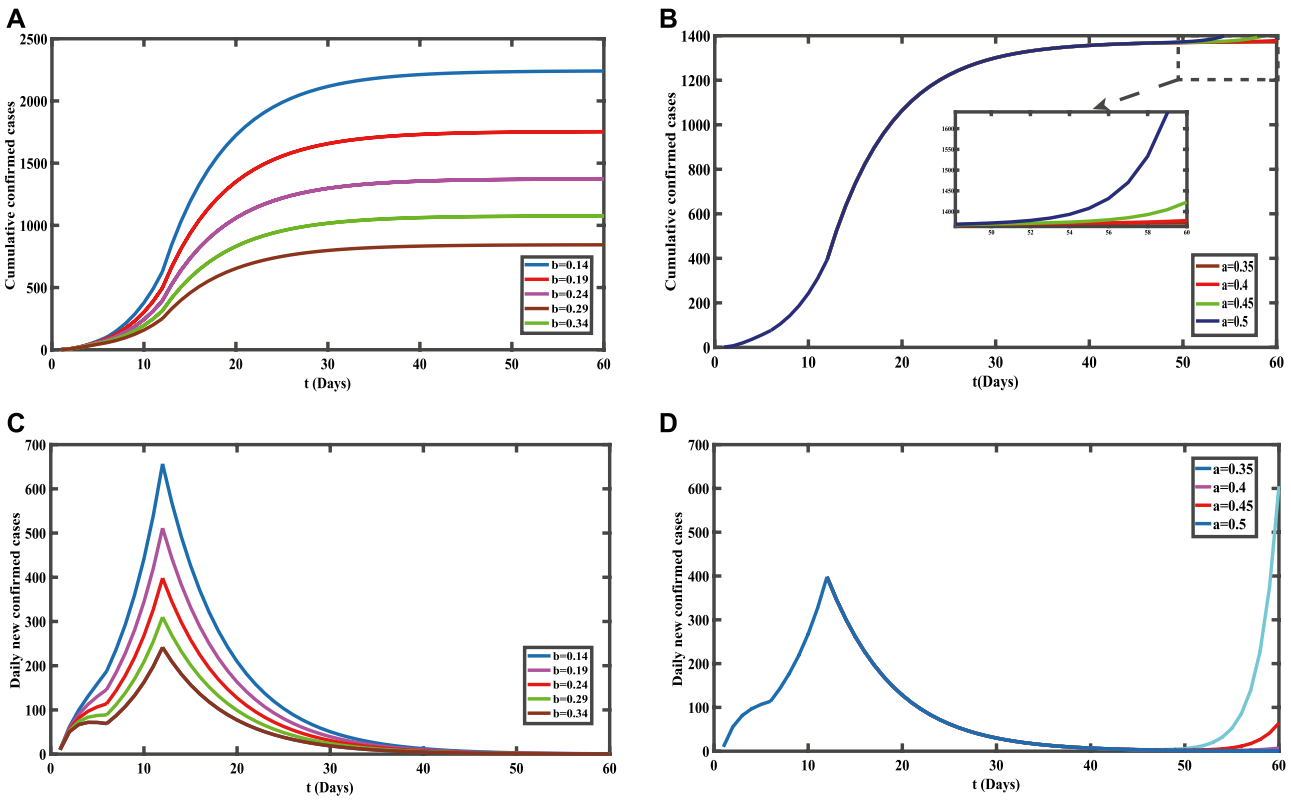


FIGURE 6 | Influence of population migration on COVID-19 cases in Guangdong Province before and after the Spring Festival. **(A)** Cumulative confirmed cases before the Spring Festival. **(B)** Cumulative confirmed cases after the Spring Festival. **(C)** Daily new confirmed cases before the Spring Festival. **(D)** Daily new confirmed cases after the Spring Festival.

immigration rate of the population before the Spring Festival. The greater the immigration rate of the population before the Spring Festival, the larger the final scale of cumulative confirmed cases (shown in **Figure 7A**) and the higher the peak value of the daily new confirmed cases (shown in **Figure 7C**).

After the Spring Festival, people began to resume work, so **Figures 7B**,**D** show the impact of the immigration rate on the confirmed cases of COVID-19. The relationship between the emigration rate and the confirmed cases is not obvious, and the final scale tends to be stable. This indicates that the scale of the emigrated population in Hunan Province after the Spring Festival has a small impact on the spread of COVID-19, which can be understood as the epidemic has been controlled.

For the transmission of COVID-19 in Hunan Province, Jia et al. established an *eSIR* model (Extended SIR Model) with time-varying transmission rates [41]. According to the model analysis, there are two turning points in the disease. And they predicted that the end of the epidemic of COVID-19 was March 3. They used the Markov chain Monte Carlo (*MCMC*) method to estimate the average basic reproduction number as 3.16 (CI: 1.73–5.25), which is roughly the same as ours. Both our study and that by Jia et al. consider the time factor. Compared with our work which reflects this factor by segmenting the dynamical modeling, Jia et al. reflected it through the transmission rate. In this paper, the turning points of the given model are January 25 and February 1, and the epidemic disappeared after February 29, which is close to the date given in Ref. 39.

4. DISCUSSION

Taking Hunan Province and Guangdong Province as examples, this paper uses dynamical systems to study the impact of population migration on the spread of COVID-19. We establish the *SEIQR* model in three stages and calculate the basic reproduction number for each stage. Subsequently, the actual data released by the Health Commission was used for fitting to prove the rationality of the model. In addition, we discussed in detail the impact of population migration around the Spring Festival on the spread of COVID-19.

Before the Spring Festival, a large-scale population moved out of Guangdong Province, and the number of people infected with COVID-19 is decreasing. After the Spring Festival, the industries began to enter the period of resuming work, and the migrant population began to move in. If the immigrant population is too large, the disease may have a second outbreak. The Guangdong Provincial Government should restrict the entry of the population reasonably. For Hunan Province, most of the migrant workers returned home for the Spring Festival. As the

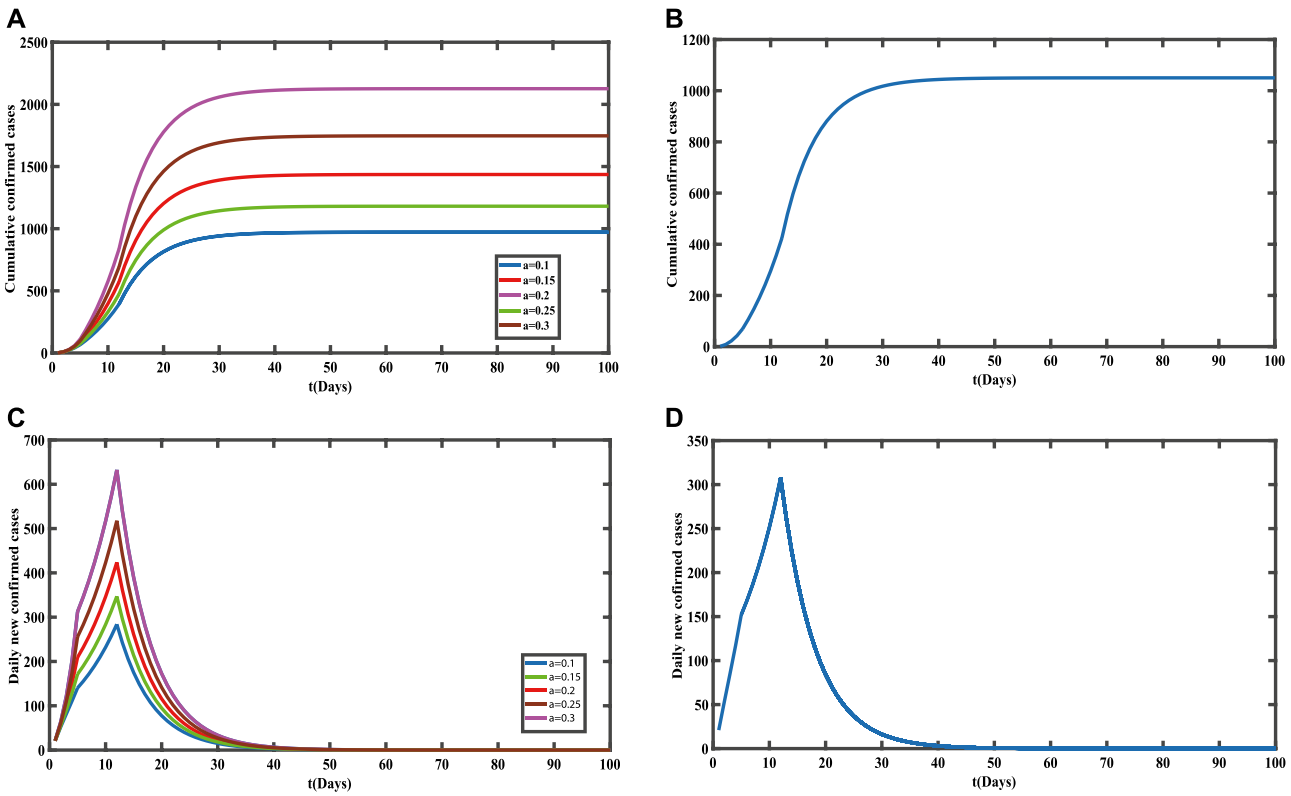


FIGURE 7 | Influence of population migration on COVID-19 cases in Hunan Province before and after the Spring Festival. **(A)** Cumulative confirmed cases before the Spring Festival. **(B)** Cumulative confirmed cases after the Spring Festival. **(C)** Daily new confirmed cases before the Spring Festival. **(D)** Daily new confirmed cases after the Spring Festival.

Spring Festival approached, the population began to immigrate in large numbers, and the number of people infected with COVID-19 continued to rise. After the Spring Festival, population emigration has little effect on the spread of COVID-19 in Hunan Province. The larger population in Guangdong Province may be the reason why the final scale of the cumulative confirmed cases in Guangdong Province is larger than that in Hunan Province. Before the Spring Festival, a large number of people immigrated to Hunan Province, which led to a large increase in the daily new confirmed cases, indicating that the population migration before the Spring Festival accelerated the spread of COVID-19. During the Spring Festival, the growth rate of daily new confirmed cases in Guangdong Province slowed down. At the beginning of February, the daily new confirmed cases in both provinces reached a peak. Subsequently, the number of daily new confirmed cases began to decline and cleared after 40 days. That is to say, the prevention and control measures made by the country in the early stage have well controlled the development of the epidemic, and the risk of people resuming work was reduced.

This work still has many shortcomings and needs to be further improved: 1) in the early stage of the spread of COVID-19, we do not consider asymptomatic infections when establishing the *SEIQR* modeling; 2) this paper studies the early stage of the epidemic when the confirmed cases may have errors because of the limited detection and medical level. The data of population migration is based on the positioning of software and the analysis of big data, which may lead to population deviation; 3) our work does not take spatial factors into consideration, nor does it study the characteristics of spatial diffusion based on the reaction-diffusion equation [42, 43]. This article analyzes the impact of population migration on the spread of COVID-19 from the perspective of dynamical modeling. Although there are some limitations, it has a certain guiding significance to control the local epidemic due to population movements.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/ Supplementary Material.

AUTHOR CONTRIBUTIONS

All authors have made great contributions to the writing of study and approved the submitted version. XG, LM, and SG established dynamical modeling. LM participated in the program design and provided valuable comments on the manuscript writing. LL collected and processed the relevant published data. SG and LL guided and revised the manuscript.

FUNDING

This study supported by the National Natural Science Foundation of China (Grant Nos. 11671241, 11801398), Program for the Outstanding Innovative Teams (OIT) of Higher Learning Institutions of Shanxi, the Natural Science Foundation for Young Scientists of Shanxi Province (Grant Nos. 201801D221003, 201801D221024), Outstanding Young Talents Support Plan of Shanxi province, and Selective Support for Scientific and Technological Activities of Overseas Scholars of Shanxi province.

REFERENCES

1. World Health Organization. WHO statement regarding cluster of Pneumonia cases in Wuhan, China. Available from: https://www.who.int/zh/ (Accessed February 11, 2020).
2. Wang L, Gao Y-H, Lou L-L, Zhang G-J. The clinical dynamics of 18 cases of COVID-19 outside of Wuhan, China. *Eur Respir J* (2020) **55**:2000398. doi:10.1183/13993003.00398-2020
3. Yang C, Ma Q, Zheng Y, Yang Y. Novel coronavirus transmission pathway. *Chin J Prev Med* (2020) **54**:374–7. doi:10.1056/NEJMoa2001316
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* (2020) **382**:1199–207. doi:10.1056/NEJMoa2001316
5. Sun G-Q, Wang S-F, Li M-T, Li L, Zhang J, Zhang W, et al. Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources. *Nonlinear Dyn* **101**, 1–13. (2020) doi:10.1007/s11071-020-05770-9
6. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020) **6**, eaba9757. doi:10.1126/science.aba9757
7. 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* (2020) **395**:1054–62. doi:10.1016/S0140-6736(20)30566-310.1016/s0140-6736(20)30566-3
8. Tian J, Wu J, Wu J, Bao Y, Weng X, Shi L, et al. Modeling analysis of COVID-19 based on morbidity data in Anhui, China. *Math Biosci Eng* (2020) **17**:2842–52. doi:10.3934/mbe.2020158
9. Feng L-X, Jing S, Jing S-L, Hu S-K, Wang D-F, Huo H-F. Modelling the effects of media coverage and quarantine on the COVID-19 infections in the UK. *Math Biosci Eng* (2020) **17**:3618–36. doi:10.3934/mbe.2020204
10. Sun Q. *Transmission dynamics of COVID-19: mathematical modelling and data fitting* (2020) Available from: https://encyclopedia.pub/1358/ (Accessed July 16, 2020)
11. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv e-prints. Available from: https://arxiv.org/abs/2002.06563v1 (Accessed July 16, 2010). doi:10.1101/2020.02.16.20023465
12. Li M-T, Sun Q, Sun G-Q, Zhang J, Zhao Y, Pei X, et al. Analysis of COVID-19 transmission in Shanxi Province with discrete time imported cases. *Math Biosci Eng* (2020) **17**:3710–20. doi:10.3934/mbe.2020208
13. He S, Tang S, Tang S, Rong L. A discrete stochastic model of the COVID-19 outbreak: forecast and control. *Math Biosci Eng* (2020) **17**:2792–804. doi:10.3934/mbe.2020153
14. Tang S, Tang B, Bragazzi NL, Xia F, Li T, He S, et al. Analysis of COVID-19 epidemic traced data and stochastic discrete transmission dynamic model. *Sci Sin Math* (2020) **50**:1–16. doi:10.1360/SSM-2020-0053
15. Wang X, Tang S, Chen Y, Feng X, Xiao Y, Xu Z. When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemic? a data-driven network modeling analysis. *Sci Sin Math* (2020) **50**:1–10. doi:10.1360/SSM-2020-0037
16. Huang S, Peng Z, Jin Z. Studies of the strategies for controlling the COVID-19 epidemic in China: estimation of control efficacy and suggestions for policy makers. *Sci Sin Math* (2020) **50**:885–98. doi:10.1360/SSM-2020-0043
17. Li Q, Xiao Y, Wu J, Tang S. Modelling COVID-19 epidemic with time delay and analyzing the strategy of confirmed cases-driven contact tracing followed by quarantine. *Acta Math Appl Sin* (2020) **43**:238–50.
18. Liu Z, Magal P, Magal P, Seydi O, Webb G. Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. *Math Biosci Eng* (2020) **17**:3040–51. doi:10.3934/mbe.2020172
19. Zhang J, Li Y, Yao M, Zhang J, Zhu H, Jin Z. Analysis of the relationship between transmission of COVID-19 in Wuhan and soft quarantine intensity in susceptible population. *Acta Math Appl Sin* (2020) **43**:162–73.
20. Musa SS, Gao D, Zhao S, Yang L, Lou Y, He D. Mechanistic modeling of the coronavirus disease 2019 outbreak in the early phase in Wuhan. *Acta Math Appl Sin* (2020) **43**:351–63.
21. Saldaña F, Flores-Arguedas H, Flores-Arguedas H, Ariel Camacho-Gutiérrez J, Barradas I Modeling the transmission dynamics and the impact of the control interventions for the COVID-19 epidemic outbreak. *Math Biosci Eng* (2020) **17**:4165–83. doi:10.3934/mbe.2020231
22. Wang L, Wang J, Wang J, Zhao H, Shi Y, Wang K, et al. Modelling and assessing the effects of medical resources on transmission of novel coronavirus (COVID-19) in Wuhan, China. *Math Biosci Eng* (2020) **17**:2936–49. doi:10.3934/mbe.2020165
23. Yan Y, Chen Y, Liu K, Luo X, Xu B, Jiang Y, et al. Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system. *Sci Sin Math* (2020) **50**:385–92. doi:10.1360/SSM-2020-002610.1360/ssm-2020-0074
24. Chang X, Liu M, Liu M, Jin Z, Wang J. Studying on the impact of media coverage on the spread of COVID-19 in Hubei Province, China. *Math Biosci Eng* (2020) **17**:3147–59. doi:10.3934/mbe.2020178
25. Zhou WK, Wang AL, Xia F, Xiao YN, Tang SY. Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak. *Math Biosci Eng* (2020) **17**:2693–707. doi:10.3934/mbe.2020147
26. Health Commission of Guangdong Province. Available from: http://wsjkw.gd.gov.cn/ (Accessed July 16, 2010).
27. Baidu.com. Available from: https://qianxi.baidu.com/ (Accessed July 16, 2010).
28. Health Commission of Hunan Province. Health Commission of Hunan Province. Information on novel coronavirus pneumonia in Hunan. Available from: http://wjw.hunan.gov.cn/ (Accessed July 16, 2010).
29. National Bureau of Statistics. National data. Available from: http://www.stats.gov.cn/ (Accessed July 16, 2010).
30. Xing Y, Song L, Sun G-Q, Jin Z, Zhang J. Assessing reappearance factors of H7N9 avian influenza in China. *Appl Math Comput* (2017) **309**:192–204. doi:10.1016/j.amc.2017.04.007
31. Sun G-Q, Xie J-H, Huang S-H, Jin Z, Li M-T, Liu L. Transmission dynamics of cholera: mathematical modeling and control strategies. *Commun Nonlinear Sci Numer Simulat* (2017) **45**:235–44. doi:10.1016/j.cnsns.2016.10.007
32. Keeling MJ, Rohani P. *Modeling infectious diseases in humans and animals* Princeton, NJ: Princeton University Press (2007) p. 15–52.
33. Anderson RM, May RM. *Infectious diseases of humans* Oxford, UK: Oxford University Press (1992)
34. Ma Z, Zhou Y, Wang W, Jin Z. *Mathematical modeling and research of infectious disease dynamics* Beijing, China: Science Press (2004) p. 1–21.
35. Li Y, Zhao S, Lou Y, Gao D, Yang L, He D. Epidemiological parameters and models of coronavirus disease 2019. *Acta Phys Sin* (2020) **69**:090202. doi:10.7498/aps.69.20200389
36. Wang Z, Xia C, Chen Z, Chen G. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern* (2020) **99**:1–9. doi:10.1109/TCYB.2019.2960605

37. Xia C, Wang Z, Zheng C, Guo Q, Shi Y, Dehmer M, et al. A new coupled disease-awareness spreading model with mass media on multiplex networks. *Inf Sci* (2019) **471**:185–200. doi:10.1016/j.ins.2018.08.050
38. Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci* (2002) **180**:29–48. doi:10.1016/S0025-5564(02)00108-6
39. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* (2020) **382**:1708–20. doi:10.1056/NEJMoa2002032
40. Hu Z, Cui Q, Han J, Wang X, Sha W E I, Teng Z. Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in Guangdong province, China. *Int J Infect Dis* (2020) **95**:231–40. doi:10.1016/j.ijid.2020.04.010
41. Jia W, Han K, Song Y, Cao W, Wang S, Yang S, et al. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Front Med* (2020) **7**:169. doi:10.3389/fmed.2020.00169
42. Guo Z-G, Sun G-Q, Wang Z, Jin Z, Li L, Li C. Spatial dynamics of an epidemic model with nonlocal infection. *Appl Math Comput* (2020) **377**:125158. doi:10.1016/j.amc.2020.125158
43. Sun G-Q, Jusup M, Jin Z, Wang Y, Wang Z. Pattern transitions in spatial epidemics: mechanisms and emergent properties. *Phys Life Rev* (2016) **19**:43–73. doi:10.1016/j.plrev.2016.08.002

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fphy.2020.603001

The Generalized-Growth Modeling of COVID-19

Ye Wu[1], *Lin Zhang*[2]*, *Wenjing Cao*[2], *Xiaofei Liu*[2] *and Xin Feng*[3,4]

[1]School of Information Engineering, Putian University, Putian, China, [2]School of Science, Beijing University of Posts and Telecommunications, Beijing, China, [3]School of Economics and Management, Yanshan University, Qinhuangdao, China, [4]School of Management Science and Engineering, Hebei GEO University, Shijiazhuang, China

The global spread of the COVID-19 pandemic is changing everything in 2020. It is of crucial importance to characterize the growth patterns during the transmission. In this paper, a generalized-growth model is established to present the evolution of the number of the total confirmed cases changing with time. Due to effective containment, the generalized growth model reveals a piecewise pattern, referred to as the sub-exponential and the sub-linear stages. Moreover, the parameters can quantify the effectiveness of the containment and the trend of resurgence in different regions all over the world. Our model provides a phenomenological approach, which is simple and transparent for better understanding of the typical patterns within the general dynamics. Our model may have implications for possible nowcasting and forecasting of the pandemic trend.

Keywords: nonlinear dynamics, complex system, COVID-19, phenomenological model, sub-exponential growth, sub-linear growth, ordinary differential equation

OPEN ACCESS

Edited by:
Aristides Moustakas,
University of Crete, Greece

Reviewed by:
Chengyi Xia,
Tianjin University of Technology, China
Chen Shen,
Yunnan University of Finance and Economics, China

***Correspondence:**
Lin Zhang
zhanglin2011@bupt.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *04 September 2020*
Accepted: *09 November 2020*
Published: *08 January 2021*

Citation:
Wu Y, Zhang L, Cao W, Liu X and Feng X (2021) The Generalized-Growth Modeling of COVID-19. Front. Phys. 8:603001. doi: 10.3389/fphy.2020.603001

1 INTRODUCTION

The global spread of the COVID-19 pandemic is changing everything in 2020. The first outbreak was in Wuhan during the Chinese New Year's Holiday in late January, and the further outbreaks of COVID-19 in Europe and the US began in March. High infectiousness and worldwide mobility make it more and more difficult to prevent the pandemic. As of October 2020, there are in total more than 40 million confirmed cases all over the world with more than 1 million deaths. The pandemic has already influenced every part of daily life, not only in the way we work, learn, and communicate but also in terms of the economic and political status of different countries and regions all over the world [1, 2]. As a starting point for better understanding and effectively preventing the further spread of COVID-19, exploring the patterns during the propagation of COVID-19 is of crucial importance. Furthermore, the effectiveness of containment measures needs a rapid and transparent response.

Several research directions related to COVID-19 have been developed since the breakout of the deadly virus. The first one is clinical analysis, focusing on the coronavirus itself [3, 4] via, for example, the estimation of the basic reproduction number R_0 [5], the effective reproduction number R_t [6], the infectiveness of the virus [1, 7], and the serial interval [8]. The second one is modeling the spreading processes by assuming some mechanisms. Different generalized SIR models are commonly used. An SIR-X model is illustrated to explain that effective containment is the reason for the subexponential growth during the early spread of COVID-19 in China [9].

The clinical analysis is helpful in understanding the virus itself, and the agent-based models are useful in understanding different factors during the spreading procedure [10, 11]. However, the trend and patterns during the epidemics is still unclear. Moreover, agent-based models are sensitive to assumptions. As time goes by, heterogeneity caused by containment measures and resurgence of the virus would make the model inconsistent and unreliable. A simple, fast, transparent, and robust

model for macroscopic analysis would therefore be more informative from a specific point of view. A generalized logistic growth model is utilized for exploring the patterns [12]. However, this unimodal growth function can only describe one wave of the pandemic with total number of infected cases estimated ahead from the model. It does not work for the second wave of COVID-19.

In this paper, a generalized growth model is used to fit the number of cumulative confirmed cases in different countries and regions to reveal the typical patterns and rate of spread of COVID-19. The heterogeneity is well described by a piecewise fitting procedure with the same growth model. As a result, the sub-exponential and sub-linear stages of the past 9 months are revealed from the real time series of different stages during the worldwide fights against COVID-19. Our findings provide a general methodology to quantify the rate of spread, the effectiveness of containment measures, and the speed of resurgence, which may be useful for further prevention decision making and effectiveness evaluation.

2 MATERIAL AND METHODS

Theoretically, the growth of the virus spread is exponential due to the sufficiency of resources. The growth therefore follows an ordinary differential equation:

$$\frac{dy}{dt} = ry(t),$$

with solution $y(t) = y(0)\, e^{rt}$, where $y(0)$ is the number of infected individuals at the beginning of dynamics, and r is the growth rate. However, outbreaks of many infective diseases show that, compared with exponential growth, sub-exponential growth is more common, such as with the growth rates of HIV/AIDS [13], Ebola [14], and foot-and-mouth disease [14]. What is more, there are several explanations for the sub-exponential growth: spatial heterogeneity, clustering of contacts [14, 15], and the heterogeneity in behavioral changes [13]. In the SIR model, the heterogeneity in time [14–16], space [13], and in combination [17] can lead to sub-exponential growth. For COVID-19, the underlying mechanisms governing sub exponential growth can be difficult to disentangle and, thus, to model.

In terms of non-pharmaceutical intervention, the growth will not be exponential and is referred to as the effectiveness of the mitigation. Moreover, the degree of sub-exponential growth (or even much slower growth) can be referred to as the effectiveness of different control measures. A phenomenological model is therefore more flexible through the introduction of a tuning parameter p that considers the sub-exponential or even much slower growth. The general-growth model is given as follows:

$$\frac{dy}{dt} = ry(t)^p,$$

where $p \neq 1$ is called the deceleration of growth. The solution of the above ordinary differential equation is

$$y(t) = \left[r(1-p)t + y(0)^{1-p}\right]^{\frac{1}{1-p}},\ p \neq 1. \quad (1)$$

When $0 < p < 1$, the model is applied to explain the sub-exponential growth at the early stage of epidemics [15]. Cases with $p > 1$ or $p < 0$ can also appear at certain stages during the process of spreading, showing the evolution of the rate of spread during the long-term pandemic of COVID-19.

In this work, regions and countries with different scales of confirmed case numbers and different containment measures were selected. Then their time series of total confirmed cases were fitted to Eq. 1, the solution of the phenomenological model, which can reproduce a variety of growth profiles. The growth can be divided into different stages of the evolution, reflecting the effectiveness of different mitigation measures. Moreover, the fitted parameters, especially the deceleration of growth parameter p, can be treated as the quantification of the spreading trends, which is comparable among different countries and regions. Generally speaking, smaller p means a slower rate of spread in different types of stages. In fact, when $p = 1$, it is the exponential growth model. The other special case is $p = 0$, which refers to linear growth with $y(t) = y(0) + rt$. When $0 < p < 1$, $y(t)$ grows sub-exponentially, or equivalently, polynominally. The smaller distance between $p \in (0, 1)$ and 1 reflects the closer proximity of sub-exponential growth to the exponential one. Moreover, when $p < 0$, the growth is sub-linear, which is much slower, and the spread dies out gradually. The larger absolute value of p in this case the spread will die out at a faster rate. Therefore, p quantifies the spread of COVID-19 in certain regions.

To differentiate different stages from each other during the pandemic spread, Eq. 1 is fitted to the real data from the very beginning of the confirmed cases in a certain region. With different containment measures, the growth patterns will be changed, and the fitting curve will deviate further from the real data if the fitting curve continues. In this situation, as soon as the fitting curve deviates from the real time series, the former fitting is stopped and a new one is restarted from the same spot. Consequently, piecewise fitting to the real time series with one general form of equation but different parameters is provided. A quantification of the rate of spread at different stages in different regions and countries is obtained. Based on the quantitative results, the evaluation of the effectiveness of mitigation measures in different countries can be obtained. Moreover, the resurgence can also be detected by the flexible and robust methodology. The provided results may shed light on nowcasting and forecasting of the spreading trend and decision making for further prevention of COVID-19.

3 RESULTS

In this paper, we use the number of cumulative confirmed cases of a certain country or region over time. The general growth model is applied to countries or regions with different scales of confirmed cases. Two levels of regions are selected, one is the provincial level (in China) and the other the national level (worldwide). The results reflect the stability and reliability of

TABLE 1 | The fitting parameters for different stages of different regions and countries. The MARE, dates, and duration are also listed for each case.

Region	*p*	*r*	*y* (0)	MARE (%)	Dates	Duration (day)
Henan	0.47	4.39	52	1.77	January 25–February 4	11
	−3.39	6.75*e* + 11	601	0.61	February 4–17	13
Anhui	0.40	5.50	11	5.57	January 24–February 6	14
	−3.10	3.87*e* + 10	589	0.62	February 6–14	9
India	0.82	0.41	684	3.26	March 17–August 23	160
Colombia	0.80	0.40	156	4.79	March 30–August 23	147
Brazil	0.80	0.67	1,260	2.21	March 22–June 7	78
	0.33	333	6.25*e* + 05	0.73	June 7–August 20	75
Mexico	0.63	2.46	314	0.90	March 31–June 20	82
	0.17	758	1.62*e* + 5	0.61	June 20–August 21	63
Russia	0.84	0.71	705	4.00	March 28–April 23	27
	0.47	33	4.90*e* + 5	1.79	April 23–May 17	25
	−0.64	3.53*e* + 07	2.46*e* + 5	0.72	May 17–August 23	99
Peru	0.59	4.89	726	4.23	April 2–June 4	64
	−0.78	6.54*e* + 7	1.73*e* + 5	0.40	June 4–July 13	39
	1.62	4.57*e* − 4	3.16*e* + 5	0.41	July 13–August 23	42
United States	0.61	16.37	2,991	1.91	March 18–April 5	19
	−0.32	2.18*e* + 6	2.46*e* + 5	0.81	April 5–June 17	74
	1.01	0.01	2.04*e* + 6	0.78	June 17–July 31	45
	−0.50	1.16*e* + 8	4.46*e* + 6	0.17	July 31–August 23	24

our phynomenological growth model. Moreover, the fitting parameters in different areas are informative, providing the rate of spread, which reflects the effectiveness of containment measures, as well as the trend of resurgence.

The piecewise fitting results are illustrated in figures. The parameters fitted to certain regions in different stages are listed in **Table 1**. For the goodness of fit in each stage, the mean absolute value of relative errors (MARE) is calculated, which is also listed in the table. The MARE is defined as

$$\mathrm{MARE} = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|y(t)-\widehat{y}(t)\right|}{y(t)},$$

where $\widehat{y}(t), t = 1, \cdots, N$ is the fitted value by the generalized growth model and N is the number of dates under consideration. The dates and duration of different stages can also be found in **Table 1**.

Due to the preventative measures, the growth trend changes as time goes by. The generalized growth model is fitted by adjusting the fitness to the real data. At the early stage, sub-exponential growth was found by several related works [12], where the generalized growth model can fit the data well. With effective containment, the growth speed will slow down, resulting in a misfitting between the real data and the fitting curve. As long as the fitting curve does not match the real data, the current fitting will be stopped, and a new one will be restarted right after the stopping spot. As a consequence, the changes from one stage to another by our fitting methodology reflect the changes of the intrinsic spreading trend of COVID-19 in certain regions. We can quantify not only the effectiveness of prevention through the generalized growth model but also the resurgence of COVID-19.

Within the fitting procedure, the key parameter is p, which presents the rate of spread at different stages in different regions, regardless the scale of confirmed case numbers. The provincial results in China can be referred to as one of the typical patterns of spread with effective prevention. However, when faced with a severe pandemic, other patterns are also revealed by use of our simple fitting model. In the following, besides the provincial results in China, the national-level results are classified according to different numbers of fitting stages, representing different types of spreading trends. For the second wave of COVID-19, our simple, flexible, and robust methodology will be helpful in quantifying the trend of spread in future stages.

Provincial level in China Two Chinese provinces, Henan and Anhui, are selected to reveal the typical growth of confirmed cases with effective prevention. As shown in **Table 1**, under strict local containment, the number of confirmed cases increases sub-exponentially at an early stage, and the parameter p is no more than 0.5. Within 2 weeks, the growth rate changes into a sub-linear one, and p is less than −3. The fitting curves are illustrated in **Figure 1**. The fitting procedure stopped at the ending stage since the increment is getting smaller. **Figure 1** shows a typical curve for a complete process of total confirmed cases. In order to get the COVID-19 pandemic under control, it is necessary that the total confirmed cases going into the sub-linear growth stage, that is, the p parameter, is negative. The ideal case for the spreading trend is changing from a sub-exponential growth into sub-linear one. At the same time, the awareness of resurgence is more important due to the high and asymptomatic infectiveness of COVID-19.

National level with one growth stage During the fitting procedure, division into different segments is necessary to consider containment measures or resurgence. However, when the containment measures are not effective enough, one stage fitting merges. **Figure 2** shows the fitting results of two typical countries: India and Colombia. The ending date is August 23, 2020, in this paper. The growth rates of these two countries are

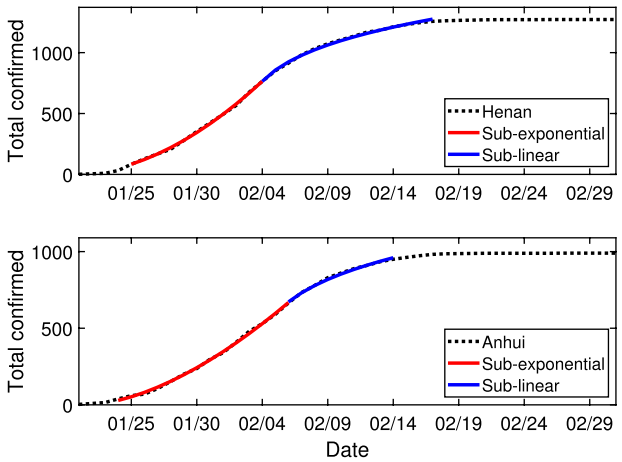


FIGURE 1 | The fitting results of Henan and Anhui provinces in China. The black dashed curves represent the real data. The red and blue segments represent the sub-exponential and sub-linear growth of the total confirmed cases over time. The parameters and dates in considerations for different cases are listed in **Table 1**.

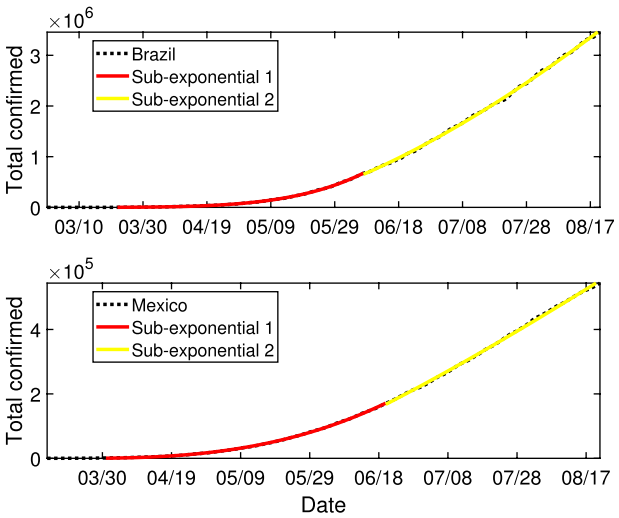


FIGURE 3 | The fitting results of Brazil and Mexico. The black dashed curves represent the real data. The red and yellow curves represent the two different sub-exponential growth segments of the total confirmed cases over time. The parameters and dates in consideration for different cases are listed in **Table 1**.

sub-exponential, with speed parameter p as 0.82 and 0.80, respectively. A larger p means closer to exponential growth, which is indicative of the rapid growth of the pandemic in these countries. The same pattern of growth lasts more than 5 months, resulting in large numbers of confirmed cases, with a scale of millions and hundreds of thousands.

Our quantitative results can tell the severity of the pandemic in these countries. The same sub-exponential growth pattern with a growth parameter p of around 0.8 had lasted more than 5 months in India and Colombia: until the end of August 2020. No effective containment is reflected by our method.

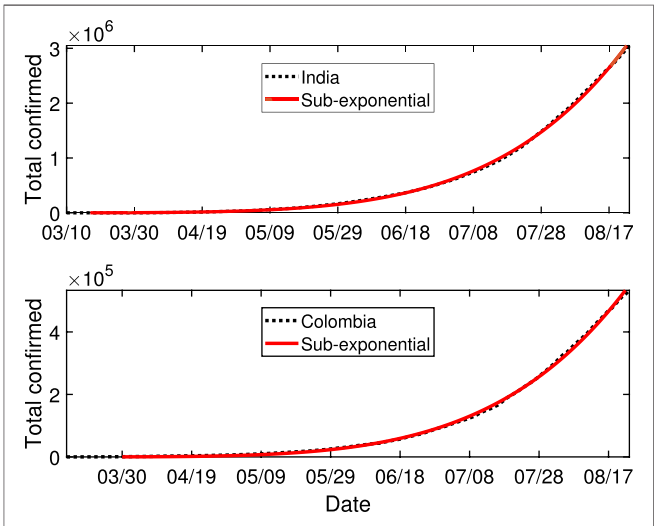


FIGURE 2 | The fitting results of India and Colombia. The black dashed curves represent the real data. The red curves represent the sub-exponential growth of the total confirmed cases changing with time. The parameters and dates in consideration for different cases are listed in **Table 1**.

National level with two growth stages Besides the severe situation with only one growth stage, some countries are showing favorable trend, resulting in two growth stages until the end of our observation date of August 23, 2020. Typical countries with two-stage fitting are illustrated: Brazil and Mexico see **Figure 3**. As shown in **Table 1**, the key parameter p gets smaller in the second sub-exponential stage. For Brazil, p changes to 0.31 from 0.82 after 78 days of growth in the first sub-exponential stage. For

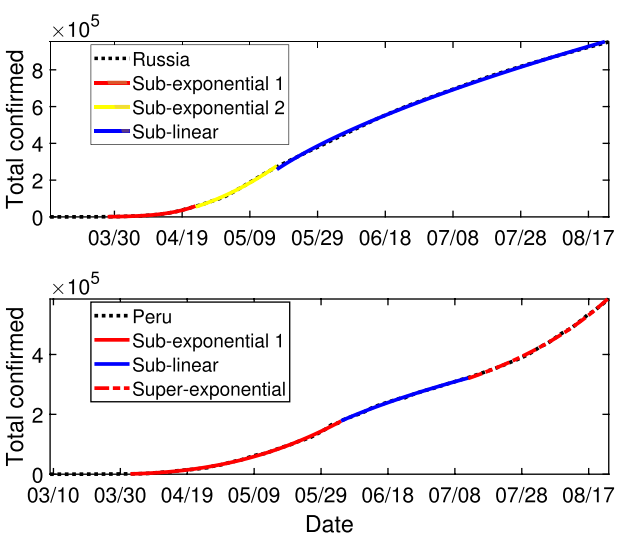


FIGURE 4 | The fitting results of Russia and Peru. The black dashed curves represent the real data. The red and yellow curves represent two different sub-exponential growth of the total confirmed cases over time. The blue curve represents the sub-linear growth. The red dashed curve represents the super-exponential growth in the third stage of Peru. The parameters and dates in consideration for different cases are listed in **Table 1**.

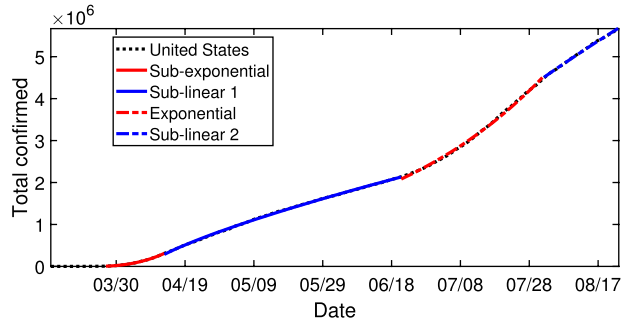


FIGURE 5 | The fitting results of United States of America. The black dashed curves represent the real data. The red curve and red dashed curve represent the sub-exponential and exponential growth of the total confirmed cases changing with time. The blue curve and blue dashed curve represent the sub-linear growth with time. The parameters and dates in consideration for different cases are listed in **Table 1**.

Mexico, p changes to 0.18 from 0.67 after 82 days of growth in the first stage.

Different from the two-stage growing pattern in provinces in China, the two stages are all sub-exponential. In the sub-exponential stage, the rate of spread is always increasing. The positive side is that the speed is slowing down, reflecting the effectiveness of the prevention measures to some extent.

National level with more than two growth stages The change from one stage to another may not always result in improved situations. More growth stages reflect the complex growing trend during the spreading of COVID-19 in certain regions.

One typical country with three stages is Russia. The fitting result is shown in **Figure 4**. It is a common rule that the growth is sub-exponential at the early stages, and this was the same in all our fittings. The second stage of Russia is still sub-exponential with smaller p from 0.84 in the first stage to 0.47 as with the cases of Brazil and Mexico. Its third stage changes into sub-linear growth with $p = -0.64$ and lasts 99 days until August 23, 2020. A slower rate than linear growth is necessary to control the spread. The three stages improve, reflecting the success of containment up to the ending date considered in this paper.

In contrast to Russia, the three stages of Peru show another growth pattern. The fitting result is also shown in **Figure 4**. Different from the situation in Russia, the second stage of Peru is already sublinear with $p = -0.78$. However, 39 days later, a severe outbreak emerged, resulting in super-exponential growth with $p = 1.62 > 1$, reflecting the rapid resurgence of the pandemic in Peru.

The differences between growth trends in Russia and Peru exemplify the difficulty of defending against the pandemic of COVID-19. There is difficulty not only in preventing the outbreak but also in maintaining the temporary victory and preventing the resurgence.

The last case comes from the United States, for which the piecewise fitting splits into four stages see **Figure 5**. The first three stages possess similar patterns to Peru, which are sub-exponential growth with $p = 0.61$, sub-linear growth with $p = -0.32$, and exponential growth with $p = 1.01$. The last stage goes back again into the sub-linear stage with $p = -0.50$ and until our considered ending date August 23, 2020. From the complicated changes in the case of the United States, we can see that it is very difficult to mitigate the pandemic of COVID-19, and, worst of all, resurgence and further waves of outbreak may last for a significant period of time.

To sum up, the growth patterns of pandemic of COVID-19 are split into two main rules: sub-exponential and sub-linear. Among all the fitting results above, the MARE is quite small, see **Table 1**, which provides the validation of the general fitting function.

The change from one stage to another is indicative of the trends of the rate of spread. Strict containment results in smaller p, and the pandemic will be under control within a shorter period. It is a common pattern that the first stage is sub-exponential, with p around 0.8, there is no containment, or the prevention is ineffective. Only extreme containment measures can control the pandemic of COVID-19 within a short space of time (say, two months or so). Effective containment can push the trend into sub-linear growth directly, as are the cases of the second stage in Peru and United States. Less effective prevention measures may get sub-exponential growth with a smaller p, as is the case for the second stage in Brazil, Mexico, and Russia. Ineffective measures make no change to the fitting curve, resulting in one stage lasting more than 5 months, as are the cases of India and Colombia. It is notable that, before the pandemic is under real control, the resurgence of the pandemic will be more severe than the early stage, as is the case with the third stage in Peru and the United States; the speed parameters p are 1.62 and 1.01, respectively, which means the super-exponential growth and exponential one for the trend of the resurgence. The result would be disastrous.

4 DISCUSSION

Starting with the observation of common sub-exponential growth in epidemic spreading of COVID-19, a phenomenological general-growth model is established to quantify the spreading speed of COVID-19 in different areas. The piecewise fitting methodology can be adapted to different trends during the pandemic, and it is simple, transparent, and robust. Our fitting results provide a reference for the rate of spread in certain countries and regions. The key parameter p not only gives the reference for speed in each stage but also a comparative reference among stages and regions.

Comprehensive fitting reveals typical patterns during the spread of COVID-19. The early stage is sub-exponential growth. Effectiveness of containment measures can be reflected by later stages. Sub-linear growth is necessary for the spreading under control. However, after a long period sub-linear growth, the resurgence may take over, and growth may thus be faster than ever.

There are various classical dynamic propagation models with considerations of heterogeneity that are used to explain the mechanism of growth patterns. Our work provides a generalized result with one consistent fitting function but different speed parameters. Our findings are informative for the general growth of the pandemic in different regions, especially the areas with severe waves during the pandemic. Compared with the fitting results of two provinces in China under strict containment, there is still a long way to go for the countries and regions currently severely affected by the pandemic. It is shown in the cases of the transmission from

Stage 2 to Stage 3 in Peru and the United States that even if the spread is under control with a sub-linear growth speed, close attention and strict containment are still necessary to prevent later outbreaks due to the highly infectious nature of the virus and its ability to exist asymptomatically.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.xlsx.

AUTHOR CONTRIBUTIONS

YW, LZ, and XF designed the analysis. LZ, WC, and XL analyzed the data. YW and LZ wrote the paper.

FUNDING

This work was jointly supported by the Fundamental Research Funds for the Central Universities (No. 2019XD-A11), the National Natural Science Foundation of China (Grant Nos. 11971074, 61671005, 61672108, 61976025).

REFERENCES

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* (2020) 395:470–3. doi:10.1016/S0140-6736(20)30185-9
2. Wang P, Zheng X, Li J, Zhu B. Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos, Solit Fractals* (2020) 139:110058. doi:10.1016/j.chaos.2020.110058
3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* (2020) 395:507–13. doi:10.1016/S0140-6736(20)30211-7
4. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* (2020) 382:1708–20. doi:10.1056/NEJMoa2002032
5. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the wuhan novel coronavirus 2019-ncov. *J Evid Base Med* (2020) 13:3–7. doi:10.1111/jebm.12376
6. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis* (2020) 92:214–7. doi:10.1016/j.ijid.2020.01.050
7. Chan JFW, Yuan S, Kok KH, To KK-W, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus inducating person-to-person transmission: a study of a family cluster. *Lancet* (2020) 395:514–23. doi:10.1016/S0140-6736(20)30154-9
8. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of covid-19 among publicly reported confirmed cases. *Emerg Infect Dis* (2020) 26:1341–3. doi:10.3201/eid2606.200357
9. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* (2020) 368:742–6. doi:10.1126/science.abb4557
10. Wang Z, Xia C, Chen Z, Chen G. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern* (2020) 99:1–9. doi:10.1109/TCYB.2019.2960605
11. Shao Q, Xia C, Wang L, Li H. A new propagation model coupling the offline and online social networks. *Nonlinear Dynam* (2019) 98:2171–83. doi:10.1007/s11071-019-05315-9
12. Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the covid-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dynam* (2020) 101:1561–81. doi:10.1007/s11071-020-05862-6
13. Szendroi B, Csányi G. Polynomial epidemics and clustering in contact networks. *Proc Biol Sci* (2004) 271(Suppl 5):S364–6. doi:10.1098/rsbl.2004.0188
14. Chowell G, Viboud C, Hyman JM, Simonsen L. The western africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates. *PLoS Curr* (2015) 7:1–10. doi:10.1371/currents.outbreaks.8b55f4bad99ac5c5db3663e916803261
15. Viboud C, Simonsen L, Chowell G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics* (2016) 15:27–37. doi:10.1016/j.epidem.2016.01.002
16. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. The basic reproductive number of ebola and the effects of public health measures: the cases of Congo and Uganda. *J Theor Biol* (2004) 229:119–26. doi:10.1016/j.jtbi.2004.03.006
17. Finkensta DT, Grenfell BT. Time series modelling of childhood diseases: a dynamical systems approach. *J Appl Stat* (2000) 49:187–205. doi:10.1111/1467-9876.00187

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fphy.2020.589963

Characterizing COVID-19 Transmission: Incubation Period, Reproduction Rate, and Multiple-Generation Spreading

Lin Zhang[1], *Jiahua Zhu*[2], *Xuyuan Wang*[1], *Juan Yang*[1], *Xiao Fan Liu*[3]* *and Xiao-Ke Xu*[4]*

[1]School of Science, Beijing University of Posts and Telecommunications, Beijing, China, [2]The First Hospital of Putian City, Putian, China, [3]Web Mining Laboratory, Department of Media and Communication, City University of Hong Kong, Hong Kong, China, [4]College of Information and Communication Engineering, Dalian Minzu University, Dalian, China

Understanding the transmission process is crucial for the prevention and mitigation of COVID-19 spread. This paper contributes to the COVID-19 knowledge by analyzing the incubation period, the transmission rate from close contact to infection, and the properties of multiple-generation transmission. The data regarding these parameters are extracted from a detailed line-list database of 9,120 cases reported in mainland China from January 15 to February 29, 2020. The incubation period of COVID-19 has a mean, median, and mode of 7.83, 7, and 5 days, and, in 12.5% of cases, more than 14 days. The number of close contacts for these cases during the incubation period and a few days before hospitalization follows a log-normal distribution, which may lead to super-spreading events. The disease transmission rate from close contact roughly decreases in line with the number of close contacts with median 0.13. The average secondary cases are 2.10, 1.35, and 2.2 for the first, second, and third generations conditioned on at least one offspring. However, the ratio of no further spread in the 2nd, 3rd, and 4th generations are 26.2, 93.9, and 90.7%, respectively. Moreover, the conditioned reproduction number in the second generation is geometrically distributed. Our findings suggest that, in order to effectively control the pandemic, prevention measures, such as social distancing, wearing masks, and isolating from close contacts, would be the most important and least costly measures.

Keywords: COVID-19, incubation period, close contacts, superspreading, effective reproduction number, spreading tree

OPEN ACCESS

Edited by:
Aristides (Aris) Moustakas, Natural History Museum of Crete; University of Crete, Greece

Reviewed by:
Gui-Quan Sun, North University of China, China
Chen Chu, Yunnan University of Finance and Economics, China
Wen-Xuan Wang, Beijing University of Posts and Telecommunications (BUPT), China

***Correspondence:**
Xiao Fan Liu
xf.liu@cityu.edu.hk
Xiao-Ke Xu
xuxiaoke@foxmail.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *31 July 2020*
Accepted: *16 November 2020*
Published: *11 January 2021*

Citation:
Zhang L, Zhu J, Wang X, Yang J, Liu XF and Xu X-K (2021) Characterizing COVID-19 Transmission: Incubation Period, Reproduction Rate, and Multiple-Generation Spreading. Front. Phys. 8:589963. doi: 10.3389/fphy.2020.589963

1 INTRODUCTION

As of July 2020, the cumulative confirmed cases of COVID-19 worldwide have exceeded 17.4 million with over 572 thousand dead. There are 22 countries with more than 100,000 confirmed cases of as of July 14, 2020. The high transmissibility of the SARS-CoV-2 virus has substantially changed people's hygiene habits, social relations, and forms of work and schooling during and after the pandemic [1]. In the absence of pharmaceutical intervention measures, public policies such as city lockdowns and workplace and school closures can mitigate the spread of disease, though with substantial economic and societal costs. The indecision regarding restarting the economy and stopping the pandemic has resulted in a wave of outbreaks in many countries [2].

Understanding the characteristics of the COVID-19 transmission process is crucial in finding a middle ground between restoring economic and societal order and controlling the pandemic. Previous research has shown that COVID-19 can be infectious pre-symptomatically [3], i.e., the virus is transmissive even without symptom onset. Finding out the incubation period's duration and the virus reproducibility during the incubation period and shortly after symptom onset but before hospitalization is thus an urgent necessity [4].

Considering the incubation period, as of Jan. 26, the mean and median were 5 and 4.75 days (obtained by 125 patients) [5]. Confirmed cases reported from Jan. 4 to Feb. 24 showed a median incubation period of 5.1 days (obtained from 181 patients) [6]. By Jan. 22, using 425 patients, the mean incubation period was 5.2 days, and [7]. Reference [8] gave a shorter incubation period of 4.2 days, inferring that COVID-19 is more infectious than initially estimated. As of Mar. 31, the mean incubation time is estimated as 8.0 with a standard deviation of 4.75 [9]. Through a renewal process, the estimated median of the incubation period is 8.1 days, which is longer than other studies [10]. The mean and median of the incubation periods were 5.84 and 5.0 days via bootstrap for groups with an age of ≥ 40, and they otherwise demonstrated a significant difference [11]. By meta-analysis, the incubation period was modeled with a lognormal distribution, and the mean and median were 5.8 and 5.1 days [12].

The transmission rate is defined as the probability that an infection occurs among susceptible people within a specific group. It is an important index for providing an indication of how social interactions are related to transmission risk. Nine reports were listed in [13], showing a rate of 35% (95% CI 27–44), depending on infection caused by different contact methods.

One of the most important indices for infectious disease is the basic reproductive number. Numerous studies are devoted to its estimate. It is estimated to be 2.2 [14], which is higher than SARS-COV and MERS-CoV [15]. More estimates for the basic reproduction number are 4.7–6.6 [8], 2.24–3.58 [16], 3.77 (95% CI 3.51–4.05) [5], and 3.60 (3.49–3.84) [17]. The effective reproduction number is changing with time; it changed from 2.35 (1.15–4.77) to 1.05 (0.41–2.39) due to lockdown in Wuhan within 1 week [18].

The best-known model within infectious disease epidemiology is the SEIR (susceptible-exposed-infectious-recovered) model with different generalization. These models are utilized at the population level for the proportion of each state at given time, aiming to investigating the strategic decisions or effectiveness of the mitigation measures. For illustration, effective containment can explains the subexponential growth in China [19], and effects of containment measures in Italy are also analyzed by an SEIR-like model [17]. More results can be found [20–27].

Clinical investigations may suffer from a limited sample size and biased sampling from the population, leading to geometrical or demographic-dependent results. Different samples and different methods also lead to different results for data analysis and estimates. Simulation of disease spread and mitigation policies require a precise setting of incubation period [19, 28]. Metapopulation disease transmission models require a prerequisite setting of the transmission rate during social gathering events to predict disease spreading range [18, 29, 30]. For a better estimate of the reproduction number, a real data sample is a crucial ingredient. However, it is difficult to collect. Considering the demand of investigating the properties and modeling of COVID-19, fine data extracted from informative line-list records can provide supporting evidence for the existing results and solid foundation for further study.

In this work, we estimate the parameters of concern from a large scale epidemiological line-list database, which contains the contact history and epidemiological timelines of 9,120 confirmed COVID-19 cases in China [31]. The duration of the incubation period and the details of close contacts and contact scenarios are extracted from the line-list. Spreading trees are reconstructed from the potential transmission pairs in the line-list data set. Hidden in the line-list records of confirmed cases, we have collected 421 chains of spreading with a total confirmed cases number of 1,140. We fit proper distributions to the incubation period as well as scale of close contact. The reproducibility is presented by the spreading tree, which can be referred to as the effective reproduction number under strict containment measures in China.

The incubation distribution is fitted by Weibull distribution with a mean and median of 7.83 and 7 days, respectively; this is in agreement with [9]. Larger data size and longer observation period tend to result in larger incubation period, which is coincidence with the long tailed nature of Weibull distribution. For the secondary attack rate, there are much fewer results due to the lack of data. We have obtained 412 close contact events to investigate the transmission rate. It is revealed that the relationship between the contact scale and transmission rate is not strongly related no matter if it is a linear or nonlinear relation. Moreover, the contact scale is fitted by Lognormal distribution, and the empirical distribution of transmission rate is also given. Finally, the reproducibility of COVID-19 under strict containment measures is investigated by the multiple-generation spreading structure, revealing the effectiveness of the containment measures in China. The key contributions of our work are those that aim for a better understanding of the properties of COVID-19 spread.

The rest of the paper is organized as follows. **Section 2** describes the data and methods. **Section 3** reports the empirical analysis and models fitted. **Section 4** discusses the implications of results and provides an explanation based on branching process and the necessity of ultra-strict prevention measures.

2 DATA AND METHODS

The line-list database used in this paper contains hand-coded information extracted from 9,120 public reported cases by mainland China health commissions from January 15 to February 29, 2020. A typically reported item is as follows:

"Patient ID: Huainan-25.

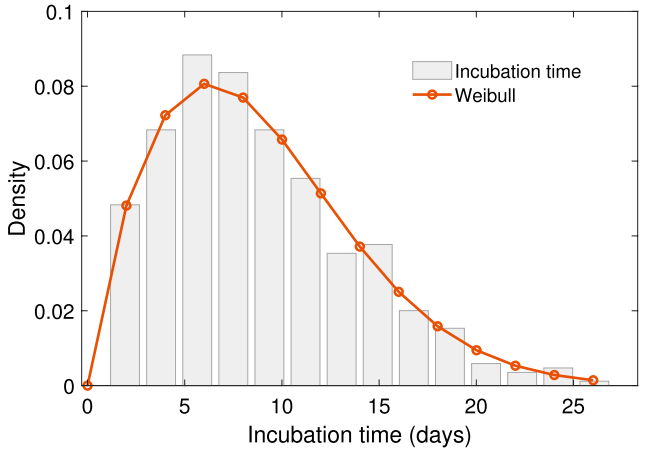


FIGURE 1 | The empirical distribution and Weibull distribution fitting of incubation time. The Weibull distribution has density function $f_W(x;A,B) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$, $x \geq 0$ with $\lambda = 9.93$, and $k = 1.79$. The K–S test is 0.17, which means that Weibull distribution is proper for the data.

The patient Huainan-25 is a 59-year-old woman who is the wife of the Huainan-26 patient. On February 12, she developed fever, muscle soreness, and other symptoms. On February 14, she went to the hospital for treatment and stayed at the hospital for observation. On February 15, her nucleic acid test was tested positive, and doctors diagnosed her as a suspected patient. Two days later, she was confirmed. Doctors have traced back 3 close contacts, all of whom have been quarantined for medical observation. During the New Year's holiday, she had close contact with her daughter, son-in-law, and granddaughter. Her son-in-law, an asymptomatic patient with a history of suspicious exposure in Hefei, stayed at a designated hospital for observation. Doctors have traced back his 46 close contacts, all of whom have been quarantined for medical observation."

The original extracted line-list database contains the epidemiology timelines, e.g., the possible date of virus exposure and date of symptom onset, for each case. We define the incubation period as the time between virus exposure and symptom onset. There are 457 cases with both dates of exposure and date of symptoms reported in the line-list database.

Close contact events are social events and scenarios such as living together, dining together, traveling together, and working together. There were 412 close contact events with the numbers of close contacts and secondary infections reported. Multiple-generation transmissions can form tree structures that originated from an initial infection. There are 421 transmission chains identified from the line-list.

3 RESULTS

3.1 Duration of the Incubation Period

The incubation period is a vital variable considering the control of the pandemic. The quarantine period of close contact people with an infected individual depends on this variable. The quarantine was usually 14 days for COVID-19. However, for strict prevention, it was suggested at the Information Office of Beijing Municipality press conference on June 28 that after the first 14 days, another 14-day quarantine is necessary in some high-risk areas.

TABLE 1 | Numbers of different types of close contact events.

Type	Number of cases	Proportion (%)
Living	386	93.69
Dining	7	1.70
Working	3	0.73
Traveling	1	0.24
Others	15	3.64

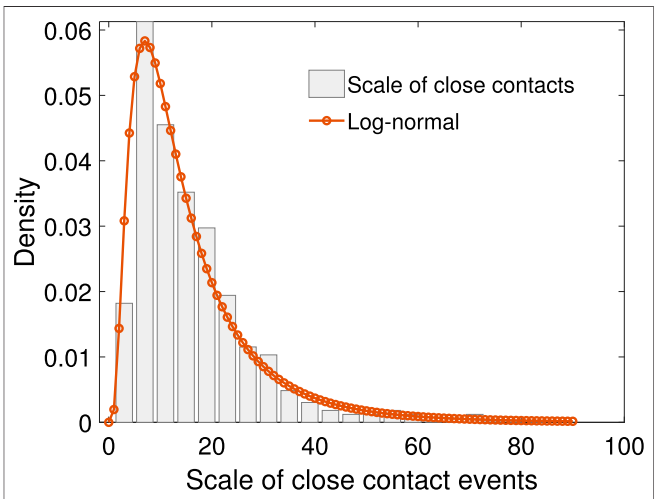


FIGURE 2 | The empirical distribution of scale of close contact events with log-normal density fitting. The density function of this log-normal distribution is $f_{LN}(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$, where the fitting parameters are $\mu = 2.495$, $\sigma = 0.745$. The p-value of the K–S test for log-normal distribution is 0.18. The notion that the scale is log-normal distributed is not rejected.

The reason why another 14 days quarantine is necessary can be found from the distribution of incubation time. The sample with 457 incubation time reveals that it is a skewed distribution, see **Figure 1**. The mean, median, and mode calculated from the sample data are 7.83, 7, and 5 days, respectively. Moreover, the empirical probability of incubation period exceeds 14 is

$$P(\text{Incubation period} \geq 14 \text{ days}) = 0.125.$$

That is to say, the chance of an asymptomatic infected individual turning into symptomatic after 14 days is about 12.5%. For strict control of COVID-19, longer quarantine is necessary. A Weibull distribution is fitted to the empirical data, with shift 1 to the right for avoiding zero. The density function is

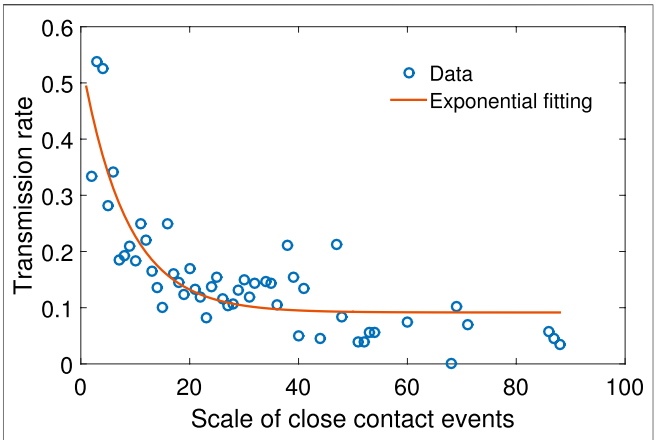


FIGURE 3 | The scatter plot of N vs. the mean of p, together with exponential function fitting. The relationship between N and p can be fitted with function $p = a * \exp(-b * N) + c$, where the fitting parameters are $a = 0.453$, $b = 0.121$, and $c = 0.092$, and the goodness of fit index is $R^2 = 0.706$.

$$f_W(x; A, B) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, x \geq 0,$$

with $\lambda = 9.93$, and $k = 1.79$. The K–S test is 0.17, which means that Weibull distribution is proper for the data.

3.2 Scale of Close Contact Events

The scale of close contact events is the number of people involved in one event of where people have gathered together in a specific way. **Table 1** shows the number of different types of social events and scenarios that can potentially facilitate disease spreading. Among the 412 close contact events, more than 93.7% happened by way of living together.

The period of our dataset is the early stage of COVID-19 spread in China. The distribution of the scale in close contact events is a natural feature seen when people are free from movement regardless of the COVID-19 pandemic. The contact scale is intrinsically positive, with a few enormously high data points typically arising. The lognormal distribution is an ideal descriptor of such data, with a positive range, right skewness, heavy right tail, and easily computed parameter estimates. Supported by the K-S test with a value of 0.18, the log-normal distribution shows the proper fitting among the positive, skewed, heavy-tailed distribution candidate. The mechanism of lognormal distributed data in ecology can be obtained by stochastic differential equation [32], which would be another topic for further investigation. The result is shown in **Figure 2**,

The density function of this log-normal distribution is

$$f_{LN}(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},$$

where the fitting parameters are $\mu = 2.495$, $\sigma = 0.745$. The p-value of the K-S test for log-normal distribution is 0.18. It is not a rejected notion that the scale is log-normal distributed. Though there are various prevention measures worldwide, various contact events result in a heterogeneous scale of close contact. The heavy-tailed nature of the close contact scale reveals a non-neglectable possibility of super-spreading events. Therefore, in order to effectively control the pandemic, maintaining social distance and wearing masks should be effective measures.

3.3 Transmission Rate and the Scale of Close Contact Events

We define the transmission rate as the number of people infected in one close contact event over the number of people in that event. **Figure 3** shows the scatter plot between the transmission rate and the scale of close contact events. It can be seen that the rate drops as the scale of events increases in a non-linear fashion.

Let p be the transmission rate and N the total number of people in the close contact events. Based on our sample, given the value of N, the mean p is calculated. The relationship between N and p can be fitted with the following exponential function:

$$p = a * \exp(-b * N) + c,$$

where the fitting parameters are $a = 0.453$, $b = 0.121$, and $c = 0.092$, and the goodness of fit index is $R^2 = 0.706$. The exponential relation reveals that a larger scale of close contact tends to smaller secondary incidence p. However, the fitting is not convinced enough. The correlation coefficient between N and p is −0.29, implying that neither a linear nor a nonlinear relation between N and p is significant. In other words, p can be treated as a natural feature of COVID-19, with weak monotonic decrease of N. The mean and median of the transmission rate is 0.20 and 0.13 with an interquartile range 0–0.3. The empirical distribution of transmission rate is also given in **Figure 4**. Protective measures to decrease the transmission rate would be the least cost ways to prevent the pandemic, such as maintaining social distance, wearing masks, and washing hands.

3.4 Spreading Tree Structures

Transmission events can create tree structures to map disease spread. There are in total 421 chains verified from the record data. Among the chains, there are 311 chains with secondary cases, out

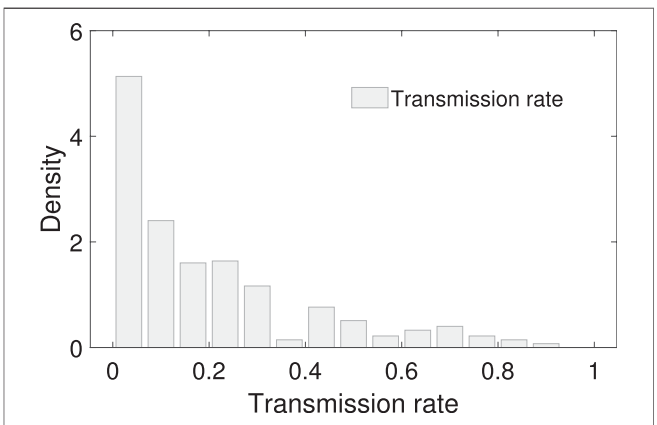


FIGURE 4 | The empirical distribution of the transmission rate. The mean and median of the transmission rate is 0.20 and 0.13 with an interquantile range 0–0.3. No proper common distribution fits the empirical distribution.

TABLE 2 | The reproduction number in each generation of transmission in the spreading tree.

Tree depth (Size)	Ratio of no child (Size)	Conditioned mean	Mean
1(421)	26.2% (110)	2.10	1.55
2(654)	93.9% (614)	1.35	0.08
3(54)	90.7% (49)	2.2	0.20
4(11)	100.0% (11)	0	0

of which there are 654 children in the second generation. However, due to effective prevention, there are only 54 and 11 children in the third and fourth generations, respectively. No fifth generation is observed in our dataset.

The reproduction number of an infection is the number of secondary infectees infected by the same confirmed individual. We define the reproduction number in each generation by dividing the number of infected people in the next generation by the present one. Based on the existence of at least one child in the next generation, the mean reproduction number in the first, second, and third generations are 2.10, 1.35, and 2.2. However, without the conditional restriction, the mean are 1.55, 0.08, and 0.2, respectively, see **Table 2**.

Using the sample of number of secondary cases caused by the 311 infectors in the first generation, empirical distribution, together with geometric fitting is shown in **Figure 5**. The geometric distribution law is $P(k \text{ secondary cases}) = p(1-p)^{k-1}$ for $k \geq 1$. The parameter is $p = 0.50$, and the K–S test value is 0.73.

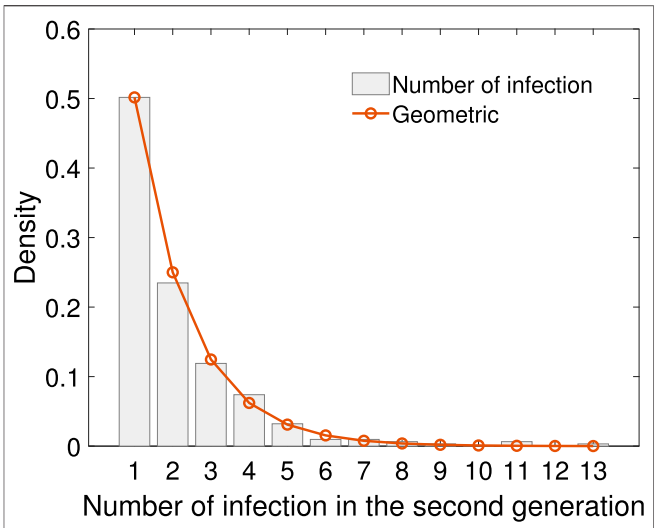


FIGURE 5 | The empirical distribution of infection numbers in the second generation with geometric fitting. The geometric distribution law is $P(k\text{secondary cases}) = p(1-p)^{k-1}$ for $k \geq 1$. The parameter is $p = 0.50$, and the K–S test value is 0.73.

4 DISCUSSION

In this study, based on the details of confirmed cases reported by the mass media, the following features are explored: the Weibull distribution of the incubation period, the Log-normal distribution of the scale of close contact events, the geometric distribution of the reproduction number in different generations of virus transmissions, and the statistical feature of secondary attack rate.

As far as we know, the distribution of the close contacts' scale is released for the first time that it is log-normal distributed due to lack of data. This heavy-tailed distribution reveals a relatively larger possibility of super spreading events comparing to light-tailed distributions. To reduce the secondary infection, it is important to take adequate measures to reduce the scale of close contact and reduce the secondary infections. Moreover, efforts should be made to trace back the close contacts to cut off the possible spreading chain in advance.

It is notable that the method here is universal to all infectious diseases. The crucial step is the line-list record of each confirmed case and the detailed transmission relationship in the spreading tree structure. For infectious diseases where only non-pharmaceutical measurement can be applied to prevent its spreading, detailed record keeping of each confirmed case and the contact history is crucial. The tree structure is good evidence for the spreading trend and helpful for the precise estimation of the effective reproductive number. Moreover, contact history is useful to nip severe infectious diseases in the bud.

Theoretically, the reproduction number, say R, is a determining index quantifying the transmissibility. To control the pandemic, R should be less than one. Borrowed from the theory of branching processes, there is a phase transition with a critical value $R = 1$. If $R < 1$, then, with a probability of one, the spread of a certain disease will die out with exponential speed. However, when $R > 1$, the rate of spread will exponentially increase. The probability of exponential increas can be obtained as the minimum nonnegative solution to the equation $f(s) = s$ for $s \in (0, 1)$, where $f(s)$ is the generating function of the reproduction number. From this point of view, the propagation of COVID-19 is an issue of "all or nothing." From this point of view, the control measures would be as strict as possible to avoid the possibility of exponential increase.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/PDGLin/COVID19_EffSerialInterval_NPI

AUTHOR CONTRIBUTIONS

LZ and JZ contributed equally as first authors. JZ, X-FL, and X-KX designed the analysis, LZ, XW, JY, and X-KX analyzed the data. LZ and X-FL wrote the paper.

FUNDING

This work was jointly supported by the Fundamental Research Funds for the Central Universities (No. 2019XD-

A11), the National Natural Science Foundation of China (Grant Nos. 11971074, 61671005, 61672108, 61976025, 61773091), the LiaoNing Revitalization Talents Program (XLYC1807106).

REFERENCES

1. Wang C, Horby PW, Hayden FG. A novel coronavirus outbreak of global health concern. *Lancet.* (2020a) 395:470–3. doi:10.1016/S0140-6736(20)30185-9
2. Wang P, Zheng X, Li J. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solit. Fractals* (2020b) 139: 110058. doi:10.1016/j.chaos.2020.110058
3. Hu Z, Xu C, Jin G. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in nanjing, China. *Sci. China Life Sci.* (2020) 63:706–11. doi:10.1007/s11427-020-1661-4
4. Du Z, Xu X, Wu Y. Serial interval of covid-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.* (2020) 26:1341–3. doi:10.3201/eid2606.200357
5. Yang Y, Lu QB, Liu MJ. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China (2020). medRxiv preprint. doi:10.1101/2020.02.10.20021675
6. Lauer SA, Grantz KH, Bi Q. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* (2020) 127:577–83. doi:10.7326/M20-0504
7. Liu Y, Gayle AA, Wilder-Smith A. The reproductive number of COVID-19 is higher compared to sars coronavirus. *J. Trav. Med.* (2020a) 6:1–4. doi:10.1093/jtm/taaa021
8. Sanche S, Lin YT, Xu C. The novel coronavirus, 2019-ncov, is highly contagious and more infectious than initially estimated (2020). arXiv preprint arXiv:2002.03268.
9. You C, Deng Y, Hu W. Estimation of the time-varying reproduction number of covid-19 outbreak in China. *Int. J. Hyg. Environ. Health* (2020) 228:1–7. doi:10.1016/j.ijheh.2020.113555
10. Qin J, You C, Lin Q. Estimation of incubation period distribution of covid-19 using disease onset forward time: a novel cross-sectional and forward follow-up study (2020). medRxiv preprint. doi:10.1101/2020.03.06.20032417
11. Men K, Wang X, Li Y. Estimate the incubation period of coronavirus 2019 (covid-19) (2020). medRxiv preprent. doi:10.1101/2020.02.24.20027474
12. McAloon C, Collins A, Hunt K Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open* 10 (2020) 1–9. doi:10.1136/bmjopen-2020-039652
13. Liu Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for sars-cov-2 (2020b). *Lancet* 395:e47. doi:10.1016/S0140-6736(20)30462-1
14. Li Q, Guan X, Wu P. Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* (2020a) 382:1199–207. doi:10.1056/NEJMoa2001316
15. Paules CI, Marston HD, Fauci AS. Coronavirus infections–more than just the common cold. *JAMA Network* (2020) 328:707–8. doi:10.1001/jama.2020.0757
16. Zhao S, Lin Q, Ran J. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis* (2020) 92:214–7. doi:10.1016/j.ijid.2020.01.050
17. Gatto M, Bertuzzo E, Maria L. Spread and dynamics of the covid-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl. Acad. Sci. USA* (2020) 117:10484–91. doi:10.1073/pnas.2004978117/-/DCSupplemental
18. Kucharski AJ, Russell TW, Diamond C. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *Lancet Infect. Dis.* (2020) 20:553–8. doi:10.1016/S1473-3099(20)30144-4
19. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* (2020) 368: 742–6. doi:10.1126/science.abb4557
20. Zhao S, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quant. Biol.* (2020) 7:1–9. doi:10.1007/s40484-020-0199-0
21. Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. *J. Med. Virol.* (2020) 92:645–59. doi:10.1002/jmv.25750
22. Li HJ, Bu Z, Wang Z, Cao J. Dynamical clustering in electronic commerce systems via optimization and leadership expansion. *IEEE Transactions on Industrial Informatics* (2020b) 16:5327–34. doi:10.1109/TII.2019.2960835
23. Ndairou F, Area I, Nieto JJ. Mathematical modeling of covid-19 transmission dynamics with a case study of wuhan. *Chaos, Solit. Fractals* (2020) 135:1–6. doi:10.1016/j.chaos.2020.109846
24. Chen YC, Luy PE, Chang CS. A time-dependent sir model for COVID-19 with undetectable infected persons (2020). arXiv:200300122
25. Giordano G, Blanchini F, Bruno R. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* (2020) 26:855–60. doi:10.1038/s41591-020-0883-7
26. Li HJ, Wang Z, Cao J. Optimal estimation of low-rank factors via feature level data fusion of multiplex signal systems. *IEEE Trans. Knowl. Data Eng.* (2020c) 13:33–9. doi:10.1109/TKDE.2020.3015914
27. Li HJ, LinWangMatjaz Perc YZ. Optimization of identifiability for efficient community detection. *New J. Phys.* 22 (2020d) 1–10. doi:10.1088/1367-2630/ab8e5e
28. Peng L, Yang W, Zhang D. Epidemic analysis of COVID-19 in China by dynamical modeling (2020).arXiv:200206563v2
29. Chinazzi M, Davis JT, Ajelli M. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020) 368: 395–400. doi:10.1126/science.aba9757
30. Hellewell J, Abbott S, Gimma A. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* (2020) 8:e488–96. doi:10.1016/S2214-109X(20)30074-7
31. Ali ST, Wang L, Lau EHY. Serial interval of sars-cov-2 was shortened over time by nonpharmaceutical interventions. *Science* (2020) 19:17–39. doi:10.1126/science.abc9004
32. Crow EL, Shimizu K. *Lognormal distribution*. New York, NY: Marcel Dekker (1987).

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 14 January 2021
doi: 10.3389/fphy.2020.588513

The Suppression of Epidemic Spreading Through Minimum Dominating Set

Jie Wang[1], *Lei Zhang*[2]*, *Wenda Zhu*[1], *Yuhang Jiang*[1], *Wenmin Wu*[1], *Xipeng Xu*[1] and *Dawei Zhao*[2]

[1]College of Science, Civil Aviation University of China, Tianjin, China, [2]Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China

COVID-19 has infected millions of people, with deaths in more than 200 countries. It is therefore essential to understand the dynamic characteristics of the outbreak and to design effective strategies to restrain the large-scale spread of the epidemic. In this paper, we present a novel framework to depress the epidemic spreading, by leveraging the decentralized dissemination of information. The framework is equivalent to finding a special minimum dominating set for a duplex network which is a general dominating set for one layer and a connected dominating set for another layer. Using the spin glass and message passing theory, we present a belief-propagation-guided decimation (BPD) algorithm to construct the special minimum dominating set. As a consequence, we could immediately recognize the epidemic as soon as it appeared, and rapidly immunize the whole network at minimum cost.

Keywords: COVID-19, epidemic spreading, minimum dominating set, multiplex network, control

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Chengyi Xia,
Tianjin University of Technology, China
Gui-Quan Sun,
North University of China, China

***Correspondence:**
Lei Zhang
zhanglei@sdas.org

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *29 July 2020*
Accepted: *19 October 2020*
Published: *14 January 2021*

Citation:
Wang J, Zhang L, Zhu W, Jiang Y, Wu W, Xu X and Zhao D (2021) The Suppression of Epidemic Spreading Through Minimum Dominating Set. Front. Phys. 8:588513. doi: 10.3389/fphy.2020.588513

INTRODUCTION

In recent years, many different epidemic types have frequently erupted and spread worldwide, leading to not only great economic losses, but also widespread social disruption [7, 11, 14, 19]. The ongoing COVID-19 pandemic in particular has infected more than 10 million people causing deaths in more than 200 countries [5, 6, 10, 11, 24]. It is therefore essential to understand the dynamic characteristics of epidemic outbreaks and to design effective strategies to restrain the large-scale spread of the epidemic [4, 9, 13, 16–18].

The most traditional method of controlling an epidemic outbreak is network immunization, which cuts the path of the spread, by immunizing parts of the network nodes [2, 20, 23]. The common solution of network immunization uses the centrality index of the network, in which the nodes with large degree, k-shell, articulation points, betweenness, or spectral coefficients are selected as the immunized nodes. Recently, the control strategy of an epidemic, based on the competing spread on top of a duplex network, has attracted a lot of attention [1, 3, 15–17]. It is known that the information propagation in a virtual social network plays an important role in controlling epidemic spreading. A representative scenario is that the infectious diseases spread in a real physical contact network of individuals, will naturally lead to the propagation of crisis awareness information on virtual social networks. The individuals who obtained the awareness information will take measures to protect themselves, resulting in the suppression of the spread of the infectious disease. This is the framework of the competing spreading process of disease and awareness information on the duplex network coupled with the physical network and virtual network. Similarly, [21] introduced the competing spreading framework of the computer virus and patch, in which the patch dissemination restrains the virus propagation while they spread through different channels [21].

However, both of these competing spreading frameworks suffer the problem that it is surely improper to assume all network nodes have the ability to monitor their neighbors and recognize the epidemic and would like to distribute the awareness information or patch.

To compensate for the shortcomings of the competing spreading framework, Takaguchi et al. proposed a novel control method which takes advantage of both traditional network immunization and competing spreading [12]. They introduced the observer node into the networks, which could monitor the state of its neighbors and could therefore recognize the epidemic if its neighbors were infected. In addition, the observer node could send awareness information to its uninfected neighbors (i.e. immunize its uninfected neighbors) if it recognized the epidemic. However, such a method focusses on a single network which ignores the fact that physical monitoring and information propagation usually takes place in different channels. As an extension, [8] introduced an improved model of deploying an observer node on multiplex networks, consisting of an epidemic spreading layer and an information spreading layer. Though the observer node-based method could monitor all the network nodes and efficiently recognize the infection, it cannot immunize all the network nodes when infection spreading is confirmed.

In this sense, Zhao et al. introduced a strategy of controlling an epidemic through the minimum dominating set (MDS) of the multiplex network [22]. They proposed a framework coupled by a duplex network (one layer is the immunization distribution network, another is the epidemic spreading network layer) and a central node and let the nodes of the MDS of the duplex network be the observer nodes. When one neighbor is infected in the epidemic spreading network layer, the observer could immediately recognize the infection and send the information to the central node. The central node would then distribute the crisis information or patches to all the observers and the observers then spread them to their neighbors in the immunization distribution network. Under the definition of MDS of the multiplex network, such a method could recognize the infection immediately as it happened and could also immediately immunize all the network nodes using the smallest number of observer nodes.

However, Zhao's method belongs to the centralization strategy. As an improvement, in this paper we propose an MDS-based decentralized strategy. We introduce a special MDS of the duplex network which is a general dominating set for one layer and a connected dominating set for another layer. Using the spin glass and message passing theory, we present a belief-propagation-guided decimation (BPD) algorithm to construct the special MDS. As a consequence, we could monitor all the network nodes and immediately recognize the infection, and rapidly immunize all the network nodes when infection spreading is confirmed, while the cost is minimized.

DEPLOYMENT OF OBSERVER NODE BASED ON SPECIAL MDS OF DUPLEX NETWORK

Different from the Zhao's method, our method only contains one duplex network where one layer is the immunization distribution network (IL), and the other is the epidemic spreading network layer (EL). Here, we introduce a novel definition of special MDS of the duplex network (see **Figure 1**). It is a minimum node set which is a general dominating set for the EL and a connected dominating set for the IL; at the same time, the size of the node set is minimum.

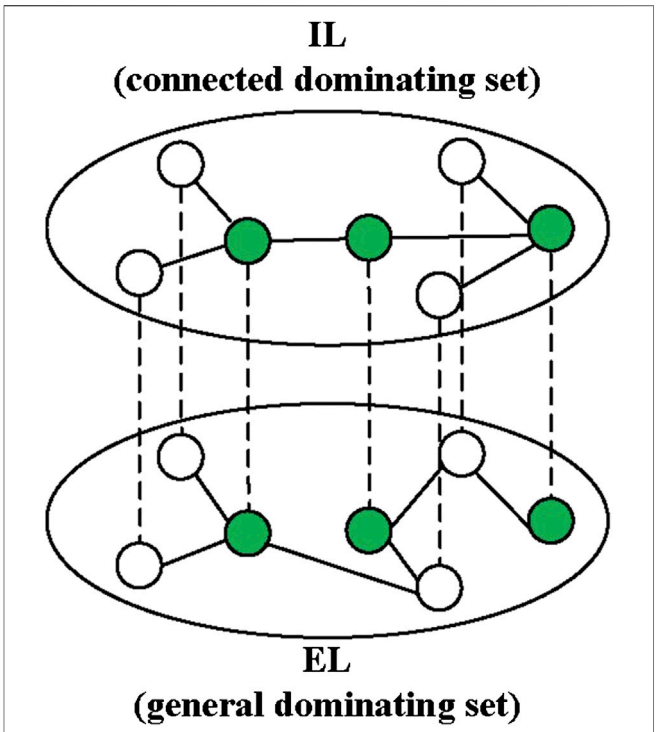


FIGURE 1 | The set of green nodes is a special MDS of the duplex network. It is a general dominating set of the EL and a connected dominating set of the IL.

To control the epidemic spreading, we let the nodes of the special MDS of the duplex network be the observer nodes. The observer nodes could monitor its neighbors' states in EL, thus could immediately recognize the infection after the neighbor was infected. In addition, the observer nodes could diffuse the information to its neighbors in IL after they recognized the infection or after receiving the information from its neighbors. Therefore, under the definition of the special MDS of the duplex network, the epidemic could be immediately recognized as soon as it appeared since all network nodes were monitored, and all network nodes could receive the related information due to the observer nodes being fully connected in IL.

CONSTRUCTION OF SPECIAL MDS OF DUPLEX NETWORK

When considering the discussion in the above section, the problem of controlling epidemic spreading is converted to construct a special MDS of the duplex network. Obviously, the construction of the special MDS is NP-hard, and so we can only try to obtain a near optimal solution. Therefore, all the MDS discussed in the rest of the paper are actually a near-optimal

special dominating set. In [22]; Zhao et al. proposed a framework of the construction method of the MDS of the multiplex network, but no detailed implementations are given. In this section, we first show the detailed procedure of the construction of general MDS of the duplex network, then a simple greedy strategy is given to construct the special MDS of the duplex network.

Construction of the General MDS of the Duplex Network

Let $G = (G^1, G^2)$ be a duplex network, where G^i denotes the network layer i. In our framework, G^1 and G^2 represent the IL and EL, respectively. The two layers have the same set V of N nodes and a distinct set of edges. If D is a dominating set of G, it must be a dominating set of each network layer of G, which means every node in $V\backslash D$ must be connected to at least one node of D in every layer. The state of node $i \in D$ is called occupied, denoted by $c_i = 1$. Otherwise $i \in V\backslash D$ is called empty but observed and denoted as $c_i = 0$. For constructing the MDS of the duplex network, we introduce the following partition function

$$Z = \sum_{\underline{c}} \prod_{i \in V} \left\{ e^{-xc_i} \left[1 - (1 - c_i) \prod_{j \in \Gamma_i^1} \left(1 - c_j\right) \right] \left[1 - (1 - c_i) \prod_{j \in \Gamma_i^2} \left(1 - c_j\right) \right] \right\} \tag{1}$$

where $\underline{c} = \{c_1, c_2, \ldots, c_N\}$ represents a configuration of node states, and Γ_i^k denotes the set of neighbors of i in G^k. x is regarded as the positive re-weighting parameter. Obviously, in **Eq. 1** $\left[1 - (1 - c_i) \prod_{j \in \Gamma_i^1} (1 - c_j)\right]\left[1 - (1 - c_i) \prod_{j \in \Gamma_i^2} (1 - c_j)\right] = 1$ implies i is occupied or observed in both layers; otherwise the term is equal to 0. Therefore, Z is only contributed by the configuration of the dominating set. In particular, only the MDS of the duplex network contributes to Z in the case of $x \to \infty$.

Solving **Eq. 1** using the message passing theory, we have the expression of marginal probability $q_i^{c_i}$ of i being in state c_i as follows

$$q_i^{c_i} = \frac{Q_i^{c_i}}{\sum_{c_i} Q_i^{c_i}} \tag{2}$$

where

$$Q_i^1 = e^{-x} \prod_{j \in \Gamma_i^1 \vee \Gamma_i^2} \sum_{c_j} q_{j \to i}^{(c_j,1)} \tag{3}$$

$$Q_i^0 = \prod_{j \in \Gamma_i^1 \vee \Gamma_i^2} \sum_{c_j} q_{j \to i}^{(c_j,0)} - \left[\prod_{j \in \Gamma_i^1} q_{j \to i}^{(0,0)} \prod_{\substack{j \in \Gamma_i^2 \\ j \notin \Gamma_i^1}} \sum_{c_j} q_{j \to i}^{(c_j,0)} + \prod_{j \in \Gamma_i^2} q_{j \to i}^{(0,0)} \prod_{\substack{j \in \Gamma_i^1 \\ j \notin \Gamma_i^2}} \sum_{c_j} q_{j \to i}^{(c_j,0)} - \prod_{j \in \Gamma_i^1 \vee \Gamma_i^2} q_{j \to i}^{(0,0)} \right]. \tag{4}$$

$\Gamma_i^1 \vee \Gamma_i^2$ represents the union of set Γ_i^1 and Γ_i^2, i.e. the set of all neighbors of nodes i in both layers. $q_{j \to i}^{(c_j,c_i)}$ denotes the joint probability of j being in state c_j and i in c_i under the assumption that node i is deleted from the network. $q_{j \to i}^{(c_j,c_i)}$ is given by

$$q_{j \to i}^{\{c_j,c_i\}} = \frac{Q_{j \to i}^{\{c_j,c_i\}}}{\sum_{c_j,c_i} Q_{j \to i}^{\{c_j,c_i\}}} \tag{5}$$

where

$$Q_{j \to i}^{\{1,0\}} = e^{-x} \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,1)} \tag{6}$$

$$Q_{j \to i}^{\{1,1\}} = Q_{j \to i}^{\{1,0\}} \tag{7}$$

$$Q_{j \to i}^{\{0,1\}} = \begin{cases} \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)} - \prod_{\substack{k \in \Gamma_j^2 \\ k \neq i}} q_{k \to j}^{(0,0)} \prod_{\substack{k \in \Gamma_j^1 \\ k \notin \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)}, & \text{if } j \in \Gamma_i^1, j \notin \Gamma_i^2 \\ \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)} - \prod_{\substack{k \in \Gamma_j^1 \\ k \neq i}} q_{k \to j}^{(0,0)} \prod_{\substack{k \in \Gamma_j^2 \\ k \notin \Gamma_j^1 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)}, & \text{if } j \notin \Gamma_i^1, j \in \Gamma_i^2 \\ \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)}, & \text{if } j \in \Gamma_i^1 \wedge \Gamma_i^2 \end{cases} \tag{8}$$

$$Q_{j \to i}^{\{0,0\}} = \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)} - \left[\prod_{\substack{k \in \Gamma_j^1 \\ k \neq i}} q_{k \to j}^{(0,0)} \prod_{\substack{k \in \Gamma_j^2 \\ k \notin \Gamma_j^1 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)} + \prod_{\substack{k \in \Gamma_j^2 \\ k \neq i}} q_{k \to j}^{(0,0)} \prod_{\substack{k \in \Gamma_j^1 \\ k \notin \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)} - \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} q_{k \to j}^{(0,0)} \right] \tag{9}$$

Eq. 9 is called the belief propagation equation which has been widely used in different kinds of optimization problems.

Based on **Eqs.1–9**, we can use the classical Belief-Propagation-Guided Decimation Algorithm (BPD) to identify the MDS of a multiplex network. As discussed in [22]; we first simplify the calculation of q_i^1.

In case j has not been occupied but observed in at least one layer, the states of its neighbors in the corresponding layer could not be considered. Therefore, $Q_{j \to i}^{\{0,1\}}$ and $Q_{j \to i}^{\{0,0\}}$ can be further simplified as

$$Q_{j \to i}^{\{0,1\}} = \prod_{\substack{k \in \Gamma_j^1 \vee \Gamma_j^2 \\ k \neq i}} \sum_{c_k} q_{k \to j}^{(c_k,0)}, \quad \text{if } \left(j \in \Gamma_i^1, j \notin \Gamma_i^2, j \in O^2\right) \text{or } \left(j \in \Gamma_i^2, j \notin \Gamma_i^1, j \in O^1\right) \tag{10}$$

$$Q_{j\to i}^{\{0,0\}}=\begin{cases}\prod\limits_{\substack{k\in\Gamma_j^1\vee\Gamma_j^2\\ k\neq i}}\sum\limits_{c_k}q_{k\to j}^{(c_k,0)}-\prod\limits_{\substack{k\in\Gamma_j^1\\ k\neq i}}q_{k\to j}^{(0,0)}\prod\limits_{\substack{k\in\Gamma_j^2\\ k\notin\Gamma_j^1\\ k\neq i}}\sum\limits_{c_k}q_{k\to j}^{(c_k,0)}, & \text{if } j\notin O^1, j\in O^2\\ \prod\limits_{\substack{k\in\Gamma_j^1\vee\Gamma_j^2\\ k\neq i}}\sum\limits_{c_k}q_{k\to j}^{(c_k,0)}-\prod\limits_{\substack{k\in\Gamma_j^2\\ k\neq i}}q_{k\to j}^{(0,0)}\prod\limits_{\substack{k\in\Gamma_j^1\\ k\notin\Gamma_j^2\\ k\neq i}}\sum\limits_{c_k}q_{k\to j}^{(c_k,0)}, & \text{if } j\in O^1, j\notin O^2\\ \prod\limits_{\substack{k\in\Gamma_j^1\vee\Gamma_j^2\\ k\neq i}}\sum\limits_{c_k}q_{k\to j}^{(c_k,0)}, & \text{if } j\in O^1\wedge O^2\end{cases}\tag{11}$$

where O^k represents the set of nodes which have been observed in G^k, and $O^1\wedge O^2$ denotes the intersection of O^1 and O^2. In addition, Q_i^0 can also be simplified if i has been observed in at least one layer

$$Q_i^0=\begin{cases}\prod\limits_{j\in\Gamma_i^1\vee\Gamma_i^2}\sum\limits_{c_j}q_{j\to i}^{(c_j,0)}-\prod\limits_{j\in\Gamma_i^1}q_{j\to i}^{(0,0)}\prod\limits_{\substack{j\in\Gamma_i^2\\ j\notin\Gamma_i^1}}\sum\limits_{c_j}q_{j\to i}^{(c_j,0)}, & \text{if } i\notin O^1, i\in O^2\\ \prod\limits_{j\in\Gamma_i^1\vee\Gamma_i^2}\sum\limits_{c_j}q_{j\to i}^{(c_j,0)}-\prod\limits_{j\in\Gamma_i^2}q_{j\to i}^{(0,0)}\prod\limits_{\substack{j\in\Gamma_i^1\\ j\notin\Gamma_i^2}}\sum\limits_{c_j}q_{j\to i}^{(c_j,0)}, & \text{if } i\in O^1, i\notin O^2\\ \prod\limits_{j\in\Gamma_i^1\vee\Gamma_i^2}\sum\limits_{c_j}q_{j\to i}^{(c_j,0)}, & \text{if } i\in O^1\wedge O^2\end{cases}\tag{12}$$

The procedure of BPD is now given by:

(1) The states of all nodes are initialized to be unoccupied and unobserved. The initial values of all joint probabilities $\{q_{j\to i}^{(c_j,c_i)}\}$ could set to be random numbers.
(2) Perform iterative calculations of the joint probabilities $\{q_{j\to i}^{(c_j,c_i)}\}$ based on **Eq. 5** for T_0 rounds. Then calculate the occupation probability $\{q_i^1\}$ for all nodes using the obtained $\{q_{j\to i}^{(c_j,c_i)}\}$ according to **Eq. 2**.
(3) Select the first r unoccupied nodes with the highest occupation probabilities occupied. Then update the states of all remaining nodes.
(4) Remove the edges between the nodes that have been observed and remove all isolated observed nodes in every layer.
(5) If the remaining network still has unobserved nodes, calculate **Eq. 5** for T_1 rounds and then calculate **Eq. 2**. Whether using **Eq. 3** and **Eq. 4** or **Eq. 3** and **Eq. 12** in **Eq. 2**, depends on whether i is unobserved or observed.
(6) Repeat operations (3)–(5) until all nodes are observed in all layers.

Greedy Construction of Special MDS of Duplex Network

In this section, we construct the special MDS of the duplex network based on the general MDS constructed via the BPD algorithm. We delete all of the nodes of layer IL which do not belong to the general MDS and denote the remaining network as IL'. If IL' is not fully connected, it must have more than one cluster. We then remove nodes form V/D into IL' to reduce the number of clusters based on the following operations.

(1) If the insertion of one node into the IL' can reduce the number of clusters, we select the node whose insertion can reduce the number of clusters to be inserted into the IL' the most.
(2) If the insertion of one node into the IL' cannot reduce the number of clusters, we select two linked nodes whose insertion can reduce the number of clusters to be inserted into the IL' the most.

The operations are performed repeatedly until the IL' is fully connected. In this way, a special MDS of the duplex network is constructed.

RESULT

In this section, we verify the effectiveness of our method by comparing it with the random algorithm and the simulated annealing algorithm.

The random algorithm is very simple. It randomly selects an unoccupied node and changes the node's state to be occupied, until a special dominating set is constructed. The simulated annealing algorithm is widely used to solve many different kinds of optimization problems. To construct a special MDS of the duplex network, we define the following energy function

$$E=u\sum_i c_i+v\sum_i f_i+w\sum_\alpha\sum_i\left(1-o_i^\alpha\right).\tag{13}$$

In the energy function, $f_i=1$ indicates that node i is occupied but has no occupied neighbor in layer IL, otherwise $f_i=0$. $o_i^\alpha=1$ indicates that node i is observed in layer α, that is it has at least one occupied neighbor in layer α, otherwise $o_i^\alpha=1$. In addition, u, v, and w are tunable parameters. At the beginning of the SA algorithm, we construct a special dominating set for the duplex network based on the random algorithm. Then we calculate the corresponding energy E based on **Eq.13**. At each time step of SA, we perform the following operations:

(1) Randomly select one node from the network. If the selected node is occupied, change its state to unoccupied. We then calculate the new energy E^{new}.
(2) Randomly select one node from the network. If the selected node is unoccupied, change its state to occupied. We then calculate the new energy E^{new}.

The operation is accepted with probability $e^{-\beta(E^{new}-E)}$, otherwise, the new occupation should be omitted. In this paper, we set $u=v=w=0.5$. β represents the inverse of the temperature. Its initial value is set to be 0.5 and then is increased by 10^{-6} at each time step. The SA algorithm terminates when β equals β_{max}. It should be noted that the SA algorithm could not ensure that the constructed dominating set is a special

TABLE 1 | The size of constructed dominating set of duplex network coupled by different network layers

Network	Layers	N	$\|E^{\alpha}\|$	Random	SA	Our
1	ER (IL)	10,000	40,085	9,784	4,462	3,407
	ER (EL)		40,130			
2	ER (IL)	10,000	40,085	9,723	4,564	3,402
	SF (EL)		39,987			
3	SF (IL)	10,000	39,986	8,810	4,669	2,384
	SF (EL)		39,987			
4	SF (IL)	10,000	39,986	9,211	4,532	3,052
	ER (EL)		40,130			

dominating set, therefore, a greedy process is still needed after the termination of the SA algorithm.

We performed experiments on four different kinds of duplex networks which are coupled by the SF network and ER random network. The results are given in **Table 1**. We can see that our method constructs the smallest special dominating set for any duplex network. Therefore, we could use the smallest observer nodes to control the epidemic spreading based on our deployment scheme.

CONCLUSION

In this paper, we designed a deployment scheme of observer nodes to control epidemic spreading based on the special MDS of the duplex network. Through the nodes of the special MDS of the duplex network, we could immediately recognize the epidemic as soon as it appeared and could immunize all network nodes—at the same time, the cost remains minimal. The deployment problem of the observer nodes is then converted to the construction problem of the special MDS of the duplex network. We used the BPD algorithm and a greedy algorithm to construct the special MDS of the duplex network. Our solution is verified to be very efficient on different kinds of duplex networks.

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

LZ and DZ proposed the framework. JW, WZ, YZ, WW, and XX developed the mathematical theory. JW and WZ performed the experiment. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (61702309), the Innovation and entrepreneurship training program of Civil Aviation University of China (201910059044), the National Key Research and Development Project (2018YFE0119700) and the Key Research and Development Project of Shandong Province (2019JZZY010132).

REFERENCES

1. Gao B, Deng Z, Zhao D. (2016) Competing spreading processes and immunization in multiplex networks. *Chaos, Solit Fractals* 93, 175–181. doi:10.1016/j.chaos.2016.10.013
2. Gao C, Liu J, Zhong N. (2010) Network immunization with distributed autonomy-oriented entities. *IEEE Trans Parallel Distr Syst* 22, 1222–1229. doi:10.1109/TPDS.2010.197
3. Granell C, Gómez S, Arenas A. (2014) Competing spreading processes on multiplex networks: awareness and epidemics. *Phys Rev E - Stat Nonlinear Soft Matter Phys* 90, 012808. doi:10.1103/PhysRevE.90.012808
4. Guo Z-G, Sun G-Q, Wang Z, Jin Z, Li L, Li C. (2020) Spatial dynamics of an epidemic model with nonlocal infection. *Appl Math Comput* 377, 125158. doi:10.1016/j.amc.2020.125158
5. Kröger M, Schlickeiser R. (2020) Gaussian doubling times and reproduction factors of the covid-19 pandemic disease. *Frontiers in Physics* 8, 276. doi:10.3389/fphy.2020.00276
6. Li H-J, Bu Z, Wang Z, Cao J. (2019a) Dynamical clustering in electronic commerce systems via optimization and leadership expansion. *IEEE Transactions on Industrial Informatics* 16, 5327–5334. doi:10.1109/TII.2019.2960835
7. Li H-J, Wang L, Zhang Y, Perc M. (2020) Optimization of identifiability for efficient community detection. *New J Phys* 22, 063035. doi:10.1088/1367-2630/ab8e5e
8. Li Z, Zhu P, Zhao D, Deng Z, Wang Z. (2019b) Suppression of epidemic spreading process on multiplex networks via active immunization. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29, 073111. doi:10.1063/1.5093047
9. Liu C, Zhou N, Zhan X-X, Sun G-Q, Zhang Z-K. (2020) Markov-based solution for information diffusion on adaptive social networks. *Appl Math Comput* 380, 125286. doi:10.1016/j.amc.2020.125286
10. Saberi M, Hamedmoghadam H, Madani K, Dolk HD, Morgan A, Morris JK, et al. (2020) Accounting for underreporting in mathematical modelling of transmission and control of covid-19 in Iran. *Frontiers in Physics* 8, 289. doi:10.3389/fphy.2020.00289
11. Sun GQ, Wang SF, Li MT, Li L, Zhang J, Zhang W, et al. (2020) Transmission dynamics of covid-19 in wuhan, China: effects of lockdown and medical resources. *Nonlinear Dyn* 101, 1981–1993. doi:10.1007/s11071-020-05770-9
12. Takaguchi T, Hasegawa T, Yoshida Y. (2014) Suppressing epidemics on networks by exploiting observer nodes. *Phys Rev E - Stat Nonlinear Soft Matter Phys* 90, 012807. doi:10.1103/PhysRevE.90.012807
13. Wang X, Jia D, Gao S, Xia C, Li X, Wang Z. (2020a) Vaccination behavior by coupling the epidemic spreading with the human decision under the game theory. *Appl Math Comput* 380, 125232. doi:10.1016/j.amc.2020.125232
14. Wang Z, Bauch CT, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, et al. (2016) Statistical physics of vaccination. *Phys Rep* 664, 1–113. doi:10.1016/j.physrep.2016.10.006
15. Wang Z, Guo Q, Sun S, Xia C. (2019) The impact of awareness diffusion on sir-like epidemics in multiplex networks. *Appl Math Comput* 349, 134–147. doi:1016/j.amc.2018.12.045
16. Wang Z, Xia C, Chen Z, Chen G. (2020b) Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Transactions on Cybernetics*. Available online at: https://ieeexplore.ieee.org/abstract/document/8957067.
17. Xia C, Wang Z, Zheng C, Guo Q, Shi Y, Dehmer M, et al. (2019) A new coupled disease-awareness spreading model with mass media on multiplex networks. *Inf Sci* 471, 185–200. doi:10.1016/j.ins.2018.08.050
18. Yin Q, Wang Z, Xia C, Dehmer M, Emmert-Streib F, Jin Z. (2020) A novel epidemic model considering demographics and intercity commuting on complex dynamical networks. *Appl Math Comput* 386, 125517. doi:10.1016/j.amc.2020.125517

19. Zhao D, Li L, Peng H, Luo Q, Yang Y. (2014a) Multiple routes transmitted epidemics on multiplex networks. *Phys Lett* 378, 770–776. doi:10.1016/j.physleta.2014.01.014
20. Zhao D, Wang L, Li S, Wang Z, Wang L, Gao B. (2014b) Immunization of epidemics in multiplex networks. *PLoS One* 9, e112018. doi:10.1371/journal.pone.0112018
21. Zhao D, Wang L, Wang Z, Xiao G. (2018) Virus propagation and patch distribution in multiplex networks: modeling, analysis, and optimal allocation. *IEEE Trans Inf Forensics Secur* 14, 1755–1767. doi:10.1109/TIFS.2018.2885254
22. Zhao D, Xiao G, Wang Z, Wang L, Xu L. (2020a) Minimum dominating set of multiplex networks: definition, application, and identification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 99, 1–15. doi:10.1109/TSMC.2020.2987163
23. Zhao D, Yang S, Han X, Zhang S, Wang Z. (2020b) Dismantling and vertex cover of network through message passing. *IEEE Transactions on Circuits and Systems II: Express Briefs* 99, 1–15. doi:10.1109/TCSII.2020.2973414
24. Zhuang Z, Cao P, Zhao S, Lou Y, Wang W, Yang S, et al. (2020) Estimation of local novel coronavirus (covid-19) cases in wuhan, China from off-site reported cases and population flow data from different sources. *Frontiers in Physics* 8, 336. doi:10.3389/fphy.2020.00336

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fphy.2020.591876

Renormalization Group Approach to Pandemics as a Time-Dependent SIR Model

Michele Della Morte[1] *and Francesco Sannino*[2,3]*

[1]IMADA & CP3-Origins, University of Southern Denmark, Odense, Denmark, [2]CP3-Origins and D-IAS, University of Southern Denmark, Odense, Denmark, [3]Dipartimento di Fisica, E. Pancini, University di Napoli, Federico II and INFN sezione di Napoli Complesso Universitario di Monte S, Napoli, Italy

We generalise the epidemic Renormalization Group framework while connecting it to a SIR model with time-dependent coefficients. We then confront the model with COVID-19 in Denmark, Germany, Italy and France and show that the approach works rather well in reproducing the data. We also show that a better understanding of the time dependence of the recovery rate would require extending the model to take into account the number of deaths whenever these are over 15% of the cumulative number of infected cases.

Keywords: Compartmental models, epidemic renormalisation group, pandemic, pandemic (COVID-19), COVID-19

OPEN ACCESS

Edited by:
Matjaž Perc,
University of Maribor, Slovenia

Reviewed by:
Muhua Zheng,
East China Normal University, China
Kimmo Tuominen,
University of Helsinki, Finland
Antonio Scala,
Institute of Complex Systems, National Research Council (CNR), Italy

***Correspondence:**
Francesco Sannino
sannino@cp3.sdu.dk

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *05 August 2020*
Accepted: *13 November 2020*
Published: *15 January 2021*

Citation:
Della Morte M and Sannino F (2021) Renormalization Group Approach to Pandemics as a Time-Dependent SIR Model. Front. Phys. 8:591876. doi: 10.3389/fphy.2020.591876

1 INTRODUCTION

Epidemic dynamics is often described in terms of a simple model introduced long time ago in [1]. Here, the affected population is described in terms of compartmentalised sub-populations that have different roles in the dynamics. Then, differential equations are designed to describe the time evolution of the various compartments. The sub-populations can be chosen to represent (S)usceptible, (I)nfected and (R)ecovered individuals (SIR model), obeying the following differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\gamma S\frac{I}{N},\\ \frac{dI}{dt} &= \gamma S\frac{I}{N} - \varepsilon I,\\ \frac{dR}{dt} &= \varepsilon I,\end{aligned} \tag{1.1}$$

with the conservation law

$$S(t) + I(t) + R(t) = N. \tag{1.2}$$

The system depends on three parameters, namely γ, ε and N. Due to the conservation law (2), only two equations are independent, so that one can drop the equation for S.

The cumulative number of infected, $\tilde{I}(t)$, that we are interested in, is related to the above sub-populations as

$$\tilde{I}(t) = I(t) + R(t). \tag{1.3}$$

We can therefore re-write the two independent SIR equations as

$$\frac{d\tilde{I}(t)}{dt} = \gamma\big(\tilde{I}(t) - R(t)\big)\left(1 - \frac{\tilde{I}(t)}{N}\right), \tag{1.4}$$

$$\frac{dR(t)}{dt} = \varepsilon\left(\tilde{I}(t) - R(t)\right). \tag{1.5}$$

Empirical modifications of the basic SIR model exist and range from including new sub-populations to generalise the coefficients γ, ε to be time-dependent in order to better reproduce the observed data (see Refs. [2–5] for recent examples).

Recently the epidemic Renormalization Group approach (eRG) to pandemics, inspired by particle physics methodologies, was put forward in [6]. The approach builds on scale transformations and invariances as introduced by Wilson in [7, 8], applied in this case to the evolution of social systems. The method has been further explored in [9], where it was demonstrated the eRG effectiveness when describing how the pandemic spreads across different regions of the world. This allowed to successfully simulate the onset of the second wave pandemic in Europe [10]. The eRG framework provided useful also to determine the impact and level of social distancing when combined with mining the Google and Apple mobility data [11]. We refer the reader to the review in [12] for a clear and comparative discussion of several approaches used in order to model the COVID-19 pandemic.

The goal of the present work is to further extend the eRG framework to properly take into account the number of recovered cases so that a better understanding of the reproduction number can also be achieved. We will start, first, by providing a map between the original eRG model and certain modified SIR models, establishing in this way and for the first time an exact connection between two powerful approaches to pandemics. We will finally test the framework via COVID-19 data.

1.1 Reviewing the Epidemic Renormalization Group

In the *epidemic Renormalization group* (eRG) approach [6], rather than the number of cases, it is convenient to discuss its logarithm, which is a more slowly varying function

$$\alpha(t) = \ln \tilde{I}(t), \tag{1.6}$$

where ln indicates the natural logarithm. The derivative of α with respect to time provides a new quantity that we interpret as the *beta-function* of an underlying microscopic model. In statistical and high energy physics, the latter governs the time (inverse energy) dependence of the interaction strength among fundamental particles. Here it regulates infectious interactions.

More specifically, as the Renormalization group equations in high energy physics are expressed in terms of derivatives with respect to the energy μ, it is natural to identify the time as $t/t_0 = -\ln(\mu/\mu_0)$, where t_0 and μ_0 are respectively a reference time and energy scale. The latter is introduced to ease the reading and notation for the physics community. We choose t_0 to be one week so that time is measured in weeks, and will drop it in the following. Thus, the dictionary between the eRG equation for the epidemic strength α and the high-energy physics analog is

$$\beta(\alpha) = \frac{d\alpha}{d\ln(\mu/\mu_0)} = -\frac{d\alpha}{dt}. \tag{1.7}$$

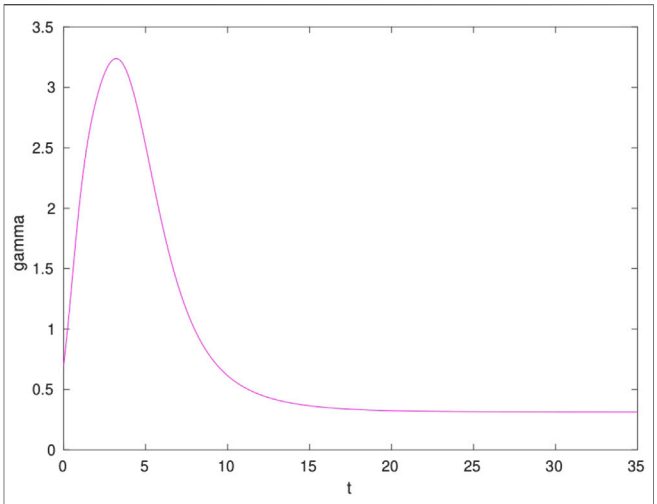


FIGURE 1 | Typical form of time-dependent γ matching the SIR system to the eRG one.

It has been shown in [6] that α captures the essential information about the infected population within a sufficiently isolated region of the world. The pandemic beta function can be parametrised as

$$-\beta(\alpha) = \frac{d\alpha}{dt} = \tilde{\gamma}\,\alpha\left(1 - \frac{\alpha}{a}\right)^n, \tag{1.8}$$

whose solution, for $n = 1$, is a familiar logistic-like function

$$\alpha(t) = \frac{ae^{\tilde{\gamma}t}}{b + e^{\tilde{\gamma}t}}. \tag{1.9}$$

The dynamics encoded in **Eq. (1.8)** is that of a system that flows from an UV fixed point at $t = -\infty$ where $\alpha = 0$ to an IR fixed point where $\alpha = a$. The latter value encodes the cumulative number of infected cases $P = \exp(a)$ in the region under study at the end of the epidemic wave. The coefficient $\tilde{\gamma}$ is the diffusion slope, while b shifts the entire epidemic curve by a given amount of time. Further details, including what parameter influences the *flattening of the curve* and location of the inflection point and its properties can be found in [6].

1.2 Connecting Epidemic Renormalization Group With SIR While Extending It

To start connecting with compartmental models we rewrite **Eq. (1.8)** as

$$\frac{d\tilde{I}}{dt} = \tilde{\gamma}\tilde{I}\ln\tilde{I}\left(1 - \frac{\ln\tilde{I}}{\ln P}\right), \tag{1.10}$$

whose solution, with the initial condition $\ln\tilde{I}_0 = \ln\tilde{I}(0) = \frac{\ln P}{b+1}$, is a logistic function written as

$$\tilde{I}(t) = \exp\left(a\frac{1}{be^{-\tilde{\gamma}t} + 1}\right). \tag{1.11}$$

In the original eRG framework the number of recovered individuals were not explicitly taken into account. This is,

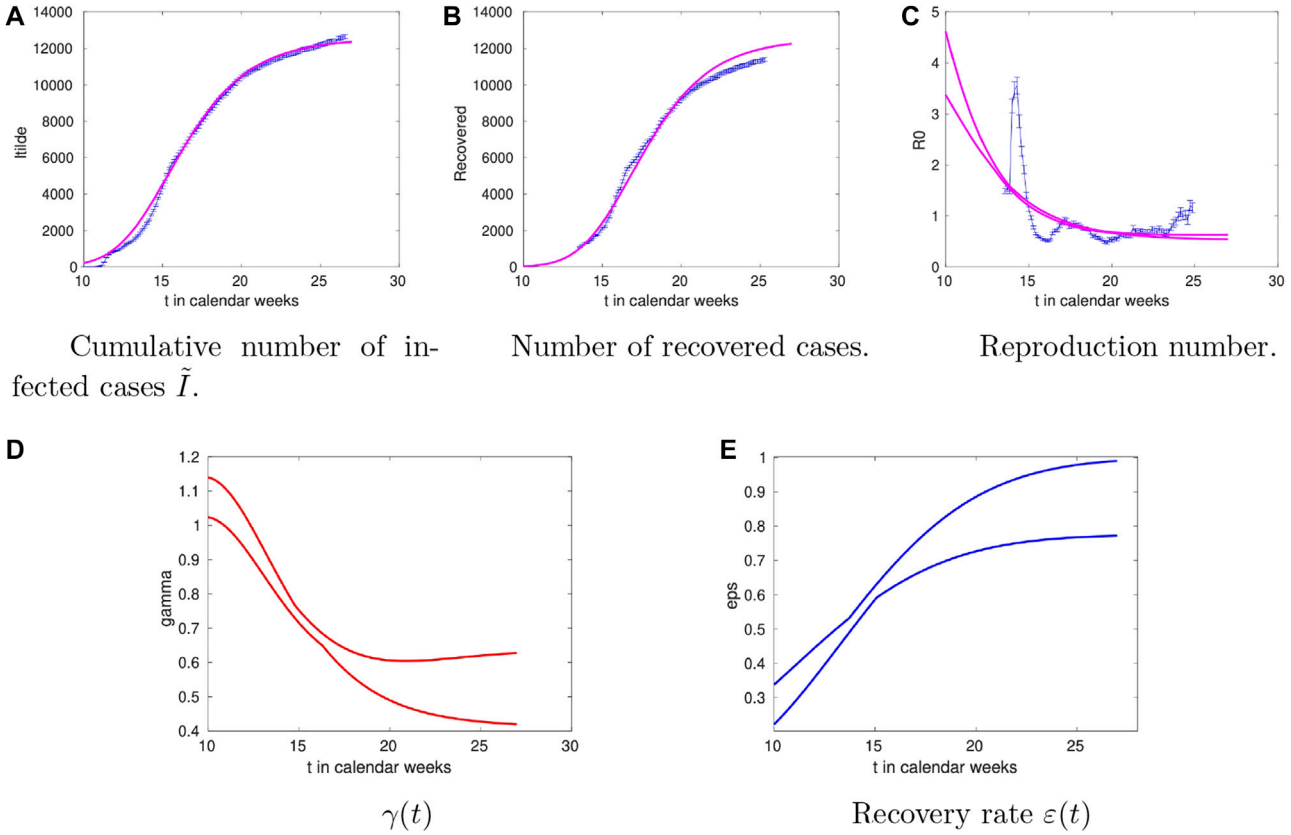


FIGURE 2 | Time dependence of $\tilde{I}(t)$, $R(t)$, $R_0(t)$, $\gamma(t)$ and $\varepsilon(t)$ within the eRG SIR model optimised to describe the Danish data for COVID-19. Solid lines are the model (shown as a 95% confidence level band for the predictions) and the dots with error-bars are the data. For the chosen confidence intervals the errors for the fit functions $\tilde{I}$ and R are not visible in this and the following figures.

however, straightforward to implement by introducing an equation for dR/dt and imposing a conservation law equivalent to the one for the SIR model. A minimal choice compatible with the conservation law is

$$\begin{aligned}\frac{dR}{dt} &= \varepsilon I,\\ \frac{dI}{dt} &= \tilde{\gamma}(I+R)\ln(I+R)\left(1-\frac{\ln(I+R)}{\ln P}\right)-\varepsilon I,\\ \frac{dS}{dt} &= -\tilde{\gamma}(I+R)\ln(I+R)\left(1-\frac{\ln(I+R)}{\ln P}\right),\end{aligned} \tag{1.12}$$

where the parameters are $\tilde{\gamma}$, P and ε. At fixed N, $\tilde{\gamma}$, P and for any value of ε, the SIR model in **(1.1)** and the eRG systems of equations match if we allow γ to be the following time-dependent function

$$\gamma(t) = \tilde{\gamma}(I(t)+R(t))\ln(I(t)+R(t))\left(1-\frac{\ln(I(t)+R(t))}{\ln P}\right)\frac{N}{I(t)S(t)}. \tag{1.13}$$

As we shall see this is a welcome feature. To better appreciate the mapping we show in **Figure 1** the time-dependent γ parameter for a hypothetical case with $N = 7$ millions, $P = 300$ thousands, $\varepsilon = 1$ and $\tilde{\gamma} = 0.7$ with initial conditions $I(0) = 3$, $R(0) = 0$ and $S(0) = N - I(0)$.

The result is a smooth function that peaks at short times and then plateaus to a fraction of $\tilde{\gamma}$. In other words the eRG naturally encodes a rapid diffusion of the disease in the initial states of the epidemic and the slow down at later times.

In terms of $\tilde{I}$ one can more compactly write the system of equations in **(1.12)** as:

$$\begin{aligned}\frac{dR}{dt} &= \varepsilon\left(\tilde{I}-R\right),\\ \frac{d\tilde{I}}{dt} &= \tilde{\gamma}\tilde{I}\ln\tilde{I}\left(1-\frac{\ln\tilde{I}}{\ln P}\right),\\ \frac{dS}{dt} &= -\frac{d\tilde{I}}{dt}.\end{aligned} \tag{1.14}$$

1.3 Reproduction Number

An important quantity for pandemics is the reproduction number related to the expected average number of infected cases due to one case. In the time-dependent generalised SIR model it is identified as:

$$R_0(t) = \frac{\gamma(t)}{\varepsilon(t)}. \tag{1.15}$$

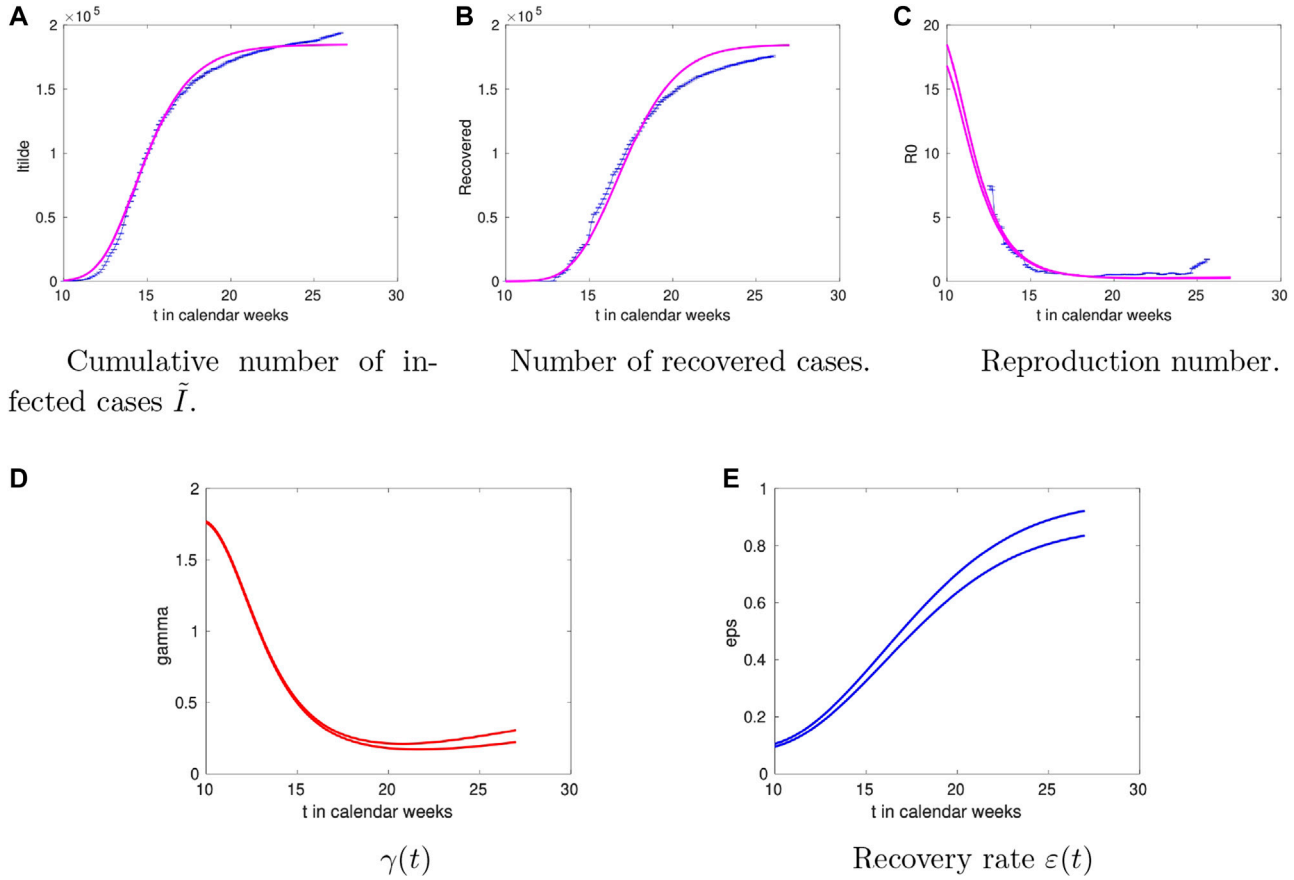


FIGURE 3 | Time dependence of $\tilde{I}(t)$, $R(t)$, $R_0(t)$, $\gamma(t)$ and $\varepsilon(t)$ within the eRG SIR model optimised to describe the German data for COVID-19. Solid lines are the model (shown as a 95% confidence level band for the predictions) and the dots with error-bars are the data.

To extract R_0 from data it is useful to recast it as:

$$R_0 = \frac{\gamma}{\varepsilon} = \frac{\frac{dI}{dt} + \frac{dR}{dt}}{\frac{dR}{dt}} \frac{N}{S} = \frac{\frac{d\tilde{I}}{dt}}{\frac{dR}{dt}} \frac{N}{S} . \tag{1.16}$$

The result holds for γ and ε generic functions of time. Here we also generalise the time-dependence of ε to be:

$$\varepsilon(t) = A\left[1 - c \cdot e^{-\frac{1}{2}\left(\frac{t-t_0}{W}\right)^2}\right] . \tag{1.17}$$

The choice above has been empirically devised to best describe the data within the current approach. As it is clear from its form this function has a dip at t_0 (possibly correlated with the peak of the newly infected cases) of width W and depth $c \cdot A$ with A the asymptotic value for $t \to \pm\infty$.

As we shall see the shape allows for a substantial increase of R_0 near the peak of the newly infected cases that could be due to a number of factors including possible health-system stress around this period.

2 TESTING THE FRAMEWORK

As a timely application we consider the COVID-19 pandemic. Here the factor N/S in **(1.16)** can be neglected as the number of total infected is at most of O(1%) of the total susceptible population and the ratio is therefore very close to unity. The reproduction number can hence be estimated as:

$$R_0(t) = \frac{\text{newly infected (t)}}{\text{newly recovered (t)}} . \tag{2.1}$$

The values for the numerator and denominator for different regions of the world can be obtained from several sources such as the World Health Organization (WHO) and Worldometers.

As testbed scenarios we consider four benchmark cases, namely Denmark, Germany, France and Italy. These countries adopted different degrees of containment measures. We find convenient to bin the data in weeks to smooth out daily fluctuations. We associate an error to both newly infected and newly recovered given by the square root of their values.

Procedurally we first fit the function $\alpha(t)$ to determine a, b and $\tilde{\gamma}$ following [6, 9]. We then solve with these, as input, the system of equations **(1.12)** for O(10^4) different choices, within reasonable ranges, of the parameters A, c, t_0 and W entering the definition of $\varepsilon(t)$, **Eq. (1.17)**. The optimal choice of such parameters is finally obtained by performing a χ^2 minimization to the data related to the recovered cases. Combining the results with $\gamma(t)$ we compare $R_0(t)$ with the actual data.

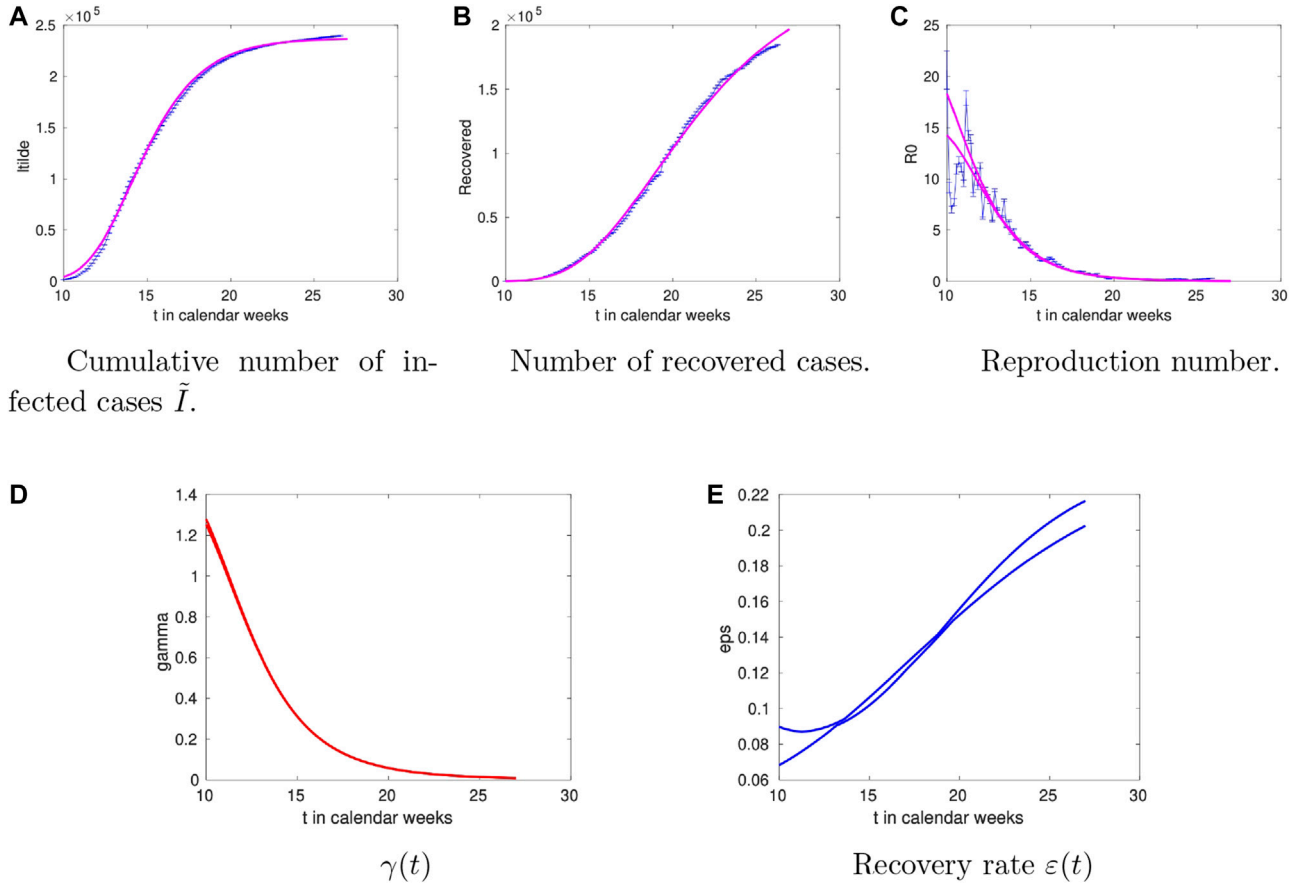


FIGURE 4 | Time dependence of $\tilde{I}(t)$, $R(t)$, $R_0(t)$, $\gamma(t)$ and $\varepsilon(t)$ within the eRG SIR model optimised to describe the Italian data for COVID-19. Solid lines are the model (shown as a 95% confidence level band for the predictions) and the dots with error-bars are the data.

For each country we show the data and the model results by grouping together five graphs in a single figure. The different panels represent $\tilde{I}(t)$, $R(t)$, $R_0(t)$, $\gamma(t)$ and $\varepsilon(t)$, all as function of the week number. Additionally the data will be reported starting some time after the outbreak. The reason being that the values of the number of recovered cases at early times is too small to be reliable and begins to be sizeable only few weeks after the outbreak. For our predictions $\varepsilon(t)$, $\gamma(t)$ and $R_0(t)$ we show bands limiting the 90% confidence level. Those are obtained shifting the data for the number of recovered cases by 1.65 standard deviations. The fitting errors for $\tilde{\gamma}$, a and b from the method in [6] can be neglected given that these parameters are highly constrained by the data.

In general we find good agreement between the data and the model with the exception of the increase in the number of infected cases occurring in the last weeks for some of the countries. Those are non-smooth events resulting from an abrupt change in the social distancing measures or from new and previously un-accounted disease hotspots. As such those events cannot be predicted by smooth models.

2.1 Denmark

The data and the model results for Denmark are shown in **Figure 2**. For the time-dependence of R_0, given in panel 2c, we observe that the model captures the variation of the data for over 10 weeks. Overall we find that the eRG model provides a reasonable description of the time dependence of the reproduction number. We observe that the recovery rate $\varepsilon(t)$ grows with time by a factor of five. There could be several factors contributing to this growth, one being a better trained health system.

2.2 Germany

For Germany the analysis is summarised in **Figure 3**. The overall trends are similar to the Danish case including the temporal trend of the recovery rate $\varepsilon(t)$.

2.3 Italy

The analysis for Italy is shown in **Figure 4**. We observe rather large values of R_0 compared to Denmark at early times and a factor of two with respect to Germany. We also observe that a good fit is obtained for $\gamma(t)$ approaching very small values at large times. This is different from Germany and Denmark, suggesting strong distancing measures being adopted by the Italian government. This seems to be further followed by a smaller value of the recovery rate, roughly about a fourth. However this last comparison is biased by the fact that the number of

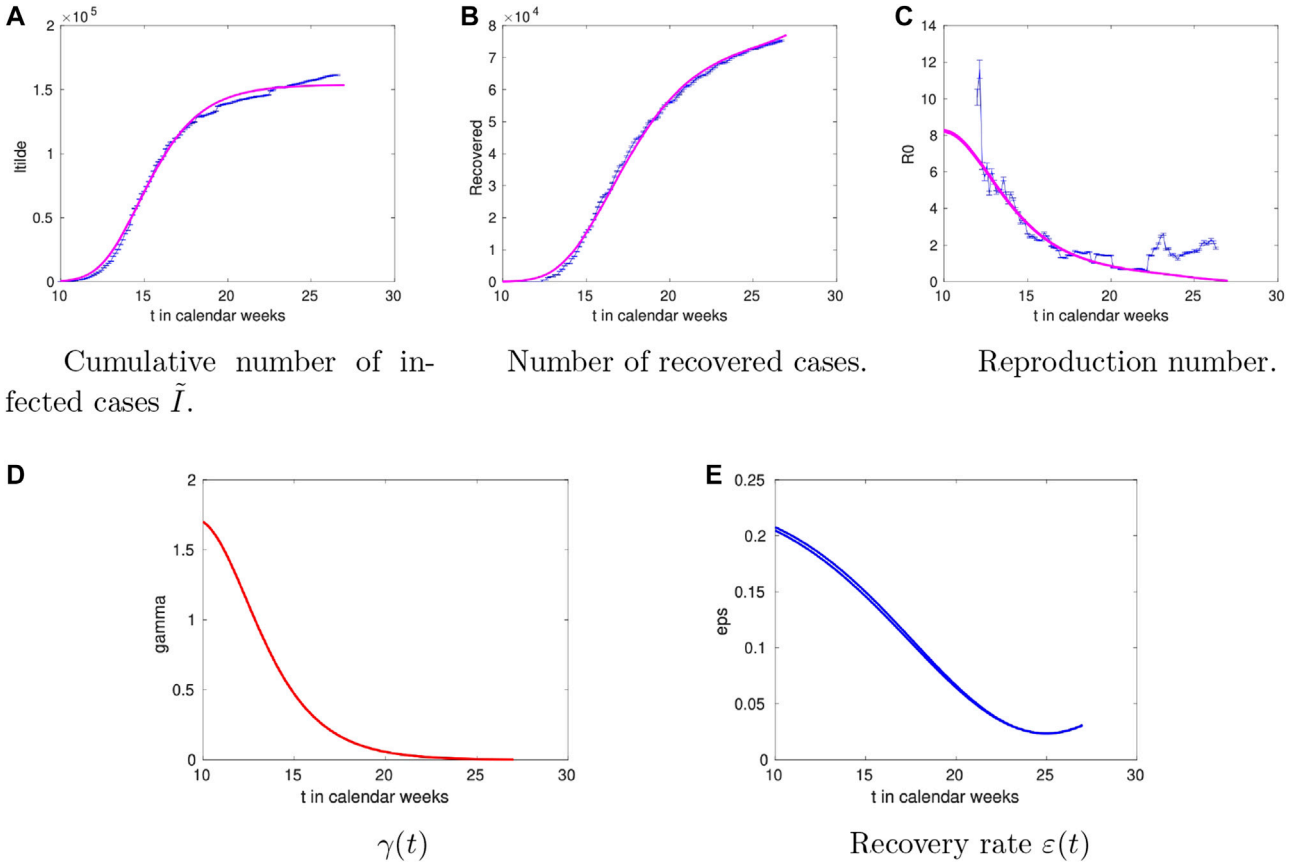


FIGURE 5 | Time dependence of $\tilde{I}(t)$, $R(t)$, $R_0(t)$, $\gamma(t)$ and $\varepsilon(t)$ within the eRG SIR model optimised to describe the French data for COVID-19. Solid lines are the model (shown as a 95% confidence level band for the predictions) and the dots with error-bars are the data.

deaths in Italy is about 15% of the number of infected cases while in Germany and Denmark it is below 5% suggesting that a more accurate description at large times would require introducing also a compartment accounting for the deaths.

2.4 France

The results for the French case can be found in **Figure 5** with the overall picture similar to the Italian one. The striking difference compared to the other countries is that the recovery rate decreases at late times. Given that the deaths in France amount to about 20% of the total infected, such difference indicates that a more complete model (including the deaths compartment) is needed.

3 CONCLUSIONS AND OUTLOOK

We generalised the epidemic Renormalization Group framework to take into account the recovered cases and to be able to determine the time dependence of the reproduction number. At the same time we show that the eRG framework can be embedded into a SIR model with time-dependent coefficients. Interestingly the resulting infection rate $\gamma(t)$ is a smooth curve with a maximum at early times while rapidly plateauing at large times. This is a welcome behaviour since it encodes the slow down in the spreading of the disease at large times coming, for example, from social distancing.

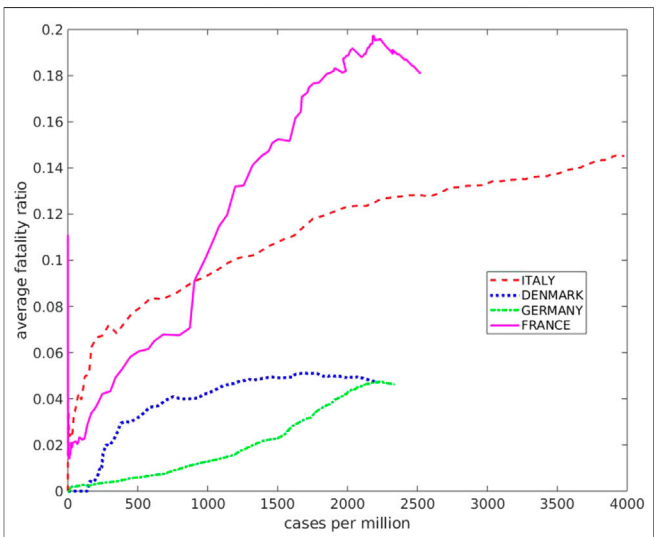


FIGURE 6 | The average fatality ratio as a function of the cumulative number of infected per million inhabitants for the countries discussed here.

We then move to confront the model to actual data by considering the spread of COVID-19 in the following countries: Denmark, Germany, Italy and France. We show that the overall approach works rather well in reproducing the data. Nevertheless the interpretation for the recovery rate is natural for Denmark and Germany while it requires to add the deaths compartment for Italy and France. The reason being that for the first two countries the number of deaths is below 5% of the number of infected cases while it is above 15% for France and Italy[1]. We therefore expect that this compartment will be relevant to include for countries with similar number of deaths. The extension of the eRG to include also the death compartment is a natural next step. Upon introducing th*D*-compartment, the system of differential equations in 1.1 (and similarly for **Eq. 1.12**) extends to

$$\begin{aligned}\frac{dS}{dt} &= -\gamma S\frac{I}{N}\,,\\ \frac{dI}{dt} &= \gamma S\frac{I}{N} - (1-\alpha)\varepsilon I - \alpha\rho I,\\ \frac{dR}{dt} &= (1-\alpha)\varepsilon I,\\ \frac{dD}{dt} &= \alpha\rho I.\end{aligned} \tag{3.1}$$

with $I(t)+R(t)+S(t)+D(t)$. becoming the new conserved quantity. Notice that two new parameters had to be introduced, i.e. the fatality ratio α and the death rate ρ. In the case of COVID-19 the data suggest that these parameters have an important time dependence, as it is manifest from **Figure 6**.

Here the running average fatality ratio (cumulative number of deaths over cumulative number of infected) is displayed, as a function of the number of infected per million inhabitants, for the countries considered here.

We are now investigating different parameterizations of these curves, however, since our main focus here is to establish an exact connection between two well known frameworks describing social dynamics (the renormalization group and compartmental models) we have restricted our attention to the simplest compartmental model, i.e., the SIR model. To our knowledge such an exact matching could not be found in previous literature. Another interesting approach could be to include the death compartment in the recovered one as put forward in [13].

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.who.int/csr/sars/country/en/ World Health Organization (WHO) and https://www.worldometers.info/coronavirus/Worldometers.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Kermack WO, McKendrick AG, A contribution to the mathematical theory of epidemics, *Proc R Soc A* (1927) 115(772):700–21. doi:10.1007/BF02464423
2. Zheng M, Zhao M, Min B, Liu Z. Synchronized and mixed outbreaks of coupled recurrent epidemics. *Sci Rep* (2017) 7:2424. doi:10.1038/s41598-017-02661-9
3. Zheng M, Wang C, Zhou J, Guan S, Zou Y, et al. Non-periodic outbreaks of recurrent epidemics and its network modelling. *Sci Rep* (2015) 5:16010. doi:10.1038/srep16010
4. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* (2020) 368(6490):489–93. doi:10.1126/science.abb3221
5. Zhang X, Ma R, Wang L, Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. *Chaos Solit Fractals* (2020) 140:110130. doi:10.1016/j.chaos.2020.109829
6. Della Morte M, Orlando D, Sannino F, Renormalization group approach to pandemics: the COVID-19 case. *Front Physiol* (2020) 8:144. doi:10.3389/fphy.2020.00144
7. Wilson K, Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. *Phys Rev B* (1971) 4:3174–83. doi:10.1103/PhysRevB.4.3174
8. Wilson K, Renormalization group and critical phenomena. II. Phase-space cell analysis of critical behavior. *Phys Rev B* (1971) 4:3184–205. doi:10.1103/PhysRevB.4.3184
9. Cacciapaglia G, Sannino F, (2005) [arXiv:2005.04956 [physics.soc-ph]]. Accepted for publication in Nature Scientific Reports.
10. Cacciapaglia G, Cot C, Sannino F, (2007) [arXiv:2007.13100] [physics.soc-ph]].
11. Cacciapaglia G, Cot C, Sannino F, (2008) [arXiv:2008.02117] [physics.soc-ph]].
12. Smit AJ, Fitchett JM, Engelbrecht FA, Scholes RJ, Dzhivhuho G, Sweijd NA, Winter is coming: a southern hemisphere perspective of the environmental drivers of SARS-CoV-2 and the potential seasonality of COVID-19, *Int J Environ Res Publ Health* (2020) 17:5634. doi:10.3390/ijerph17165634
13. Scala A, Flori A, Spelta A, Brugnoli E, Cinelli M, Quattrociocchi W, et al. Time, space and social interactions: exit mechanisms for the Covid-19 epidemics *Sci Rep* (2020) 10:13764. doi:10.1038/s41598-020-70631-9

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

The reviewer KT declared a past co-authorship with one of the authors FS to the handling editor.

frontiers in Physics

ORIGINAL RESEARCH
published: 18 January 2021
doi: 10.3389/fphy.2020.601459

Dynamical Analysis of a Mathematical Model of COVID-19 Spreading on Networks

Wang Li[1], *Xinjie Fu*[2], *Yongzheng Sun*[1] *and Maoxing Liu*[2]*

[1]School of Mathematics, China University of Mining and Technology, Xuzhou, China, [2]School of Science, North University of China, Taiyuan, China

In this article, an *SEAIRS* model of COVID-19 epidemic on networks is established and analyzed. Following the method of the next-generation matrix, we derive the basic reproduction number R_0, and it shows that the asymptomatic infector plays an important role in disease spreading. We analytically show that the disease-free equilibrium E_0 is asymptotically stable if $R_0 \leq 1$; moreover, the effects of various quarantine strategies are investigated and compared by numerical simulations. The results obtained are informative for us to further understand the asymptomatic infector in COVID-19 propagation and get some effective strategies to control the disease.

Keywords: COVID-19 epidemic, networks, asymptomatic, the basic reproduction number, quarantine

OPEN ACCESS

Edited by:
Chengyi Xia,
Tianjin University of Technology, China

Reviewed by:
Chuansheng Shen,
Anqing Normal University, China
Weigang Sun,
Hangzhou Dianzi University, China

***Correspondence:**
Maoxing Liu
liumaoxing@126.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *01 September 2020*
Accepted: *25 November 2020*
Published: *18 January 2021*

Citation:
Li W, Fu X, Sun Y and Liu M (2021) Dynamical Analysis of a Mathematical Model of COVID-19 Spreading on Networks. Front. Phys. 8:601459. doi: 10.3389/fphy.2020.601459

1 INTRODUCTION

Coronaviruses are a group of enveloped viruses with a positive-sense, single-stranded RNA and viral particles resembling a crown from which the name derives. They belong to the order of Nidovirales, family of Coronaviridae, and subfamily of Orthocoronavirinae. A recent coronavirus outbreak has started since December 2019. So far, the new virus has infected more than 25 million people and killed at least 842,000 of them [1–3].

In order to better study the spread and control of infectious diseases, several classical mathematical models have been discussed to understand the transmission mechanism and dynamics of the disease, such as the SI model, SIR model, SEIR model, and SEIRS model [4, 5]. Different factors are considered in different models, and the purpose of these models is finding the transmission mechanism and dynamics of the diseases for controlling the diseases. This process needs a contact between them. It is found that the best pattern of contact is contact in the networks [6–11], and study epidemic model on networks has been a hot field in mathematical biology [6–20,35].

To find the transmission mechanism, we have to know who is the infector [27]. An infector is an individual who carries the virus and can export the virus. Sometimes, an individual with the infected state may not show any symptoms of disease, and the state can be defined as asymptomatically infected. An individual with the infected state who shows symptoms of disease can be defined as symptomatically infected. Exposed state and asymptomatically infected state are more dangerous than symptomatically infected state. When a susceptible individual contacts a symptomatic one, he(she) will usually do something possible to prevent the disease spreading.

In the disease spreading process, a symptomatic infector will be restricted by individual behavior or public behavior, but a symptomatic one has a higher efficiency than an asymptomatic one. A class of works studied the role of symptomatic infectors and

asymptomatic infectors in diseases infections, such as dengue virus [21, 22], malaria [23], and *Norovirus* [24–27]. The results of these works all show that the asymptomatic cases cannot be neglected. Recently, several mathematic models within asymptomatic infection have been studied, such as *SAIR* model [21], CI model [28], *SEIADR* model [29], and $SEII_eQR$ model [30]. But mathematical models within asymptomatic infection on networks are still relatively rare. When the COVID-19 pandemic begin to outbreak, quarantining certain individuals may be the most efficient way to stop the outbreak of disease, but the choice of which is an important step in the control process, and thus we also discuss the effect of different quarantine strategies.

In this article, we consider an *SEAIRS* epidemic model on scale-free networks to study the spread of COVID-19. Further, we consider the effect of the asymptomatic state changing into the symptomatic state. By studying the dynamics of the model, we try to find what role the three infectors, E, A, and I, play in disease spreading, especially for the asymptomatic infectors, E and A. Next, in **Section 2**, we establish an *SEAIRS* COVID-19 model on network. In **Section 3**, we study the dynamics of the model. In **Section 4**, we study and compare the effects of various control strategies, and carry out numerical simulations to illustrate the theoretical results in **Section 5**. Finally, conclusions and some discussions are given in **Section 6**.

2 THE MODEL

In the spread process, each node has five states: susceptible (S); exposed (E); asymptomatically infected (A), those who are infectious but have no symptoms; symptomatically infected (I), those who are infectious but have symptoms; and removed (R). The disease transmission flow is depicted in **Figure 1**. A susceptible individual can be infected by contact at rate β_1 if there are infected individuals with A in its neighbors, or be infected by contact at rate β_2 if there are symptomatically infected individuals I. The exposed individual will become A and I with rate $p\alpha$ and $(1-p)\alpha$, and A will become I individual at rate q. The infected A and I will become a recovered individual at rate μ_1 and μ_2 due to treatment, respectively. A recovered individual will become S at rate δ. All parameters are nonnegative.

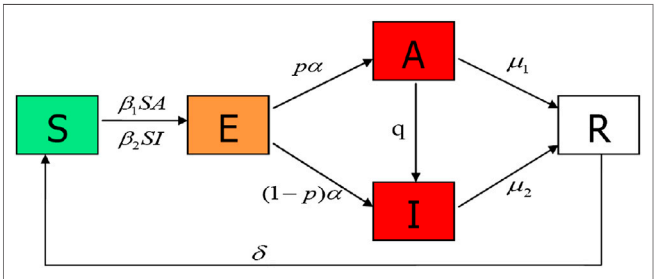


FIGURE 1 | State-transition rules of the SEAIRS model of COVID-19.

In this article, we consider an *SEAIRS* epidemic model on scale-free networks as follows:

$$\begin{cases} \dfrac{dS_k(t)}{dt} = -\beta_1 k S_k(t)\theta_1(t) - \beta_2 k S_k(t)\theta_2(t) + \delta R_k(t), \\ \dfrac{dE_k(t)}{dt} = \beta_1 k S_k(t)\theta_1(t) + \beta_2 k S_k(t)\theta_2(t) - \alpha E_k(t), \\ \dfrac{dA_k(t)}{dt} = p\alpha E_k(t) - qA_k(t) - \mu_1 A_k(t), \\ \dfrac{dI_k(t)}{dt} = (1-p)\alpha E_k(t) + qA_k(t) - \mu_2 I_k(t), \\ \dfrac{dR_k(t)}{dt} = \mu_1 A_k(t) + \mu_2 I_k(t) - \delta R_k(t). \end{cases} \tag{2.1}$$

Following [6], $\langle k\rangle = \sum_k kP_k$ describes the average degree and $P_k(k = 1, 2, \ldots, n)$ describes the degree distribution; the probability $\theta_i(i = 1, 2, 3)$ is a link pointing to an exposed individual, asymptomatic infected individual, and symptomatic infected individual. Here, N_k is a constant that stands for the number of nodes with degree k. Then, $N_k = S_k(t) + E_k(t) + A_k(t) + I_k(t) + R_k(t)$, $(k = 1, 2, \ldots, n)$ and $\sum_k N_k = N$. We have

$$\theta_1(t) = \frac{1}{\langle k\rangle}\sum_k kP_k\frac{A_k(t)}{N_k}, \tag{2.2}$$

$$\theta_2(t) = \frac{1}{\langle k\rangle}\sum_k kP_k\frac{I_k(t)}{N_k}. \tag{2.3}$$

For the practice, the initial condition for model **(2.1)** satisfies the following:

$$\begin{cases} 0 \le S_k(0), E_k(0), A_k(0), I_k(0), R_k(0) \le N_k, \\ S_k(0) + E_k(0) + A_k(0) + I_k(0) + R_k(0) = N_k, \end{cases} \quad (k = 1, 2, \ldots, n). \tag{2.4}$$

3 DYNAMICAL BEHAVIORS OF THE MODEL

In this section, we study the stability of disease-free equilibrium of model **(2.1)**. First, we derive the basic reproduction number of the model.

3.1 Equilibria and Basic Reproduction Number

Lemma 3.1. Suppose that $S_k(t), E_k(t), A_k(t), I_k(t), R_k(t)$ is a solution of model **(2.1)** satisfying initial conditions of **Eq. (2.4)**, then

$$\Omega = \{(S_k, E_k, A_k, I_k, R_k) | 0 \le S_k, E_k, A_k, I_k, R_k \le N_k, k = 1, 2, \ldots, n\}$$

is a positive invariant for model **(2.1)**.

The proof of this lemma is similar to that in [31]. We ignore it here. The stability analysis and numerical simulations are investigated in the positive invariant Ω.

It is easy to find that model **(2.1)** has a disease-free equilibrium E_0, it is the only susceptible state. $E_0 = (S_1^0, 0, 0, 0, 0, S_2^0, 0, 0, 0, 0, \cdots, S_n^0, 0, 0, 0, 0)$, where $S_k^0 = N_k$, $k = 1, 2, \ldots, n$. Following the method of the next-generation matrix [32], the basic reproduction number can be calculated

by $R_0 = \rho(FV^{-1})$. Here, F is the matrix of the rate of generation of new infections, and V is the matrix of the rate of transfer of individuals out of the two compartments. F is given by

$$F = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}_{3n\times 3n}$$

and

$$A_{ij} = \begin{pmatrix} 0 & \frac{\beta_1 S_i^0 P_j}{\langle k\rangle N_j} ij & \frac{\beta_2 S_i^0 P_j}{\langle k\rangle N_j} ij \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In addition, $V = \mathrm{diag}(B_1, B_2, \ldots, B_n)$, where

$$B_k = \begin{pmatrix} \alpha & 0 & 0 \\ -p\alpha & q+\mu_1 & 0 \\ -(1-p)\alpha & -q & \mu_2 \end{pmatrix},$$

and

$$B_k^{-1} = \begin{pmatrix} \frac{1}{\alpha} & 0 & 0 \\ \frac{p}{q+\mu_1} & \frac{1}{q+\mu_1} & 0 \\ \frac{q+(1-p)\mu_1}{\mu_2(q+\mu_1)} & \frac{q}{\mu_2(q+\mu_1)} & \frac{1}{\mu_2} \end{pmatrix},$$

so, $V^{-1} = \mathrm{diag}(B_1^{-1}, B_2^{-1}, \ldots, B_n^{-1})$, and the basic reproduction number R_0 can be obtained:

$$R_0 = \rho(FV^{-1}) = \frac{\mu_2 p\beta_1 + [q+(1-p)\mu_1]\beta_2}{\mu_2(q+\mu_1)} \frac{\langle k^2\rangle}{\langle k\rangle}. \tag{3.1}$$

Let J_0 be the Jacobian matrix of the middle three equations of model **(2.1)**, then $J_0 = F - V$ and $s(J_0) = \max\{Re\lambda$: λ is an eigenvalue of $J_0\}$, $R_0 < 1 \Leftrightarrow s(J_0) < 0$, $R_0 > 1 \Leftrightarrow s(J_0) > 0$. [34]

Remark 3.1. Let $R_{0,a} = \frac{p\beta_1}{q+\mu_1}\frac{\langle k^2\rangle}{\langle k\rangle}$, $R_{0,d} = \frac{[q+(1-p)\mu_1]\beta_2}{\mu_2(q+\mu_1)}\frac{\langle k^2\rangle}{\langle k\rangle}$, then $R_0 = R_{0,a} + R_{0,d}$, if $q = 0$, this result is corresponding to the result in [27].

Remark 3.2. If we do not consider the asymptomatic infected state, that is $\beta_1 = 0$, $\mu_1 = \mu_2 = \mu$, $p = q = 0$, then $R_0 = \frac{\beta_2}{\mu}\frac{\langle k^2\rangle}{\langle k\rangle}$, which is corresponding to the result in [6, 33].

Lemma 3.2. If and only if $R_0 > 1$, there is a unique endemic equilibrium E_+.

Proof. We consider the right side of the model **(2.1)** to be equal to zero. Then, we obtain

$$\begin{cases} -\beta_1 kS_k\theta_1 - \beta_2 kS_k\theta_2 + \delta R_k = 0, \\ \beta_1 kS_k\theta_1 + \beta_2 kS_k\theta_2 - \alpha E_k = 0, \\ p\alpha E_k - qA_k - \mu_1 A_k = 0, \\ (1-p)\alpha E_k + qA_k - \mu_2 I_k = 0, \\ \mu_1 A_k + \mu_2 I_k - \delta R_k = 0. \end{cases}$$

Following the above equation, using $\theta = \beta_1\theta_1^* + \beta_2\theta_2^*$, we can find

$$E_k^* = \frac{k\theta}{\alpha}S_k^*, A_k^* = \frac{pk\theta}{q+\mu_1}S_k^*, I_k^* = \frac{[q+(1-p)\mu_1]k\theta}{\mu_2(q+\mu_1)}S_k^*, R_k^* = \frac{k\theta}{\delta}S_k^*. \tag{3.2}$$

According to the condition $N_k^* = S_k^* + E_k^* + A_k^* + I_k^* + R_k^*$, we can get

$$S_k^* = \frac{\mu_2\alpha(q+\mu_1)\delta N_k^*}{\delta\mu_2\alpha(q+\mu_1) + \{\mu_2\alpha(q+\mu_1) + \delta\mu_2(q+\mu_1) + \delta\mu_2 p\alpha + \delta\alpha[q+(1-p)\mu_1]\}k\theta}. \tag{3.3}$$

Substituting E_k^*, A_k^*, and I_k^* of **Eq.(3.2)** and **Eq.(3.3)** into θ, we can obtain $\theta f(\theta) = 0$, where

$$f(\theta) = 1 - \frac{A_1}{\langle k\rangle}\sum_k \frac{k^2 P_k}{A_2 + A_3 k\theta},$$

and $A_1 = \delta\alpha\{\mu_2 p\beta_1 + [q+(1-p)\mu_1]\beta_2\}$, $A_2 = \delta\mu_2(q+\mu_1)\alpha$, $A_3 = \mu_2(q+\mu_1)\alpha + \delta\{\mu_2(q+\mu_1) + \mu_2 p\alpha + \alpha[q+(1-p)\mu_1]\}$. It is easy to find that $\theta f(\theta) = 0$ has a trivial solution with no condition, and model **(2.1)** has a disease-free equilibrium. In addition,

$$\frac{df(\theta)}{d\theta} = \frac{A_1 A_3}{\langle k\rangle}\sum_k \frac{k^3 P_k}{(A_2 + A_3 k\theta)^2} > 0,$$

$$f(\beta_1+\beta_2) = 1 - \frac{A_1}{\langle k\rangle}\sum_k \frac{k^2 P_k}{A_2 + A_3 k(\beta_1+\beta_2)}$$

$$> 1 - \frac{A_1}{\langle k\rangle}\sum_k \frac{k^2 P(k)}{A_3 k(\beta_1+\beta_2)} = 1 - \frac{A_1}{A_3(\beta_1+\beta_2)}$$

$$= 1 - \frac{\delta\alpha\{\mu_2 p\beta_1 + [q+(1-p)\mu_1]\beta_2\}}{\mu_2(q+\mu_1)\alpha + \delta\{\mu_2(q+\mu_1) + \mu_2 p\alpha + \alpha[q+(1-p)\mu_1]\}}\frac{1}{\beta_1+\beta_2}$$

$$> 1 - \frac{\delta\alpha\{\mu_2 p\beta_1 + [q+(1-p)\mu_1]\beta_2\}}{\mu_2(q+\mu_1)\alpha + \delta\{\mu_2(q+\mu_1) + \mu_2 p\alpha + \alpha[q+(1-p)\mu_1]\}} > 0.$$

Thus, $\theta f(\theta) = 0$ has a nontrivial solution if and only if $f(0) < 0$, that is if and only if $R_0 > 1$, there is a unique endemic equilibrium.

3.2 Stability of Disease-free and Endemic Equilibrium

Theorem 3.1. If $R_0 \le 1$, E_0 of model **(2.1)** is locally asymptotically stable; if $R_0 > 1$, E_0 is unstable, and there is a unique endemic equilibrium E_+, and E_+ is locally asymptotically stable.

Proof. From the results of Lemma 3.1, Lemma 3.2, and the basic reproduction number, we only need to prove the case of $R_0 = 1$. We consider the following model:

$$\begin{cases} \frac{dS_k}{dt} = -\beta_1 kS_k\theta_1 - \beta_2 kS_k\theta_2 + \delta R_k, \\ \frac{dE_k}{dt} = \beta_1 kS_k\theta_1 + \beta_2 kS_k\theta_2 - \alpha E_k, \\ \frac{dA_k}{dt} = p\alpha E_k - qA_k - \mu_1 A_k, \\ \frac{dI_k}{dt} = (1-p)\alpha E_k + qA_k - \mu_2 I_k. \end{cases} \tag{3.4}$$

The Jacobian matrix of model **(3.4)** at disease-free equilibrium is given by

$$J = \begin{pmatrix} J_{11} & J_{12} & \cdots & J_{1n} \\ J_{21} & J_{22} & \cdots & J_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ J_{n1} & J_{n2} & \cdots & J_{nn} \end{pmatrix}_{4n\times 4n}$$

with

$$J_{ij} = \begin{pmatrix} -\delta & -\delta & -\frac{\beta_1 P_j N_i}{\langle k\rangle N_j} ij - \delta & -\frac{\beta_2 P_j N_i}{\langle k\rangle N_j} ij - \delta \\ 0 & -\alpha & \frac{\beta_1 P_j N_i}{\langle k\rangle N_j} ij & \frac{\beta_2 P_j N_i}{\langle k\rangle N_j} ij \\ 0 & p\alpha & -q-\mu_1 & 0 \\ 0 & (1-p)\alpha & q & -\mu_2 \end{pmatrix}.$$

The characteristic equation of the disease-free equilibrium is

$$(\lambda+\mu_2)^{n-1}(\lambda+q+\mu_1)^{n-1}(\lambda+\delta)^n(\lambda+\alpha)^{n-1}(\lambda^3+H_2\lambda^2 +H_1\lambda+H_0)=0,$$

where

$$H_2 = \alpha+q+\mu_1+\mu_2>0,$$

$$H_1 = \alpha(q+\mu_1)+\alpha\mu_2+\mu_2(q+\mu_1)-[p\alpha\beta_1+(1-p)\alpha\beta_2]\frac{\langle k^2\rangle}{\langle k\rangle},$$

$$H_0 = \mu_2\alpha(q+\mu_1)-\alpha\{\mu_2 p\beta_1+[q+(1-p)\mu_1]\beta_2\}\frac{\langle k^2\rangle}{\langle k\rangle},$$

when $R_0 = 1$, $H_0 = 0$, and

$$\alpha\mu_2 p\beta_1\frac{\langle k^2\rangle}{\langle k\rangle}<\mu_2\alpha(q+\mu_1)\Rightarrow\alpha p\beta_1\frac{\langle k^2\rangle}{\langle k\rangle}<\alpha(q+\mu_1),$$

$$(1-p)\alpha\mu_1\beta_2\frac{\langle k^2\rangle}{\langle k\rangle}<\mu_2\alpha(q+\mu_1)\Rightarrow(1-p)\alpha\beta_2\frac{\langle k^2\rangle}{\langle k\rangle}<\mu_2(\alpha+q+\mu_1).$$

Then, we obtain $H_1>0$, and the characteristic equation can be written as

$$\lambda(\lambda+\mu_2)^{n-1}(\lambda+q+\mu_1)^{n-1}(\lambda+\delta)^n(\lambda+\alpha)^{n-1}(\lambda^2+H_2\lambda+H_1)=0,$$

where the eigenvalues of J are all negative except zero eigenvalue when $R_0 = 1$.

4 QUARANTINE STRATEGIES

Quarantine is helpful in controlling diseases. In this section, we will discuss three different quarantine strategies for COVID-19 disease model on networks.

4.1 Proportional Quarantine

In this case, for fixed spreading rates β_1 and β_2, let $\omega(0<\omega<1)$ is the density of quarantine nodes in the network. At the mean-field level, the presence of proportional quarantine will effectively reduce the spreading rate. Thus, we can approximately use $\beta_1(1-\omega)$ to substitute β_1 and use $\beta_2(1-\omega)$ to substitute β_2; thus, the system **(2.1)** becomes

$$\begin{cases} \frac{dS_k(t)}{dt} = -\beta_1(1-\omega)kS_k(t)\theta_1(t)-\beta_2(1-\omega)kS_k(t)\theta_2(t)+\delta R_k(t) \\ \frac{dE_k(t)}{dt} = \beta_1(1-\omega)kS_k(t)\theta_1(t)+\beta_2(1-\omega)kS_k(t)\theta_2(t)-\alpha E_k(t) \\ \frac{dA_k(t)}{dt} = p\alpha E_k(t)-qA_k(t)-\mu_1 A_k(t) \\ \frac{dI_k(t)}{dt} = (1-p)\alpha E_k(t)+qI_k(t)-\mu_2 I_k(t) \\ \frac{dR_k(t)}{dt} = \mu_1 A_k(t)+\mu_2 I_k(t)-\delta R_k(t) \end{cases} \tag{4.1}$$

Using the same method in **Section 3**, we obtain a self-consistency equation as follows:

$$\theta = \frac{\delta\alpha(1-\omega)A_1}{\langle k\rangle}\sum_k\frac{\langle k^2\rangle P_k}{A_2+A_3k\theta}\theta\equiv\tilde{f}(\theta).$$

By arguments similar to those in **Section 3**, the epidemic threshold $\widetilde{R_w}$ is determined by the following inequality:

$$\frac{d\tilde{f}(\theta)}{d\theta}\Big|_{\theta=0}>1;$$

therefore, it can be shown that

$$\widetilde{R_w} = \frac{\mu_2 p(1-\omega)\beta_1+[q+(1-p)\mu_1](1-\omega)\beta_2}{\mu_2(q+\mu_1)}\frac{\langle k^2\rangle}{\langle k\rangle}, \tag{4.2}$$

that is,

$$\widetilde{R_w} = (1-\omega)R_0. \tag{4.3}$$

Note that in **(4.3)**, when $\omega = 0$, that is, if no quarantine were done, then $\widetilde{R_w} = R_0$; when $0<\omega<1$, $\widetilde{R_w}<R_0$, that is, the quarantine scheme is effective; while as $\omega\to 1$, $\widetilde{R_w}\to 0$, that is, in the case of a full quarantine, it would be impossible for the epidemic to spread in the network.

4.2 Targeted Quarantine

While proportional quarantine schemes are effective, there may be more efficient schemes due to the heterogeneous nature of scale-free networks. We introduce an upper threshold κ such that all nodes with connectivity $k>\kappa$ are immunized, that is, we define the quarantine rate ω_k by

$$\omega_k = \begin{cases} 1, & k>\kappa, \\ c, & k=\kappa, \\ 0, & k<\kappa. \end{cases}$$

where $0<c<1$ and $\sum_k\omega_k P(k)=\overline{\omega}$, where $\overline{\omega}$ is the average quarantine rate. The epidemic dynamics model is

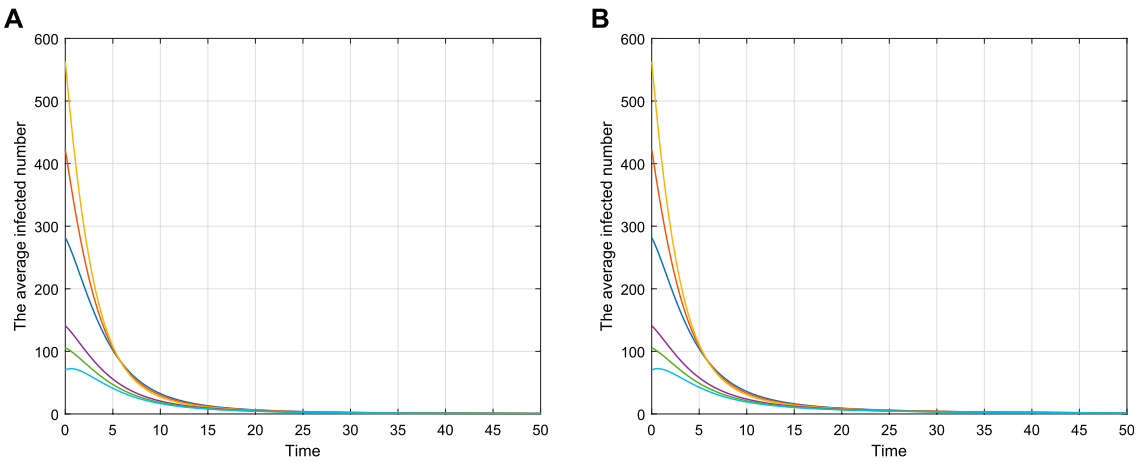


FIGURE 2 | Asymptotical stability of the model.

TABLE 1 | Parameters of **Figure 2**

Figure 2	β_1	β_2	δ	α	p	q	μ_1	μ_2	R_0
A	0.08	0.08	0.4	0.4	0.5	0.5	0.2	0.4	0.8193 < 1
B	0.0905	0.10003	0.4	0.4	0.5	0.5	0.2	0.4	1

TABLE 2 | Parameters of **Figure 4**

	β_1	β_2	Δ	α
Figure 4	0.0020178	0.033949	0.12178	0.1055744
	p	q	μ_1	μ_2
Figure 4	0.40043	0.040403	0.0119,536	0.20955

$$\begin{cases} \dfrac{dS_k(t)}{dt} = -\beta_1(1-\omega_k)kS_k(t)\theta_1(t) - \beta_2(1-\omega_k)kS_k(t)\theta_2(t) + \delta R_k(t) \\ \dfrac{dE_k(t)}{dt} = \beta_1(1-\omega_k)kS_k(t)\theta_1(t) + \beta_2(1-\omega_k)kS_k(t)\theta_2(t) - \alpha E_k(t) \\ \dfrac{dA_k(t)}{dt} = p\alpha E_k(t) - qA_k(t) - \mu_1 A_k(t) \\ \dfrac{dI_k(t)}{dt} = (1-p)\alpha E_k(t) + qI_k(t) - \mu_2 I_k(t) \\ \dfrac{dR_k(t)}{dt} = \mu_1 A_k(t) + \mu_2 I_k(t) - \delta R_k(t) \end{cases} \tag{4.4}$$

This leads to

$$\theta = \frac{\delta\alpha A_1}{\langle k\rangle}\sum_k \frac{\langle k^2\rangle P_k(1-\omega_k)}{A_2 + A_3 k\theta}\theta \equiv \tilde{f}(\theta).$$

Therefore, the epidemic threshold

$$\widehat{R_w} = \frac{\mu_2 p\beta_1 + [q + (1-p)\mu_1]\beta_2}{\mu_2(q+\mu_1)}\frac{\langle k^2\rangle - \langle k^2\omega_k\rangle}{\langle k\rangle}. \tag{4.5}$$

Note that $\langle k^2\omega_k\rangle = \overline{\omega}\langle k^2\rangle + \sigma'$, where $\sigma' = \langle(\omega_k - \overline{\omega}) \times [k^2 - \langle k^2\rangle]\rangle$ is the covariance of ω_k and k^2. There may be κ (usually big enough) where $\sigma' < 0$, but for appropriately small κ, $\omega_k - \overline{\omega}$ and $k^2 - \langle k^2\rangle$ have the same signs except for some k where $\omega_k - \overline{\omega}$ and/or $k^2 - \langle k^2\rangle$ is zero; Therefore $\sigma' > 0$ for appropriate κ. Then,

$$\widehat{R_w} < \frac{1-\overline{\omega}}{1-\omega}\widetilde{R_w}.$$

If we set $\overline{\omega} = \omega$, then,

$$\widehat{R_w} < \widetilde{R_w}\,(0 < \omega < 1),$$

which means the targeted quarantine scheme is more efficient than the uniform quarantine scheme for the same average quarantine rate.

4.3 High-Risk Quarantine

If a neighbor of a susceptible individual is infected, vaccinating this susceptible individual is

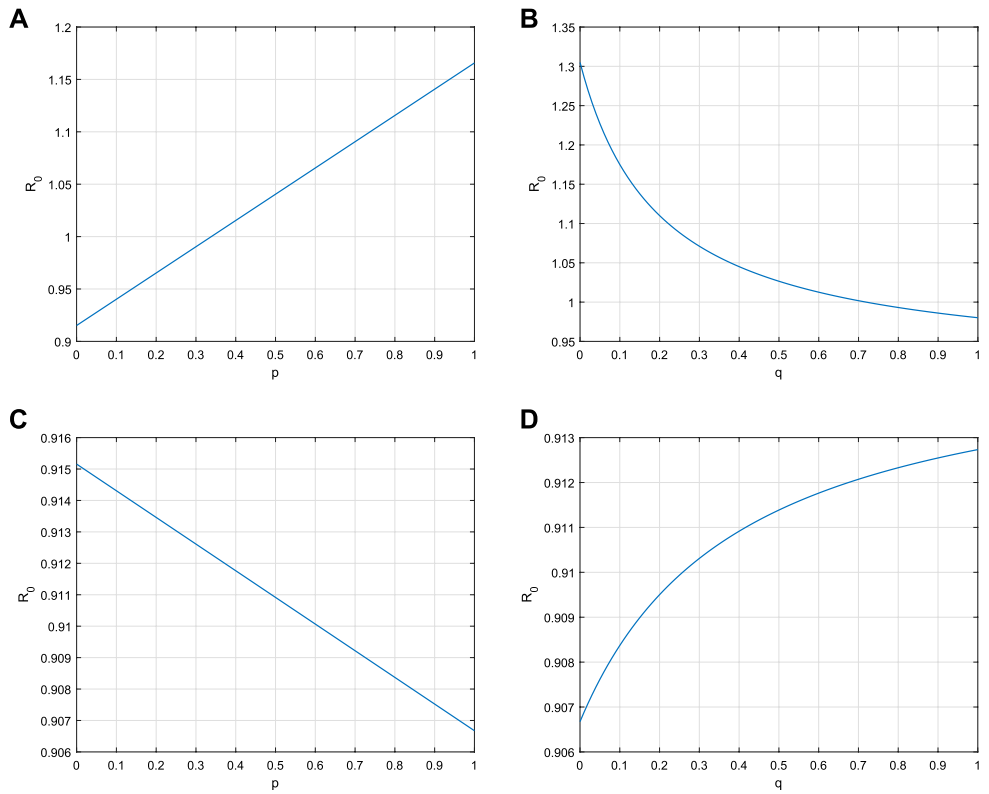


FIGURE 3 | Sensitivity of model parameters.

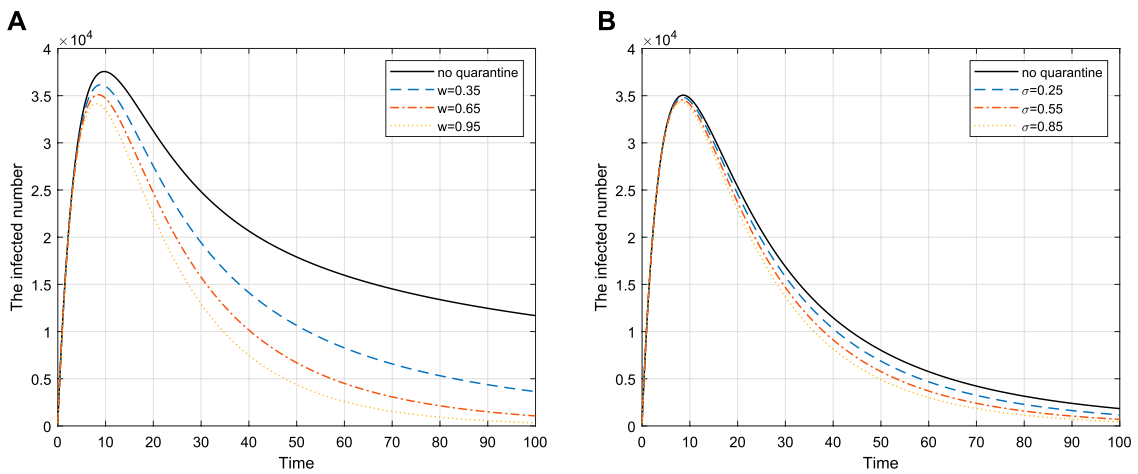


FIGURE 4 | Effectiveness of quarantine for different ω and σ respectively.

called high-risk immunity [37]. We introduce the parameter σ, that is, the probability that a node is quarantined per unit time, and $\Gamma = \frac{kP_k}{\langle k \rangle N}$, which denotes the probability that any given node is the neighbor of some specific nodes. The epidemic dynamics model is

$$\begin{cases} \frac{dS_k(t)}{dt} = -\beta_1(1-\sigma\Gamma)kS_k(t)\theta_1(t) - \beta_2(1-\sigma\Gamma)kS_k(t)\theta_2(t) + \delta R_k(t) \\ \frac{dE_k(t)}{dt} = \beta_1(1-\sigma\Gamma)kS_k(t)\theta_1(t) + \beta_2(1-\sigma\Gamma)kS_k(t)\theta_2(t) - \alpha E_k(t) \\ \frac{dA_k(t)}{dt} = p\alpha E_k(t) - qA_k(t) - \mu_1 A_k(t) \\ \frac{dI_k(t)}{dt} = (1-p)\alpha E_k(t) + qI_k(t) - \mu_2 I_k(t) \\ \frac{dR_k(t)}{dt} = \mu_1 A_k(t) + \mu_2 I_k(t) - \delta R_k(t) \end{cases} \tag{4.6}$$

This leads to

$$\theta = \frac{\delta\alpha A_1}{\langle k\rangle}\sum_k \frac{\langle k^2\rangle P_k(1-\sigma\Gamma)}{A_2 + A_3 k\theta}\theta \equiv \overline{f}(\theta).$$

And the epidemic threshold

$$\overline{R}_0 = \frac{\mu_2 p\beta_1 + [q + (1-p)\mu_1]\beta_2}{\mu_2(q+\mu_1)}\left(\frac{\langle k^2\rangle}{\langle k\rangle} - \frac{\sigma\langle k^3 P_k\rangle}{\langle k\rangle^2 N}\right) < R_0.$$

This means that the implementation of high-risk quarantine is effective, and the longer the period of $\sigma\langle k^3 P_k\rangle$, the more effective it is.

5 NUMERICAL SIMULATIONS

In this section, we present numerical simulations of model **(2.1)**, **(4.1)** and **(4.6)** to illustrate the results in **Section 3 and 4**. Without loss of generality, we consider the dynamical process on scale-free networks with $P_k = ck^{-\gamma}$ and $\sum_{k=1}^{n} P_k = 1$, $\gamma = 3$, $n = 200$. The average infected number is $A(t) + I(t) = \sum_{k=1}^{n} A_k(t)P_k + \sum_{k=1}^{n} I_k(t)P_k$. The parameters of the model are estimated by MCMC using the number of cases in Wuhan (http://wjw.wuhan.gov.cn/ztzl_28/fk/yqtb/).

Figure 2 shows the asymptotical stability of model **(2.1)** with different values, and the parameter values are shown in **Table 1**. **Figures 2A,B** show the time series of the average infected number $I(t)$ with different initial values, which is in accord with the asymptotical stability of the disease-free equilibrium.

The sensitivity analysis of the basic reproduction number R_0 can be performed by model parameters. It is easy to obtain that

$$\frac{\partial R_0}{\partial \beta_1} = \frac{p}{q+\mu_1}\frac{\langle k^2\rangle}{\langle k\rangle} > 0, \quad \frac{\partial R_0}{\partial \beta_2} = \frac{q + (1-p)\mu_1}{(q+\mu_1)\mu_2}\frac{\langle k^2\rangle}{\langle k\rangle} > 0,$$
$$\frac{\partial R_0}{\partial \mu_1} = \frac{-\beta_2 pq - \beta_1\mu_2 p}{\mu_2(q+\mu_1)^2}\frac{\langle k^2\rangle}{\langle k\rangle} < 0, \quad \frac{\partial R_0}{\partial \mu_2} = -\frac{\beta_2[q + (1-p)\mu_1]}{\mu_2^2(q+\mu_1)}\frac{\langle k^2\rangle}{\langle k\rangle} < 0,$$
$$\frac{\partial R_0}{\partial p} = \frac{\beta_1\mu_2 - \beta_2\mu_1}{\mu_2(q+\mu_1)}\frac{\langle k^2\rangle}{\langle k\rangle}, \quad \frac{\partial R_0}{\partial q} = \frac{p(\beta_2\mu_1 - \beta_1\mu_2)}{\mu_2(q+\mu_1)^2}\frac{\langle k^2\rangle}{\langle k\rangle}. \tag{5.1}$$

Equation (5.1) shows R_0 is linearly positively correlated with β_1 and β_2; R_0 is negatively correlated with μ_1 and μ_2; the correlation between R_0 and p or q is determined by $\beta_2\mu_1 - \beta_1\mu_2$. As can be seen from **Figure 3** that the monotonicity of R0 increases or decreases with respect to p and q, respectively. In **Figure 4**, the parameters of the model are estimated by MCMC using the number of cases in Wuhan and the parameter values are shown in **Table 2**. **Figure 4** shows the change of total infection density with time under different ω and σ. It can be seen that with the increase of ω and σ, the total infection density is becoming lower as ω and σ increase.

6 CONCLUSION AND DISCUSSION

In this article, we have generalized a traditional homogeneous epidemic model with asymptomatic infectors to a network case to study the spread of COVID-19. As one can see, homogeneous epidemic models can be considered as a special case of network epidemic models, and network epidemic models are more realistic and refined description of disease propagation in population. Considering two cases of infected state (show symptoms and show no symptoms), we investigated an *SEAIRS* model of COVID-19 on scale-free networks to approach the disease progress and calculated the basic reproduction number R_0. By theoretical analysis, we obtained the asymptotical stability of the equilibria: if $R_0 \le 1$, E_0 is asymptotically stable. Then, we analyzed the different quarantine strategies in the model. At last, numerical simulations illustrated the theoretical analysis. The results show that the asymptomatic infectors may infect health individuals as well as symptomatic infectors.

The *SEAIRS* model is an extension of the *SEAIRS* model, and it is a case of the $S \to S$ model. The $S \to S$ model, $S \to I$ model, and $S \to R$ model differ in terms of pathological mechanisms. It is difficult to say which one is better than others except introducing a special disease. Within $S \to S$ models, formally, either the *SIS* model or *SIRS* model can be a special case of the *SEIRS* model. The parameters β_1 and β_2 denoted in the model are constant. In fact, many infectious diseases show seasonal phenomenon, such as measles, chickenpox, and mumps [36], and each individual has a periodic order; the parameters above should have some periodic property.

At the same time, we have also discussed proportional, targeted quarantine and high-risk quarantine schemes for network models. The result of Leung et al. [27] by illustration is suitable for *Norovirus* and measles. Different from their work, the formulaic result seems to be unsuitable for any kind of disease, for none of the parameters in the model is assigned to a specific value. In fact, for different disease, the parameters in the model may present special values. COVID-19 is now spreading faster abroad. Susceptible people will also be infected by asymptomatic patients. If not controlled, the epidemic will spread more widely and the number of infected people will be more difficult to control. The epidemic situation in China was obviously well controlled after the implementation of the strategy of lockdown, and other quarantine strategies. Therefore, the model and quarantine strategy in this article can also be applied to other countries. And, we hope that the results will be helpful to study and control the spread of COVID-19 disease.

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

ML and YS designed the study, and WL and ML carried out the analysis and contributed to writing the article. XF and YS performed numerical simulations.

FUNDING

Research Project Supported by Fund for Shanxi 1331KIRT and the outstanding youth fund of North University of China.

REFERENCES

1. *Report on novel coronavirus pneumonia epidemic in real time*. (2020) https://voice.baidu.com/act/newpneumonia/newpneumonia/.
2. Tang B, Wang X, Li Q, Bragazzi NL, Tang SY, Xiao YN et al, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Comput Math* (2020) 9:462. doi:10.3390/jcm9020462
3. Tang B, Bragazzi NL, Li Q, Tang S, Xiao Y and Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infect Dis Model* (2020) 5:248–55. doi:10.1016/j.idm.2020.02.001
4. Hethcote HW. Three basic epidemiological models. *Appl Math Ecol* (1989) 355: 119–44.
5. Li MY, Muldowney JS and van den Driessche P. Global stability of SEIRS models in epidemiology. *Can Appl Math Q* (1999) 7:409–25.
6. Pastor-Satorras R and Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett* (2001) 86:3200–3. doi:10.1103/PhysRevLett.86.3200
7. Butts CT. Revisiting the foundations of network analysis. *Science* (2009) 325(5939):414–6. doi:10.1126/science.1171022
8. Newman M. *Networks: an introduction*. New York, NY, USA: Oxford University Press (2010).
9. Vespignani A. Predicting the behavior of techno-social systems. *Science* (2009) 325(5939):425–8. doi:10.1126/science.1171990
10. Vespignani A. Modeling dynamical processes in complex socio-technical systems. *Nat Phys* (2012) 8:32–9. doi:10.1038/nphys2160
11. Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A. Epidemic processes in complex networks. *Rev Mod Phys* (2015) 87:925–79. doi:10.1017/cbo9780511791383.010
12. Wang B, Han Y and Tanaka G. Interplay between epidemic spread and information propagation on metapopulation networks. *J Theor Biol* (2017) 420:18–25. doi:10.1016/j.jtbi.2017.02.020
13. Liu G, Liu Z and Jin Z. Dynamics analysis of epidemic and information spreading in overlay networks. *J Theor Biol* (2018) 444:28–37. doi:10.1016/j.jtbi.2018.02.010
14. Zhang X, Shan C, Jin Z and Zhu H. Complex dynamics of epidemic models on adaptive networks. *J Differ Equ* (2019) 266(1):803–32. doi:10.1016/j.jde.2018.07.054
15. Massad E, Ma S, Chen M, Struchiner C, Stollenwerk N and Mara A. Scale-free network of a dengue epidemic. *Appl Math Comput* (2008) 195:376–81. doi:10.1016/j.amc.2007.04.102
16. Mishra BK and Pandey SK. Dynamic model of worms with vertical transmission in computer network. *Appl Math Comput* (2011) 217: 8438–46. doi:10.1016/j.amc.2011.03.041
17. Zhang JP and Jin Z. The analysis of an epidemic model on networks. *Appl Math Comput* (2011) 217:7053–64. doi:10.1007/978-3-662-47824-0_3
18. Nian F and Yao S. The epidemic spreading on the multi-relationships network. *Appl Math Comput* (2018) 339:866–73. doi:10.1016/j.amc.2018.07.030
19. Wang ZS, Guo QT, Sun SW and Xia CY. The impact of awareness diffusion on SIR-like epidemics in multiplex networks. *Appl Math Comput* (2019) 349: 134–47. doi:10.1016/j.chaos.2019.109425
20. Wang ZS, Xia CY, Chen ZQ and Chen GR. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern* (2020) doi:10.1109/TCYB.2019.2960605
21. Grunnill M. An exploration of the role of asymptomatic infections in the epidemiology of dengue viruses through susceptible, asymptomatic, infected and recovered (SAIR) models. *J Theor Biol* (2018) 439:195–204. doi:10.1016/j.jtbi.2017.12.009
22. Mishra A and Gakkhar S. Modeling of Dengue with impact of asymptomatic infection and ADE factor. *Differ Equ Dyn Syst* (2018) 19:1–17. doi:10.1007/s12591-018-0436-4
23. Beretta E, Capasso V and Garao DG. A mathematical model for malaria transmission with asymptomatic carriers and two age groups in the human population. *Math Biosci* (2018) 300:87–101. doi:10.1016/j.mbs.2018.03.024
24. Gallimore CI, Cubitt D, du Plessis N and Gray JJ. Asymptomatic and symptomatic excretion of noroviruses during a hospital outbreak of gastroenteritis. *J Clin Microbiol* (2004) 42:2271–4. doi:10.1128/jcm.42.5.2271-2274.2004
25. García C, DuPont HL, Long KZ, Santos JI and Ko G. Asymptomatic norovirus infection in Mexican children. *J Clin Microbiol* (2006) 44:2997–3000. doi:10.1128/JCM.00065-06
26. Bucardo F, Nordgren J, Carlsson B, Kindberg E, Paniagua M, Möllby R, et al, and Svensson L. Asymptomatic norovirus infections in Nicaraguan children and its association with viral properties and histo-blood group antigens. *Pediatr Infect Dis J* (2010) 29:934–9. doi:10.1097/INF.0b013e3181ed9f2f
27. Leung KY, Trapman P and Britton T. Who is the infector? Epidemic models with symptomatic and asymptomatic cases. *Math Biosci* (2018) 301:190–8. doi:10.1016/j.mbs.2018.04.002
28. Chisholm RH, Campbell PT, Wu Y, Tong SYC, McVernon J and Geard N Implications of asymptomatic carriers for infectious disease transmission and control. *R Soc Open Sci* (2018) 5:172341. doi:10.1098/rsos.172341
29. De la Sen M, Alonso-Quesada S, Ibeas A and Nistal R. On an SEIADR epidemic model with vaccination, treatment and dead-infectious copses removal controls. *Math Comput Simul* (2019) 163:47–79. doi:10.3390/ijerph16152689
30. Ma Y, Liu M, Hou Q and Zhao J. Modelling seasonal HFMD with the recessive infection in Shandong, China. *Math Biosci Eng* (2013) 10:1159–71. doi:10.3934/mbe.2013.10.1159
31. Cao J, Wang Y, Alofi A, Mazrooei AA and Elaiw A. Global stability of an epidemic model with carrier state in heterogeneous networks. *IMA J Appl Math* (2015) 80:1025. doi:10.1093/imamat/hxu040
32. van den Driessche P and Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci* (2002) 180:29–48. doi:10.1016/s0025-5564(02)00108-6
33. Liu J and Zhang T. Epidemic spreading of an SEIRS model in scale-free networks. *Commun Nonlinear Sci Numer Simulat* (2011) 16:3375–84. doi:10.1063/1.5001176
34. Zhao XQ and Jing ZJ. Global asymptotic behavior in some cooperative systems of functional differential equations. *Can Appl Math Q* (1996) 4:421–44.
35. Barthélemy M, Barrat A, Pastor-Satorras R and Vespignani A. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys Rev Lett* (2004) 92:178701. doi:10.1103/PhysRevLett.92.178701
36. Yorke JA and London WP. Recurrent outbreaks of measles, chickenpox and mumps. II. Systematic differences in contact rates and stochastic effects. *Am J Epidemiol* (1973) 98:469–82. doi:10.1093/oxfordjournals.aje.a121575
37. Nian F and Wang X. Efficient immunization strategies on complex networks. *J Theor Biol* (2010) 264:77–83. doi:10.1016/j.jtbi.2010.01.007

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 20 January 2021
doi: 10.3389/fphy.2020.593421

Epidemics Forecast From SIR-Modeling, Verification and Calculated Effects of Lockdown and Lifting of Interventions

R. Schlickeiser[1,2]* *and M. Kröger*[3]*

[1]*Institut für Theoretische Physik, Lehrstuhl IV: Weltraum- und Astrophysik, Ruhr-Universität Bochum, Bochum, Germany,* [2]*Institut für Theoretische Physik und Astrophysik, Christian-Albrechts-Universität zu Kiel, Kiel, Germany,* [3]*Polymer Physics, Department of Materials, ETH Zurich, Zurich, Switzerland*

Due to the current COVID-19 epidemic plague hitting the worldwide population it is of utmost medical, economical and societal interest to gain reliable predictions on the temporal evolution of the spreading of the infectious diseases in human populations. Of particular interest are the daily rates and cumulative number of new infections, as they are monitored in infected societies, and the influence of non-pharmaceutical interventions due to different lockdown measures as well as their subsequent lifting on these infections. Estimating quantitatively the influence of a later lifting of the interventions on the resulting increase in the case numbers is important to discriminate this increase from the onset of a second wave. The recently discovered new analytical solutions of Susceptible-Infectious-Recovered (SIR) model allow for such forecast. In particular, it is possible to test lockdown and lifting interventions because the new solutions hold for arbitrary time dependence of the infection rate. Here we present simple analytical approximations for the rate and cumulative number of new infections.

Keywords: coronavirus (2019-nCoV), statistical analysis, pandemic spreading, time-dependent infection rate, parameter estimation

OPEN ACCESS

Edited by:
Sen Pei,
Columbia University, United States

Reviewed by:
Ming Tang,
East China Normal University, China
Jiannan Wang,
Beihang University, China

***Correspondence:**
R. Schlickeiser
rsch@tp4.ruhr-uni-bochum.de
M. Kröger
mk@mat.ethz.ch

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *10 August 2020*
Accepted: *23 October 2020*
Published: *20 January 2021*

Citation:
Schlickeiser R and Kröger M (2021) Epidemics Forecast From SIR-Modeling, Verification and Calculated Effects of Lockdown and Lifting of Interventions. Front. Phys. 8:593421. doi: 10.3389/fphy.2020.593421

1 INTRODUCTION

The Susceptible-Infectious-Recovered (SIR) model has been developed nearly hundred years ago [1, 2] to understand the time evolution of infectious diseases in human populations. The SIR system is the simplest and most fundamental of the compartmental models and its variations [3–17]. The considered population of $N \gg 1$ persons is assigned to the three compartments s (susceptible), i (infectious), or r (recovered/removed). Persons from the population may progress with time between these compartments with given infection ($a(t)$) and recovery rates ($\mu(t)$) which in general vary with time due to non-pharmaceutical interventions taken during the pandemic evolution.

Let $I(t) = i(t)/N$, $S(t) = s(t)/N$ and $R(t) = r(t)/N$ denote the infected, susceptible and recovered/removed fractions of persons involved in the infection at time t, with the sum requirement $I(t) + S(t) + R(t) = 1$. In terms of the reduced time $\tau(t) = \int_0^t d\xi a(\xi)$, accounting for arbitrary but given time-dependent infection rates, the SIR-model equations are [1, 2, 18]

$$\frac{dI}{d\tau} = j - KI, \quad \frac{dS}{d\tau} = -j, \quad \frac{dR}{d\tau} = KI \tag{1}$$

in terms of the time-dependent ratio $K(t) = \mu(t)/a(t)$ of the recovery and infection rates and the medically interesting daily rate of new infections

$$\dot{J}(t) = a(t)j(\tau) = \dot{\tau}j(\tau), \tag{2}$$

where the dot denotes a derivative with respect to t.

For the special and important case of a time-independent ratio $K(t) = k = \text{const.}$ new analytical results of the SIR-model (1) have been recently derived [19] – hereafter referred to as paper A. The constant k is referred to as the inverse basic reproduction number $k = 1/R_0$. The new analytical solutions assume that the SIR equations are valid for all times $t \in [-\infty, \infty]$, and that time $t = \tau = 0$ refers to the "observing time" when the existence of a pandemic wave in the society is realized and the monitoring of newly infected persons $\dot{J}(t)$ is started. In paper A it has been shown that, for arbitrary but given infection rates $a(t)$, apart from the peak reduced time τ_0 of the rate of new infections, all properties of the pandemic wave as functions of the reduced time are solely controlled by the inverse basic reproduction number k. The dimensionless peak time τ_0 is controlled by k and the value $\varepsilon = -\ln S(0)$, indicating as only initial condition at the observing time the fraction of initially susceptible persons $S(0) = e^{-\varepsilon}$. This suggests to introduce the relative reduced time $\Delta = \tau - \tau_0$ with respect to the reduced peak time. In real time t the adopted infection rate $a(t)$ acts as second parameter, and the peak time t_m, where $\dot{J}(t)$ reaches its maximum must not coincide with the time, where the reduced j reaches its maximum, i.e., $\tau_m \equiv \tau(t_m) \neq \tau_0$, in general.

2 RESULTS AND DISCUSSION

According to paper A the three fractions of the SIR-model

$$\begin{aligned} &S(\tau) = 1 - J(\tau), \qquad I(\tau) = \frac{j(\tau)}{1 - J(\tau)}, \\ &R(\tau) = -k\ln[1 - J(\tau)] \end{aligned} \tag{3}$$

can be expressed in terms of the cumulative number $J(\tau)$ and differential daily rate $j(\tau) = dJ/d\tau$ of new infections. The cumulative number satisfies the nonlinear differential equation

$$j(\tau) = \frac{dJ}{d\tau} = (1 - J)[J + k\ln(1 - J)] \tag{4}$$

Two important values are $J_0(k) = J(\tau_0)$, where j attains its maximum with $(dj/dJ)_{J_0} = 0$, and the final cumulative number $J_\infty(k)$ at $\tau = t = \infty$, when the second bracket on the right-hand side of the differential **Eq. 4** vanishes, i.e., $J_\infty + k\ln(1 - J_\infty) = 0$. The two transcendental equations can be solved analytically in terms of Lambert's W function, as shown in paper A. In the present manuscript we are going to avoid Lambert's function completely, and instead use the following approximants (**Figure 1A**)

$$J_0(k) = (3 - k)(1 - k)\left(1 + k + k^2\right)/6, \tag{5}$$

$$J_\infty(k) = 1 - \exp[-(1 - k)(1 + \kappa)/k], \tag{6}$$

$$\kappa(k) = (4 - k)k/3 \tag{7}$$

Without any detailed solution of the SIR-model equations the formal structure of **Eqs 3** and **4** then provides the final values $I_\infty = j_\infty = 0$, $R_\infty = J_\infty$, and $S_\infty = 1 - J_\infty$. We list these values together with κ in **Table 1**. We emphasize that the final cumulative number J_∞, determined solely by the value of k, remains unchanged (**Table 1**). With NPIs one can only flatten and distort the epidemics curve (compared to the case of no NPIs taken) but not change the final cumulative number.

2.1 New infections

The exact solution of the differential **Eq. 4** is given in inverse form by (**Appendix A**)

$$\tau = \int_{1-e^{-\varepsilon}}^{J} \frac{dy}{(1 - y)[y + k\ln(1 - y)]}, \tag{8}$$

which can be integrated numerically (subject to numerical precision issues), replaced by the approximant presented in paper A (involving Lambert's function), or semi-quantitatively captured by the simple approximant to be presented next. The solution $J(\Delta)$ as a function of the relative reduced time $\Delta = \tau - \tau_0$, with the reduced peak time approximated by

$$\tau_0 = \frac{1 - k}{f_m(k)}\left[\ln\frac{J_0}{1 - J_0} - \ln\left(e^{\varepsilon} - 1\right)\right], \tag{9}$$

corresponding to $J = J_0$ in **Eq. 8**, and where $f_m(k) = 1 - k + \ln k$, is reasonably well captured by (**Appendix C**)

$$\begin{aligned} J(\Delta) = \frac{1}{2}[1 + \tanh Y_1(\Delta)]\Theta[\Delta_s(k) - \Delta] + \left\{1 - \frac{1 - J_\infty}{2}[1\right. \\ \left. + \coth Y_2(\Delta)]\right\}\Theta[\Delta - \Delta_s(k)] \end{aligned} \tag{10}$$

with the Heaviside step function $\Theta(x) = 1(0)$ for $x \geq (<)0$. In **Eq. 10**

$$\begin{aligned} Y_1 &= \frac{1}{2}\left[\frac{f_m(k)(\Delta - \Delta_s)}{1 - k} + \ln\frac{1 - k}{k}\right], \\ Y_2 &= \frac{1}{2}\left[E_0(k)(\Delta - \Delta_s) + \ln\frac{k}{(1 - k)\kappa}\right], \end{aligned} \tag{11}$$

with

$$\Delta_s = \frac{1 - k}{f_m(k)}\ln\frac{(1 - k)(1 - J_0)}{kJ_0}, \qquad E_0(k) = \left[\frac{k}{(1 - k)\kappa} - 1\right]f_m(k) \tag{12}$$

also tabulated in **Table 1**. We note that $\Delta_s(k)$ is always positive. **Figure 2** shows the approximation (**Eq. 10**) for the cumulative number as a function of the relative reduced time Δ for different values of k. For a comparison with the exact variation obtained by the numerical integration of **Eq. 8** see **Appendix C**. The agreement is remarkably well with maximum deviations less than 30 percent. The known limiting case of $k = 0$ is captured exactly by the approximant (**Appendix D**).

For the corresponding reduced differential rate $j(\Delta)$ in reduced time we use the right hand side of **Eq. 4** with $J = J(\Delta)$ from **Eq. 10**, cf. **Figure 3**. Note, that this j is not

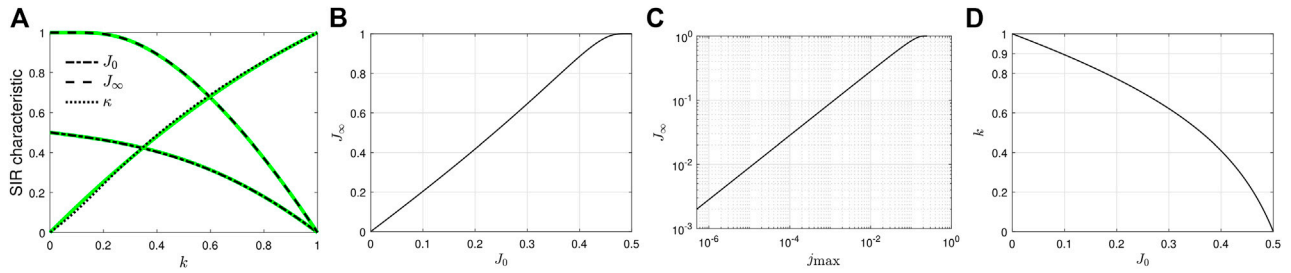


FIGURE 1 | (A) Approximants J_0, J_∞, and κ (thick green) used in this manuscript, cf. **Eqs (5)–(7)**, compared with the exact functions [19] (thin black). **(B)** J_∞ vs. J_0 and **(C)** J_∞ vs. $j_{\max}$ for the SIR model. **(D)** k as function of J_0 according to **Eq. 14**. For $J_0 \leq 0.1$ this is well approximated by $k \approx 1 - J_0$, where J_0 can be replaced by the cumulative fraction of infected people at the time of the maximum in the daily number of newly infected people.

TABLE 1 | Exact parameter values depending on the inverse basic reproduction number k.

k	κ	J_0	J_∞	S_∞	f_m	E_0	Δ_s	$j_{\max}$
0.00	0.00	0.500	1.000	0.000	1.000	0.0000	∞	0.2500
0.05	0.05	0.492	1.000	0.000	0.800	0.0000	3.5339	0.2327
0.10	0.11	0.483	1.000	0.000	0.670	0.0003	3.0441	0.2156
0.15	0.17	0.473	0.999	0.001	0.565	0.0049	2.7698	0.1986
0.20	0.24	0.462	0.993	0.007	0.478	0.0173	2.5745	0.1819
0.25	0.31	0.450	0.980	0.020	0.403	0.0348	2.4181	0.1653
0.30	0.37	0.436	0.959	0.041	0.339	0.0535	2.2835	0.1490
0.35	0.43	0.421	0.930	0.070	0.283	0.0709	2.1621	0.1330
0.40	0.49	0.403	0.893	0.107	0.234	0.0857	2.0491	0.1174
0.45	0.54	0.384	0.848	0.152	0.191	0.0971	1.9418	0.1022
0.50	0.59	0.363	0.796	0.203	0.153	0.1050	1.8386	0.0876
0.55	0.64	0.339	0.739	0.261	0.121	0.1093	1.7386	0.0736
0.60	0.69	0.313	0.676	0.324	0.094	0.1099	1.6416	0.0603
0.65	0.73	0.283	0.607	0.393	0.070	0.1071	1.5477	0.0479
0.70	0.78	0.251	0.533	0.467	0.050	0.1009	1.4571	0.0365
0.75	0.82	0.217	0.454	0.546	0.034	0.0914	1.3701	0.0263
0.80	0.86	0.179	0.371	0.629	0.022	0.0788	1.2871	0.0174
0.85	0.89	0.138	0.284	0.716	0.012	0.0633	1.2084	0.0102
0.90	0.93	0.095	0.193	0.807	0.005	0.0449	1.1343	0.0047
0.95	0.97	0.049	0.098	0.902	0.001	0.0237	1.0648	0.0012
1.00	1.00	0.000	0.000	1.000	0.000	0.0000	1.0000	0.0000

identical with the one obtained via $j = dJ/d\tau$, because J does not solve the SIR equations exactly. The peak value $j_{\max}$ in the reduced time rate occurs when $J = J_0$ and is thus determined by $j_{\max} = (1 - J_0)(1 - J_0 - k)$, also tabulated in **Table 1**.

As can be seen in **Figure 3** the rate of new infections (**Eq. 12**) is strictly monoexponentially increasing $j(\Delta \ll 0) \simeq e^{\Gamma_1 \Delta}$ with $\Gamma_1(k) = f_m(k)/(1-k)$ well before the peak time, and strictly monoexponentially decreasing well above the peak time $j(\Delta \gg 0) \propto e^{-\Gamma_2 \Delta}$ with the $\Gamma_2 = (1 - J_\infty)\Gamma_1/\kappa$. These exponential rates exhibit a noteworthy property and correlation in reduced time:

$$\frac{\Gamma_2}{\Gamma_1} = \frac{1 - J_\infty}{\kappa} \tag{13}$$

The SIR parameter k affects several key properties of the differential and cumulative fractions of infected persons. If the maximum $\dot{J}(t_m)$ of the measured daily number of newly infected persons has passed already, we find it most convenient to estimate k from the cumulative value $J(t_m)$ at this time t_m. While the maximum of $\dot{J}(t)$ must not occur exactly at $\tau(t_m) = \tau_0$ (**Appendix F**), we can still use J_0 as an approximant for the value of $J(t_m)$ and the relationship between J_0 and k can be inverted to read (**Appendix E**)

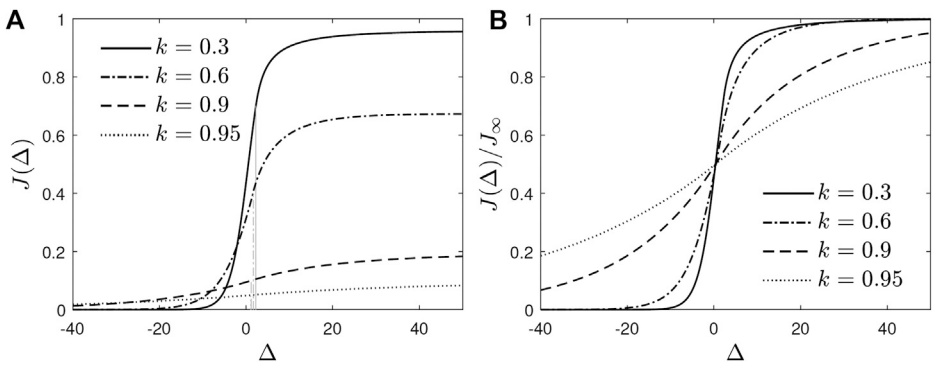


FIGURE 2 | (A) Cumulative number $J(\Delta)$ for different values of k according to **Eq. 10**. The vertical gray lines starting at the Δ-axis indicate the respective values of $\Delta_s(k)$. **(B)** Same as in **(A)**, divided by the final J_∞.

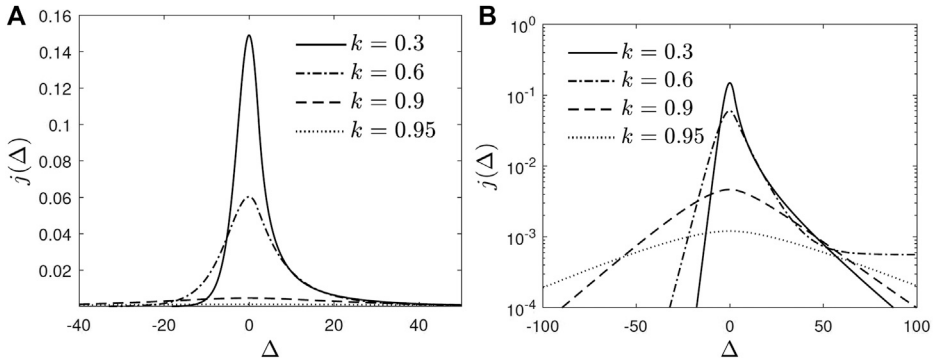


FIGURE 3 | Reduced differential rate $j(\Delta)$ of newly infected fraction corresponding to the cumulative $J(\Delta)$ shown in **Figure 2**. **(A)** linear scale, **(B)** semilogarithmic scale.

$$k = \frac{2(1-J_0)-1}{1+\ln(1-J_0)} = 1 - J_0 - \frac{J_0^2}{2} + O(J_0^3) \tag{14}$$

The dependency of k on J_0 is shown in **Figure 1C**. With the so-obtained value for k at hand, the infection rate $a(t_m)$ at peak time can be inferred from $a(t_m) = \dot{J}(t_m)/j_{\max}(k)$. It provides a lower bound for a_0.

A major advantage of the new analytical solutions in paper A and here is their generality in allowing for arbitrary time-dependencies of the infection rate $a(t)$. Such time-dependencies result at times greater than the observing time $t = 0$ from non-pharmaceutical interventions (NPIs) taken after the pandemic outbreak [20] such as case isolation in home, voluntary home quarantine, social distancing, closure of schools and universities and travel restrictions including closure of country borders, applied in different combinations and rigor [21] in many countries. These NPIs lead to a significant reduction of the initial constant infection rate a_0 at later times. It is also important to estimate the influence of a later lifting of the NPIs on the resulting increase in the case numbers in order to discriminate this increase from the onset of a second wave. Especially in the papers by Dehning et al. [17], Flaxman et al. [22] and those reviewed by Estrada [4] the influence of NPIs on the time evolution of the Covid-19 pandemics has been studied using numerical solutions of the SIR-model equations. Our analytical study presented here is superior to these results from numerical simulations as its predictions are particularly robust for the late forecast of the pandemic wave.

2.2 Modeling in Real Time of Lockdowns

The corresponding daily rate $\dot{J}(t)$ and cumulative number $J(t)$ of new infections in real time t for given time-dependent infection rates $a(t)$ are $J(t) = J(\tau(t))$ and $\dot{J}(t)$ given by **Eq. 2**. From a medical point of view the daily rate $\dot{J}(t)$ is most important as it determines also i) the fatality rate [23] $d(t) \simeq f\dot{J}(t - t_d)$ with the fatality percentage $f \simeq 0.005$ in countries with optimal medical services and hospital capacities and the delay time of $t_d \simeq 7$ days, ii) the daily number of new seriously sick persons [24] NSSPs = $2f\dot{J}(t - t_d)$ needing access to breathing apparati, and iii) the day of maximum rush to hospitals $t_r = t_m + t_d$. In countries with poor medical and hospital capacities and/or limited access to them the fatality percentage is significantly higher by a factor h which can be as large as 10.

To calculate the rate and cumulative number in real time according to **Eq. 2** we adopt as time-dependent infection rate the integrable function known from shock wave physics

$$a_{\rm LD}(t) = \frac{a_0}{2}\left[1 + q - (1-q)\tanh\frac{t-t_a}{t_b}\right] \simeq \begin{cases} a_0 & \text{for } t \ll t_a \\ qa_0 & \text{for } t \gg t_a, \end{cases} \tag{15}$$

which implies

$$\tau_{\rm LD}(t) = \frac{a_0}{2}\left[(1+q)t - (1-q)t_b \ln\left(\frac{\cosh\frac{t-t_a}{t_b}}{\cosh\frac{t_a}{t_b}}\right)\right] \simeq \begin{cases} a_0 t & \text{for } t \ll t_a \\ qa_0 t & \text{for } t \gg t_a \end{cases} \tag{16}$$

The time-dependent lockdown infection rate (**Eq. 15**) is characterized by four parameters: i) the initial constant infection rate a_0 at early times $t \ll t_a$, ii) the final constant infection rate $a_1 = qa_0$ at late times $t \gg t_a$ described by the quarantine factor $q = a_1/a_0 \le 1$, first introduced in Refs. 21 and 24, iii) the time t_a of maximum change, and iv) the time t_b regularizing the sharpness of the transition. The latter is known to be about $t_b \simeq 7$–14 days reflecting the typical 1–2 weeks incubation delay. Consequently, the parameter q mainly affects the amplitude $\dot{J}$ shown in the left columns of **Figures 5** and **6** (note that we also plotted the case of no NPIs taken (i.e., $q = 1$) for comparison). Alternatively, the transition time t_b controls the rapidness of the transition in the fraction of infected persons per day and therefore the widespread.

Moreover, the initial constant infection rate a_0 characterizes the Covid-19 virus: if we adopt the German values $a_0 \simeq 58\ \text{days}^{-1}$ and $t_b \simeq 11$ determined below, with the remaining two parameters q and t_a we can represent with the chosen functional form **Eq. 15** four basic types of reductions: 1) drastic (small $q \ll 1$) and rapid (t_a small), 2) drastic (small $q \ll 1$) and late (t_a large), 3) mild (greater q) and rapid (t_a small), and 4) mild (greater q) and late (t_a large). The four types are exemplified in **Figure 4**.

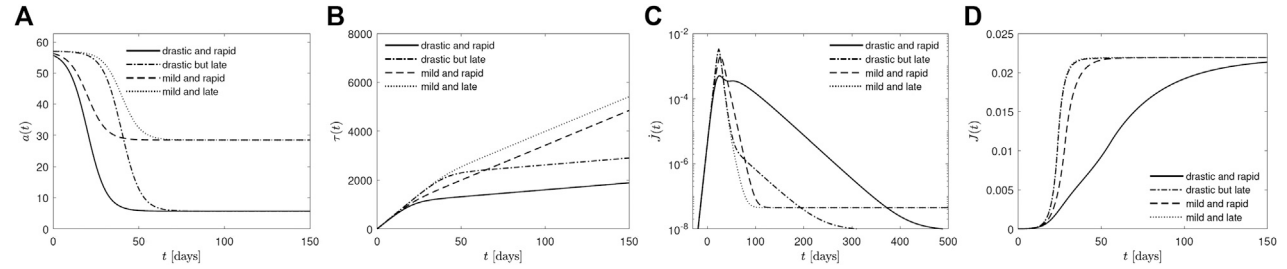


FIGURE 4 | (A) Infection rate $a(t)$, **(B)** reduced time $\tau(t)$, **(C)** daily rate of new infections $\dot{J}(t)$, **(D)** cumulative fraction $J(t)$ of infected persons. In each panel we consider four basic types of reductions: 1) drastic (small $q = 0.1$) and rapid ($t_a = 20$), 2) drastic (small $q = 0.1$) and late ($t_a = 40$), 3) mild ($q = 0.5$) and rapid ($t_a = 20$), and 4) mild ($q = 0.5$) and late ($t_a = 40$). Remaining parameters due to Germany: $t_b = 11$ days, $a_0 = 57$ days^{-1}, and $k = 0.989$. Both curves in **(C)** for the late cases reach the same value at the maximum, which is plausible as the curve remains unaffected at the time of the maximum. The rapid cases tend to lower the maximum amplitude already at the time of the maximum, and thus tend to decrease it compared with the late cases.

2.3 Verification and Forecast

In countries where the peak of the first Covid-19 wave has already passed such as e.g. Germany, Switzerland, Austria, Spain, France and Italy, we may use the monitored fatality rates and peak times to check on the validity of the SIR model with the determined free parameters. However, later monitored data are influenced by a time varying infection rate $a(t)$ resulting from non-pharmaceutical interventions (NPIs) taken during the pandemic evolution. Only at the beginning of the pandemic wave it is justified to adopt a time-independent injection rate $a(t) \simeq a_0$ implying $\tau = a_0 t$. Alternatively, also useful for other countries which still face the climax of the pandemic wave, it is possible to determine the free parameters from the monitored cases in the early phase of the pandemic wave. We illustrate our parameter estimation using the monitored data from Germany with a total population of 83 million persons ($P = 8.3 \times 10^7$).

In Germany the first two deaths were reported on March 9 so that $\varepsilon = 4.8 \times 10^{-6}$ corresponding to about 400 infected people 7 days earlier, on March 2 ($t = 0$). The maximum rate of newly infected fraction, $\dot{J}_{\max} \simeq 380/fP$, occurred $t_m = 37$ days later, consistent with a peak time of fatalities on 16 April 2020. At peak time the cumulative death number was $D_m = 3820/P$ corresponding to $J_m = D_m/f = 200 D_m = 0.009$. This implies $k \approx 1 - J_0 \approx 0.991$ according to **Eq. 14** (and not far from the value $k = 0.989$ to be determined from the fit shown in **Figure 5**). From the initial exponential increase of daily fatalities in Germany we extract $\Gamma_1(k) a_0 \simeq 0.28$, corresponding to a doubling time of $\ln(2)/\Gamma_1 a_0 \simeq 2.3$ days, as we know $\Gamma_1(k) = f_m(k)/(1-k) \simeq 0.0046$ already from the above k. The quantity a_m we can estimate from the measured $\dot{J}_{\max}$, as $\dot{J}_{\max} = a_m j_{\max}$ and $j_{\max}(k) \approx 4.2 \times 10^{-5}$. Using the mentioned value for $\dot{J}_{\max}$, we obtain $a_m \approx 22$/days as a lower bound for a_0.

With these parameter values the entire following temporal evolution of the pandemic wave in Germany can be predicted as function of t_b and q. To obtain all parameters consistently, we fitted the available data to our model without constraining any of the parameters (**Figure 5**). This yields for Germany $k \simeq 0.989$, $t_a \simeq 21$ days, $q \simeq 0.15$, $a_0 \simeq 58$ days, and $t_b = 11$ days. The obtained parameters allow us to calculate the dimensionless peak time $\tau_m \simeq 1353$, the dimensionless time $\tau_0 \simeq 1390$, as well as $J_m \simeq 0.009$, $J_0 \simeq 0.011$, $\Gamma_1 \simeq 0.0056$ and $\Gamma_2 \simeq 0.0055$.

We note that the value of $k = 0.989$ implies for Germany that $J_\infty(0.989) = 0.022$ according to **Figure 1**, so that at the end of the first Covid-19 wave in Germany 2.2% of the population, i.e., 1.83 million persons will be infected. This number corresponds to a final fatality number of $D_\infty = 9146$ persons in Germany.

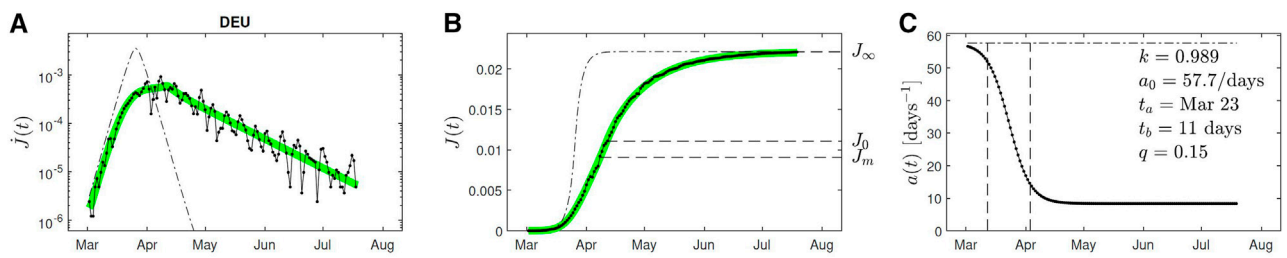


FIGURE 5 | (A) Measured data $\dot{J}(t)$ of new daily infected fraction (black circles) for Germany (DEU) compared with the model $\dot{J}(t) = a(t) j(\tau(t))$ outlined here (green). Shown for comparison is the case where no NPIs are imposed ($q = 1$, black dot-dashed). **(B)** The measured cumulative fraction $J(t)$ (black circles) together with the model prediction (green), and the reference $q = 1$ case. Also depicted are J_0 and the value at peak time, J_m. **(C)** The infection rate $a(t)$ corresponding to the curves shown in **(A)** and **(B)**. Model parameters mentioned in the figure; a dropped from an initial value of $\simeq 58$/days down to 7.8/days during the 2nd half of march. This realized case can be directly compared with the four hypothetic cases shown in **Figure 4**. For details on how to obtain the parameters see **Appendix G**.

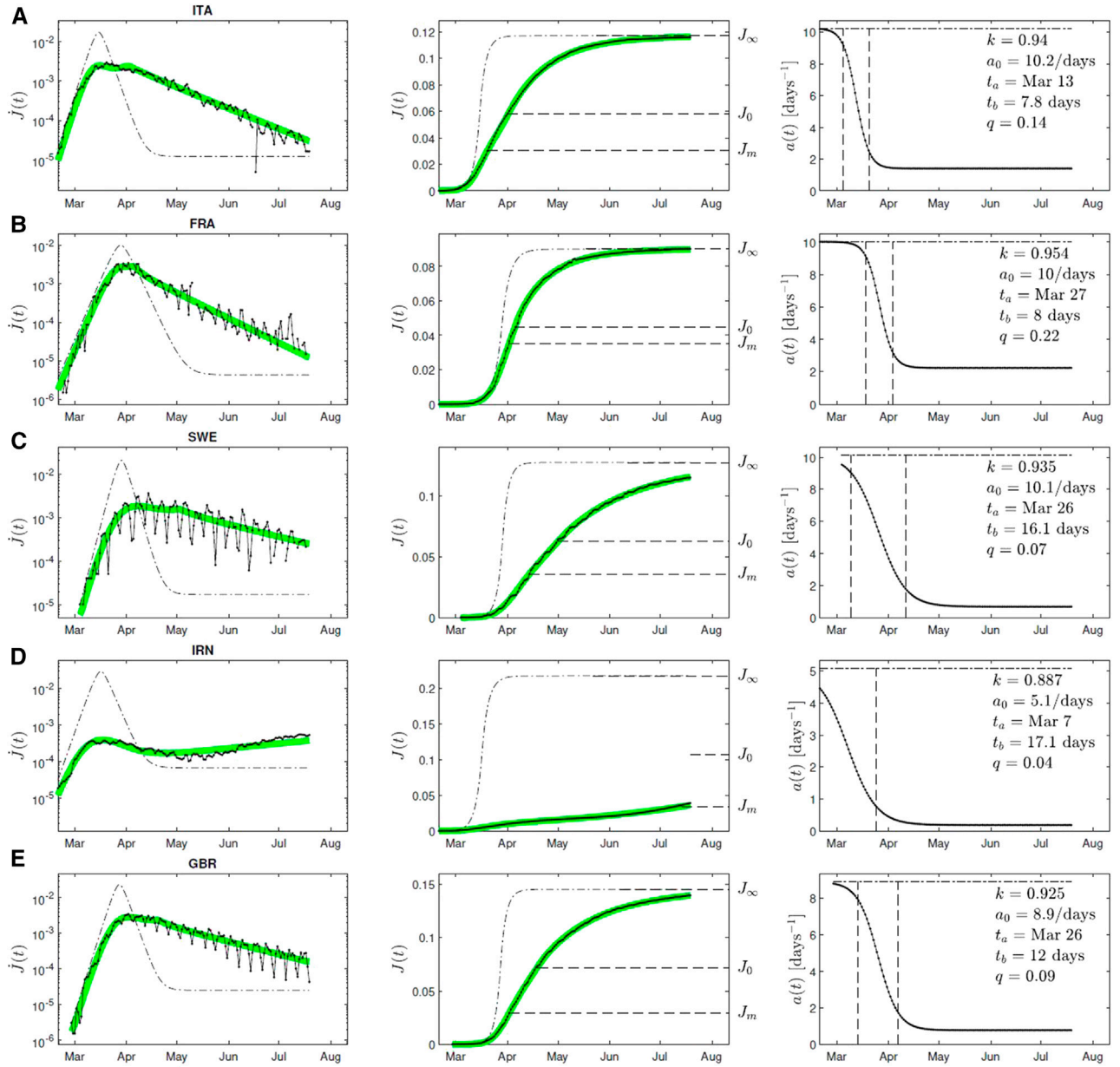


FIGURE 6 | Same as **Figure 5** for other countries: **(A)** Italy (ITA), **(B)** France (FRA), **(C)** Sweden (SWE), **(D)** Iran (IRN), **(E)** Great Britain (GBR).

Of course, these numbers are only valid estimates if no efficient vaccination against Covid-19 will be available.

An important consequence of the small quarantine factor $q = 0.15$ is the implied flat exponential decay after the peak. Because $\Gamma_1 \simeq \Gamma_2$ for $k = 0.989$, the exponential decay is by a factor q smaller than the exponential rise prior the climax, i.e.,

$$\dot{J}(t \gg t_m) \propto e^{-\Gamma_2 a_0 q t} = e^{-\frac{\Gamma_1 (1-J_\infty) a_0 q t}{\kappa}} \simeq e^{-t/21.8 \text{ days}} \tag{17}$$

Equation 17 yields a decay half-live of $\ln(2) \times 21.8$ days $\simeq 15$ days to be compared with the initial doubling time of $\simeq 2.3$ days. For Germany we thus know that their lockdown was drastic and rapid: the time $t_a \simeq$ March 23 is early compared to the peak time $t_m \simeq$ April 8 resulting in a significant decrease of the infection rate with the quarantine factor $q \simeq a_m/a_0 = 0.15$. In **Figure 5** we calculate the resulting daily new infection rate as a function of the time t for the parameters for Germany, and compare with the measured data. In the meantime, the strict lockdown interventions have been lifted in Germany: this does not effect the total numbers J_∞ and D_∞ but it should reduce the half-live decay time further.

We also performed this parameter estimation for other countries with sufficient data. For some of them data is visualized in **Figure 6**, parameters for the remaining countries are tabulated in **Tables 2** and **3**. Most importantly, with the

TABLE 2 | Model parameters and model implications. The columns are as follows: country (α_3 code), population P in millions (M), outbreak time defining $t = 0$, fitted time t_a, estimated time t_0 corresponding to the reduced peak time τ_0 of $j(\tau)$, fitted SIR parameter k, fitted initial infection rate a_0, fitted parameter t_b, fitted quarantine factor q, estimated doubling time t_2 characterizing the early exponential increase, estimated decay half life t_2' characterizing the late exponential decrease, estimated unreported number of infections per reported number, estimated final fraction J_∞ of infected population, final number of estimated fatalities $D_\infty P = J_\infty P f$. We use $f = 0.005$ as the probability to decease from a Covid-19 infection (reported plus unreported).

country	P/M	$t = 0$	t_a	t_m	t_0	k	a_0	t_b	q	t_2	t_2'	dark	J_∞	$D_\infty P$
AFG	34.66	Mar 18	Apr 3	Jul 7	Jun 28	0.995	68.9	36	0.16	4.0	24.8	7.1	1.00%	1725
ALB	2.88	Mar 11	Mar 22	Jul 17	Jul 19	0.988	24.4	1	0.14	4.7	34.2	6.8	2.41%	346
AND	0.08	Mar 19	Mar 25	Mar 23	Apr 4	0.931	7.6	0	0.24	2.6	12.0	11.8	13.50%	52
ARG	43.85	Mar 6	Apr 4	Jul 13	Jul 19	0.987	21.8	12	0.20	4.8	24.1	4.6	2.63%	5774
ARM	2.93	Mar 22	Jun 9	Jul 1	Jul 19	0.897	1.1	14	0.30	11.4	43.7	4.0	19.78%	2893
AUT	8.75	Mar 9	Mar 11	Mar 31	Apr 5	0.992	149.8	19	0.05	1.1	23.4	7.2	1.66%	725
BEL	11.35	Mar 4	Mar 28	Apr 3	Apr 9	0.911	6.1	7	0.23	2.5	12.3	30.8	17.30%	9822
BFA	18.65	Mar 14	Mar 20	Mar 19	Apr 5	0.999	1754.5	0	0.25	2.8	10.9	10.1	0.06%	53
BGR	7.13	Mar 7	Mar 31	Jul 15	Jul 19	0.974	10.1	28	0.12	5.3	47.3	7.8	5.10%	1828
BLR	9.51	Mar 25	Mar 19	Jun 25	Jul 19	0.975	40.9	32	0.03	1.4	55.1	1.6	4.91%	2344
BOL	10.89	Mar 23	Apr 12	Jun 24	Jul 3	0.968	7.9	29	0.31	5.5	18.4	8.7	6.30%	3428
BRA	207.65	Mar 11	Mar 15	May 27	Jul 13	0.919	12.1	46	0.02	1.4	64.9	8.3	15.83%	164365
CAF	4.60	May 24	May 27	Jun 15	Jun 14	0.999	123.9	11	1.87	9.3	5.0	2.6	0.24%	55
CHE	8.37	Mar 1	Mar 22	Mar 23	Apr 3	0.976	22.0	5	0.25	2.6	10.9	11.8	4.70%	1969
CHL	17.91	Mar 16	Jun 7	Jul 9	Jul 8	0.910	2.0	11	0.22	7.5	38.3	5.5	17.43%	15610
CHN	1378.67	Jan 15	Feb 1	Apr 9	Feb 16	0.999	1544.6	9	0.09	2.6	28.1	10.9	0.07%	4765
COL	48.65	Mar 15	Apr 4	Jul 16	Jul 19	0.899	4.0	7	0.19	3.3	20.1	8.4	19.48%	47385
CUB	11.48	Mar 19	Mar 22	Apr 16	Apr 15	0.999	631.5	21	0.25	2.9	11.5	7.1	0.15%	86
CZE	10.56	Mar 17	Mar 25	Apr 1	Apr 11	0.997	235.1	8	0.08	1.7	21.4	5.3	0.69%	365
DEU	82.67	Mar 2	Mar 23	Apr 8	Apr 11	0.989	57.7	11	0.15	2.2	15.1	9.0	2.21%	9146
DNK	5.73	Mar 8	Mar 23	Mar 28	Apr 8	0.989	52.6	6	0.16	2.4	15.3	9.2	2.15%	615
DOM	10.65	Mar 12	Mar 23	Jul 17	Jul 19	0.747	2.6	16	0.04	1.9	76.4	4.0	45.91%	24442
DZA	40.61	Mar 6	Mar 31	Apr 2	Jul 19	0.977	18.4	9	0.05	3.3	65.2	10.0	4.48%	9105
ECU	16.39	Mar 7	Mar 7	Apr 25	Jul 8	0.935	15.5	40	0.01	1.3	119.5	14.9	12.72%	10422
EGY	95.69	Mar 6	Apr 3	Jun 8	Jun 22	0.994	52.5	5	0.22	4.4	19.8	10.5	1.20%	5724
ESP	46.44	Feb 26	Mar 17	Mar 25	Mar 31	0.937	12.4	6	0.14	1.7	13.4	21.9	12.27%	28498
ETH	102.40	Mar 29	Mar 29	Jul 19	Jul 10	0.999	215	4	1.29	21	16.2	4.7	0.06%	315
FRA	66.90	Feb 19	Mar 27	Apr 2	Apr 5	0.954	10.0	8	0.22	3.0	14.2	28.5	9.01%	30142

exception of the six countries ARM, DOM, IRN, PAN, PER, SMR we found values of $k > 0.9$ for all other countries investigated corresponding to basic reproduction numbers $R_0 = 1/k < 1.11$. These values are significantly smaller than the estimates of $R_0 \in [2.4, 5.6]$ in the mainstream literature on Covid-19 [4, 22]. Part of these significant differences may be explained by the different definitions of R_0.

While the inverse basic reproduction number $k = 1/R_0 = \mu(t)/a(t)$ in the SIR-model is clearly defined as the ratio of the recovery to infection rate, there are alternative definitions of the basic reproduction number R_0 using the effective reproduction factor $R(t)$. As discussed in detail in Sect. 4 of Ref. 25 $R(t)$ has to be calculated from the convolution

$$R(t) = \frac{c(t)}{\int_0^\infty ds\, W(s) c(t-s)} \tag{18}$$

of the number of daily cases $c(t)$ with the serial interval distribution $W(t)$ describing the probability for the time lag between a person's infection and the subsequent transmission of the virus to a second person. As different choices of the serial interval distribution are used in the literature this leads to differences in the calculated associated effective reproduction factors $R(t)$. As R_0 is identical to the value $R(t_0)$ at the starting time of the outbreak it is not clear in the moment that this R_0 will be identical to the $1/k$ of the SIR model [26].

2.4 Summary and Conclusion

In this work we derived for the first time an analytical approximation for the solution for the SIR-model equations with an accuracy better than 30 percent. The explicit approximation refers to the fraction of newly infected persons per day $\dot{J}$ as a function of the relative reduced time with respect to the reduced peak time. This closed form of the analytical solution only depends on a single parameter k, the ratio of infection to recovery rates. We assume that this ratio is independent of time. As $\dot{J}$ can be directly compared with the monitored death and infection rates in different countries, we see no advantage in using the more complicated SEIR-model which currently does not allow for a closed analytical solution. An analytic solution of the SIR model with an accuracy better than 5% is available as well from our yet unpublished work where we did not consider time-dependent $a(t)$, but it has the disadvantage that it involves Lambert's function.

For the first time in the history of SIR-research (these equations have been discovered 93 years ago!) we thus have derived an analytical solution which can be applied successfully to all accumulated data of virus diseases in the world. Being of analytic form it is superior to all existing numerical simulations and results in the literature. We also discovered for the first time how to extract the value of k from the monitored data which is highly nontrivial. We applied this new method to the data taken

TABLE 3 | Continuation of **Table 2**.

country	P/M	$t = 0$	t_a	t_m	t_0	k	a_0	t_b	q	t_2	t_2'	dark	J_∞	$D_\infty P$
GAB	1.98	Apr 16	Apr 30	Jun 8	Jun 6	0.997	64.6	0	0.39	7.6	19.4	1.6	0.56%	56
GBR	65.64	Feb 29	Mar 26	Apr 2	Apr 19	0.925	8.9	12	0.09	2.0	26.1	30.9	14.54%	47719
GHA	28.21	Mar 16	Mar 21	Jun 12	Jun 18	0.999	464.9	0	0.22	4.4	19.9	1.2	0.13%	190
GRC	10.75	Mar 7	Mar 24	Mar 27	Apr 7	0.998	191.9	6	0.16	3.8	24.6	10.1	0.38%	203
GTM	16.58	Mar 28	Jun 11	Jul 7	Jul 6	0.985	9.7	10	0.48	9.4	19.9	9.0	3.03%	2513
HND	9.11	Mar 22	Mar 28	Jun 24	Jul 14	0.978	26.0	17	0.11	2.5	23.5	6.6	4.28%	1952
HRV	4.17	Mar 19	Apr 11	Apr 11	Apr 28	0.996	72.3	17	0.04	4.8	122.5	6.3	0.79%	165
IND	1324.17	Mar 6	Mar 21	Jun 8	Jun 28	0.997	189.9	36	0.12	2.4	20.8	5.9	0.61%	40316
IRL	4.77	Mar 7	Mar 13	Apr 17	Apr 16	0.962	21.2	29	0.11	1.7	16.9	13.7	7.41%	1768
IRN	80.28	Feb 12	Mar 8	Jul 14	Jul 19	0.819	3.1	17	0.04	2.3	84.0	11.5	33.82%	135755
ITA	60.60	Feb 15	Mar 13	Mar 20	Apr 1	0.940	10.2	8	0.14	2.2	17.6	28.7	11.70%	35442
KOR	25.37	Feb 14	Feb 16	Mar 14	Mar 21	0.999	989.6	13	0.07	1.2	17.4	4.3	0.22%	285
KWT	4.05	Apr 6	May 5	May 22	Jun 6	0.986	21.8	20	0.09	4.5	51.0	1.5	2.83%	573
LBN	6.01	Mar 4	Mar 31	Jul 19	Jul 19	0.999	95.5	2	0.09	9.9	105.3	3.6	0.29%	88
LUX	0.58	Mar 11	Mar 19	Apr 4	Apr 4	0.981	29.5	7	0.21	2.4	12.1	4.0	3.82%	111
MAR	35.28	Mar 10	Mar 24	Mar 28	Apr 17	0.999	628.8	13	0.03	2.5	82.8	3.6	0.18%	316
MDA	3.55	Mar 20	Mar 25	Jun 1	Jun 15	0.973	25.9	23	0.06	1.9	33.7	7.0	5.40%	960
MEX	127.54	Mar 13	Mar 24	Jul 19	Jun 24	0.954	16.2	42	0.06	1.8	31.1	25.4	8.99%	57326
MKD	2.08	Mar 16	Mar 31	Jun 25	Jul 19	0.927	5.4	9	0.11	3.4	33.1	10.1	14.29%	1488
MRT	4.30	May 6	Jun 1	Jun 3	Jun 7	0.996	66.4	4	0.29	5.1	17.9	5.3	0.82%	175
MYS	31.19	Mar 10	Mar 17	Mar 15	Mar 30	0.999	2064.2	7	0.11	1.7	16.2	2.8	0.08%	123
NGA	185.99	Mar 23	Apr 25	Jun 9	Jun 24	0.999	405.4	21	0.16	5.2	33.1	4.7	0.13%	1214
NLD	17.02	Mar 1	Mar 22	Mar 31	Apr 8	0.963	15.4	5	0.20	2.4	12.6	23.7	7.26%	6182
NPL	28.98	May 10	Jun 2	May 29	Jun 20	0.999	953.4	3	0.34	7.7	22.5	0.5	0.04%	55
PAK	193.20	Mar 11	Apr 8	Jun 11	Jun 13	0.997	101.1	17	0.22	4.0	18.2	4.4	0.69%	6677
PAN	4.03	Mar 15	Mar 26	Jul 10	Jul 19	0.848	3.5	10	0.09	2.5	35.5	4.8	28.80%	5808
PER	31.78	Mar 13	Mar 6	Jul 16	Jul 19	0.703	3.2	47	0.03	1.3	68.9	10.1	52.66%	83666
PHL	103.32	Mar 5	Mar 23	Jul 5	Jul 19	0.996	142.4	19	0.04	2.5	58.0	5.7	0.78%	4024
POL	37.95	Mar 6	Apr 2	Apr 17	Jun 5	0.993	63.1	20	0.05	3.4	67.0	8.3	1.30%	2474
PRT	10.33	Mar 11	Mar 14	Mar 31	Apr 22	0.982	86.4	16	0.02	0.9	38.8	7.1	3.67%	1896
ROU	19.71	Mar 15	Mar 17	Jul 14	Jul 19	0.909	13.8	21	0.02	1.1	79.1	11.7	17.60%	17343
RUS	144.34	Mar 18	Mar 18	May 22	Jul 5	0.984	60.7	44	0.02	1.4	61.8	3.4	3.12%	22543
SEN	15.41	Mar 28	May 9	Jun 8	Jun 24	0.998	79.9	1	0.55	11.5	21.0	4.3	0.30%	232
SMR	0.03	Mar 3	Mar 13	Mar 13	Mar 19	0.867	2.9	0	0.41	3.5	10.4	12.0	25.32%	42
SOM	15.01	Apr 6	Apr 19	Apr 15	May 4	0.999	591.7	0	0.24	3.7	15.2	6.0	0.13%	95
SRB	7.06	Mar 15	Mar 25	Jul 19	Jul 19	0.931	10.1	15	0.03	1.9	73.7	5.1	13.50%	4763
SWE	9.90	Mar 7	Mar 26	Apr 14	May 1	0.935	10.1	16	0.07	2.1	33.6	14.7	12.75%	6311
TCD	15.48	Apr 21	Apr 28	Apr 30	May 4	0.999	1387.0	3	0.22	2.1	9.2	16.9	0.10%	75
THA	68.86	Mar 17	Mar 29	Mar 26	Apr 1	0.999	3463.0	0	0.56	4.8	8.6	3.6	0.02%	58
TUN	11.40	Mar 15	Mar 30	Mar 26	Apr 1	0.999	659.7	6	0.29	4.7	16.4	7.3	0.09%	50
TUR	79.51	Mar 12	Mar 17	Apr 9	May 6	0.990	144	21	0.01	1.0	108.3	5.1	1.97%	7851
USA	323.13	Feb 24	Mar 23	Apr 10	May 16	0.938	10.2	23	0.03	2.1	81.6	7.8	12.18%	196763
ZAF	55.91	Mar 22	May 18	Jul 14	Jul 19	0.968	6.5	21	0.32	6.6	21.1	3.7	6.34%	17728

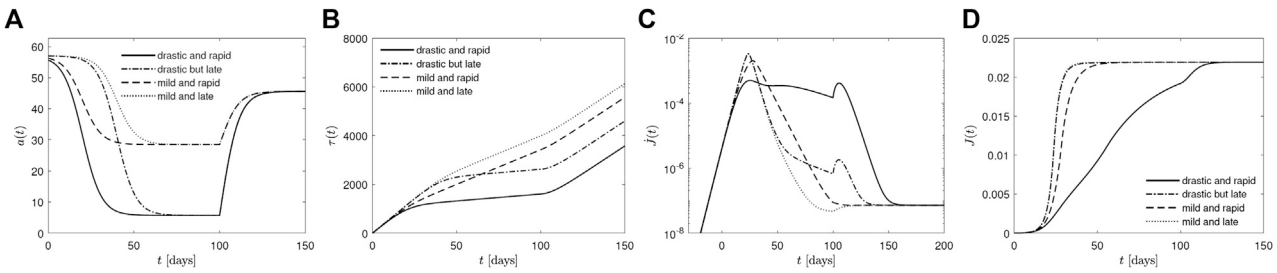


FIGURE 7 | Same as **Figure 4** but with incomplete lifting ($\eta = 0.8$) hundred days after breakout ($t_s = 100$ days). **(A)** infection rate $a(t)$, **(B)** reduced time $\tau(t)$, **(C)** daily rate of new infections $\dot{J}(t)$, **(D)** cumulative fraction $J(t)$ of infected persons. In each panel we consider four basic types of reductions: 1) drastic (small $q = 0.1$) and rapid ($t_a = 20$), 2) drastic (small $q = 0.1$) and late ($t_a = 40$), 3) mild ($q = 0.5$) and rapid ($t_a = 20$), and 4) mild ($q = 0.5$) and late ($t_a = 40$). Remaining parameters due to Germany: $t_b = 11$ days, $a_0 = 57$ days^{-1}, and $k = 0.989$.

for the Covid-19 pandemic waves in many countries. Our work includes an estimate on the effects of non-pharmaceutical interventions in these countries. This is possible as our analytical solution holds for arbitrary but given time dependencies $a(t)$ of the infection rates.

An example, on how lockdown lifting can be modeled is described in **Appendix H**. The situation is depicted in **Figure 7**. The lifting will increase $a(t)$ from its present value up to a value that might be close to the initial a_0. While the dynamics is altered, the final values remain unaffected by the dynamics, except, if the first pandemic wave is followed by a 2nd one. The values for J_∞ provided in **Tables 2** and **3** provide a hint on how likely is a 2nd wave. These values correspond to the population fraction that had been infected already. While this fraction is extremely large in Peru (53%), it is still below 1% in several of the larger countries. The tables also report the unreported number of infections per reported number (column "dark"), estimated from the number of fatalities, reported infections, and the death probability f.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://pomber.github.io/covid19/timeseries.json.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc A* (1927) 115:700–21.
2. Kendall DG. Deterministic and stochastic epidemics in closed populations. In: Conference third Berkeley symposium on mathematical statistics and probability; 1955 Jul–Aug. Vol. 4. Berkeley, CA: University of California Press (1956) p. 149–65.
3. Hethcode HW. The mathematics of infectious diseases. *SIAM Rev* (2000) 42: 599–653. doi:10.1137/s0036144500371907
4. Estrada E. Covid-19 and sars-cov-2. modeling the present, looking at the future. *Phys Rep* (2020) 869:1–51. doi:10.1016/j.physrep.2020.07.005
5. O'Regan SM, Kelly TC, Korobeinikov A, O'Callaghan MJA, Pokrovskii AV. Lyapunov functions for SIR and SIRS epidemic models. *Appl Math Lett* (2010) 23:446–8. doi:10.1016/j.aml.2009.11.014
6. Satsuma J, Willox R, Ramani A, Grammaticos B, Carstea AS. Extending the SIR epidemic model. *Phys Stat Mech Appl* (2004) 336:369–75. doi:10.1016/j.physa.2003.12.035
7. Cadoni M. How to reduce epidemic peaks keeping under control the time-span of the epidemic. *Chaos Solitons Fractals* (2020) 138:109940. doi:10.1016/j.chaos.2020.109940
8. Cadoni M, Gaeta G. Size and timescale of epidemics in the sir framework. *Phys D* (2020) 411:132626. doi:10.1016/j.physd.2020.132626
9. Chekroun A, Kuniya T. Global threshold dynamics of an infection age-structured SIR epidemic model with diffusion under the Dirichlet boundary condition. *J Diff Equ* (2020) 269:117–48. doi:10.1016/j.jde.2020.04.046
10. Imron C, HariyantoYunus M, Surjanto SD, Dewi NAC. Stability and persistence analysis on the epidemic model multi-region multi-patches. *J Phys Conf Ser* (2019) 1218: 012035. doi:10.1088/1742-6596/1218/1/012035
11. Karaji PT, Nyamoradi N. Analysis of a fractional SIR model with general incidence function. *Appl Math Lett* (2020) 108:106499. doi:10.1016/j.aml.2020.106499
12. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of Covid-19. *Appl Intell* (2020) 50, 3913–25. doi:10.1007/s10489-020-01770-9
13. Samanta S, Sahoo B, Das B. Dynamics of an epidemic system with prey herd behavior and alternative resource to predator. *J Phys Math Theor* (2019) 52: 425601. doi:10.1088/1751-8121/ab264d
14. Sene N. SIR epidemic model with mittag-leffler fractional derivative. *Chaos Solitons Fractals* (2020) 137:109833. doi:10.1016/j.chaos.2020.109833
15. Simon M. SIR epidemics with stochastic infectious periods. *Stoch Process their Appl* (2020) 130:4252–4274. doi:10.1016/j.spa.2019.12.003
16. Tian C, Zhang Q, Zhang L. Global stability in a networked SIR epidemic model. *Appl Math Lett* (2020) 107:106444. doi:10.1016/j.aml.2020.106444
17. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilcek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* (2020) 369:eabb9789. doi:10.1126/science.abb9789
18. Barmparis GD, Tsironis GP. Estimating the infection horizon of covid-19 in eight countries with a data-driven approach. *Chaos Solitons Fractals* (2020) 135:109842. doi:10.1016/j.chaos.2020.109842
19. Kröger M, Schlickeiser R. Analytical solution of the SIR-model for the temporal evolution of epidemics. part a: time-independent reproduction factor. *J Phys A* (2020) 53:505601. doi:10.20944/preprints202007.0416.v1
20. Ferguson N, Laydon D, Nedjati-Gilani G. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. London: Imperial College London (2020) doi:1025561/77482
21. Schlickeiser R, Kröger M. Dark numbers and herd immunity of the first covid-19 wave and future social interventions. *Epidem Int J* (2020) 4:000152. doi:10.23880/eij-16000152
22. Flaxman S, Mishra S, Mishra S, Gandy A, Unwin HJT, Mellan TA, et al. Estimating the effects of non-pharmaceutical interventions on Covid-19 in Europe. *Nature* (2020) 584:257–261. doi:10.1038/s41586-020-2405-7
23. Schüttler J, Schlickeiser R, Schlickeiser F, Kröger M. Covid-19 predictions using a gauss model, based on data from april 2. *Physics* (2020) 2:197–212. doi:10.3390/physics2020013
24. Schlickeiser R, Schlickeiser F. A Gaussian model for the time development of the sars-cov-2 corona pandemic disease. predictions for Germany made on 30 March 2020. *Physics* (2020) 2:164–70. doi:10.3390/physics2020010
25. Kröger M, Schlickeiser R. Gaussian doubling times and reproduction factors of the Covid-19 pandemic disease. *Front Phys* (2020) 8:276. doi:10.3389/fphy.2020.00276
26. Schlickeiser R, Kröger M. First consistent determination of the basic reproduction number for the first Covid-19 wave in 71 countries from the SIR-epidemics model with a constant ratio of recovery to infection rate. *Global J Front Res F* (2020) 20:37–43. doi:10.3929/ethz-b-000456421

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.

APPENDIX A: NON-PARAMETRIC SOLUTION OF THE SIR MODEL

We start from the Eq. 19 from part A

$$\tau = \int_{\varepsilon}^{G} \frac{dx}{1 - e^{-x} - kx} \tag{A1}$$

and substitute

$$y = 1 - e^{-x},\ x = -\ln(1-y),\ \frac{dx}{dy} = \frac{1}{1-y} \tag{A2}$$

Consequently, as the cumulative number of new infections is given (see Eq. 37 from part A) by

$$\tau = \int_{\psi}^{J} \frac{dy}{(1-y)f(y)},\quad f(y) = y + k\ln(1-y) \tag{A3}$$

with the abbreviation $\psi = J(0) = 1 - e^{-\varepsilon}$ for the initial value. This inverse relation $\tau(J)$ is the general solution of the SIR-model for constant k. It is not in parametrized form.

APPENDIX A.1: Maximum of *j*

Taking the derivative of Eq. 37 from part A with respect to τ we obtain

$$1 = \frac{j}{(1-J)[J + k\ln(1-J)]} \tag{A4}$$

or the exact SIR relation

$$j = (1-J)[J + k\ln(1-J)] \tag{A5}$$

Equation A5 provides

$$\frac{dj}{dJ} = 1 - k - 2J - k\ln(1-J) \tag{A6}$$

The maximum value $j_{\max}$ occurs for $(dj/dJ)_{J_0} = 0$ providing

$$1 - J_0 = \frac{k\ln(1-J_0) + k + 1}{2} \tag{A7}$$

Setting $1 - J_0 = e^{-X}$ yields

$$e^{-X} = -\frac{k}{2}\left(X - \frac{k+1}{k}\right), \tag{A8}$$

which is of the form of Eq. G1 from part A, and solved in terms of the non-principal Lambert function W_{-1} as

$$X = \frac{k+1}{k} + W_{-1}(\alpha_0),\quad \alpha_0 = -\frac{2}{ke}e^{-1/k,} \tag{A9}$$

so that

$$J_0 = 1 - e^{-\frac{1+k}{k} - W_{-1}(\alpha_0)} = 1 + \frac{k}{2}W_{-1}(\alpha_0) \tag{A10}$$

The maximum value is then given by

$$j_{\max} = j(J_0) = \frac{k^2}{4}\{[1 + W_{-1}(\alpha_0)]^2 - 1\} \tag{A11}$$

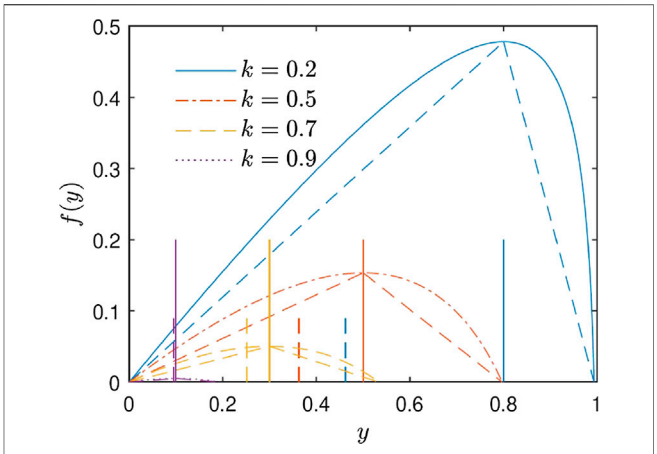


FIGURE B1 | Comparison of the approximation (B2) with the exact curve for $f(y)$ for different values of k. Vertical solid lines mark the position of the maximum of the function.

and this can also be written as $j_{\max} = (1-J_0)(1-J_0-k)$ with J_0 from **Eq. A10**. According to **Eq. 8** the reduced peak time in the dimensionless rate of new infections is then given by

$$\tau_0 = \int_{\psi}^{J_0} \frac{dy}{(1-y)f(y)}, \tag{A12}$$

which is the only quantity depending besides on k also on ε via $\psi = 1 - e^{-\varepsilon}$. In order to have our approximation depending only on k we therefore introduce the relative reduced time with respect to the peak reduced time

$$\Delta = \tau - \tau_0 = \int_{J_0}^{J} \frac{dy}{(1-y)f(y)} \tag{A13}$$

which is still exact, independent of ε and only determined by the value of k.

APPENDIX B: APPROXIMATING THE FUNCTION *f* (y)

The function $f(y)$ defined in **Eq. A3** vanishes for $y_c + k\ln(1-y_c) = 0$, or $1 - y_c = e^{-y_c/k}$ with the solution

$$y_c(k) = J_\infty(k) = (1-k)(1+\kappa) \tag{B1}$$

where κ was already stated in the introduction. According to **Eq. A13** the value J_∞ corresponds to $\Delta = \tau = \infty$, so the maximum value of the cumulative number of new infections is $J_{\max} = J_\infty$.

Moreover, the function $f(y)$ attains its maximum value $f_m(k) = f(y = 1-k) = 1 - k + k\ln k$ at $y_m = 1 - k$. As approximation we use

$$f(y) \simeq \begin{cases} f_1(y) & \text{for } y \le 1-k \\ f_2(y) & \text{for } (1-k) \le y \le J_\infty \end{cases} = f_m \begin{cases} \dfrac{y}{1-k} & \text{for } y \le 1-k \\ \dfrac{J_\infty - y}{J_\infty - (1-k)} & \text{for } 1-k \le y \le J_\infty \end{cases} \tag{B2}$$

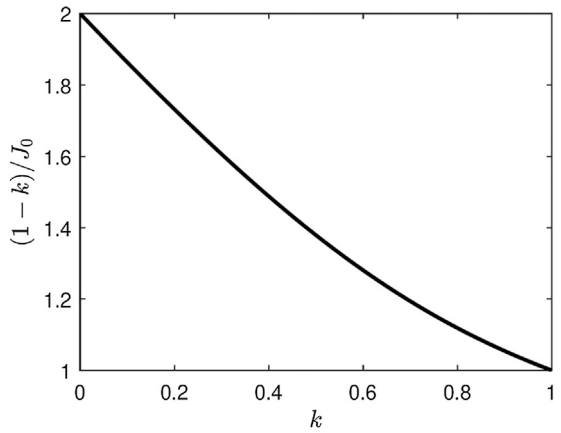


FIGURE C1 | The ratio $(1-k)/J_0(k)$ as a function of k.

which is shown in **Figure B1** in comparison with the function $f(y)$. The agreement is reasonably well with maximum deviations less than 30%.

APPENDIX C: APPROXIMATIONS FOR $J(\tau)$

Figure C1 demonstrates that $J_0(k)$ is always smaller than $1-k$. In order to calculate the integral in **Eq. A13** with the approximation **Eq. B1** we then have to investigate two cases: 1) For $J_0 < 1-k$ and $J < 1-k$ only the function f_1 contributes and

$$\Delta_{(J<1-k,J_0<1-k)} = \int_{J_0}^{J} \frac{dy}{(1-y)f_1(y)} \tag{C1}$$

2) For $J \geq 1-k > J_0$ both functions f_1 and f_2 contribute and

$$\begin{aligned}\Delta_{(J\geq 1-k>J_0)} &\simeq \int_{J_0}^{1-k} \frac{dy}{(1-y)f_1(y)} + \int_{1-k}^{J} \frac{dy}{(1-y)f_2(y)} \\ &= \Delta_s + \int_{1-k}^{J} \frac{dy}{(1-y)f_2(y)}\end{aligned} \tag{C2}$$

with

$$\begin{aligned}\Delta_s &= \int_{J_0}^{1-k} \frac{dy}{(1-y)f_1(y)} = \frac{1-k}{f_m(k)} \int_{J_0}^{1-k} \frac{dy}{(1-y)y} = \frac{1-k}{f_m(k)} \ln \frac{\frac{1}{J_0}-1}{\frac{1}{1-k}-1} \\ &= \frac{1-k}{f_m(k)} \ln \frac{(1-J_0)(1-k)}{kJ_0}\end{aligned} \tag{C3}$$

denoting the relative time corresponding to the value $J = 1-k$. We consider each case in turn.

Appendix C.1 Case (1): $J \leq 1-k$, $J_0 < 1-k$

Here **Eq. C1** provides

$$\frac{f_m \Delta}{1-k} = \int_{J_0}^{J} \frac{dy}{(1-y)y} = \ln \frac{\frac{1}{J_0}-1}{\frac{1}{J}-1}, \tag{C4}$$

so that the difference of **Eqs C3** and **C4** yields

$$\frac{f_m(\Delta - \Delta_s)}{1-k} = \ln \frac{kJ}{(1-k)(1-J)}, \tag{C5}$$

or after inversion

$$J(\tau) = \left[1 + \frac{k}{1-k} e^{-\frac{f_m(\Delta-\Delta_s)}{1-k}}\right]^{-1} \tag{C6}$$

Appendix C.2 Case (2): $J \geq 1-k > J_0$

Here **Eq. C2** with **Eq. C3** yields

$$\begin{aligned}f_m \Delta &\simeq (1-k) \int_{J_0}^{1-k} \frac{dy}{(1-y)y} + [J_\infty - (1-k)] \int_{1-k}^{J} \frac{dy}{(1-y)(J_\infty - y)} \\ &= f_m \Delta_s + (J_\infty - 1 + k) \int_{1-k}^{J} \frac{dy}{(1-y)(J_\infty - y)},\end{aligned} \tag{C7}$$

so that

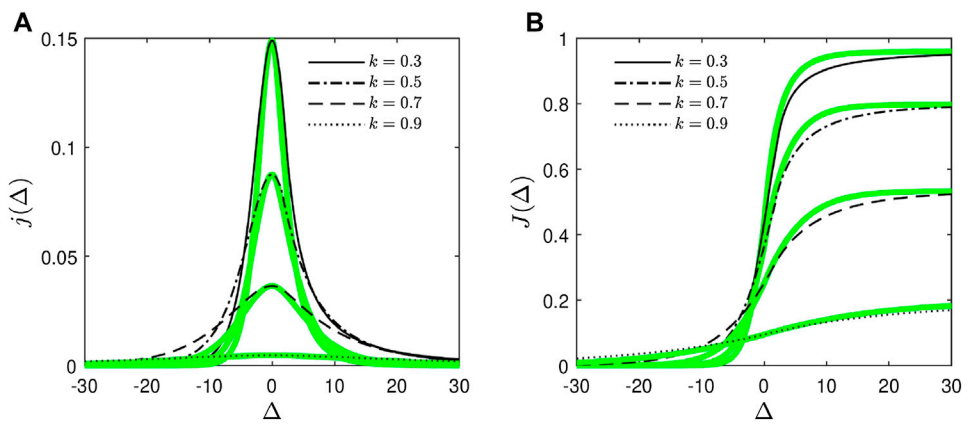


FIGURE C2 | Comparison for **(A)** $j(\Delta)$ and **(B)** $J(\Delta)$ between exact solution of the SIR model (green) and the approximant used here (black) for various k. Our approximant [19] in terms of Lambert's function is shown as well, but coincides with the exact solution (green) in this representation.

$$\frac{f_m(\Delta-\Delta_s)}{J_\infty-(1-k)}=\int_{1-k}^{J}\frac{dy}{(1-y)(J_\infty-y)}=-\frac{1}{1-J_\infty}\ln\frac{1-\frac{1-J_\infty}{1-J}}{1-\frac{1-J_\infty}{k}}$$
$$=-\frac{1}{1-J_\infty}\ln\frac{k(J_\infty-J)}{(J_\infty-1+k)(1-J)} \tag{C8}$$

After straightforward but tedious algebra we obtain

$$\frac{J_\infty-J}{1-J}=\frac{J_\infty-(1-k)}{k}e^{-E}, \tag{C9}$$

$$E(\Delta)=\frac{1-J_\infty}{J_\infty-(1-k)}f_m(\Delta-\Delta_s) \tag{C10}$$

and consequently

$$J(\Delta)=1+\frac{J_\infty-1}{1-\frac{(1-k)\kappa}{k}e^{-E(\Delta)}} \tag{C11}$$

Using the identities $2(1+e^{-2Y})^{-1}=1+\tanh Y$ and $2(1-e^{-2Y})^{-1}=1+\coth Y$ we combine the results **Eqs C6** and **C11** to the analytical approximation of the SIR-model equations at all reduced times, stated in **Eqs 10–12** above. A comparison with the exact numerical solution of the SIR model is provided in **Figure C2**. The corresponding $j(\Delta)$ is obtained from $J(\Delta)$ via **Eqs A5**.

APPENDIX D: SI-LIMIT $k = 0$

In the limit $k = 0$ **Eq. A7** provides $J_0(k=0)=1/2$ so that with $\lim_{k\to 0} f_m(k=0)=1$ the time scale (**Eq. C3**) becomes

$$\Delta_s(k=0)=\lim_{k\to 0}\ln\frac{1-k}{k}=\infty \tag{D1}$$

With this result

$$Y_1(k=0)=\frac{\Delta}{2}-\frac{1}{2}\lim_{k\to 0}\left[\Delta_s+\ln\frac{k}{1-k}\right]$$
$$=\frac{\Delta}{2}-\frac{1}{2}\lim_{k\to 0}\left[\ln\frac{1-k}{k}+\ln\frac{k}{1-k}\right]=\frac{\Delta}{2} \tag{D2}$$

Consequently, the cumulative number **Eq. 10** and the rate **Eq. 4** in this case for all times correctly reduce to

$$J(\Delta,k=0)=\frac{1}{2}\left[1+\tanh\frac{\Delta}{2}\right],\qquad j(\Delta,k=0)=\frac{1}{4\cosh^2(\Delta/2)} \tag{D3}$$

APPENDIX E: RELATIONSHIP BETWEEN J_0 AND K

Here we prove **Eq. 14**. According to paper A the quantity J_0 is given by $J_0=1-e^{-G_0}$ with

$$G_0=\frac{1+k}{k}+W_{-1}\left(\frac{-2e^{-1/k}}{ke}\right)=-\ln(1-J_0) \tag{E1}$$

where e denotes Euler's number and W_{-1} the non-principal solution of Lambert's equation $z=We^W$. **Equation E1** is of the form $x=r+c^{-1}W(ce^{-cr}/\beta)$ upon identifying $c=1$, $r=1/k$, $\beta=-ke/2$, and $x=-[1+\ln(1-J_0)]$. From paper A we thus know that $e^{-cx}=\beta(x-r)$ holds, or equivalently

$$(1-J_0)e=-\frac{ke}{2}\left[-1-\ln(1-J_0)-\frac{1}{k}\right]=\frac{ke}{2}[1+\ln(1-J_0)]+\frac{e}{2} \tag{E2}$$

This is readily solved for k, and thus proves **Eq. 14**.

APPENDIX F: TIME OF MAXIMUM IN THE MEASURED DIFFERENTIAL RATE $\dot{J}(t)$

One has $J(t)=J(\tau(t))$ and $\dot{J}(t)=\dot{\tau}(t)J'(\tau(t))=a(t)j(\tau(t))$ since $j=J'$ if we let the prime denote a derivative with respect to τ. The maximum in $\dot{J}(t)$ thus fulfills

$$0=\ddot{J}(t_m)=\dot{a}(t_m)j(\tau(t_m))+a^2(t_m)j'(\tau(t_m)) \tag{E1}$$

or equivalently,

$$0=\left[\frac{d\ln j}{d\tau}+\frac{\dot{a}}{a^2}\right]_{t=t_m} \tag{E2}$$

From part A we know that

$$\frac{d\ln j}{d\tau}=1-2J-k\ln(1-J)-k \tag{E3}$$

and our $J_0=J(\tau_0)$ solves $1=2J_0+k\ln(1-J_0)+k$. That is, $j'(\tau_0)=0$. If a does not depend on time, $\tau_0=\tau(t_m)=a_0t_m$, but this is not generally the case. To find t_m and $\tau_m\equiv\tau(t_m)$ one has to solve **Eq. E1**, or **Eq. E2**. **Equation E2** with **Eq. E3** is solved by

$$J_m=J(\tau_m)=1+\frac{k}{2}W_{-1}(\alpha_m) \tag{E4}$$

with

$$\alpha_m=-\frac{2e^{-(1+C_m)/k}}{ek},\quad C_m=-\frac{\dot{a}(t_m)}{a^2(t_m)} \tag{E5}$$

The corresponding j is, according to **Eq. 4**,

$$j(\tau_m)=(1-J_m)[J_m+k\ln(1-J_m)] \tag{E6}$$

The smaller C_m, the closer is J_m to J_0.

APPENDIX G: FITTING THE DATA

As discussed in length in paper A we base our analysis of existing data on the reported cumulative number of deaths, $D(t)$, from which we estimate the cumulative number of infections $J(t)=D(t-t_d)/f=200D(t-t_d)$ with $t_d=7$ days. From the cumulative value $J_m=J(t_m)$ at the time t_m of the maximum in $\dot{J}(t)$ we estimate k via **Eq. 14** upon assuming $J_m\approx J_0$. Similarly, $a_m=a(t_m)$ is estimated from $a_m=\dot{J}(t_m)/j_{\max}(k)$. These t_m, k, a_m are not the final values, but provide starting values which are then used in the minimization of the deviation between measured and modeled $J(t)$. The minimization is performed assuming the time-dependent $a(t)$ parameterized by **Eq. H1** involving parameters $t_a>0$, $t_b>0$, $a_0>0$, $q\in[0,1]$, $q<\eta\in[0,1]$ and $t_s>t_m$. While $\tau(t)$ is given by the integrated $a(t)$, we use three strategies to model $J(t)$: i) the numerical solution of the SIR model, ii) the

approximant $G(\tau)$ and $J(\tau) = 1 - e^{-G(\tau)}$ developed in part A, and iii) the approximant $J(\Delta)$ given by **Eq. 10** with $\Delta = \tau - \tau_m$ and τ_m specified by **Eq. 9**. Because the numerical solution (i) is extremely well approximated by (ii), and (ii) and (iii) compared to (i) not prone to numerical instabilities at small and large Δ, we present results only for method (iii), as they can be readily reproduced by a reader without Lambert's function at hand.

APPENDIX H: MODELING OF LOCKDOWN LIFTING

Similarly to the lockdown modeling a later lifting of the NPIs can be modeled by adopting the infection rate

$$a(t) = a_{\rm LD}(t)\,\Theta(t_s - t) + a_{\rm stop}(t)\,\Theta(t - t_s) \tag{H1}$$

where t_s denotes the stop time of the lockdown still represented by the infection rate **Eq. 15**, and where $a_{\rm LD}$ is given by **Eq. 15**. The infection rate after t_s is assumed to be

$$a_{\rm stop}(t) = a_0\left[q_s + (\eta - q_s)\tanh\frac{t - t_s}{t_b}\right] \simeq \begin{cases} q_s a_0 & \text{for } t = t_s \\ \eta a_0 & \text{for } t \gg t_s, \end{cases} \tag{H2}$$

with $q_s = a_{\rm LD}(t_s)/a_0$ the quarantine factor reached at the time t_s of lifting. Together with the reduced time given by **Eq. 16** we now find

$$\tau(t) = \tau_{\rm LD}(t)\,\Theta(t_s - t) + \tau_{\rm stop}(t)\,\Theta(t - t_s) \tag{H3}$$

and

$$\tau_{\rm stop}(t) = \tau_{\rm LD}(t_s) + q_s(t - t_s)a_0 + (\eta - q_s)a_0 t_b \ln\left[\cosh\left(\frac{t - t_s}{t_b}\right)\right] \tag{H4}$$

with $\tau_{\rm LD}(t)$ from **Eq. 16**. For the four basic types of **Figure 4** we demonstrate in **Figure 7** the effect of incomplete lifting.

frontiers in Physics

ORIGINAL RESEARCH
published: 21 January 2021
doi: 10.3389/fphy.2020.602722

Estimating Parameters of Two-Level Individual-Level Models of the COVID-19 Epidemic Using Ensemble Learning Classifiers

Zeyi Liu[1,2], *Rob Deardon*[2,3], *Yanghui Fu*[1], *Tahsin Ferdous*[2], *Tony Ware*[2] *and Qing Cheng*[1]*

[1]College of Systems Engineering, National University of Defense Technology, Changsha, China, [2]Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada, [3]Department of Production Animal Health, University of Calgary, Calgary, AB, Canada

The ongoing COVID-19 pandemic has led to a serious health crisis, and information obtained from disease transmission models fitted to observed data is needed to inform containment strategies. As the transmission of virus varies from city to city in different countries, we use a two-level individual-level model to analyze the spatiotemporal SARS-CoV-2 spread. However, inference procedures such as Bayesian Markov chain Monte Carlo, which is commonly used to estimate parameters of ILMs, are computationally expensive. In this study, we use trained ensemble learning classifiers to estimate the parameters of two-level ILMs and show that the fitted ILMs can successfully capture the virus transmission among Wuhan and 16 other cities in Hubei province, China.

Keywords: COVID-19 epidemic, individual-level model, SARS-CoV-2 transmission, spatiotemporal analysis, ensemble learning classifiers

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Qianqian Tong,
University of Connecticut, United States
Yanmi Wu,
Zhengzhou University, China

***Correspondence:**
Qing Cheng
sgggps@163.com

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *04 September 2020*
Accepted: *12 November 2020*
Published: *21 January 2021*

Citation:
Liu Z, Deardon R, Fu Y, Ferdous T, Ware T and Cheng Q (2021) Estimating Parameters of Two-Level Individual-Level Models of the COVID-19 Epidemic Using Ensemble Learning Classifiers. Front. Phys. 8:602722. doi: 10.3389/fphy.2020.602722

INTRODUCTION

The COVID-19 epidemic [1, 2] has caused the most serious threat to global health since the early 20th century; the exponential spread of the SARS-CoV-2 virus around the world has caused over 26 million confirmed cases and 860 thousand deaths worldwide as reported by the John Hopkins University COVID-19 web dashboard (https://coronavirus.jhu.edu/map.html) at the time of writing [3]. The spread of the SARS-CoV-2 virus, which causes COVID-19, has varied considerably in different areas, in part depending on the control different measures taken. Intensive testing, tracing, and isolation of infected cases have enabled control of transmission in some places, such as China and Singapore [4]. At the opposite extreme, many countries lack the testing and public health resources to take similar measures to control the COVID-19 epidemic, which can result in unhindered spread. Between these extremes, many countries have taken measures that facilitate "social distancing", such as closing schools and workplaces and limiting the size of gatherings. In order to analyze the dynamics of COVID-19 outbreak, we build an epidemic model based on the individual-level model (ILM) of Deardon et al (2010) [5] to catch the spread of SARS-CoV-2 virus within and among cities.

The individual-level model (ILM) framework enables us to express the probability of a susceptible individual being infected at a point in discrete time, as a function of their interactions with the surrounding infectious population, while also allowing the incorporation of the effect of individually varying risk factors. Here, we consider an extension of the Deardon et al (2010) framework of ILMs, to allow the probability of infection to depend upon two levels of transmission dynamics. The first is a

within-city (or region) level; the second is a between-city level. Infectious diseases are generally modeled through compartmental frameworks, and here we place our ILMs within the susceptible-infectious-removed (SIR) framework [6]. In the SIR framework, infected individuals become instantly infectious upon exposure, with no dormant or latent period. Since, in reality, infection is not observed instantaneously, and many infected individuals are not recorded at all in the data, we add an "observation model" which ties the epidemic generating model above to the observed data. This consists of a geometric distribution-based "delay model" and a "reporting model" which assumes that the probability of a true case being reported follows a Bernoulli distribution.

ILMs are intuitive and flexible due to being expressed in terms of individual interactions [7–9], but the cost of computation to parameterize them using observed data is often expensive, especially when dealing with a disease spreading in large populations. Traditional parameter estimation methods, such as Bayesian Markov chain Monte Carlo, have an associated high computation cost. Recent works by Nsoesie et al (2011) [10], Pokharel et al (2014) [11], and Augusta et al (2019) [12] have shown how to bypass the likelihood calculations by using machine learning classifiers to fit ILMs to data. In this work, we develop this approach to explore the use of ensemble learning classifiers to accurately and efficiently find the parameters for our two-level ILM, which incorporates a delay and reporting mechanism.

GENERATING MODEL

In this section, we present the two-level epidemic ILM [13] and observation model (delay model and reporting model) which ties the epidemic model to observed data. We denote the set of individuals who are susceptible, infectious, or removed at time t in city/region k as $S_{k,t}$, $I_{k,t}$, or $R_{k,t}$, respectively. Note, for given t, these sets are mutually exclusive, so individuals cannot be in multiple states, or multiple cities. Here, we assume time is discretized so that time point t, for $t = 1, 2, \ldots n$, represents a continuous time interval $[t, t + 1)$.

Two-Level Individual-Level Model

The number of newly infectious persons in city k at time point $t + 1$ is given by

$$\boldsymbol{I}_{\boldsymbol{k},\boldsymbol{t}+1}^{\boldsymbol{new}} = \boldsymbol{binomial}\left(\left|\boldsymbol{S}_{\boldsymbol{k},\boldsymbol{t}}\right|; \boldsymbol{P}_{\boldsymbol{k},\boldsymbol{t}}\right) \tag{1}$$

where $\left|S_{k,t}\right|$ is the number of susceptible individuals within city k at time t, and $P_{k,t}$ is the probability of each susceptible individual in the k^{th} city being infected at time t. Here, $P_{k,t}$ is given by

$$\boldsymbol{P}_{\boldsymbol{k},\boldsymbol{t}} = 1 - \boldsymbol{e}^{\left[-\boldsymbol{\alpha}_0\left|\boldsymbol{I}_{\boldsymbol{k},\boldsymbol{t}}\right| - \boldsymbol{\alpha}_1 \sum_{\boldsymbol{j} \in (1,2,\ldots,\boldsymbol{n})/\boldsymbol{k}} \boldsymbol{d}_{\boldsymbol{k},\boldsymbol{j}}^{-\boldsymbol{\beta}}\left|\boldsymbol{I}_{\boldsymbol{j},\boldsymbol{t}}\right|\right]} \tag{2}$$

where n is the number of cities in the population; $\left|I_{k,t}\right|$ is the number of infectious individuals within city k at time t; α is a parameter representing the risk of infection within cities; and α_1 and β are parameters representing the risk of infection between cities, with β capturing the decay rate of a power-law distance-based kernel, $d_{k,j}^{-\beta}$. Note, decreasing β will lead to a lower rate of decay in the infection kernel and thus more long-distance infections.

In the two-level ILM-SIR model, the transitions from susceptible to infectious and from infectious to recovered are treated as events of interest. In this work, the number of time points (days) between I and R is referred to as the infectious period, denoted by γ. The constant infectious period expresses the number of days over which an infectious individual is capable of transmitting the disease.

Observation Model

Here, we consider adding an observation model which ties epidemics generated by epidemic model to observed data.

Delay Model

Since there is a delay between infection and observation of that infection (reporting), we use a delay model to better represent reality. Specifically, given true infection times $\tau_i \in Z^+$ for each infected individual i, we let the potential observation time for individual i to be

$$\boldsymbol{\tau}_{\boldsymbol{i}}^{\boldsymbol{D}} = \boldsymbol{\tau}_{\boldsymbol{i}}^{\boldsymbol{I}} + \boldsymbol{z}_{\boldsymbol{i}}, \text{ where } \boldsymbol{z}_{\boldsymbol{i}} \sim \boldsymbol{Geometric}\left(\boldsymbol{P}^{\boldsymbol{D}}\right) \tag{3}$$

where P^D is the delay rate parameter. Note τ_i^D is a potential observation time, since case i may not be observed at all.

Reporting Model

The second component of the observation model, the "reporting model", accounts for asymptomatic, or otherwise unreported, cases of COVID-19. Here, we assume the probability of observing a case and it being recorded in the data (at time τ_i^D) follows a Bernoulli distribution, such that

$$\boldsymbol{\delta}_{\boldsymbol{i}} \sim \boldsymbol{Bernoulli}\left(\boldsymbol{P}^{\boldsymbol{R}}\right), \boldsymbol{\delta}_{\boldsymbol{i}} = \begin{cases} \mathbf{1}, & \text{infected } i \text{ is reported} \\ \mathbf{0}, & \text{infected } i \text{ is unreported} \end{cases} \tag{4}$$

where P^R is the reporting rate parameter.

ENSEMBLE LEARNING CLASSIFIERS

In supervised learning algorithms, the goal is to learn a stable classification (or regression) model that performs well across a wide range of data scenarios. Often, however, this is a difficult goal to achieve. Ensemble learning is the process by which multiple models are strategically "learned" and combined to solve a computational intelligence problem. Ensemble learning is primarily used to provide for an improved performance over any single model, or to reduce the likelihood of the selection of a poor single model. Bagging, boosting, and stacking are common ensemble learning algorithms. Note, here, we are concerned with classification rather than regression problems.

Bagging

Bagging, which stands for bootstrap aggregating, is one of the earliest, most intuitive, and perhaps the simplest ensemble based algorithms, with a surprisingly good performance [14]. A diversity of classifiers in bagging is obtained by using

bootstrapped replicas of the training data. That is, different training data subsets are randomly drawn—with replacement—from the entire training dataset. Each training data subset is used to train a different classifier of the same type. Individual classifiers are then combined by taking a simple majority vote of their decisions. For any given instance, the class chosen by the greatest number of classifiers is the ensemble decision.

The random forest is a bagging method for trees, later extended to incorporate random selection of features to help control variance [15, 16].

Boosting

Similar to bagging, boosting also creates an ensemble of classifiers by resampling the data, which are then combined by majority voting. However, in boosting, resampling is strategically geared to provide the most informative training data for each consecutive classifier.

The gradient boosted decision tree (GBDT) method [17–19] uses decision trees as the base learner and sums the predictions of a series of trees. At each step, a new decision tree is trained to fit the residuals between ground truth and the current prediction. Many improvements have since been proposed. XGBoost [20] uses a second-order gradient to guide the boosting process and improve the accuracy. LightGBM [21] aggregates gradient information in histograms to significantly improve the training efficiency; it splits the tree leaf-wise with the best fit, whereas other boosting algorithms split the tree depth-wise or level-wise rather than leaf-wise. AdaBoost, short for Adaptive Boosting, can be used in conjunction with many other types of learning algorithms to improve performance; the output of the other learning algorithms (weak learners) is combined into a weighted sum that represents the final output of the boosted classifier. Finally, CatBoost [22] proposed a novel strategy to deal with categorical features.

Stacking

Stacking, sometimes called stacked generalization, is also an ensemble learning method that combines multiple classification (or regression) models via a metaclassifier or a metaregressor. The base level models are trained based on a complete training set; then the metamodel is trained on the outputs of the base level model as features. Stacking involves training a learning algorithm to combine the predictions of several other learning algorithms. Stacking typically yields performance better than any single one of the trained models [23]. It has been successfully used on both supervised learning tasks (regression, classification, and distance learning) and unsupervised learning (density estimation).

EXPERIMENT

Typically, ILMs are fitted to data using computationally intensive techniques such as Bayesian Markov chain Monte Carlo methods. In order to avoid this computational expense, in this study we use the method of Pokharel et al (2014) to fit our models to data.

TABLE 1 | The parameters of ILM generating model.

	α_0	α_1	β	γ	p^D	p^R
1	0.001	0.8	3	7	0.5	0.6
2	0.005	0.1	2	3	0.3	0.8
3	0.005	0.2	1	6	0.8	0.4
4	0.01	0.5	2	9	0.7	0.5
5	0.05	0.9	1	8	0.8	0.4
6	0.05	0.1	3	6	0.8	0.5
7	0.1	0.6	2	4	0.6	0.8
8	0.2	0.3	1	6	0.7	0.9
9	0.2	0.1	3	4	0.6	0.9
10	0.3	0.5	1	6	0.5	0.5
11	0.4	0.7	3	6	0.3	0.5
12	0.5	0.1	2	3	0.8	0.3

Broadly this method involves defining a set of candidate generating models, each with different parameter values. Then, epidemics are repeatedly generated from the candidate models and summarized. Here, we summarize the epidemics using the number of observed cases per day, what we term the "epidemic curve". These epidemic curve summary statistics form the training set used to build a classifier mapping the epidemic curve (input features) to the generating model (class). The classifier can then be used to identify the most likely generating model for future observed summaries of epidemic data sets; several ensemble learning classifiers such as random forest, XGBoost, LightGBM, AdaBoost, CatBoost, and stacking are used to seek the best fitted parameter for the two-level ILM model. Here, we verify the accuracy of these ensemble learning classifiers by testing their performance on data simulated from the two-level ILM model. In *Real Data Case Study: COVID-19 in Hubei Province, China*, we will use such classifiers to estimate the parameters that give the model of best fit when applied to COVID-19 data from Hubei province, China.

Simulation Study

We now recap the parameters we need to identify: α_0, β, and α_1 are the parameters of the two-level epidemic model in **Eq. 2**; t_{stop} is the interval of time from the initial infection (unknown) to the day when the epidemic stopped by external intervention (lockdown) in Hubei province; γ is the infectious period, assumed to be constant for all individuals; p^D is the rate parameter of the delay model; and p^R is the rate parameter of the reporting model. Here we assume that when day $t = t_{stop}$ is reached, the rate of new infections becomes negligible, and so $\alpha_0 = \alpha_1 = 0$. Thus, newly observed cases after day $t = t_{stop}$ result from earlier infections becoming observed through the delay model.

In the simulation experiment, we suppose there are total 100 cities, where (x,y) coordinates are simulated uniformly across a 100×100 unit. Further, the population of each city is set at 1,000. Each candidate two-level ILM is used to generate 100 epidemics, summarized as epidemic curves. To initialize the SIR model, we set the value of I_0 to 50 in the one city (chosen randomly), where the disease originates. Here, a maximum of 12 candidate epidemic generating models are considered in each analysis with parameters shown in **Table 1**. Sets of generated epidemic

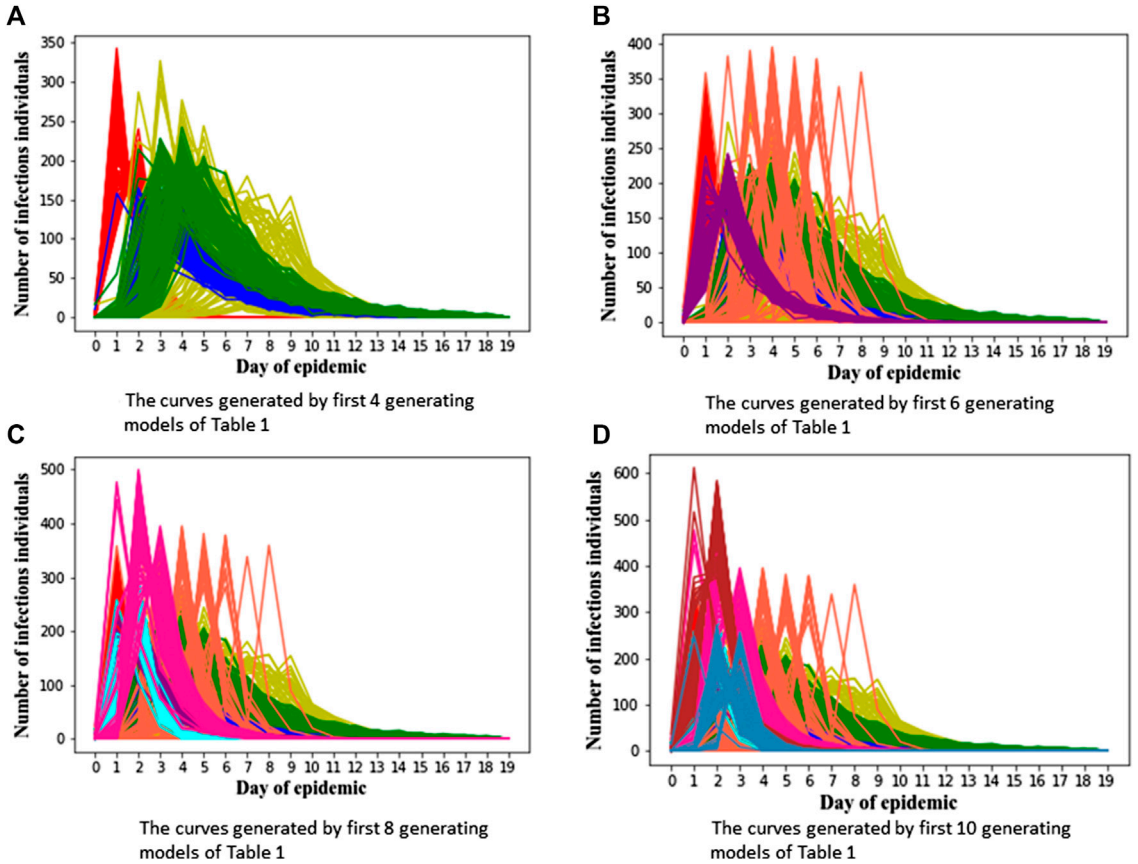


FIGURE 1 | Epidemic curves generated under various training sets.

TABLE 2 | The accuracy of different classifiers for training sets of 70 curves per generating model.

Methods sets	Random forest	XGBoost	LightGBM	AdaBoost	CatBoost	Stacking
4	0.98	0.98	0.98	0.98	0.98	0.99
6	0.98	0.96	0.97	0.99	0.99	0.99
8	0.94	0.91	0.94	0.94	0.94	0.93
10	0.95	0.94	0.95	0.96	0.96	0.95
12	0.97	0.97	0.96	0.98	0.98	0.98

curves are shown in **Figure 1**, with different colors denoting different epidemic generating models.

In order to verify the performance of classifiers, we have carried out five classification tasks, each consisting of different numbers of epidemic generating models. These tasks consisted of the first 4, 6, 8, 10, and 12 models of **Table 1**, respectively. The generated data is randomly divided into a training set and test set, with 70% of the data for the training set, and the rest for the test set. The results of classification are shown in **Tables 2, 3**.

We can see **Figure 1** has epidemic curves that overlap substantially. However, all classifiers achieve quite high accuracy. We also repeated the simulation study using 200 curves per epidemic generating model (140 training; 60 test). The results are shown in **Table 3**, and we can see that accuracy increases when we have larger training sets.

The details of super-parameters of classifiers used to get the best classification score are as follows. The parameters in AdaBoost classifier are as follows: the max depth is 6, the number of estimators is 1,000, and the learning rate is 0.008; the parameters in CatBoost classifier are as follows: the depth is 6, the iteration is 1,200, and the learning rate is 0.05; the parameters in LightGBM classifier are as follows: the max depth is 8, the number of estimators is 150, and the learning rate is 0.05; the parameters in random forest classifier are as follows: the max depth is 8, the number of estimators is 300; the parameters in

TABLE 3 | The accuracy of different classifiers for training sets of 140 curves per generating model.

Methods sets	Random forest	XGBoost	LightGBM	AdaBoost	CatBoost	Stacking
4	1.00	1.00	0.99	1.00	1.00	1.00
6	0.98	0.98	0.98	0.98	0.98	0.98
8	0.97	0.95	0.96	0.96	0.96	0.96
10	0.97	0.96	0.96	0.97	0.97	0.97
12	0.96	0.96	0.96	0.97	0.97	0.97

TABLE 4 | Adjusted population of 17 cities in Hubei.

City	Population in central urban area (10,000)	Central urban area (km^2)	Density ratio	Adjusted population (10,000)
Wuhan	918	917.5	1	918
Huangshi	85.99	260.36	0.3303	28.4
Shiyan	73.44	28.09	2.6145	192.01
Yichang	93.13	35.11	2.6525	247.03
Xiangyang	132.57	78.11	1.6972	225
Ezhou	45.97	208.95	0.22	10.11
Jingmen	57.17	30.83	1.8544	106.02
Xiaogan	57.38	109.02	0.5263	30.2
Jingzhou	87.05	52.62	1.6543	144.01
Huanggang	33.8	102	0.3314	11.2
Xianning	41.32	105.91	0.3901	16.12
Suizhou	50.25	26.58	1.8905	95
Xiantao	40.5	90.12	0.4494	18.2
Qianjiang	43.74	103.98	0.4207	18.4
Tianmen	29.92	49.19	0.6083	18.2
Enshi	28.39	20.28	1.3999	39.74
Shennongjia	3.6	1.62	2.2222	8

XGBoost classifier are as follows: the max depth is 10, the number of estimators is 250, and the learning rate is 0.1.

Real Data Case Study: COVID-19 in Hubei Province, China

We now consider training a classifier to find the parameters of best fit for the two-level ILM for COVID-19 data from China. As the first reported COVID-19 cases happened in the city of Wuhan, we choose the Hubei province in China as the example in this study. There are in total 17 cities in Hubei province, and we utilize information on the population of each city and the distance between cities for the two-level epidemic ILM.

For the distance between each city, we use the center of each city as its coordinate point, and then we choose to use the shortest road traffic distance based on Baidu map (https://map.baidu.com). Rather than using the total population size to calculate the terms $|S_{k,t}|$ and $|I_{k,t}|$, we consider using the population in the central urban area (where most citizens live). Further, we scaled the population of each city based on the ratio of population density in central urban area relative to that in Wuhan. Thus, the population measure of each of the other 16 cities in Hubei will be greater or less than one depending on whether their population density in central urban area is greater or less than that of Wuhan. These measures of population density are shown in **Table 4** (adjusted population column).

The reported case data had some anomalies, and so some preprocessing was carried out. For example, the number of new cases in Wuhan on February 12, 2020, was recorded as more than ten thousand, which is much larger than on other days. Also, there were two other cities which had a two-day spike of an excessively large magnitude. We believe these spikes represent retrospectively found cases, which should have been recorded as cases on earlier days, "dumped into the data" on those "spike" days. Further, some values in the reported data were negative because the health agencies subtracted retrospectively discovered false positives from the date on which the false positives were discovered, rather than the day on which they were initially recorded. For the large one-day spike, we took the average of values of three days before and three days after the spike and used this average value to replace the spike case count. Then we "scattered" the excess cases onto past days, at a rate proportional to the previously observed cases recorded on each day. The two-day spikes were replaced in a similar manner, the difference being that in this case two spikes were replaced by the average value. For negative values, we simply replaced them with zero. Our preprocessed data can be found in the **Supplementary Tables S1, S2**.

We build our classifier in the following way. To begin, we consider epidemic generating models that are relatively spaced out in the parameter space. We initially set the range of parameter α_0 from 5×10^{-8} to 5×10^{-6}, the range of α_1 from 0.0005 to 0.05, the range of β from 1 to 9, the range of t_{stop} from 40 to 100, the

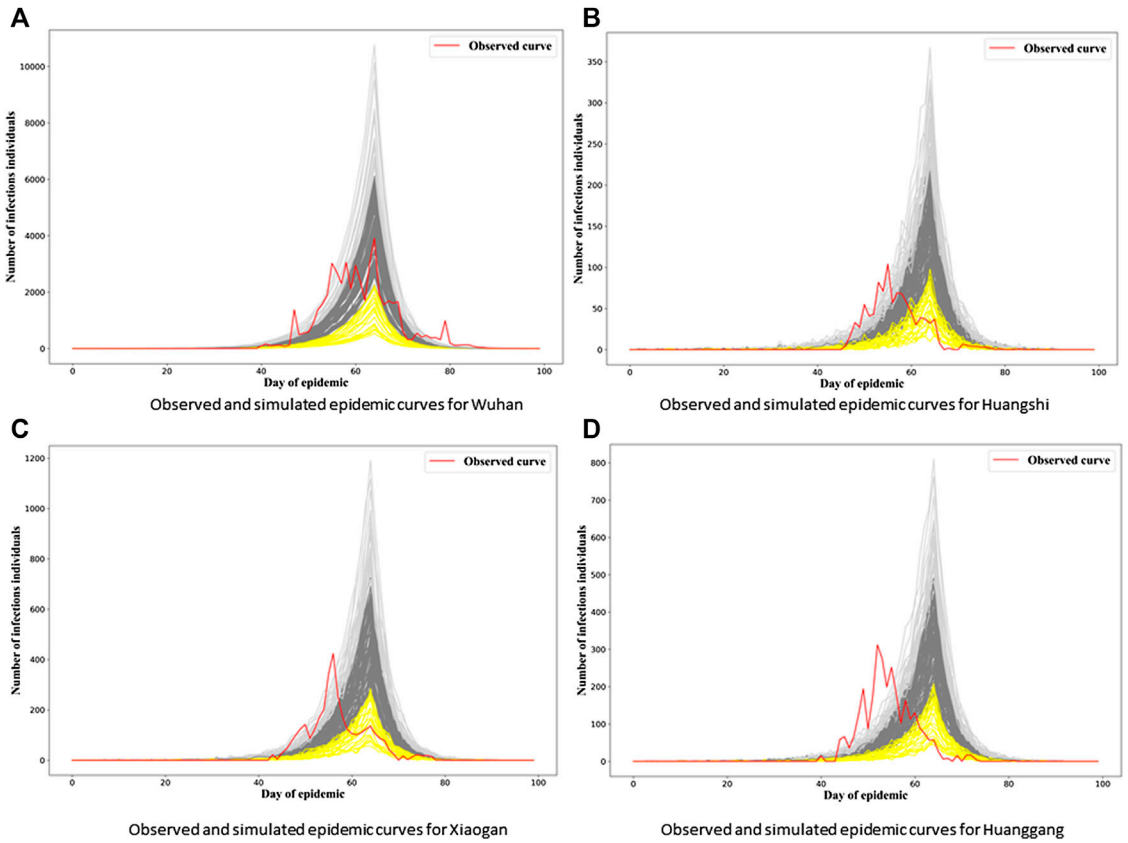


FIGURE 2 | Observed and simulated epidemic curves under final model for four cities in Hubei. Real data shown in red; light gray and yellow curves represent 15% strongest and weakest epidemics simulated from the final model, respectively; remaining dark gray curves represent remaining 70% of simulated epidemics.

range of γ from 10 to 30, the range of p^R from 0.1 to 0.7, and the range of p^D from 0.4 to 1. Specifically, for α_0, we consider values of 5×10^{-8}, 5×10^{-7}, and 5×10^{-6}; for α_1, we set the values to be 0.0005, 0.005, and 0.05. We set the step size of β to be 4 (i.e., we considered values 1,5, and 9), the step size of t_{stop} to be 30, the step size of γ to be 10, and the step size of p^D and p^R to be 0.3. Thus, we have 3^7 epidemic generating models. We build our classifier based upon epidemics generated by each model and then use the Adaboost classifier to identify which of the candidate models is the most likely generating model for the real data.

After the first round of classification, the most likely generating model is found to be one with parameters: $\alpha_0 = 5 \times 10^{-8}$, $\alpha_1 = 0.005$, $\beta = 1$, $t_{stop} = 70$, $\gamma = 20$, $p^D = 0.4$, and $p^R = 0.7$. Next, a less spaced-out set of parameter values are considered to define the generating models. We set the range of α_0 from 2×10^{-8} to 8×10^{-8}, the range of α_1 from 0.002 to 0.008, the range of β from 0.5 to 3.5, the range of t_{stop} from 55 to 85, the range of γ from 15 to 25, the range of p^D from 0.2 to 0.6, and the range of p^R from 0.5 to 0.9. We set the step size of α_0 to be 3×10^{-8}, the step size of α_1 to be 0.003, the step size of β to be 1.5, the step size of t_{stop} to be 15, the step size of γ to be 5, the step size of p^D to be 0.2, and the step size of p^R to be 0.2. The parameters of the most likely generating model we get in this second round are $\alpha_0 = 2 \times 10^{-8}$, $\alpha_1 = 0.002$, $\beta = 2$, $t_{stop} = 70$, $\gamma = 15$, $p^D = 0.4$, and $p^R = 0.7$. As the approach continues, parameters of candidate generating models obviously get closer and closer to each other, with generated curves overlapping more and more and classification becoming less well defined. After six rounds of classification, we converge on estimates of the parameters of the most likely epidemic generating model for the real data: $\alpha_0 = 2.1 \times 10^{-8}$, $\alpha_1 = 0.002$, $\beta = 2.5$, $t_{stop} = 64$, $\gamma = 16$, $p^D = 0.25$, and $p^R = 0.85$.

Recall that t_{stop} denotes the interval of time from the initial infection to the day when the epidemic was severely curtailed due to lockdown. Given that the report of first suspected cases was in early December 2019, the estimate of $t_{stop} = 64$ matches quite closely to real circumstances (note, Wuhan had severe restrictions on Jan 23, 2020, and the other cities in Hubei were similarly "closed down" within two or three days).

Figures 2A to **Figure 2D** show generated epidemic curves from our chosen model for four cities in Hubei; one is Wuhan and the other three are chosen arbitrarily. Given the stochasticity inherent in our model, and the complex population structure the epidemic is being transmitted/simulated through, we get a lot of variability in the epidemic curves generated. We can see that the "fitted" two-level ILM model captures the dynamics of the SARS-

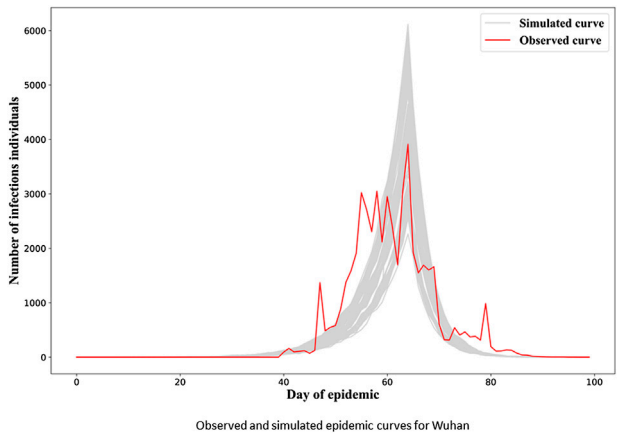


FIGURE 3 | Observed and simulated epidemic curves under final model for Wuhan. Real data shown in red; dark gray curves represent 70% of curves simulated from the final model closest to the true curve.

CoV-2 spread reasonably well, especially in Wuhan. However, there is a tendency for the epidemic peak under our model to be overestimated and arrive a little late in other cities.

Of course, our model is relatively simple, assuming homogeneity between cities in terms of both the transmission process (after accounting for population dynamics) and the observation model. Our suspicion is that the observation model may be a major issue here. For example, note that the delay mechanism is mimicking both a biological process (the incubation period of the disease) and a bureaucratic process (diagnosis and processing and publication of numbers of cases per day). It therefore seems perfectly plausible that the delay between infection and reporting of cases could differ between different jurisdictions, in this case, cities.

Since the epidemic observed in Wuhan was by far the most substantial, it makes sense that the Wuhan data would be driving the inference process for the final model. It therefore makes sense that the model ends up parameterized in such a way that the data in Wuhan are mimicked well by the fitted model, and the other cities less so. Also, since Wuhan was the first city infected, it also makes sense that the delay between infection and reporting would be larger for that city and others, since it was operating with less information than other cities which had to deal with their infections a little later on.

In **Figure 3**, we see the "gray epidemic curves" (the 70% closest to the observed epidemic) for Wuhan. **Figures 4**, **5** show these "gray curves" for other cities. We can see that in Huangshi in

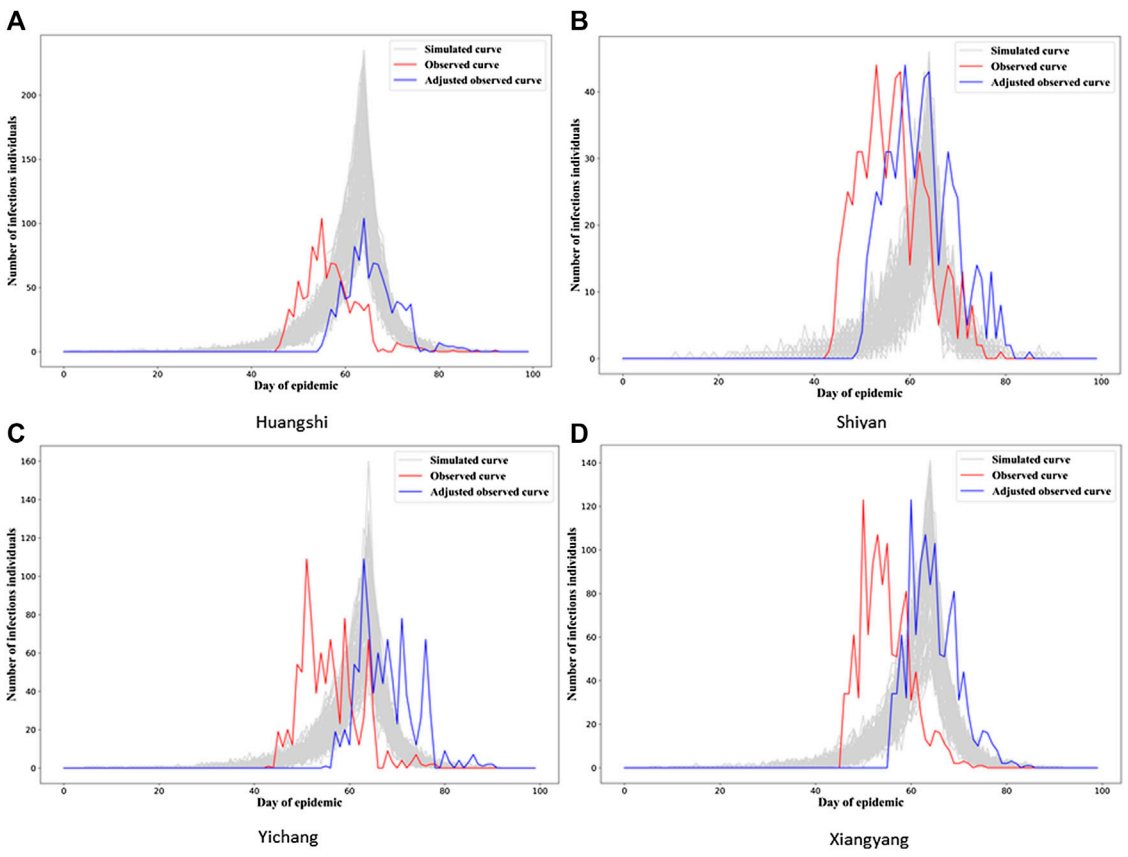


FIGURE 4 | Observed and simulated epidemic curves under final model for four cities in Hubei (other 12 cities in **Figures 5**, **6**; for Wuhan, see **Figure 3**). Real data shown in red; shifted real data accounting for potential differences in delays to reporting in different cities shown in blue; dark gray curves represent 70% of curves simulated from the final model closest to the true curve.

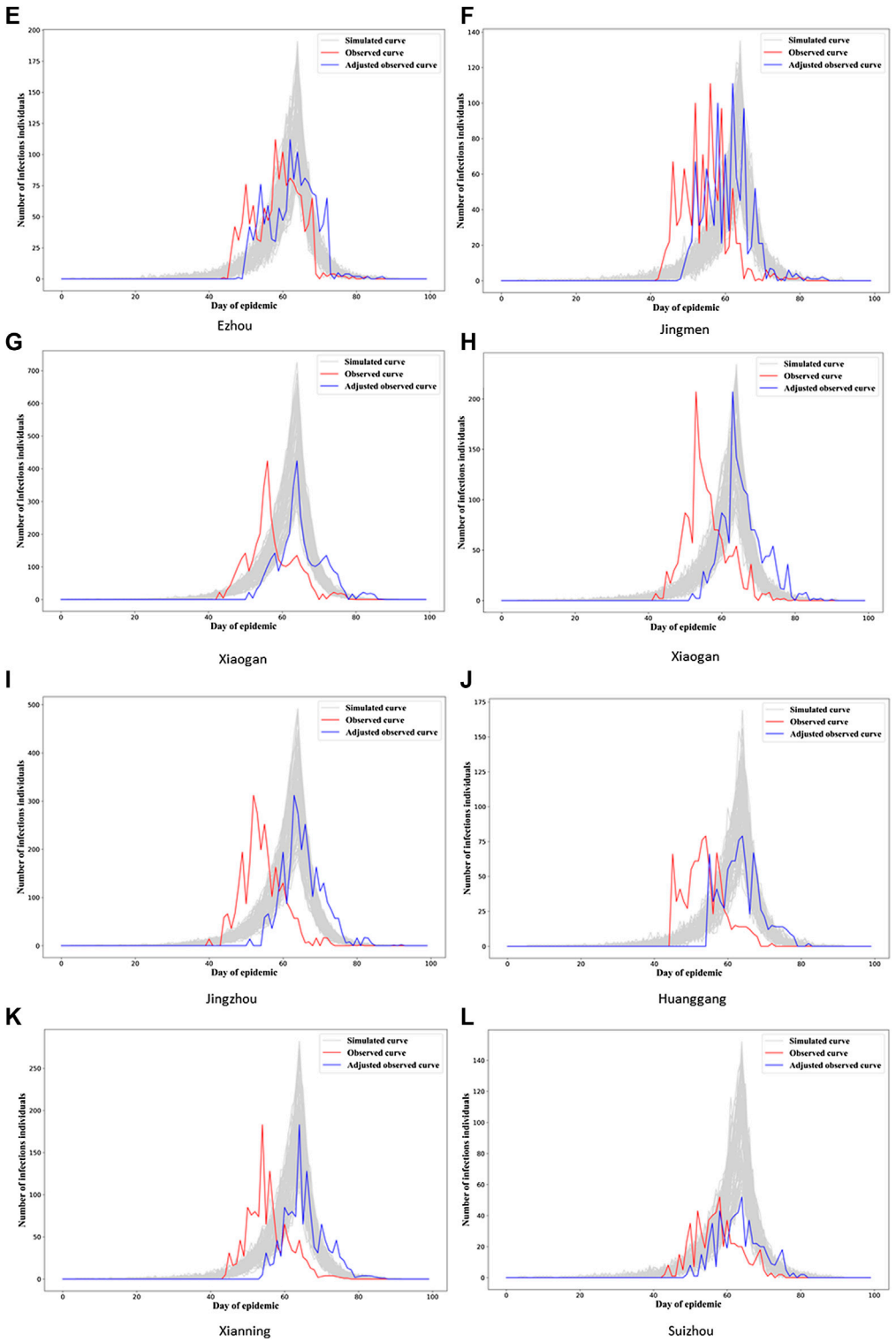


FIGURE 5 | Observed and simulated epidemic curves under final model for eight cities in Hubei (other four cities in **Figure 6**; for Wuhan, see **Figure 3**). Real data shown in red; shifted real data accounting for potential differences in delays to reporting in different cities shown in blue; dark gray curves represent 70% of curves simulated from the final model closest to the true curve.

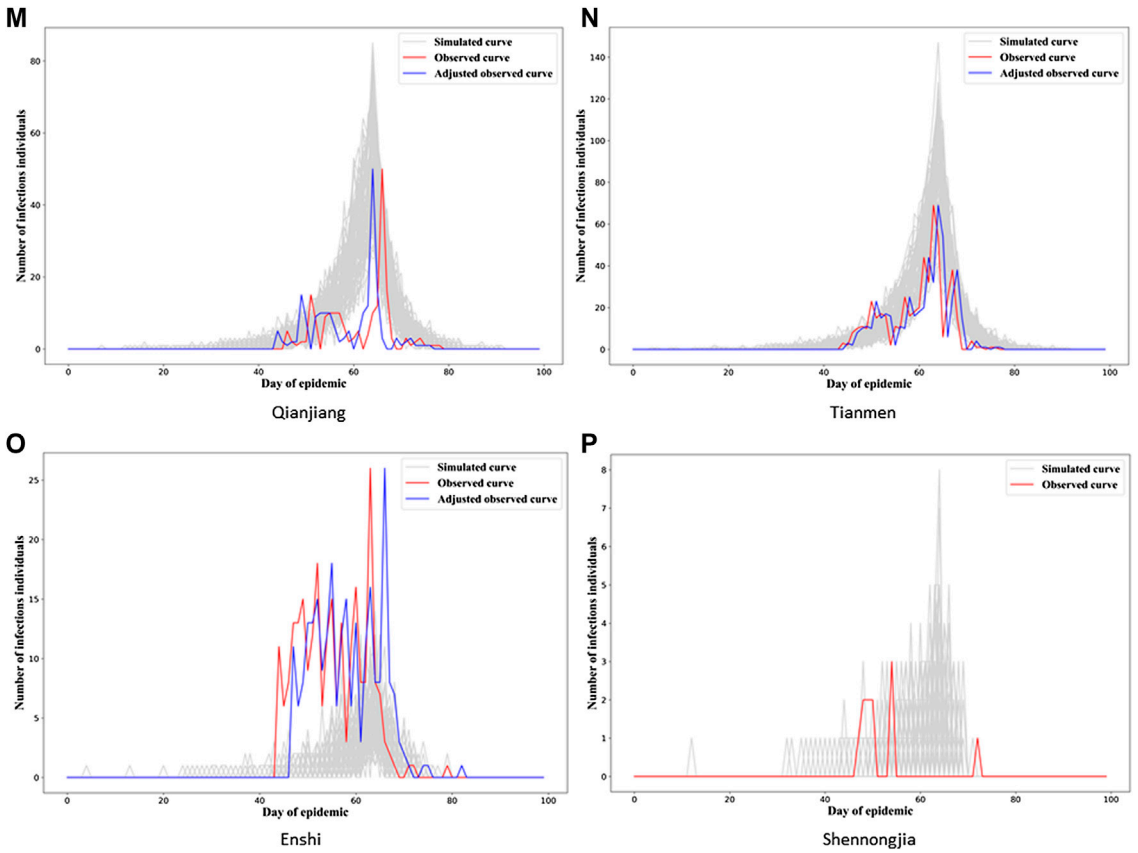


FIGURE 6 | Observed and simulated epidemic curves under final model for four cities in Hubei (for Wuhan, see **Figure 3**). Real data shown in red; shifted real data accounting for potential differences in delays to reporting in different cities shown in blue; dark gray curves represent 70% of curves simulated from the final model closest to the true curve.

Figure 4A, if we shift the epidemic curve based on the observed cases (shown in red) by a few days (shown in blue), the epidemic curves produced by the fitted model much better match the observed curve. Throughout **Figures 4–6**, we see a similar pattern for the other cities in the province.

This would imply that the next step we might want to take in refining our model is to allow for heterogeneity in the observation model parameters between cities, probably starting with the delay mechanism rate parameter, p^D.

Overall, these results show that the two-level ILM model fitted using an ensemble classifier can reasonably well reflect the spread of SARS-CoV-2 among cities; this is especially true for Wuhan, and we can see the potential for better capturing the dynamics of COVID-19 transmission among other cities through further model development.

CONCLUSION

We construct a statistical inference framework that allows us to fit a two-level individual-level epidemic model to data. We use several ensemble learning classifiers to successfully estimate model parameters, avoiding the high computation costs exhibited by traditional methods of inference. The simulation study shows good performance of the fitted model, and we successfully fit our model to real data on COVID-19 transmission among 17 cities in Hubei province, China.

FUTURE WORK

In this study, we focus on analyzing the transmission of SARS-CoV-2 among 17 cities in Hubei province, China. It would certainly be of interest to see how our model performs on COVID-19 data from different countries and indeed data on other diseases. Here, also we choose to model disease transmission within an SIR framework. In many scenarios, more complex compartmental frameworks such as SEIR or SIRS would be more appropriate. It would therefore be desirable to test if classification-based inference for our model works similarly well, as well as considering such frameworks for COVID-19 transmission itself. An additional limiting factor was that we made the simplifying assumption that the infectious period was the same for all individuals. This is obviously not true in practice, and so it would be desirable to relax this assumption.

There are also likely other risk factors, in addition to population size/density, we might want to include within models of COVID-19 spread to improve them. These could include demographic descriptors of age distribution within cities, knowledge about traffic flows between cities, and socioeconomic covariates.

As discussed previously, we might well wish to allow the parameters of the observation model to vary between cities, to allow for differences in recording and reporting procedures. In addition, we may want to think about allowing these parameters to change over time in line with jurisdictional government policy.

Finally, we will explore more classification methods, such as deep learning methods, to attempt to build even more accurate classifiers, especially for more complex models and populations structures.

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

ZL and RD designed and managed the project, ZL, YF, TF, and RD conceived and designed the experiments, ZL, YF, and QC analyzed the data, ZL, RD, TW, and QC wrote the manuscript.

FUNDING

This study is supported by the National Natural Science Foundation of China (82041020, 91846301), Natural Science Foundation of Hunan Province (2020JJ5678), and China Postdoctoral Science Foundation (2019M653922). ZL is supported by the Graduate Research and Innovation Projects of Hunan Province (CX20200069).

ACKNOWLEDGMENTS

ZL is supported by the CSC Scholarship offered by the China Scholarship Council.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.602722/full#supplementary-material.

REFERENCES

1. Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. *N Engl J Med* (2020) 382:1268–9. doi:10.1056/NEJMe2002387
2. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health* (2020) 25(3):278–80.
3. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* (2020) 20(5):533–4. doi:10.1016/S1473-3099(20)30120-1
4. Cheng Q, Liu Z, Cheng G, Huang J. Heterogeneity and effectiveness analysis of COVID-19 prevention and control in major cities in China through time-varying reproduction numbers estimation (2020). Research Square [Preprint]. Available at: https://www.researchsquare.com/article/rs-20987/v1 (Accessed April 2, 2020).
5. Deardon R, Brooks SP, Grenfell BT, Keeling MJ, Tildesley MJ, Savill NJ, et al. Inference for individual-level models of infectious diseases in large populations. *Stat Sin* (2010) 20(1):239–61.
6. Anderson RM, Anderson B, May RM. *Infectious diseases of humans: dynamics and control*. Oxford University Press (1991) 122. p.
7. Gibson GJ. Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *J Roy Stat Soc: Series C (Applied Statistics)* (1997) 46(2):215–33.
8. Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, et al. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* (2001) 294(5543):813–7. doi:10.1126/science.1065973
9. Neal PJ, Roberts GO. Statistical inference and model selection for the 1861 Hagelloch measles epidemic. *Biostatistics* (2004) 5(2):249–61. doi:10.1093/biostatistics/5.2.249
10. Nsoesie EO, Beckman R, Marathe M, Lewis B. Prediction of an epidemic curve: a supervised classification approach. *Stat Commun Infect Dis* (2011) 3(1):5. doi:10.2202/1948-4690.1038
11. Pokharel G, Deardon R. Supervised learning and prediction of spatial epidemics. *Spat Spatiotemporal Epidemiol* (2014) 11:59–77. doi:10.1016/j.sste.2014.08.003
12. Augusta C, Deardon R, Taylor G. Deep learning for supervised classification of spatial epidemics. *Spat Spatiotemporal Epidemiol* (2019) 29:187–98. doi:10.1016/j.sste.2018.08.002
13. Ferdous T. *On the effect of ignoring within-unit infectious disease dynamics when modelling spatial transmission*. [Master's thesis]. Calgary (AB): University of Calgary (2019).
14. Breiman L. Bagging predictors. *Mach Learn* (1996) 24(2):123–40.
15. Ho TK. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition; 1995 August 14–16; Syndey, Australia. IEEE (1995). p. 278–82.
16. Amit Y, Geman D. Shape quantization and recognition with randomized trees. *Neural Comput* (1997) 9(7):1545–88.
17. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* (2001) 1189–232.
18. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* (2002) 38(4):367–78.
19. Li P. Robust logitboost and adaptive base class (abc) logitboost. arXiv [Preprint] (2012). Available at: https://arxiv.org/abs/1203.3491 (Accessed March 15, 2012).
20. 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; San Francisco, CA; August 13–17, 2016 (2016). p. 785–94.
21. 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). p. 3146–54.
22. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* (2018). p. 6638–48.
23. Wolpert DH. Stacked generalization. *Neural Network* (1992) 5(2):241–59. doi:10.1016/S0893-6080(05)80023-1

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 25 January 2021
doi: 10.3389/fphy.2020.602075

The Connectedness of the Coronavirus Disease Pandemic in the World: A Study Based on Complex Network Analysis

Sha Zhu[1], *Meng Kou*[2], *Fujun Lai*[3]*, *Qingxiang Feng*[4]* *and Guorong Du*[5]

[1]School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, China, [2]School of Aerospace Engineering,Beijing Institute of Astronautical System Engineering, Beijing, China, [3]School of Finance, Yunnan University of Finance and Economics, Kunming, China, [4]Institute of Guangdong Hong Kong and Macao Development Studies, Sun Yat-sen University, Guangzhou, China, [5]Chinese Academy of Finance and Development,Central University of Finance and Economics, Beijing, China

Since the coronavirus disease (COVID-19) pandemic started at the beginning of 2020, it has seriously affected various countries' economic and social development and accelerated the economic recession worldwide. Therefore, the connectedness of the global COVID-19 network across countries is studied in this article. Based on COVID-19 correlations in 122 countries, we construct a complex network of COVID-19 from January 19, 2020, to August 15, 2020. We then deconstruct the overall global network connectedness and analyze the connectedness characteristics. Moreover, we empirically investigate the network connectedness influencing factors by using various countries' macroeconomic and social data. We find that the global COVID-19 pandemic network has some prominent complex network properties, such as low path length, high clustering, and good community structure. Furthermore, population density, economic size, trade, government spending, and quality of medical treatment are significant macrofactors affecting COVID-19 connectedness in different countries.

Keywords: COVID-19, pandemic, connectedness, complex network, economic and social factors

OPEN ACCESS

Edited by:
Hui-Jia Li,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Ju Xiang,
Changsha Medical University, China
Chengyi Xia,
Tianjin University of Technology, China

***Correspondence:**
Fujun Lai
lfjlfj999@163.com
Qingxiang Feng
fengqx8@mail.sysu.edu.cn

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *02 September 2020*
Accepted: *07 December 2020*
Published: *25 January 2021*

Citation:
Zhu S, Kou M, Lai F, Feng Q and Du G (2021) The Connectedness of the Coronavirus Disease Pandemic in the World: A Study Based on Complex Network Analysis. Front. Phys. 8:602075. doi: 10.3389/fphy.2020.602075

INTRODUCTION

The COVID-19 global pandemic has changed the world remarkably. In just a few months, from limited local transmission in several countries, COVID-19 evolved into a multicountry spread and raged in more than 100 countries and regions across five continents. The COVID-19 pandemic has disrupted and changed the lifestyle of millions of people around the world and has had a profound impact on international relations and the economy. At present, increasing attention is being paid to the prevention and control of the COVID-19 pandemic. The pandemic is well under control in some countries, such as China and Germany, but most countries, including the United States, still do not have effective ways to control COVID-19. The fight against the COVID-19 pandemic has become a top priority for governments in many countries and regions, and the public health sector has paid unprecedented attention to protecting people's lives.

The rapid spread of the COVID-19 pandemic is more than just a medical problem. It is also linked to an infection network among countries worldwide. The complex network science of infectious diseases can reveal necessary information about disease transmission behavior, as many studies have shown. First, the research has revealed the typical characteristics of disease transmission behavior, such as the critical transmission threshold. The virus can spread in the network only when the transmission probability is more significant than a specific value called the propagation threshold of

the network (Moore and Newman [1]). Moreover, nonrandom networks, such as small-world and community structure networks, are critical to disease transmission. Many real networks have small-world characteristics in that they tend to have small average distances and large clustering coefficients. In such a network, diseases more easily infect others. In addition, scholars [2] have found that a complex network with a community structure can promote or effectively inhibit the spread of diseases.

Furthermore, the key nodes in a network have a profound impact on disease transmission. Outbreaks of infectious diseases often spread rapidly from a small number of nodes to the entire network. Scientists propose indicators such as page rank, node degree, and k-shell to evaluate the network science characteristic [3]. In disease research, the information of these key nodes can reveal the path of disease transmission. Finally, network science information can be formulated to control the spread of diseases and reduce harmful transmission processes. In recent research, Wang et al. [4] use a two-layer network model method to find that people who received positive prevention information had a lower risk of disease infection than those who received negative prevention information.

Therefore, this study establishes a complex network of COVID-19 infections worldwide and investigates the influencing factors of the pandemic's connectedness across countries. We integrate three methods, i.e., complex networks, visualization principles, and regression models, to examine the global infection path of the COVID-19 pandemic and explore possible causes for the rapid COVID-19 spread. First, we calculate the correlation of COVID-19 infection growth rates between each pair of countries and obtain a global correlation matrix for COVID-19 infections using the sample from January 19, 2020, to August 15, 2020. Second, we establish a global infection complex network using a global correlation matrix to outline global COVID-19 pandemic relations and study the global infection complex network properties. Third, based on the properties of our complex network, we further investigate how economic and social factors (such as economic size, population density, international trade, government expenditure, inflation, and medical level) affect the network connectedness of the global COVID-19 infection complex network, shedding light on the prevention and control of the COVID-19 pandemic for policymakers.

The main conclusions are as follows. First, the complex network of the global COVID-19 pandemic has some prominent complex network properties, such as small-world network and community structure characteristics. In our network, there are a total of 958 edges in 122 countries around the world. The average path length between any two countries is approximately 3, the clustering coefficient is 0.58, and the network can be divided into several modularity classes. Second, in terms of intercountry connectedness, we further examine important topological characteristics in our network, e.g., degree, triangles, centrality, modularity class, hub, and authority. We use visual figures to demonstrate these features. The deeper the color is, the greater the connectedness in the visual figure is. This suggests that our network reflects the actual situation of the COVID-19 pandemic. Moreover, based on similarity, we divide the complex network into ten modularity classes. On the one hand, the community classification shows that the United States, Russia, and Colombia, which are suffering severely from the COVID-19 pandemic, are in one community, and China and Japan, which have better epidemic control, are in another community. On the other hand, the community classification is closely related to each country's geographical location. Finally, the analysis of influencing factors reveals that countries with considerable economic size, high population density, and high government expenditure have a higher network connectedness value. In contrast, countries with adequate import capacity and better medical levels can reduce the connectedness. In short, these empirical results can clearly explain the reasons for the rapid global spread of the COVID-19 pandemic.

The contributions of this study are as follows. First, the global COVID-19 pandemic network is constructed based on the number of infected people, and the topological information and connectedness characteristics of the network are obtained, to some extent explaining the spread of the COVID-19 pandemic from the perspective of a complex network. The global COVID-19 pandemic network is a complex network with the characteristics of short average distance and large clustering coefficient and has a good community structure division. Second, integrating the visualization method into the global COVID-19 pandemic network provides visual results of network connectedness, helping us determine the potentially dangerous countries. Finally, we identify economic and social factors affecting COVID-19 network connectedness, including population density, GDP, exports, government spending, and life expectancy. To reduce network connectedness, the government should implement measures to reduce population density, increase medical equipment imports, improve the medical treatment level, and effectively allocate government expenditures.

LITERATURE REVIEW

Vespignani, a professor who is famous for modeling the global spread of epidemics, proposes that there are inherent limitations to predictions in complex sociotechnical systems. However, mathematical and computational models have successfully forecast the size of epidemics and have been used to communicate the risks of uncurbed infectious disease outbreaks in recent years [5]. Since the beginning of the COVID-19 outbreak, many mathematical models have been proposed to describe the pandemic's spread. Some of them forecast the future dynamics of COVID-19. Perc et al. [6] forecast COVID-19 by using a simple iteration method that needs only the daily values of confirmed cases as input. Other researchers focus on diagnosing the spread of COVID-19 and analyzing the factors of disease transmission. Liu et al. [7] propose a susceptible-asymptomatic-infected-removed (SAIR) model on social networks to describe the spread of COVID-19 and analyze the outbreak based on epidemic data from Wuhan. Kraemer et al. [8] show that China's drastic control measures substantially

mitigated the spread of COVID-19. Tsiotas and Magafas [9] build on the visibility graph algorithm to study the Greek COVID-19 infection curve as a complex network. These studies consider the spread of COVID-19 to be complicated, and diagnosing the scale and characteristics of the COVID-19 pandemic is conducive to governments effectively blocking the spread of this epidemic in their nations. Hence, we believe that it is very effective and important to prevent and control the further spread of COVID-19 by studying the connectedness characteristics of the pandemic complex network.

Studying the connectedness characteristics in a complex network through theoretical analysis and numerical simulation is helpful in understanding the microscopic mechanism of the spread process of an epidemic to discover practical ways to control the spread. Previous research has shown that a real social network is a typical complex network with small-world, scale-free, and high clustering characteristics. Small-world characteristics accelerate the spread of an epidemic, scale-free features facilitate the spread of an epidemic, and high clustering characteristics cause community clustering in epidemic transmission [5, 10–17].

Hence, some scholars have combined complex network theory with epidemiology. Ni [18] builds a dynamic model of an epidemic and shows that infectious diseases spread much faster in public networks than in in-home networks. In the beginning, the number of infected people in the network increases exponentially, and the spread speed obviously increases with increasing local structural scale. Chen et al. [19] simulate the propagation of the H1N1 virus based on complex networks. The simulation results show that the targeted immunization strategy can effectively inhibit the spread of diseases. Wang et al. [3] deeply investigates influential spreaders' identification in complex networks based on various centrality indices. Wang et al. [4] use a two-layer network model method to find that people who received positive prevention information had a lower risk of disease infection than those who received negative prevention information.

All the abovementioned studies try to provide feasible ways to curb the spread of the epidemic by using mathematical models or complex network tools. They can help us to work with limited data to solve problems in a constantly changing environment. However, at present, there are few studies that examine the global network of COVID-19. Therefore, determining the complicated connectedness of the COVID-19 complex network can fill gaps in the literature and effectively help the government identify good policies to prevent the COVID-19 pandemic in a timely manner. This study attempts to show the global situation of the COVID-19 pandemic based on a complex network. It profoundly analyzes the significant economic and social factors that affect the connectedness of the COVID-19 network to powerfully supplement the above literature.

DATA

The growth rate of COVID-19-infected people in each country is calculated using the COVID-19 infection sample from January 19, 2020, to August 15, 2020. The data came from the CSMAR database. To analyze the influencing factors of global COVID-19 infection network connectedness, we use each country's economic and social macrodata in 2018 as samples. Economic and social variables include GDP, total exports, total imports, inflation rate, government expenses, population density, and life expectancy. All these data are from the World Bank database. From them, we take logarithms of GDP, total exports, total imports, government expenses, and population density to reduce the absolute values.

METHODOLOGY

Construction of COVID-19 Pandemic Network

In this study, we build a COVID-19 infection complex network among 122 countries. Any node in our network is a specific country. Moreover, the possible connection between any two countries is called an edge in the network. These nodes and connected edges constitute the entire network of global COVID-19 infections. The network may rely more on certain countries, as many paths pass through them. Once this kind of country is identified and excluded, the network may collapse. Therefore, identifying these import infection nodes can enable governments to better fight the pandemic.

The key step in constructing the COVID-19 pandemic network is measuring the correlation of the growth rate of COVID-19-infected people among countries. The specific construction process of the COVID-19 pandemic network is as follows.

First, we exclude countries with fewer than 2,000 infected people and then calculate the growth rate of infected people in the remaining 122 countries individually.

$$g_{i,t} = \left(\frac{\text{infected}_{i,t}}{\text{infected}_{i,t-1}} - 1 \right) \times 100\% \quad i = 1, 2, \ldots, 122, \tag{1}$$

where $g_{i,t}$ is the infected persons growth rate of country i in month t, $\text{infected}_{i,t}$ is the number of infected persons of country i in month t, and $\text{infected}_{i,t-1}$ is the number of infected persons in country i in month $t-1$.

Second, we compute the correlation coefficients of the infected persons' growth rate between any two countries. The correlation coefficient is formulated as follows:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}, \tag{2}$$

where $\text{Cov}(X, Y)$ is the covariance of X and Y, $\text{Var}[X]$ is the variance of X, and $\text{Var}[Y]$ is the variance of Y.

Third, we need to set a threshold for correlation coefficients to define the existence of the possible connection. To fulfill this purpose, we take half of the mean of each country's maximum correlation coefficient (approximately 0.4) as the threshold value for the existence of connectedness. If the correlation coefficient between the two countries is higher than the threshold, a line is drawn in

the network to connect these two countries, indicating a connection between them.

This method can test the network connectedness between all countries, draw undirected connecting lines in the network, and finally create a COVID-19 pandemic infection network.

Complex Network Characteristics

Centrality Analysis

Centrality is an important indicator of each node's status and role in the network, and it can be used to judge whether an individual node is influential [3]. The higher a node's centrality is, the more significant its influence in the network is, suggesting that it can influence other nodes more prominently. In complex network analysis, there are many ways to characterize the centrality of a node. The most important are betweenness centrality, closeness centrality, eigenvector centrality, and page rank.

1) Betweenness centrality

In a network, a certain node's degree may be very small, but it may be the intermediate node between any two other nonadjacent nodes. If this node is removed, the connectivity between the nonadjacent nodes will be interrupted. Therefore, the node plays a vital role in the network. For such a node, we can define betweenness.

The core idea is that the interaction between two nonadjacent nodes depends on other nodes in the network, especially those on the shortest path. They exert a certain control over the two nonadjacent nodes. For example, if node A is located on multiple shortest paths between other nonadjacent nodes, node A plays a greater role and has greater betweenness centrality.

Freeman [20] suggests that the shortest path between nonadjacent nodes v_j and v_l will pass some nodes in a network. If v_i is passed by many other shortest paths, then it is important in the network. Moreover, the importance or influence of v_i for v_j and v_l can be represented by B_i, defined as follows:

$$B_i = \sum_{1 \le j \le N} \left[\frac{n_{jl}(i)}{n_{jl}} \right], \tag{3}$$

where n_{jl} is the number of shortest paths between nodes v_j and v_l; $n_{jl}(i)$ is the shortest path between nodes v_j and v_l, which passes through node v_i; and N is the total number of nodes in the network.

Then, the betweenness centrality of v_i is defined as follows:

$$C_B(v_i) = \frac{2B_i}{[(N-1)(N-2)]}. \tag{4}$$

2) Closeness centrality

Sabidussi [21] proposes that the closeness centrality of v_i measures the extent of proximity to other nodes through the shortest path. The closer the node is, the more important it is in the network. The closeness centrality value is the reciprocal of the average distance between node v_i and all other nodes that v_i can reach. Specifically, closeness centrality is defined as follows:

$$C_i^c = \frac{1}{L_i} = \frac{n-1}{\sum_{j \in \Gamma, j \ne i} d_{ij}}, \tag{5}$$

where Γ is the set of nodes that v_i can reach, L_i is the average distance to all other reachable nodes of v_i, and d_{ij} is the distance between v_i and v_j. The closeness indicator describes the difficulty of reaching other nodes in the network through a certain node v_i. It reflects the ability of node v_i to exert influence on other nodes in the network. Therefore, the closeness centrality can effectively reflect the global structure of the network.

3) Eigenvector centrality

Bonacich [22] suggests that eigenvector centrality is another important centrality measurement method for nodes. For node v_i, let the centrality score x_i be proportional to the sum of the centrality scores of all nodes connected to it.

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{ij} x_j, \tag{6}$$

where N is the total number of nodes and λ is the maximum eigenvalue. Each score of the eigenvector must be positive.

4) Page rank

Page rank [23] can be understood as a random walk on a graph. It can be defined as follows:

$$x_k = \sum_{i=1}^{N} a_{ik} x_i = (1-d) + d \sum_{i:b_{ik}=1} \frac{x_i}{r_i}, \tag{7}$$

where x_k is the page rank value of node k; d is the model parameter, usually d = 0.85; and a_{ik} is the probability of transition from page i to page j.

Small-World Network

The small-world property is an important topological feature of the network. Watts and Strogatz [10] propose the small-world network model (WS small-world network). They argue that some network systems are large in size but have a relatively small distance between any two nodes. A typical small-world network has a shorter average distance and a larger clustering coefficient. For example, there are no more than six people between any two strangers in a small-world social network.

1) Clustering coefficient

In a WS small-world network, the clustering coefficient quantifies the tendency of the nodes of a complex graph to cluster. In terms of a certain node v in the WS small-world network, v_i and v_j are two neighboring nodes of node v. If we reconnect v_i and v_j by random probability p, then the probability of v_i and v_j still being neighboring nodes is $(1-p)$, and $(1-p)^3$ is the probability that there is still an edge between v_i and v_j. Watts

and Strogatz [10] and Barrat and Weigt [24] argue that the WS clustering coefficient can be calculated as follows:

$$C_{WS}(p) = \frac{3(k-1)}{2(2k-1)}(1-p)^3. \quad (8)$$

The initial network is a one-dimensional lattice of N sites, with periodic boundary conditions (i.e., a ring), each vertex being connected to its $2k$ nearest neighbors. The vertices are then visited one after the other; each link connecting a vertex to one of its k nearest neighbors in the clockwise sense is left in place with probability $(1-p)$ and with probability p is reconnected to a randomly chosen other vertex.

2) Average path length

The average distance between any two nodes would also show interesting properties in a given complex network. The diameter D of a complex network is defined as the maximum distance between any pair of nodes.

Watts and Strogatz [10] argue that in terms of random networks, almost all networks have the same diameter. This means that the Nth-order random network with probability P has a tiny change in diameter. Watts [25] suggests that the diameter D is usually concentrated as follows:

$$D = \frac{\ln N}{\ln\langle k\rangle} \approx \frac{\ln N}{\ln(pN)}. \quad (9)$$

Supposing L_{ER} is the average path length of the ER random network, for a randomly selected node in the ER random network, there are approximately $\langle k\rangle^{L_{ER}}$ other nodes in the network with a distance from the selected node that is equal or very close to L_{ER}. Therefore, $L_{\mathrm{ER}} \propto \frac{\ln N}{\ln\langle k\rangle}$. It is a typical small-world property that the average path length is proportional to the logarithmic growth function. Since $\ln N$ grows slowly with N, a large random network may have small average path lengths.

Modularity Class

Newman [26, 27] proposes modularity, which enables a clear evaluation indicator to measure the segmentation of network communities. In other words, modularity refers to the difference between an ordinary network and a random network under a specific community division criterion. Because the random network does not have a community structure, the larger the difference is, the better the community division will be.

Newman [27] defines modularity as follows:

$$Q = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\left(s_i s_j + 1\right) = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)s_i s_j, \quad (10)$$

where m is the total number of edges in the network, A is the corresponding adjacency matrix of the network, and $A_{ij} = 1$ represents the existence of an edge between node i and node j. Otherwise, there is no edge connection. k_i is the degree of node i, s_i is the label of node i belonging to a community, and $\frac{1}{2}(s_i s_j + 1) = 1$ if and only if $s_i = s_j$; otherwise, $\frac{1}{2}(s_i s_j + 1) = 0$.

EMPIRICAL RESULTS AND DISCUSSION

COVID-19 Pandemic Network

In this section, we present our COVID-19 pandemic network. We first calculate the daily growth rate of the infected population in 122 countries individually and then calculate the correlation coefficient (ρ_{ij}) of the growth rate between any two countries. Furthermore, the threshold value for the existence of connectedness is set as 0.4 since the half mean of each country's maximum correlation coefficient is 0.38. If the correlation coefficient between country i and country j is higher than the threshold value, i.e., $\rho_{ij} > 0.4$, then the element $\{a_{ij}\}$ in the connectedness matrix is 1; otherwise, it is 0. If $a_{ij} = 1$, there is an edge between country i and country j, and if $a_{ij} = 0$, there is no edge between country i and country j. In addition, the value of the element on the diagonal of the connectedness matrix is set as 0, which indicates that there is no link between any country and itself. We obtain our COVID-19 pandemic connectedness matrix through the above calculation method (as the number of countries involved is 122, we cannot present the 122 × 122 matrix here). Because a zero growth rate will affect the similarity calculation, we excluded 91 countries and regions with fewer than 2,000 infected people.

Finally, the COVID-19 pandemic network obtained in this study is a complex network without considering the direction, and it is an undirected network with 122 nodes and 958 connected edges.

Using the software Gephi, we generate a visual graph of our network. The results are shown in **Figure 1**. Taking China, the United States, and Brazil as examples, China has connectedness with Singapore, Japan, and Thailand; the United States has connectedness with Australia, Kenya, Colombia, Pakistan, Dominica, Turkey, and Luxembourg; and Brazil has connectedness with 16 countries, including Austria, Portugal, Denmark, Saudi Arabia, and France.

Then, we use Gephi to obtain the network topology characteristic values, which is the basis for the next step of the analysis. **Figures 2** and **3** show the COVID-19 pandemic network parameters. **Figure 2** shows the degree distribution in our COVID-19 pandemic network. The overall distribution of the graph shows that the COVID-19 pandemic network is relatively dispersed. **Figure 3** shows the clustering coefficient graph. The larger the clustering coefficient is, the higher the extent of the connection between neighbors is. Our COVID-19 pandemic network has a clustering coefficient of 0.58. This large clustering coefficient indicates that the COVID-19 pandemic network is a very close world.

Other overall characteristics of the COVID-19 pandemic network are shown in the following table.

Table 1 shows that the COVID-19 pandemic network has 122 nodes, and the number of edges between nodes is 985 with an average degree of 7.85, average path length of 3.104, clustering coefficient of 0.58, and diameter of 8. Moreover, the clustering coefficient is large, so the network has close-world network characteristics.

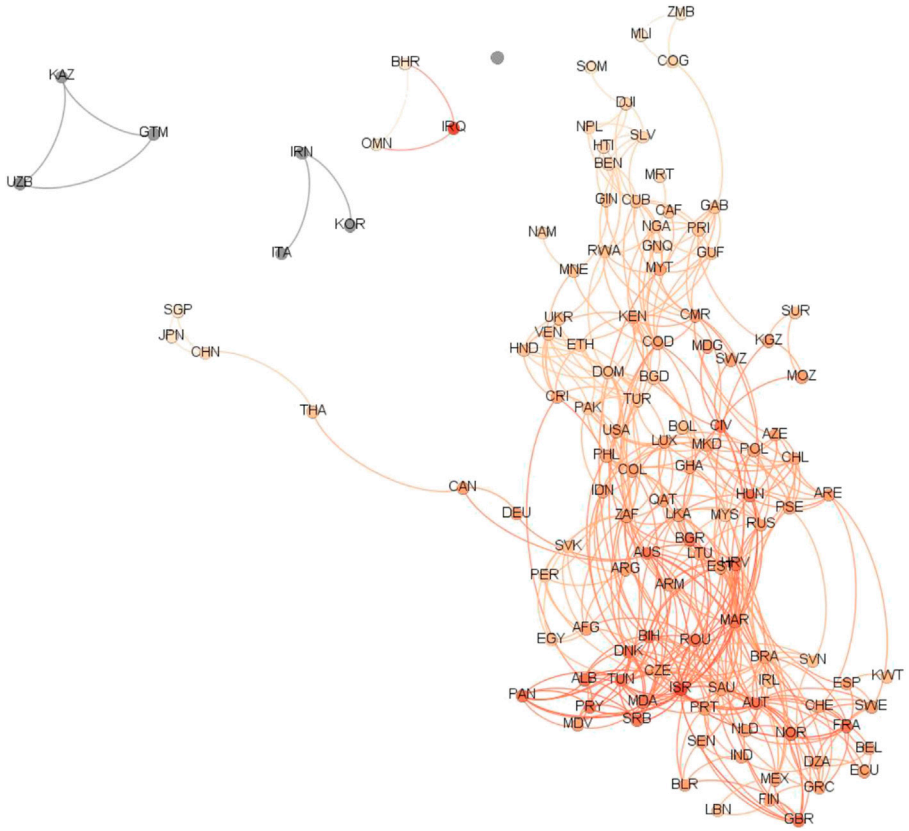


FIGURE 1 | The global COVID-19 pandemic network. Note: This is the global COVID-19 pandemic network. It excludes countries with fewer than 2,000 infected people. When the correlation of the growth rate of infected people is larger than 0.4, there is a connected edge between countries.

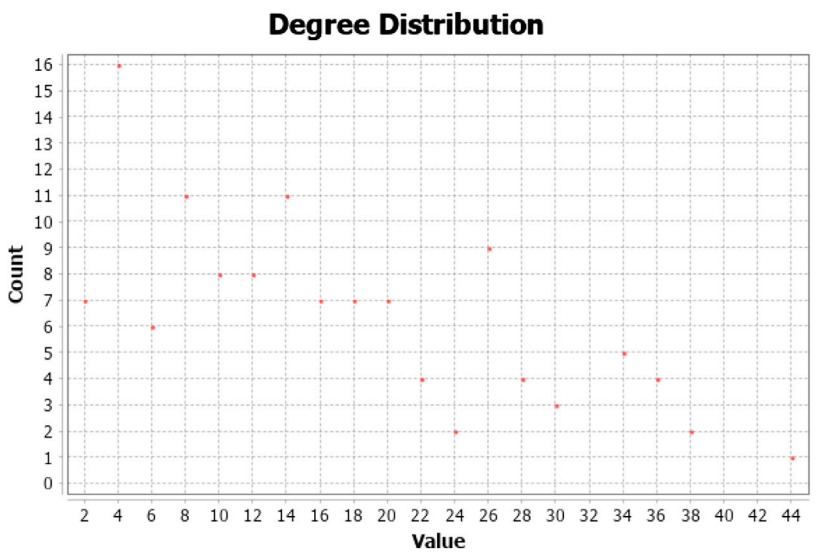


FIGURE 2 | The degree distribution of the global network. Note: **Figure 2** represents the degree distribution. The vertical and horizontal axes of each subplot are node counts and parameter distribution values, respectively. The horizontal axis represents the degree value of each country. The vertical axis represents the number of countries. For example, value = 2 and count = 7 means that there are seven countries, and they all have a degree of two.

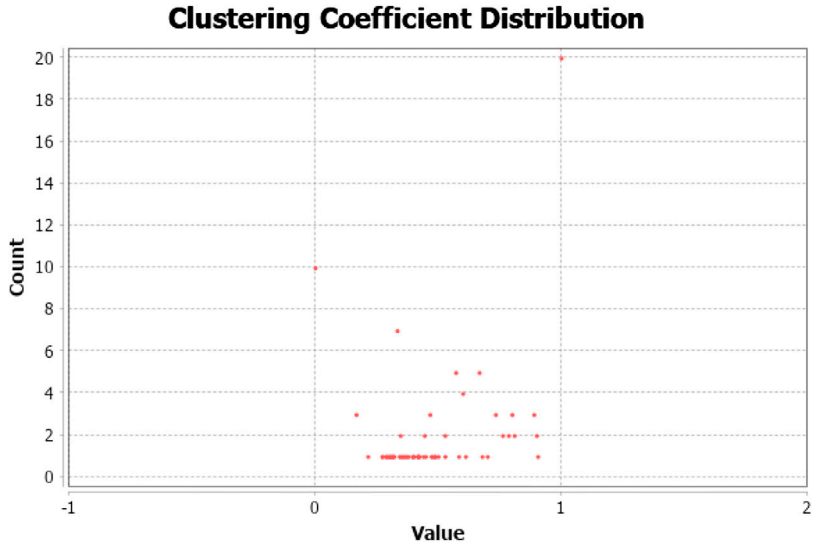


FIGURE 3 | The clustering coefficient distribution of the global network. Note: **Figure 3** represents the distribution of clustering coefficients in various countries. The vertical and horizontal axes of each subplot are node counts and parameter distribution values, respectively. The horizontal axis represents the clustering coefficient value of each country. The vertical axis represents the number of countries.

TABLE 1 | Topological characteristics of the COVID-19 pandemic network.

Network	N	E	<d>	K	C	D
COVID-19 pandemic	122	958	7.85	3.104	0.58	8

Table 1 *presents the topological information of the COVID-19 pandemic network. N is the number of nodes. E is the number of edges. <d> is the average degree of the whole network. K represents the average path length of the network. C is the clustering coefficient of the network. D is the diameter of the network.*

In addition, the small-world network has the characteristics of a small path length and a high clustering coefficient between nodes. Our global COVID-19 pandemic network's average path length is 3.104, and its clustering coefficient is 0.58, indicating that the global COVID-19 pandemic network has some characteristics of a small-world network.

Connectedness Characteristics of the Global COVID-19 Pandemic Network

In this section, we further analyze some important characteristics of the global COVID-19 pandemic network. Specifically, we focus on the degree, triangles, centrality, modularity class, and authority and then provide the visual results.

Figure 4 represents the degree of each country in the global COVID-19 pandemic network. The degree of country i is the total number of edges directly connected to that country. The deeper a country's color is, the greater its degree is. As shown in **Figure 4**, Brazil has the deepest color on the American continents; in Europe, Austria has the greatest degree; in Asia, Israel and Sri Lanka have the highest number of edges; in Oceania, the color of Australia is the deepest; and in Africa, the countries with the greatest degree are South Africa and Morocco. The above results indicate that these countries have the greatest degrees with other countries and a prominent ability to spread COVID-19 to other countries. Thus, more attention should be paid to their capacity and measures for COVID-19 pandemic prevention and control.

Figures 5–8 show the centrality analysis results, reflecting the relative importance of each node in the network. **Figure 5** shows betweenness centrality. The countries with the deepest colors are the Democratic Republic of the Congo, Australia, and Croatia, which are located in Africa, Oceania, and Europe, respectively. This finding indicates that these three countries play an important betweenness role in the COVID-19 pandemic network. Additionally, Canada and Cuba in North America are also significant. We may more easily halt the global spread of COVID-19 if we control the pandemic in these three countries first.

Figure 6 shows eigenvector centrality. The three countries with the deepest colors are Moldova, Sri Lanka, and Austria, located in Europe, Asia, and Europe, respectively. This finding shows that the three countries have the highest centrality scores and are the nodes with the greatest importance in the COVID-19 pandemic network. Again, Austria shows its importance in the COVID-19 pandemic network, indicating that the effective control of COVID-19 in Austria has a significant role in halting the global spread of the COVID-19 pandemic.

Figure 7 represents closeness centrality. The three countries with the deepest colors are Kazakhstan, Iraq, and Uzbekistan, all located in Asia. This means that Kazakhstan, Iraq, and Uzbekistan may be important for Asian COVID-19 control. **Figure 8** shows harmonic closeness centrality. The empirical results are the same as those shown in **Figure 7**.

Figure 9 shows the analysis results of authority, reflecting the influence of nodes in the network. The three countries with the deepest colors are Moldova, Israel, and Austria, located in

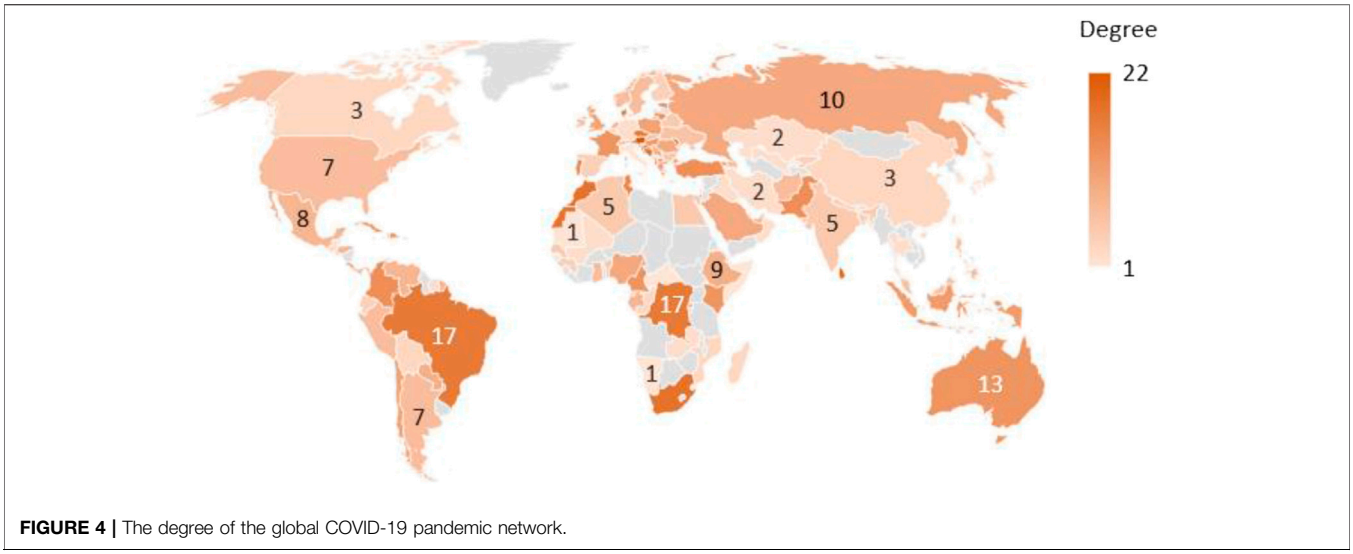


FIGURE 4 | The degree of the global COVID-19 pandemic network.

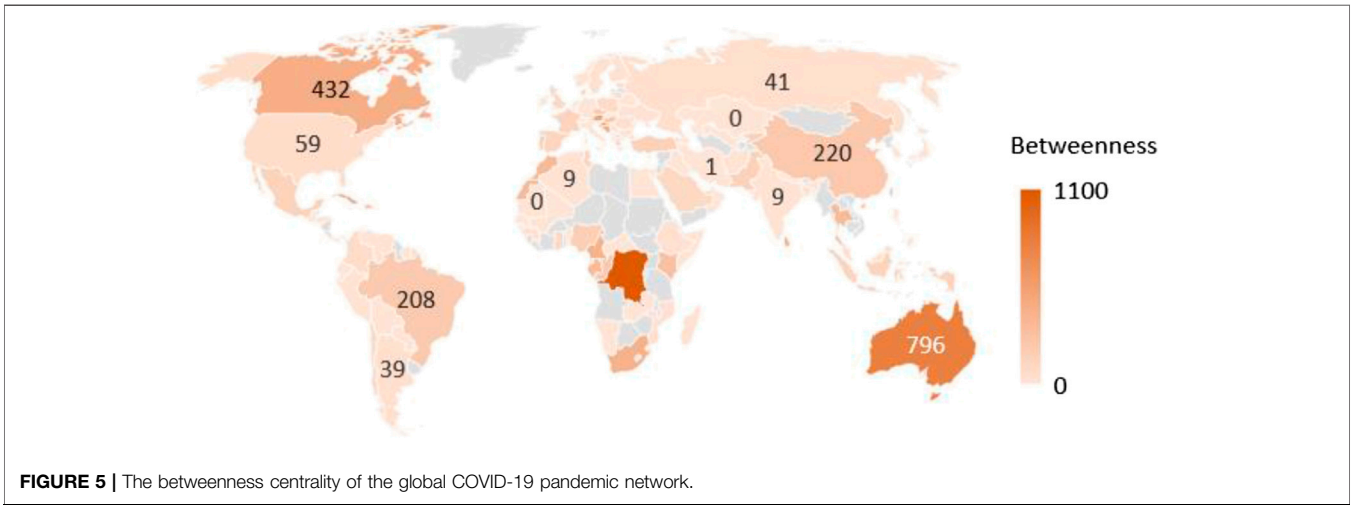


FIGURE 5 | The betweenness centrality of the global COVID-19 pandemic network.

Europe, Asia, and Europe, respectively. This indicates that the infections in Moldova, Israel, and Austria are highly dangerous, as these three countries may have high virus spread capacity.

Figure 10 shows modularity-class analysis results, describing the community situation in the network. The results show that the modularity coefficient is 0.526, and the COVID-19 pandemic network is divided into 10 communities, representing community structures. Community 0 contains 8 countries, including Palestine, Poland, and the United Arab Emirates. Community 1 contains Cuba, the Democratic Republic of the Congo, and 23 other countries. Community 2 contains Mexico, Moldova, Sri Lanka, Israel, and so on. Community 3 contains 13 countries, including China, Canada, Britain, India, Germany, Australia, and Japan. Community 4 contains 24 countries, including France, Brazil, Austria, and South Africa. Community 5 contains Bahrain, Iraq, and Oman. Community 6 contains 11 countries, including Bangladesh, Kyrgyzstan, and Ghana. The countries in Community 7 are Uzbekistan, Kazakhstan, and Guatemala. The countries in Community 8 are Iran, South Korea, and Italy. Community 9 contains 16 countries, including the United States, the Russian Federation, and Colombia.

The Determinants of Connectedness

According to the above analysis, the connectedness of the global COVID-19 infection network varies across different countries. Therefore, what factors influence the network connectedness of COVID-19 infection? Hereafter, this study further explores how economic factors, population density, medical level, and other social factors influence the connectedness of COVID-19 by using a quantitative regression method to provide a theoretical basis and guidance for policymakers' prevention and control measures.

We build the econometric model as follows:

$$\begin{aligned}\text{connectedness}_i = {} & \beta_0 + \beta_1 \text{LnDensity} + \beta_2 \text{LnGDP} + \beta_3 \text{LnExport} \\ & + \beta_4 \text{LnImport} + \beta_5 \text{Inflation} + \beta_5 \text{Expense} \\ & + \beta_6 \text{LifeExpectancy} + \varepsilon_i,\end{aligned}$$

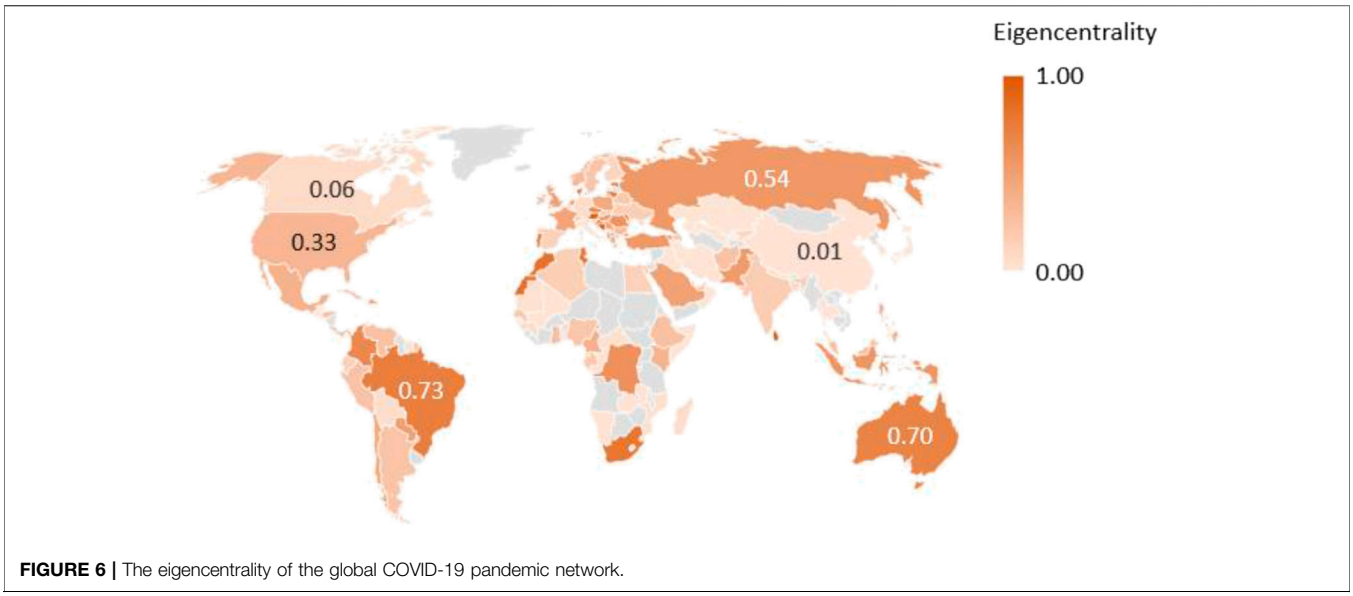


FIGURE 6 | The eigencentrality of the global COVID-19 pandemic network.

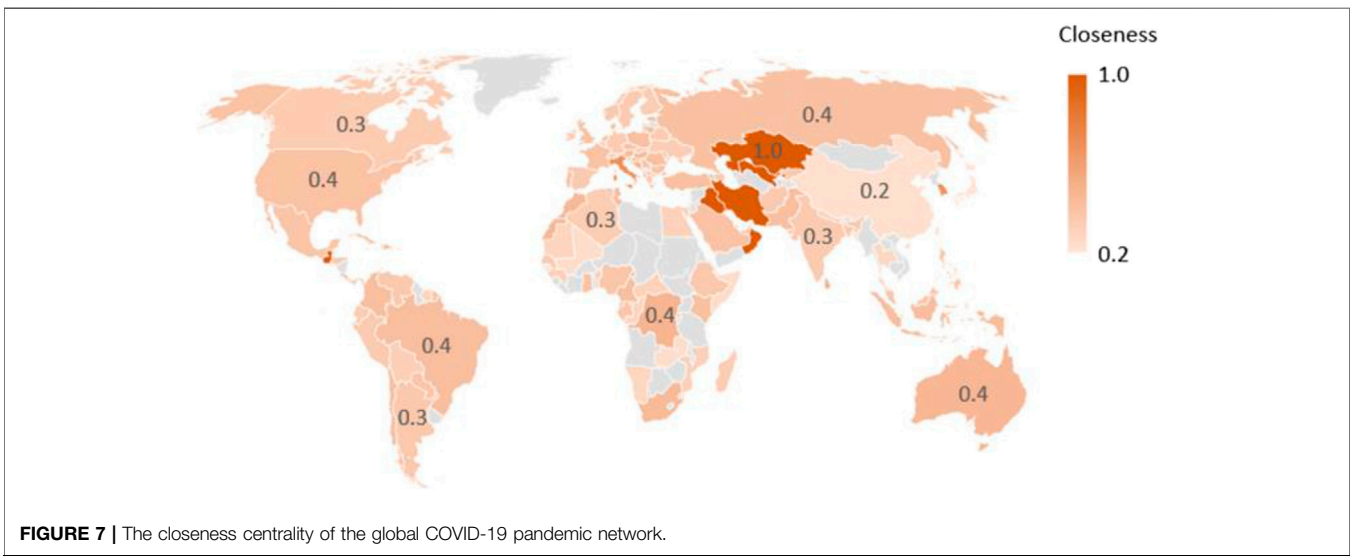


FIGURE 7 | The closeness centrality of the global COVID-19 pandemic network.

where connectedness$_i$ represents degree, eigenvector centrality, betweenness centrality, closeness centrality, harmonic centrality, page rank, authority, hub, and triangles separately. LnDensity means the logarithms of population density. LnGDP is the logarithms of economic size. LnExport is the logarithms of total exports. LnImport is the logarithms of total imports. Inflation indicates monetary policy. Expense represents government expenditure. Life represents life expectancy, reflecting the quality of medical treatment.

Table 2 shows the regression results between the characteristics of the COVID-19 pandemic network and other possible determinant factors. In columns (1)–(7), we investigate how the degree, eigenvector centrality, betweenness centrality, closeness centrality, harmonic closeness centrality, page rank, authority, hub, and triangles are affected by possible influencing factors.

In columns (1) and (2), LnGDP and Expense are significantly positive at the 10% confidence level, indicating that economic level and government expenditure will significantly increase degree and eigenvector centrality. In column (3), LnGDP is also significantly positive at the 5% confidence level, indicating that the economic scale significantly increases betweenness centrality, consistent with the result of degree and eigenvector centrality. Inflation is significantly negative at the 5% confidence level, suggesting that it significantly reduces betweenness centrality. In column (5), the estimated coefficient of LnGDP is still significantly positive at 10%, suggesting that economic size will significantly increase harmonic closeness centrality. In column (6), LnDensity, LnGDP, and Expense are all significantly positive at the 5% confidence level, indicating that population density, economic scale, and government expenditure

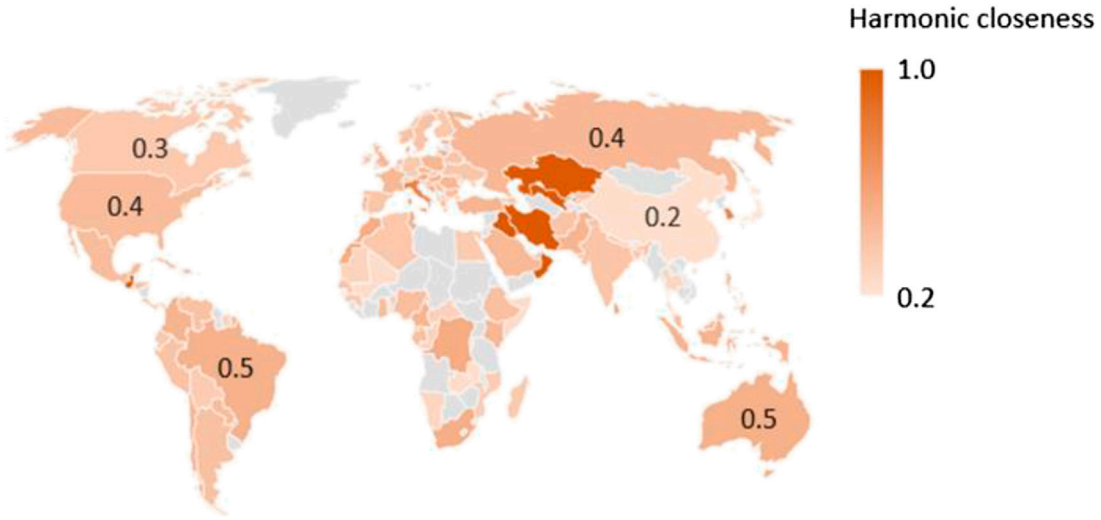


FIGURE 8 | The harmonic closeness of the global COVID-19 pandemic network.

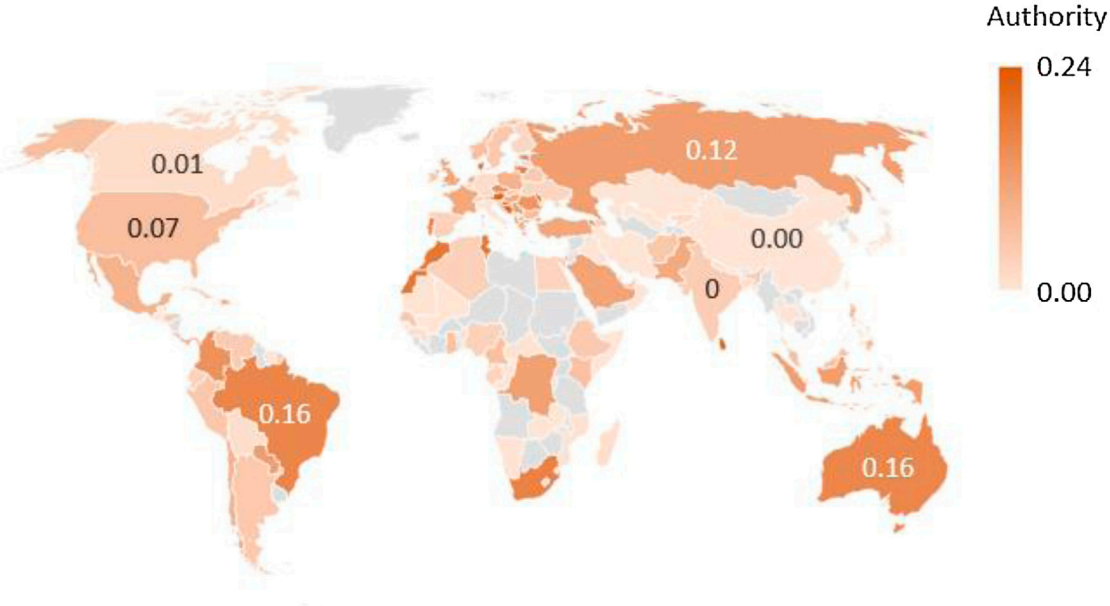


FIGURE 9 | The authority of the global COVID-19 pandemic network.

increase page rank. However, LnImport is significantly negative at 5%, indicating that imports significantly reduce page rank. In columns (7) and (8), Expense has a significantly positive coefficient at 1%, meaning that the amount of government expenditure will significantly increase authority and hub. In column (9), LnGDP and Expense are significantly positive at 5%, indicating that the size of the economy and government spending significantly increase triangles.

Overall, the economic development level and government spending show significant positive effects on COVID-19 infection connectedness. If policymakers hope to curb the spread of COVID-19 effectively, they should give priority to the COVID-19 pandemic, appropriately reduce economic activities, and avoid crowd gatherings to reduce the risk of infection. Moreover, government expenditures need to be allocated more effectively. Controlling and preventing the spread of COVID-19 should be the first goal of government expenditures. More government resources should be used to purchase medical devices and improve medical staff protection and treatment. The United States is a case in point. As the United States has prioritized the economy in its COVID-19 prevention policy, a large amount of fiscal spending has been used to subsidize economic activities rather than medical resources. As a result, the number of

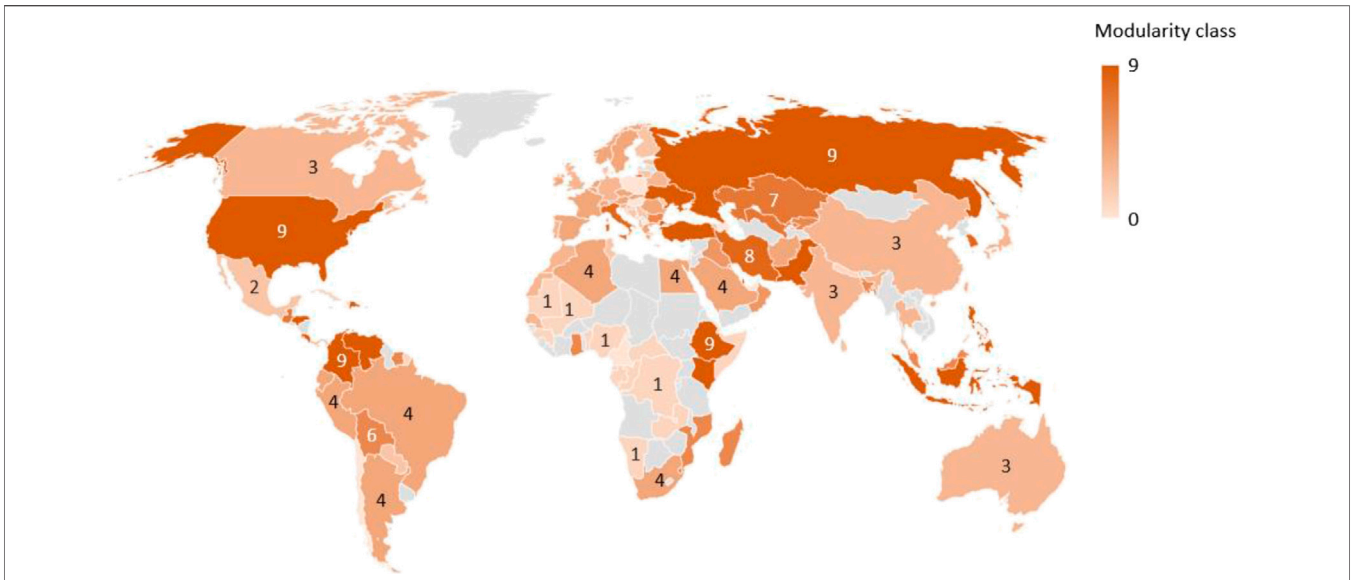


FIGURE 10 | The modularity class of the global COVID-19 pandemic network.

TABLE 2 | Determinants of connectedness: global.

Variables	(1) Degree	(2) Eigen	(3) Betweenness	(4) Closeness	(5) Harmonic	(6) Page rank	(7) Authority	(8) Hub	(9) Triangles
LnDensity	0.7450	0.0217	−13.9320	−0.0020	0.0024	0.0007**	0.0036	0.0036	2.1949
	(1.46)	(0.75)	(−0.53)	(−0.12)	(0.14)	(2.03)	(0.55)	(0.55)	(1.50)
LnGDP	2.9796**	0.1207*	72.1867**	0.0440	0.0551*	0.0025***	0.0249	0.0249	8.1021**
	(2.28)	(1.76)	(2.39)	(1.60)	(1.81)	(3.51)	(1.63)	(1.63)	(2.08)
LnExport	0.2037	−0.0009	−10.2764	−0.0181	−0.0111	0.0004	−0.0014	−0.0014	0.3559
	(0.10)	(−0.01)	(−0.20)	(−0.52)	(−0.27)	(0.31)	(−0.06)	(−0.06)	(0.05)
LnImport	−4.1171	−0.1654	−73.2675	−0.0328	−0.0537	−0.0036**	−0.0341	−0.0341	−12.4868
	(−1.50)	(−1.12)	(−1.06)	(−0.80)	(−1.05)	(−2.36)	(−1.03)	(−1.03)	(−1.48)
Inflation	−0.0531	−0.0037	−6.4161**	−0.0000	−0.0001	−0.0001*	−0.0010	−0.0010	−0.0766
	(−0.78)	(−1.03)	(−2.39)	(−0.02)	(−0.02)	(−1.78)	(−1.29)	(−1.29)	(−0.33)
Expense	0.2523***	0.0126***	2.4979	−0.0001	0.0012	0.0001***	0.0029***	0.0029***	0.7041***
	(3.53)	(3.47)	(1.06)	(−0.03)	(0.55)	(3.39)	(3.48)	(3.48)	(3.25)
Life	0.0605	0.0067	0.6182	0.0031	0.0033	0.0000	0.0019	0.0019	0.5413
	(0.54)	(1.12)	(0.14)	(1.28)	(1.29)	(0.46)	(1.38)	(1.38)	(1.47)
Constant	7.0192	0.0728	217.2144	0.2184	0.2472	0.0080*	−0.0029	−0.0029	8.3716
	(1.00)	(0.21)	(0.92)	(1.48)	(1.45)	(1.95)	(−0.04)	(−0.04)	(0.41)
Observations	79	79	79	79	79	79	79	79	79
R-squared	0.228	0.212	0.102	0.039	0.057	0.232	0.218	0.218	0.232

*This table constructs regression models for global determinants of connectedness. Dependent variables are network connectedness characteristics. Independent variables are economic and social factors. Values in parentheses are the T statistical values; ***, **, and * indicate significance at the levels of 1, 5, and 10%, respectively.*

confirmed COVID-19 cases in the United States is approaching 9 million.

The Determinants of Connectedness: Asia and Europe

To further analyze how economic and social factors affect Asian and European countries' connectedness characteristics differently, this study divides the samples into Asian and European countries.

Table 3 shows the empirical results in Asia. In Asia, LnDensity, LnGDP, and LnExport are significantly positive at the 10% confidence level for degree, eigenvector, authority, hub, and triangles, indicating that population density, economic activities, and export activities show significantly positive effects in the Asian COVID-19 pandemic network. However, LnImport is persistently significantly negative at the 10 or 5% confidence level for degree, eigenvector, page rank, authority, hub, and triangles, showing that imports can effectively reduce the characteristics of the COVID-19 pandemic network in Asia.

TABLE 3 | Determinants of connectedness: Asia.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Degree	Eigen	Betweenness	Closeness	Harmonic	Page rank	Authority	Hub	Triangles
LnDensity	3.1811*	0.1962**	21.1112	−0.0711	−0.0586	0.0013*	0.0438**	0.0438**	10.8935**
	(2.28)	(2.87)	(0.69)	(−1.01)	(−0.88)	(1.92)	(2.87)	(2.87)	(2.80)
LnGDP	6.5357*	0.3872*	62.9644	0.0091	0.0281	0.0038*	0.0874*	0.0874*	21.9928*
	(2.00)	(2.30)	(0.79)	(0.09)	(0.26)	(1.96)	(2.29)	(2.29)	(2.24)
LnExport	10.9008	0.6387*	50.8330	0.0305	0.0631	0.0059	0.1464*	0.1464*	35.9940*
	(1.78)	(2.16)	(0.37)	(0.16)	(0.31)	(1.36)	(2.19)	(2.19)	(1.96)
LnImport	−21.0506**	−1.2356**	−149.3194	−0.0607	−0.1177	−0.0118*	−0.2827**	−0.2827**	−70.3091**
	(−2.43)	(−3.04)	(−0.74)	(−0.25)	(−0.46)	(−1.92)	(−3.06)	(−3.06)	(−2.81)
Inflation	0.0684	−0.0058	−7.8613	0.0080	0.0072	0.0000	−0.0025	−0.0025	0.4325
	(0.19)	(−0.34)	(−0.82)	(0.52)	(0.48)	(0.19)	(−0.65)	(−0.65)	(0.41)
Expense	0.1327	0.0092	3.9181	−0.0034	−0.0019	0.0000	0.0023	0.0023	0.3640
	(0.46)	(0.67)	(0.74)	(−0.35)	(−0.20)	(0.22)	(0.74)	(0.74)	(0.43)
Life	−0.3162	−0.0248	−2.2921	0.0116	0.0087	0.0001	−0.0051	−0.0051	−0.6678
	(−0.69)	(−1.13)	(−0.22)	(0.56)	(0.42)	(0.44)	(−1.05)	(−1.05)	(−0.50)
Constant	66.9493*	4.1986**	655.5296	0.2487	0.5011	0.0241	0.9375**	0.9375**	188.7532
	(1.88)	(2.42)	(0.83)	(0.16)	(0.29)	(1.08)	(2.39)	(2.39)	(1.77)
Observations	16	16	16	16	16	16	16	16	16
R-squared	0.545	0.632	0.250	0.311	0.292	0.520	0.640	0.640	0.611

*Note: This table constructs regression models for the determinants of connectedness in Asia. Dependent variables are network connectedness characteristics. Independent variables are economic and social factors. Values in parentheses are the T statistical values; ***, **, and * represent significance at the levels of 1, 5, and 10%, respectively.*

TABLE 4 | Determinants of connectedness: Europe.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Degree	Eigen	Betweenness	Closeness	Harmonic	Page rank	Authority	Hub	Triangles
LnDensity	0.9020	0.0689	28.5923	0.0345	0.0392	0.0007	0.0157	0.0157	3.5322
	(0.60)	(0.89)	(0.72)	(1.67)	(1.68)	(0.69)	(0.87)	(0.87)	(0.74)
LnGDP	0.2223	0.0478	60.4216	0.0681	0.0740	0.0009	0.0140	0.0140	−0.1817
	(0.09)	(0.36)	(0.73)	(1.37)	(1.39)	(0.56)	(0.45)	(0.45)	(−0.02)
LnExport	0.5186	0.1156	12.5290	0.0634	0.0832	−0.0007	0.0137	0.0137	−7.0843
	(0.07)	(0.30)	(0.06)	(0.89)	(0.95)	(−0.14)	(0.15)	(0.15)	(−0.28)
LnImport	−2.1137	−0.2598	−113.1319	−0.1503	−0.1819	−0.0010	−0.0503	−0.0503	2.1277
	(−0.21)	(−0.49)	(−0.36)	(−1.25)	(−1.30)	(−0.14)	(−0.41)	(−0.41)	(0.06)
Inflation	−1.0538**	−0.0673***	−25.0360*	−0.0101**	−0.0120**	−0.0007**	−0.0168***	−0.0168***	−4.0368**
	(−2.37)	(−2.85)	(−1.85)	(−2.22)	(−2.15)	(−2.26)	(−3.01)	(−3.01)	(−2.49)
Expense	0.2490**	0.0105*	6.4767	0.0034*	0.0045**	0.0002**	0.0023	0.0023	0.6057*
	(2.26)	(1.83)	(1.54)	(1.99)	(2.37)	(2.60)	(1.69)	(1.69)	(1.73)
Life	−0.5094*	−0.0381**	−8.2352	−0.0066**	−0.0073**	−0.0003	−0.0088**	−0.0088**	−1.7204*
	(−1.90)	(−2.62)	(−1.06)	(−2.61)	(−2.34)	(−1.48)	(−2.57)	(−2.57)	(−1.77)
Constant	62.6958**	4.5161***	1,092.8542	0.9114***	1.0772***	0.0359**	1.0523***	1.0523***	217.3826**
	(2.78)	(3.93)	(1.45)	(4.58)	(4.53)	(2.30)	(3.94)	(3.94)	(2.71)
Observations	33	33	33	33	33	33	33	33	33
R-squared	0.326	0.398	0.232	0.301	0.335	0.330	0.410	0.410	0.409

*Note: This table constructs regression models for the determinants of connectedness in Europe. Dependent variables are network connectedness characteristics. Independent variables are economic and social factors. Values in parentheses are the T statistical values; ***, **, and * indicate significance at the levels of 1, 5, and 10%, respectively.*

In summary, for Asian countries, population density and economic activities significantly increase the connectedness of the COVID-19 pandemic network, which is consistent with the full sample results. As most Asian countries have large populations and high residential density, it is very important and influential for them to decrease economic activities and strengthen quarantine measures appropriately. In addition, more substantial imports would reduce the connectedness of the COVID-19 pandemic network in Asian countries. We suspect that one of the reasons is that most Asian countries' medical level is still relatively low. More imports of materials and medical equipment can effectively reduce domestic resource pressure, thus reducing connectedness.

Table 4 shows the empirical results for Europe. In Europe, the coefficients of Expense for degree, eigenvector, closeness, harmonic closeness, and page rank are significantly positive at

10 or 5%, meaning that government expenses in Europe may be a burden and that governments do not effectively use their resources to control the COVID-19 pandemic. The estimated coefficients of inflation for all characteristics are all significantly negative except for betweenness and page rank. All the estimated coefficients of Life are significantly negative at 10 or 5%, indicating that as countries with higher life expectancy are more likely to have a better medical level; life expectancy is negatively associated with the connectedness of the COVID-19 pandemic network.

Compared with Asian countries, economic activity, population density, and government expenditure do not significantly increase the connectedness of the COVID-19 pandemic network in Europe. However, life expectancy can effectively reduce network connections. As European countries have more capital intensive and high value-added industries compared with Asia, the infection possibility of COVID-19 is relatively low in those industries. We believe that cutting down economic activities can not effectively prevent and control the spread of COVID-19 pandemic in Europe. Moreover, as the population density in Europe is much lower than that in Asia, it is difficult for the government of European countries to lower their population density furthermore. Therefore, population density is not significantly associated with the connectedness of the COVID-19 pandemic in Europe. Compared with Asian countries, most European countries have a higher economic development level, so their government expenditure is generally used for medical expenditure. Government expenditure can effectively strengthen a country's medical resources. Therefore, life expectancy in Europe is higher, and life expectancy is negatively associated with the connectedness of the COVID-19 network, suggesting that the European medical level is useful for preventing and controlling COVID-19.

Due to differences in economic and social characteristics, policymakers should use different prevention and control measures for different countries. Each country should implement prevention policy measures according to its local conditions. Only in this way can we effectively control the further spread of COVID-19.

CONCLUSION

Based on the growth pattern of the number of COVID-19 infections worldwide, this study constructed a complex network among 122 countries. Complex network analysis was used to deconstruct the overall connectedness of COVID-19 infection networks, visualize the connectedness characteristics among countries in the network, and further explore the influencing factors of the global COVID-19 infection network based on macrodata such as the economy, population, and life expectancy.

Our research shows that, first, in terms of overall connectedness, the global COVID-19 pandemic network has small-world network, close-world network, and community structure features. There are a total of 958 edges in 122 countries worldwide, which are divided into ten communities. This network has a small average path length (3.14) and a high clustering coefficient (0.58), and its diameter is 8. Second, in terms of intercountry connectedness, some countries have a significant influence on the global COVID-19 pandemic network, such as Austria (Europe), Croatia (Europe), Moldova (Europe), Brazil (South America), Canada (North America), Iran (Asia), Kazakhstan (Asia), Uzbekistan (Asia), Republic of the Congo (Africa), and Austria (Oceania). More attention should be paid to these countries for further control of COVID-19.

In addition, we divide the network into ten communities. The United States, Russia, and Colombia, which are suffering more severely from COVID-19, are in a community, while China and Japan, which have better control of COVID-19, are in another community.

Finally, in the influencing factor analysis, our study finds that population density, economic size, exports, and government expenditure significantly increased the connectedness of the COVID-19 infection network. However, imports and life expectancy significantly reduced the connectedness of the COVID-19 infection network. This result indicates that countries with large economic size, high population density, and high government expenditure more easily suffer from and spread COVID-19. In contrast, countries with adequate import capacity and a better medical level can reduce the network connectedness.

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

SZ performed the coding, data analysis, and original draft. MK gave formal writing-review and editing. FL proposed the conceptualization and methodology and modified final draft. QF performed the funding acquisition, original draft preparation, and data preparation. GD performed the data collection and original arrangement.

FUNDING

This study was funded by the Scientific Research Fund Project of Yunnan Education Department (No.2021J0586), the Fundamental Research Funds for the Central Universities. (No.20wkpy85), and the China Postdoctoral Science Foundation (No. 2019M663356).

REFERENCES

1. Moore C, Newman ME. Epidemics and percolation in small-world networks. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* (2000) 61(5): 5678–82. doi:10.1103/physreve.61.5678
2. Stegehuis C, Van Der Hofstad R, Van Leeuwaarden JS. Epidemic spreading on complex networks with community structures. *Sci Rep* (2016) 6(1):29748–7. doi:10.1038/srep29748
3. Wang J, Li C, Xia C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Appl Math Comput* (2018) 334: 388–400. doi:10.1016/j.amc.2018.04.028
4. Wang Z, Xia C, Chen Z, Chen G. Epidemic propagation with positive and negative preventive information in multiplex networks. *IEEE Trans Cybern* (2020) [Epub ahead of print]. doi:10.1109/TCYB.2019.2960605
5. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. *Rev Mod Phys* (2015) 87(3):925. doi:10.1103/RevModPhys.87.925
6. Perc M, Gorišek Miksić N, Slavinec M, Stožer A. Forecasting Covid-19. *Front Phys* (2020) 8:127. doi:10.3389/fphy.2020.00127
7. Liu C, Wu X, Niu R, Wu X, Fan R. A new SAIR model on complex networks for analysing the 2019 novel coronavirus (COVID-19). *Nonlinear Dynam* (2020) 1–11. doi:10.1007/s11071-020-05704-5
8. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* (2020) 368(6490):493–7. doi:10.1126/science.abb4218
9. Tsiotas D, Magafas L. *The effect of anti-COVID-19 policies to the evolution of the disease: a complex network analysis to the successful case of Greece* (2020) arXiv preprint arXiv:2004.06536.
10. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* (1998) 393(6684):440–2. doi:10.1038/30918
11. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett* (2001) 86(14):3200–3. doi:10.1103/PhysRevLett.86.3200
12. Newman ME. Communities, modules and large-scale structure in networks. *Nat Phys* (2012) 8(1):25–31. doi:10.1038/nphys2162
13. Wang P, Tian C, Lu J-a. Identifying influential spreaders in artificial complex networks. *J Syst Sci Complex* (2014) 27(4):650–665. doi:10.1007/s11424-014-2236-4
14. Estrada E, Hatano N, Benzi M. The physics of communicability in complex networks. *Phys Rep* (2012) 514(3):89–119. doi:10.1016/j.physrep.2012.01.006
15. Liu Y-Y, Barabási A-L. Control principles of complex systems. *Rev Mod Phys* (2016) 88(3):035006. doi:10.1103/RevModPhys.88.035006
16. Liao H, Mariani MS, Medo M, Zhang Y-C. Ranking in evolving complex networks. *Phys Rep* (2017) 689:1–54. doi:10.1016/j.physrep.2017.05.001
17. Hu J, Chengyi X, Huijia L, Peican Z, Wenjun X. Properties and structural analyses of USA's regional electricity market: a visibility graph network approach. *Appl Math Comput* (2020) 385:125434. doi:10.1016/j.amc.2020.125434
18. Ni S. *Research on modeling of infectious disease spreading based on complex Network theory*. Beijing, China: School of Aeronautics and Astronautics, Tsinghua University (2009).
19. Chen D, Huang S, Shang M. Complex network model and application in epidemic spreading and controlling. *Comput Sci* (2011) 38(6):118–21. doi:10.3969/j.issn.1002-137X.2011.06.028
20. Freeman LC. A set of measures of centrality based on betweenness. *Sociometry* (1977) 40:35–41. doi:10.2307/3033543
21. Sabidussi G. The centrality of a graph. *Psychometrika* (1966) 31(4):581–603. doi:10.1007/BF02289527
22. Bonacich P. Some unique properties of eigenvector centrality. *Soc Network* (2007) 29(4):555–64. doi:10.1016/j.socnet.2007.04.002
23. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: bringing order to the web. Stanford InfoLab (1999) 9(1):1–14.
24. Barrat A, Weigt M. On the properties of small-world network models. *Eur Phys J B* (2000) 13(3):547–60. doi:10.1007/s100510050067
25. Watts DJ. Networks, dynamics, and the small-world phenomenon. *Am J Sociol* (1999) 105(2):493–527. doi:10.1086/210318
26. Newman ME. Fast algorithm for detecting community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* (2004) 69(6):066133. doi:10.1103/PhysRevE.69.066133
27. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci USA* (2006) 103(23):8577–82. doi:10.1073/pnas.0601602103

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.

frontiers in Applied Mathematics and Statistics

METHODS
published: 25 January 2021
doi: 10.3389/fams.2020.611805

Causal Analysis of Health Interventions and Environments for Influencing the Spread of COVID-19 in the United States of America

Zhouxuan Li[1], *Tao Xu*[1], *Kai Zhang*[1], *Hong-Wen Deng*[2], *Eric Boerwinkle*[1] *and Momiao Xiong*[1]*

[1]School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States, [2]Tulane Center of Biomedical Informatics and Genomics, Deming Department of Medicine, School of Medicine, Tulane University, New Orleans, LA, United States

OPEN ACCESS

Edited by:
Zhanwei Du,
University of Texas at Austin, United States

Reviewed by:
Ting Tian,
Sun Yat-sen University, China
Alessandro Rovetta,
Mensana srls, Italy

***Correspondence:**
Momiao Xiong
Momiao.Xiong@uth.tmc.edu

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *29 September 2020*
Accepted: *03 December 2020*
Published: *25 January 2021*

Citation:
Li Z, Xu T, Zhang K, Deng H-W, Boerwinkle E and Xiong M (2021) Causal Analysis of Health Interventions and Environments for Influencing the Spread of COVID-19 in the United States of America. Front. Appl. Math. Stat. 6:611805. doi: 10.3389/fams.2020.611805

Given the lack of potential vaccines and effective medications, non-pharmaceutical interventions are the major option to curtail the spread of COVID-19. An accurate estimate of the potential impact of different non-pharmaceutical measures on containing, and identify risk factors influencing the spread of COVID-19 is crucial for planning the most effective interventions to curb the spread of COVID-19 and to reduce the deaths. Additive model-based bivariate causal discovery for scalar factors and multivariate Granger causality tests for time series factors are applied to the surveillance data of lab-confirmed Covid-19 cases in the US, University of Maryland Data (UMD) data, and Google mobility data from March 5, 2020 to August 25, 2020 in order to evaluate the contributions of social-biological factors, economics, the Google mobility indexes, and the rate of the virus test to the number of the new cases and number of deaths from COVID-19. We found that active cases/1,000 people, workplaces, tests done/1,000 people, imported COVID-19 cases, unemployment rate and unemployment claims/1,000 people, mobility trends for places of residence (residential), retail and test capacity were the popular significant risk factor for the new cases of COVID-19, and that active cases/1,000 people, workplaces, residential, unemployment rate, imported COVID cases, unemployment claims/1,000 people, transit stations, mobility trends (transit), tests done/1,000 people, grocery, testing capacity, retail, percentage of change in consumption, percentage of working from home were the popular significant risk factor for the deaths of COVID-19. We observed that no metrics showed significant evidence in mitigating the COVID-19 epidemic in FL and only a few metrics showed evidence in reducing the number of new cases of COVID-19 in AZ, NY and TX. Our results showed that the majority of non-pharmaceutical interventions had a large effect on slowing the transmission and reducing deaths, and that health interventions were still needed to contain COVID-19.

Keywords: COVID-19, causal inference, time series, control of the spread, transmission dynamics, public health interventions

INTRODUCTION

As of August 25, 2020, the number of cumulative cases of COVID-19 in the US exceeded 5,727,107 and included 170,305, deaths (John Hopkins Coronavirus Resource Center, https://coronavirus.jhu.edu/MAP.HTML), thus causing a devastating public health and economic crisis. Since the number of new cases in the US remains high (36,339 in the US on August 25, 2020) ((John Hopkins Coronavirus Resource Center, https://coronavirus.jhu.edu/MAP.HTML), curbing the spread of COVID-19 is urgently needed [1]. There is increasing recognition that many geographic, economic and environmental factors contribute to the outbreak of COVID-19. In the absence of vaccines and specifically effective medications, non-pharmaceutical public health interventions and personal hygiene practices are the only options to slow the spread of COVID-19 [2, 3]. The effects of the different factors and intervention measures on the spread of COVID-19 vary. Identifying key factors that most contribute to the rapid spread of COVID-19, and accurately estimating the potential impact of different non-pharmaceutical measures for containing COVID-19 are crucial for planning the most effective interventions to curb the spread of COVID-19 [4].

The widely used statistical methods for COVID-19 epidemiological factor analysis and evaluation of intervention measures include correlation analysis [2, 4–6], regression [7, 8], logistic regression [9] and a transmission dynamic model coupled with a linear model [10]. The most examined scalar factors consist of underlying health conditions such as high blood pressure, diabetes, stroke, cardiac or kidney diseases, and aging individuals [2, 11, 12], atmospheric temperature [5], age, gender, ethnicity, and population density [2, 13], airflow [2], and socioeconomics such as median income [9, 14].

The most explored non-pharmaceutical public health interventions and digital technologies for curbing the spread of COVID-19 include social distancing, case isolation and quarantine as well as closuring borders, schools travel restrictions, use of face-masks, and testing [15–18] and population surveillance, case identification, contact tracing, mobility data collection, and communication technology, which utilize billions of mobile phones and large online datasets to provide information for the evaluation of intervention strategies and to strengthen the curb of the spread of COVID-19 [17–19].

Although association analysis is of great importance for curbing the spread of COVID-19, association measures dependence between two variables or two sets of variables in the data, and use the dependence for prediction and evaluation of the effects of environmental, social-economic factors and public health interventions on the spread of COVID-19 [20, 21]. It is well recognized that association analysis is not a direct method to discover the causal mechanism of complex diseases. Association analysis may detect superficial patterns between intervention measures and transmission variables of COVID-19. Its signals provide limited information on the causal mechanism of the transmission dynamics of COVID-19 [22]. Association analysis has been a major paradigm for statistical evaluation of the effects of influencing factors and health interventions on the spread of COVID-19 [23]. Understanding the transmission mechanism of COVID-19 based on association analysis remains elusive. The question to uncover the transmission mechanisms of COVID-19 is causal in nature.

Distinguishing causation from association is an age-old problem. Methods for causation analysis that is one of the most challenging problems in science and technology need to be developed as an alternative to association analysis [24]. A number of researchers have performed causal analysis of COVID-19 to evaluate the causal effects of mobility, awareness, and temperature [22], social distancing [21], mobility [25], herd immunity [26], and mask use [27]. However, most causal analysis of COVID-19 have treated time series data as pseudo-cross-sectional data. In some cases, causal analysis of COVID-19 treated the data as time series; time series was assumed stationary. In practice, the number of new cases and the number of deaths from COVID-19 were nonstationary time series in most cases [28]. The environmental, social-economic and geographic factors, and intervention measures include two types of data: scalar variables and time series (stationary or nonstationary) variables.

The purpose of this paper is to develop a general framework for the causal analysis of COVID-19 in the US. The number of new cases and deaths from COVID-19 are taken as response variables. The factors and intervention measures are taken as potential causal variables. If the factor and intervention variables are scalar variables, the additive noise models (ANMs) [29] are used to test for causation between the response variable and potential causal variable where the number of new cases or deaths should be averaged over time. Most intervention measures are time series data. An essential difference between time series and cross-sectional data is that the time series data have temporal order, but cross-sectional data do not have any order. As a consequence, the causal inference methods for cross sectional data cannot be directly applied to time series data. Basic tools in statistical analysis are the raw of large numbers and the central limit theorem. Applications of these tools usually assume that all moment functions are constant. When the moment functions of the time series vary over time, the raw of large numbers and the central limit theorem cannot be applied. In order to use basic probabilistic and statistical theories, the nonstationary time series must be transformed to stationary time series [30].

A widely used concept of causality for time series data is Granger causality [31, 32]. Underlying the Granger causality is the following two principles:

(1) Effect does not precede the cause in time;
(2) The effect series contains unique causal series information which is not present elsewhere.

The multivariate linear Granger causality test will be used to test causality between the number of new cases and deaths from COVID-19 and environmental, economic and intervention time

series variables [33]. The proposed ANMs and multivariate linear Granger causality analysis methods are applied to the surveillance data of lab-confirmed COVID-19 cases in the US, UMD data, and Google mobility data from March 5, 2020 to August 25, 2020 in order to evaluate the contributions of social-biological factors, economics, the Google mobility indexes, and the rate of virus testing to the number of the new cases and number of deaths from COVID-19. Data and software for implementing the algorithms for causal analysis can be downloaded from our website https://sph.uth.edu/research/centers/hgc/software/xiong/.

MATERIAL AND METHODS

Nonlinear Additive Noise Models for Bivariate Causal Discovery

The ANMs are used for identifying causal effect of a factor or an intervention measure on the number of new cases or an intervention measure [29, 34]. Assume no confounding, no selection bias, and no feedback. Let Y be the average number of new cases or new deaths from COVID-19 and X be a scalar factor or an intervention measure such as gender, population density, ethnic group, among others. Consider a bivariate additive noise model $X \to Y$ where Y is a nonlinear function of X and independent additive noise E_Y:

$$Y = f_Y(X) + E_Y, \quad X \coprod E_Y, \tag{1}$$

where X and E_Y are independent. Then, the density $P_{X,Y}$ is said to be induced by the additive noise models (ANM) from X to Y [35]. In some cases, we may have the following alternative direction of the ANMs: $Y \to X$:

$$X = f_X(Y) + E_X, \quad Y \coprod E_X, \tag{2}$$

where Y and E_X are independent. If the density $P_{X,Y}$ is induced by the ANM $X \to Y$, but not by the ANM $Y \to X$, then the ANM $X \to Y$ is identifiable.

Assume that $n + m$ state data were sampled. Divide the dataset into a training data set by specifying $D_1 = \{Y_n, X_n\}$, $Y_n = [y_{1,\ldots,} y_n]^T$, $X_n = [x_1, \ldots, x_n]^T$ for fitting the model and a test data set $D_2 = \{\tilde{Y}_m, \tilde{X}_m\}$, $\tilde{Y}_m = [\tilde{y}_1, \ldots, \tilde{y}_m]^T$, $\tilde{X}_m = [\tilde{x}_1, \ldots, \tilde{x}_m]^T$ for testing the independence, where n is not necessarily equal to m.

Procedures for using the ANM to assess causal relationships between two variables are summarized below [34].

Step 1. Regress Y on X using the training dataset D_1 and non-parametric regression methods:

$$Y = \widehat{f}_Y(X) + E_Y \tag{3}$$

Step 2. Calculate the residual $\widehat{E}_Y = Y - \widehat{f}_Y(X)$ using the test dataset D_2 and test whether the residual $\widehat{E}_Y$ is independent of causal X to assess the ANM $X \to Y$.

Step 3. Repeat the procedure to assess the ANM $\to X$.

Step 4. If the ANM in one direction is accepted and the ANM in the other is rejected, then the former is inferred as the causal direction.

There are many non-parametric methods that can be used to regress Y on X or regress X on Y. For example, we can use neural networks [36], smoothing spline regression methods [37], B-spline [38] and local polynomial regression [39]. In this paper, the smoothing spline regression method was used to fit the regression models.

Covariance can be used to measure association but cannot be used to test independence between two variables with a non-Gaussian distribution (https://en.wikipedia.org/wiki/Correlation_and_dependence). A covariance operator that is a generalization of the finite dimensional covariance matrix to infinite dimensional feature space can be used to test for independence between two variables with arbitrary distributions. Specifically, we will use the Hilbert-Schmidt norm of the cross-covariance operator or its approximation, the Hilbert-Schmidt independence criterion (HSIC) to measure the degree of dependence between the residuals and potential causal variable and test for their independence [35, 40].

The covariance operator can be defined as

$$Cov\left(f(X), g(Y)\right) == \langle f, C(X,Y)g\rangle_{\mathcal{F}},$$

where f, g are any nonlinear functions and $C(X,Y)$ is the covariance operator and $\langle . \rangle_{\mathcal{F}}$ is an inner product in the $\mathcal{F}$ Hilbert space. The covariance operator is positive symmetric, which implies linearity and continuity. In addition, the covariance operator maps spaces in their dual spaces [41]. The Hilbert-Schmidt norm of the covariance operator can be used as criterion for assessing independence between two random variables and is called the Hilbert-Schmidt independence criterion (HSIC). The Hilbert-Schmidt norm of the centered covariance operator is defined as

$$HSIC^2(X,Y) = \|C_{XY}\|^2_{HS},$$

where $\|.\|_{HS}$ is the Hilbert-Schmidt norm.

We know that (Wang et al., 2018). $HSIC^2(X,Y) = 0$ if and only if X and Y are independent. $HSIC^2(X,Y)$ can be approximated by.

$$HSIC^2(X,Y) = \frac{1}{n^2}\mathrm{tr}(KHLH),$$

where n is a sample size, K and L = are $n \times n$ dimensional kernel matrices and $H = I - \frac{1_n 1_n^T}{n}$. We used the Gaussian kernel: $k(x,y) = e^{\frac{x-y^2}{2\sigma^2}}$, $\sigma > 0$ and polynomial kernel. The order- d polynomial kernel is defined as $k(x,y) = (x^T y + c)^d$. The fourth and sixth order polynomial kernels were used in this analysis. To test independence between the potential cause X and residual E_Y , we calculated $HSIC^2(X, E_Y)$ as follows.

$$HSIC^2(X,E_Y) = \frac{1}{n^2}\mathrm{tr}\left(K_X H K_{E_Y}\right),$$

$$K_{E_Y} = \begin{bmatrix} k_{E_Y}(\varepsilon_1, \varepsilon_1) & \cdots & k_{E_Y}(\varepsilon_1, \varepsilon_n) \\ \vdots & \vdots & \vdots \\ k_{E_Y}(\varepsilon_n, \varepsilon_1) & \cdots & k_{E_Y}(\varepsilon_n, \varepsilon_n) \end{bmatrix}$$

$$\text{and } K_X = \begin{bmatrix} k_X(x_1, x_1) & \cdots & k_X(x_1, x_n) \\ \vdots & \vdots & \vdots \\ k_X(x_n, x_1) & \cdots & k_X(x_n, x_n) \end{bmatrix},$$

$$\varepsilon_i = \widehat{E}_{Y(i)} = Y_i - \widehat{f}_Y(X_i),\ i = 1, \ldots, n,$$

In summary, the general procedure for testing independence between the average number of new cases or new deaths and the scalar factor or intervention measure is given as follows [34, 35]:

Step 1: Divide a data set into a training data set $D_n = \{Y_n, X_n\}$ for fitting the model and a test data set $D_m = \{\tilde{Y}_n, \tilde{X}_n\}$ for testing the independence.
Step 2: Use the training data set and any non-parametric regression methods

(a) Regress Y on X: $Y = f_Y(X) + E_Y$,
(b) Regress X on Y: $X = f_X(y) + E_X$.

Step 3: Use the test data set and any non-parametric regression methods that fits the test data set $D_m = \{\tilde{Y}_n, \tilde{X}_n\}$ to predict residuals:

(a) $\widehat{E}_{Y_X} = \tilde{Y} - \tilde{f}_Y(\tilde{X})$,
(b) $\widehat{E}_{X_Y} = \tilde{X} - \widehat{f}_X(\tilde{X})$.

Step 4: Calculate the dependence measures $HSIC^2(E_Y, X)$ and $HSIC^2(E_X, Y)$.
Step 5: Infer causal direction:

$$X \to Y \text{ if } HSIC^2(E_Y, X) < HSIC^2(E_X, Y); \quad (4)$$

$$Y \to X \text{ if } HSIC^2(E_Y, X) > HSIC^2(E_X, Y); \quad (5)$$

If $HSIC^2(E_Y, X) = HSIC^2(E_X, Y)$, then causal direction is undecided.

We do not have closed analytical forms for the asymptotic null distribution of the HSIC and hence it is difficult to calculate the p-values of the independence tests. To solve this problem, the permutation/bootstrap approach can be used to calculate the p-values of the causal test statistics. The null hypothesis is H_0 : no causations $X \to Y$ and $Y \to X$ (Both X and E_Y are dependent, and Y and E_X are dependent).

Calculate the test statistic

$$T_C = \left|HSIC^2(E_Y, X) - HSIC^2(E_X, Y)\right|. \quad (6)$$

Assume that the total number of permutations is n_p. For each permutation, we fix x_i, $i = 1, \ldots, m$ and randomly permutate y_i, $i = 1, \ldots, m$. Then, fit the ANMs and calculate the residuals E_X^i, E_Y^i, $i = 1, \ldots, n$ and test statistic T_c. Repeat above procedures n_p times. The p-values are defined as the proportions of the statistic T_C (computed on the permuted data) greater than or equal to $\widehat{T}_C$ (computed on the original test data D_m).

Each state was a sample. Since the sample sizes were small (only 50), the p-value for declaring significance was 0.05 without Bonferroni correction for multiple comparison.

Multivariate Linear Granger Causality Test

Before performing multivariate linear Grander causality test, we first need to transform nonstationary time series to stationary time series.

Consider an m-variable VAR with p lags:

$$Y_t = \mu + \sum_{i=1}^{p} A_i Y_{t-i} + \varepsilon_t, \quad (7)$$

where Y_t is a m dimensional vector, μ is a mean, the A_i $(i = 1, \ldots, p)$ are $m \times m$ coefficient matrices and m dimensional residual vector ε_t is assumed to have mean zero $(E[\varepsilon_t] = 0)$, with no autocorrelation $(E[\varepsilon_t \varepsilon_{t-s}^T] = 0)$, but can be correlated across equations $(E[\varepsilon_t \varepsilon_t^T] = \Sigma)$.

Vector error correction model (VECM) consists of first differences of cointegrated $I(1)$ variables, their lags, and error correction terms:

$$\Delta Y_t = \mu + \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \varepsilon_t, \quad (8)$$

where matrixes Π and Φ_i $(i = 1, \ldots, p-1)$ are functions of matrices A_i $(i = 1, \ldots, p)$.

When two non-stationary variables are cointegrated, the VAR model should be augmented with an error correction term for testing the Granger causality (Engle and Granger, 1987).

The VECM can be reduced to

$$\Delta Y_t = A_0 + A(L)\Delta Y_{t-1} + \alpha(ECM_{t-1}) + \varepsilon_t, \quad (9)$$

where

$$\Delta Y_t = \begin{bmatrix} \Delta Y_{1t} \\ \vdots \\ \Delta Y_{mt} \end{bmatrix},\ A_0 = \begin{bmatrix} A_{10} \\ \vdots \\ A_{m0} \end{bmatrix},\ A(L) = \begin{bmatrix} A_{11}(L) & \cdots & A_{1m}(L) \\ \vdots & \vdots & \vdots \\ A_{m1}(L) & \cdots & A_{mm(L)} \end{bmatrix},$$

$$\alpha = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1k} \\ \vdots & \vdots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mk} \end{bmatrix},$$

$$ECM_{t-1} = \beta^T Y_{t-1} + Ct,\ Y_{t-1} = \begin{bmatrix} Y_{1,t-1} \\ \vdots \\ Y_{m,t-1} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \vdots & \vdots \\ \beta_{m1} & \cdots & \beta_{mk} \end{bmatrix},\ C = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}.$$

Consider two non-stationary time series, $X_t = [X_{1t}, \ldots, X_{m,t}]^T$ and $Y_t = [Y_{1t}, \ldots, Y_{mt}]^T$. Let $m = m_1 + m_2$.

Suppose that X_t and Y_t are cointegrated with the residuals ECM_t. The VECM model for testing the Granger causality is given by

$$\begin{bmatrix} \Delta X_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} A_x \\ A_y \end{bmatrix} + \begin{bmatrix} A_{xx}(L) & A_{xy}(L) \\ A_{yx}(L) & A_{yy}(L) \end{bmatrix} \begin{bmatrix} \Delta X_{t-1} \\ \Delta Y_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} ECM_{t-1} + \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \quad (10)$$

where A_x and A_y are m_1 and m_2 dimensional vectors of intercept terms, respectively, $A_{xx}(L)$, $A_{xy}(L)$, $A_{yx}(L)$ and $A_{yy}(L)$ are $n_1 \times n_1$, $n_1 \times n_2$, $n_2 \times n_1$ and $n_2 \times n_2$ dimensional matrices of lag polynomials, respectively, α_x and α_y are n_1 and n_2 dimensional coefficient vectors for the error correction term ECM_{t-1}, respectively. The lag length was selected using the two-stage procedure [42].

There are four different cases of causal relationships between two vectors of time series X_t and Y_t [33].

(1) If $A_{xy}(L)$ is significantly different from the zero, while $A_{yx}(L)$ shows no significantly different from zero, then there exists a unidirectional Ganger causality from time series Y_t to X_t;
(2) If $A_{yx}(L)$ is significantly different from zero, while $A_{xy}(L)$ shows no significantly difference from zero, then there exists a unidirectional Ganger causality from X_t to Y_t;
(3) If both coefficients $A_{xy}(L)$ and $A_{yx}(L)$ are significantly different from zero, then there exists bidirectional Granger causality between X_t and Y_t;
(4) If both coefficients $A_{xy}(L)$ and $A_{yx}(L)$ are not significantly different from zero, then X_t and Y_t are not rejected to be independent.

The four statements imply that Ganger causal relationships between X_t and Y_t depend on the coefficients $A_{xy}(L)$ and $A_{yx}(L)$. Therefore, the null hypotheses for testing the Ganger causality between X_t and Y_t are.

(1) $H_0^1 : A_{xy}(L) = 0$,
(2) $H_0^2 : A_{yx}(L) = 0$, and
(3) Both H_0^1 and $H_0^2 : A_{xy}(L) = 0$ and $A_{yx}(L) = 0$.

Likelihood ratio tests for multivariate Granger causality are given by.

(1) The likelihood ration statistics for testing the null hypothesis: $H_0^1 : A_{xy}(L) = 0$ is

$$T_{ml} = (T - (np+1))\left(\log|\widehat{\Sigma}_0| - \log|\widehat{\Sigma}|\right), \quad (11)$$

which is asymptotically distributed as a central $\chi^2_{(n_1 \times n_2 \times p)}$ under the hull hypothesis H_0^1.

(2) The likelihood ration statistics for testing the null hypothesis: $H_0^2 : A_{yx}(L) = 0$ is

$$T_{ml} = (T - (np+1))\left(\log|\widehat{\Sigma}_0| - \log|\widehat{\Sigma}|\right),$$

which is asymptotically distributed as a central $\chi^2_{(n_2 \times n_1 \times p)}$ under the hull hypothesis H_0^2.

(3) The likelihood ration statistics for testing the null hypothesis: H_0^1 and $H_0^2 : A_{xy}(L) = 0$ and $A_{yx}(L) = 0$ is

$$T_{ml} = (T - (np+1))\left(\log|\widehat{\Sigma}_0| - \log|\widehat{\Sigma}|\right),$$

which is asymptotically distributed as a central $\chi^2_{(2n_2 \times n_1 \times p)}$ under the hull hypothesis H_0^1 and H_0^2.

The total number of variables to be tested was 18. The p-value for declaring significance after Bonferroni correction was 0.0028.

Data Collection

Data on the number of new cases and new deaths of COVID-19 across the 50 states in the US were obtained from John Hopkins Coronavirus Resource Center (https://coronavirus.jhu.edu/MAP.HTML). Google mobility indexes were downloaded from Google COVID-19 Community Mobility Reports (https://www.google.com/covid19/mobility/). Comprehensive data and insights on COVID-19's impact on mobility, economy, and society were downloaded from the University of Maryland COVID-19 Impact Analysis Platform (https://data.covid.umd.edu) [43, 44]. All data were collected from March 5, 2020 to August 25, 2020.

RESULTS

Test for Scalar Potential Causes

The scalar variables tested for causation of the new cases and deaths from COVID-19 in the US included the number of contact tracing workers per 100,000 people, percent of population above 60 years of age, median income, population density, percentage of African Americans, percentage of Hispanic Americans, percentage of males, employment density, number of points of interests for crowd gathering per 1,000 people, number of staffed hospital beds per 1,000 people, and number of ICU beds per 1,000 people. The number of new cases and deaths were averaged over time. Each state was a sample. Since the sample sizes were small, the p-value for declaring significance was 0.05 without Bonferroni correction for multiple comparison. The p-values for testing 11 scalar potential causes of the number of new cases and deaths from COVID-19 in the US using both Gaussian kernel and polynomial kernels were summarized in **Table 1** where minimum of three p-values using Gaussian kernel and fourth order and sixth order polynomial kernels were listed and only one causal direction was observed. We observed from **Table 1** that population density (minimum of p-value < 0.0002, which was due to Gaussian kernel), percentage of males (minimum of p-value < 0.03, which was due to Gaussian kernel) and Percentage of Hispanic Americans (minimum of p-value < 0.0325, which was due to sixth order polynomial kernel) showed significant evidence of causing the spread of COVID-19. Employment density (minimum of p-values < 0.0223, which was due to sixth polynomial kernel), Percentage of African American (minimum of p-values < 0.024, which is due to Gaussian kernel), population density (minimum of p-values < 0.025, which was due to Gaussian kernel) and Percentage of males (minimum of p-values <0.0377, which was due to fourth order polynomial) showed significant evidence of causing deaths due to COVID-19. percentage of Hispanic Americans (minimum of

TABLE 1 | *p*-values for testing 10 scalar potential causes of the number of new cases and deaths of COVID-19 in the US.

Risk factor	*p*-value							
	Gaussian kernel		Polynomial (order 6)		Polynomial (order 4)		Minimum	
	New cases	Deaths	New cases	Deaths	New cases	Deaths	New cases	Deaths
Percent of population above the age of 60 years	0.4634	0.1530	0.3472	0.1130	0.2532	0.1547	0.2532	0.1130
Median income	0.1109	0.0760	0.1202	0.0654	0.1713	0.1832	0.1109	0.0654
Percentage of African Americans	0.6526	0.0240	0.9326	0.0710	0.3459	0.0343	0.3459	0.0240
Percentage of Hispanic Americans	0.0575	0.0640	0.0325	0.0520	0.0422	0.1473	0.0325	0.0520
Percentage of males	0.0300	0.1440	0.0403	0.0542	0.0401	0.0377	0.0300	0.0377
Population density	0.0002	0.0250	0.0327	0.0359	0.0265	0.0952	0.0002	0.0250
Employment density	0.4571	0.0590	0.3536	0.0223	0.3485	0.0248	0.3485	0.0223
Number of staffed hospital beds per 1,000 people	0.6732	0.3130	0.5352	0.3224	0.6506	0.2374	0.5352	0.2374
Number of ICU beds per 1,000 people	0.5134	0.4860	0.7516	0.4438	0.6114	0.9037	0.5134	0.4438
Number of contact tracing workers per 100,000 people	0.4203	0.8190	0.4102	0.7215	0.3907	0.7336	0.3907	0.7215

p-value < 0.052, which was due to sixth order polynomial kernel) were close to significance level 0.05 for causing death.

Population density was an important risk factor for both the spread and death from COVID-19. High density resulted in closer contact, stronger interaction among residents and lower social distancing, which facilitated the spread and increased the death rate from COVID-19 [45–48]. However, our results were contradictory with the conclusion of Hamidi et al. [49]. Some literature also confirmed that high proportion of African Americans caused a high rate of deaths [48, 50, 51]. Our results concluded that percentage of Hispanic Americans was a risk factor for the spread and a weak risk factor for death from COVID-19, while the literature showed stronger evidence that Hispanic communities were highly vulnerable to COVID-19 [52].

The second most significant demographic risk factor for the spread of COVID-19 was percentage of males. We found higher COVID-19 morbidity in males than females. However, we did not find higher COVID-19 mortality in males than females.

It was reported that higher COVID-19 mortality in males than females can be due to the following factors [53]. The first factor was higher expression of angiotensin-converting enzyme-2 (ACE two; receptors for coronavirus) in males than females. The second factor was sex-based immunological differences due to sex hormone and the X chromosome.

Test for Granger Causality

Daily mobility and social distancing data from a COVID-19 impacted the analysis platform, including four categories: category A: mobility and social distancing, category B: COVID and health, category C: economic impact, and category D: vulnerable population. A total of 12 temporal metrics in four categories and 12 metrics from the COVID-19 impact analysis platform, six daily Google Community Mobility indexes and protest attendee data that captured real-time trends in movement patterns for each state in the US were included in the analysis to test for Granger causality between these risk factors, health intervention measures and the number of new cases and deaths from COVID-19 across 50 states in the US [44, 54]. The total number of variables to be tested was 18. The p-value for declaring significance after Bonferroni correction was 0.0028.

All 18 metrics except for protest attendee showed high significance in causing a reduction of the new cases of COVID-19 in 19 less affected states: VT, WY, ME, AK, NH, WV, ND, SD, NM, RI, DE, KY, KS, CT, CO, IA, WA, WI, and MS. Most of these states were less populated. However, although CA was most affected and the most populated state, all 18 metrics except protest attendance showed a strong significance in causing rapid spread of COVID-19 (**Table 2** and **Supplementary Table S1**). To provide complete causal testing information, we listed the values of statistics for testing 18 temporal potential causes of the number of new cases of COVID-19 across 50 states in the US in **Supplementary Table S2**.

All 18 metrics showed no significance in causing reduction of the new cases of COVID-19 in Florida. The majority of the 18 metrics did not demonstrate evidence that they can significantly mitigate the spread of COVID-19 in most the affected states such as TX, NY, GA, IL, AZ, NJ, NC, and TN. These 10 states were in the top largest states by population in the US. Public health intervention measures such as closing schools and businesses, avoiding public gatherings, restricting traffic, placing residents to stay-at-home and adherence to guidelines were less well implemented or difficult to implement homogeneously due to large populations and geographical areas [55]. These results also explained why the number of new cases of COVID-19 in these states was high and confirmed by several studies [56–58].

Table 3 summarized the ranges of p-values, **Supplementary Tables S3, S4** summarized all p-values and values of statistics for testing 18 temporal potential causes of the number of new deaths from COVID-19 across 50 states in the US, respectively. All 18 metrics except for protest attendance showed high significant evidence for causing a reduction of new deaths across 50 states except for Michigan (MI) in the US. Our results suggested that a cascade of causes led to the COVID-19 tragedy in the US.

Table 4 listed the most significant risk factor for the new cases of COVID-19 in each of the 50 states in the US. Active Cases/1,000 People, workplaces, number of tests completed/1,000 people, imported COVD cases, unemployment rate and unemployment claims/1,000 people, mobility trends for places of residence (residential), retail and recreation, mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters (retail) and test capacity

TABLE 2 | *p*-values testing 18 temporal potential causes of the number of new cases of COVID-19 in the top 10 most affected states and bottom 10 less affected states in the US.

State	CA	FL	TX	NY	GA	IL	AZ	NJ	NC	TN
Number of cumulative cases	673,095	605,502	586,730	430,774	258,354	224,887	199,273	190,021	157,741	145,417
Miles/person	7.7E-08	3.1E-02	2.3E-03	9.1E-02	2.8E-04	4.9E-04	7.3E-05	4.7E-03	9.2E-06	5.9E-06
Population	1.6E-06	3.6E-02	5.1E-03	6.1E-02	3.2E-04	5.2E-05	1.9E-04	1.2E-03	2.5E-04	2.9E-05
% Change in consumption	1.4E-05	5.5E-02	4.7E-03	1.8E-03	8.7E-04	1.8E-04	2.2E-04	9.6E-07	1.1E-04	1.1E-04
Social distancing index	8.0E-06	9.1E-02	8.8E-03	4.0E-02	3.4E-04	4.6E-04	5.9E-04	1.5E-05	3.5E-03	2.7E-04
Unemployment claims	4.3E-05	6.8E-02	1.7E-02	1.1E-03	2.3E-03	8.0E-04	1.1E-03	2.2E-07	4.2E-03	3.9E-04
Unemployment rate	3.7E-08	2.6E-02	4.4E-03	1.7E-01	4.2E-04	5.3E-06	2.4E-04	2.8E-03	1.5E-04	4.8E-05
% Working from home	5.2E-07	3.7E-02	2.4E-03	3.4E-02	2.4E-04	6.6E-06	2.2E-05	2.8E-04	1.2E-05	2.0E-06
Active cases/1,000 people	2.3E-18	7.2E-02	3.1E-06	5.8E-01	2.8E-10	2.3E-05	7.1E-07	9.5E-04	7.5E-14	3.4E-12
Testing capacity	9.0E-05	7.2E-02	1.3E-02	3.0E-02	2.4E-03	2.1E-05	4.1E-04	1.2E-04	1.7E-03	1.8E-04
Tests done/1,000 people	1.2E-14	3.6E-02	1.1E-05	1.0E-01	2.5E-09	1.5E-03	7.0E-05	3.8E-03	1.3E-07	3.7E-11
Imported COVID cases	1.3E-16	4.5E-01	1.1E-04	2.6E-01	1.7E-07	1.3E-03	5.1E-04	5.9E-04	1.7E-06	4.4E-12
Retail	1.3E-05	9.6E-02	2.2E-02	2.9E-02	4.4E-03	5.9E-04	1.4E-03	3.4E-07	4.5E-03	6.1E-04
Grocery	1.5E-05	1.1E-01	4.1E-03	2.0E-01	5.8E-04	1.5E-05	2.7E-04	9.0E-06	2.6E-04	4.5E-05
Parks	2.4E-08	8.2E-02	6.3E-03	4.1E-02	3.1E-02	6.4E-02	5.6E-03	4.0E-03	8.6E-03	1.7E-05
Transit	5.8E-06	6.4E-02	7.3E-03	2.0E-02	3.3E-04	2.2E-07	3.0E-04	9.2E-06	1.2E-03	3.3E-04
Workplaces	1.5E-05	9.6E-03	2.6E-03	3.5E-03	2.1E-05	5.8E-08	9.3E-06	2.2E-05	1.2E-04	6.1E-05
Residential	7.0E-05	1.7E-02	7.2E-03	9.6E-05	3.6E-04	1.7E-07	6.5E-05	1.1E-06	2.8E-04	1.5E-04
Protest	3.6E-01	9.6E-01	6.6E-01	7.7E-01	8.0E-03	5.1E-01	3.8E-01	1.7E-01	4.1E-01	2.3E-03
State	**SD**	**ND**	**WV**	**NH**	**HI**	**MT**	**AK**	**ME**	**WY**	**VT**
Number of cumulative cases	11,559	10,229	9,395	7,150	6,984	6,624	5,666	4,368	3,634	1,572
Miles/person	3.7E-08	1.1E-04	1.3E-09	3.8E-14	9.5E-03	5.8E-04	2.8E-05	8.2E-11	2.7E-13	4.8E-14
Population	4.2E-08	5.4E-04	6.6E-08	1.7E-18	2.0E-02	1.2E-03	5.4E-04	1.7E-12	1.7E-11	6.8E-16
% Change in consumption	2.1E-08	1.1E-04	4.6E-10	3.6E-19	3.1E-02	1.4E-04	7.2E-05	1.2E-11	4.2E-10	7.5E-19
Social distancing index	5.7E-05	9.8E-03	5.0E-06	2.8E-19	3.3E-02	6.5E-03	2.4E-03	1.6E-08	1.9E-09	1.6E-19
Unemployment claims	5.0E-06	4.3E-03	1.8E-06	3.6E-16	3.1E-02	3.3E-03	1.4E-03	1.9E-07	1.0E-08	2.1E-32
Unemployment rate	3.8E-08	1.3E-03	2.5E-06	4.8E-26	2.9E-02	2.7E-03	2.8E-04	7.1E-18	8.0E-12	3.8E-14
% Working from home	1.5E-09	1.3E-05	1.5E-12	1.4E-16	2.3E-03	1.1E-04	8.3E-04	1.3E-11	2.4E-15	1.8E-16
Active cases/1,000 people	3.9E-09	3.3E-13	1.5E-16	6.2E-27	1.3E-08	4.8E-09	3.2E-07	6.8E-13	2.1E-21	6.9E-15
Testing capacity	1.1E-07	3.4E-05	5.4E-07	1.3E-22	2.1E-02	3.6E-03	2.0E-03	1.7E-14	6.0E-10	1.5E-23
Tests done/1,000 people	4.5E-07	4.6E-11	2.1E-16	1.9E-12	3.8E-05	1.9E-11	6.3E-09	7.4E-07	1.5E-19	5.7E-13
Imported COVID cases	2.7E-07	6.8E-09	9.6E-13	9.0E-32	7.4E-06	2.7E-08	2.0E-06	1.3E-14	5.4E-19	2.5E-13
Retail	1.3E-07	6.1E-03	4.4E-06	4.8E-19	2.9E-02	2.7E-04	3.1E-04	2.0E-09	5.4E-14	1.4E-20
Grocery	1.3E-04	3.7E-04	1.1E-07	1.7E-22	2.0E-02	2.3E-04	7.5E-04	4.3E-10	3.3E-13	5.6E-20
Parks	3.9E-06	4.9E-04	1.6E-06	3.9E-12	1.9E-02	1.6E-07	2.8E-04	7.8E-07	1.6E-19	3.1E-12
Transit	3.3E-06	3.2E-03	7.4E-09	1.3E-22	2.2E-02	1.6E-04	5.6E-04	6.3E-11	2.0E-13	3.0E-19
Workplaces	3.9E-12	2.3E-04	1.2E-07	1.4E-23	1.8E-02	4.6E-04	4.7E-04	9.8E-15	1.9E-09	2.2E-19
Residential	3.5E-09	1.6E-03	2.0E-06	3.0E-24	1.6E-02	1.4E-03	6.3E-04	4.9E-12	5.6E-08	3.1E-22
Protest	2.4E-01	2.6E-01	9.7E-01	1.1E-03	1.3E-02	2.1E-01	9.6E-03	8.0E-02	8.3E-01	1.6E-01

Retail: Retail and recreation, mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters.
Grocery: Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.
Parks: Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens.
Transit: Transit stations, mobility trends for places like public transport hubs such as subway, bus, and train stations.
Workplaces: Mobility trends for places of work.
Residential: Mobility trends for places of residence.
Test Rate: Ratio of the number of individuals who have taken the virus test over the total population in the region.
Attendee: Number of attendees in the protest.

were the most significant risk factors for the new cases of COVID-19 in 23, 7, 6, 5, 4, 2, 1 and 1 states of the US, respectively.

Table 5 summarized the most significant risk factor for the deaths from COVID-19 in each of the 50 states in the US. Active Cases/1,000 people, workplaces, residential, unemployment rate, imported COVID cases, unemployment claims/1,000 people, transit, test done/1,000 people, grocery, testing capacity, retail, percentage of change in consumption, percentage of working from home were the most significant risk factor for the deaths of COVID-19 in 17, 10, 4, 4, 3, 2, 2, 2, 1, 1, 1, 1 states, respectively. We also observed that the number of protest attendees showed mild significant evidence to cause increasing the number of new cases of COVID-19 in KY (p-value < 0.00012), KS (p-value < 0.00026), NH (p-value < 0.00108), MA (p-value < 0.0016) and TN (p-value < 0.0024) or to cause more deaths from COVID-19 in OR (p-value < 5.11 E-05), TX (p-value < 0.00017), ME (p-value < 0.00028), KS (p-value < 0.00061), MI (p-value < 0.0015), OH (p-value < 0.0021) and NC (p-value < 0.0023).

In some cases, two causal directions may occur. To examined whether the COVID-19 cases and deaths caused daily mobility and social distancing metrics to change, we summarized the values of statistics and p-values for testing COVID-19 new case

TABLE 3 | Ranges of *p*-values for testing 18 temporal potential causes of the number of new deaths from COVID-19 across 50 states in the US.

State	*p*-value		State	*p*-value	
	Lower bound	Upper bound		Lower bound	Upper bound
AK	1.47E-18	3.95E-24	MT	7.66E-15	4.41E-25
AL	7.45E-11	1.10E-18	NC	2.93E-08	8.95E-18
AR	1.17E-12	1.54E-30	ND	2.54E-17	1.04E-20
AZ	4.35E-11	1.29E-27	NE	5.11E-15	2.62E-23
CA	1.10E-04	1.54E-13	NH	6.67E-15	3.41E-28
CO	1.41E-21	6.44E-26	NJ	1.80E-03	1.96E-09
CT	6.73E-07	3.13E-16	NM	7.20E-07	1.15E-16
DE	1.20E-03	1.35E-06	NV	9.00E-09	1.45E-17
FL	2.50E-04	1.47E-15	NY	4.30E-02	1.52E-05
GA	6.53E-07	5.81E-17	OH	1.82E-08	2.67E-17
HI	1.77E-15	1.14E-20	OK	4.29E-07	1.89E-19
IA	6.78E-06	8.06E-15	OR	3.37E-15	4.50E-26
ID	1.96E-08	7.11E-18	PA	3.86E-15	1.11E-23
IL	3.81E-05	1.34E-12	RI	4.26E-09	2.16E-22
IN	1.78E-07	3.39E-17	SC	8.97E-08	6.75E-25
KS	8.74E-30	1.42E-40	SD	5.04E-18	6.01E-25
KY	1.89E-13	1.40E-24	TN	7.28E-09	2.06E-24
LA	7.71E-11	2.38E-20	TX	6.22E-07	2.59E-23
MA	1.25E-03	3.92E-04	UT	3.90E-13	8.12E-29
MD	3.80E-02	7.63E-06	VA	3.39E-17	2.61E-16
ME	2.29E-15	4.29E-25	VT	3.39E-07	1.23E-29
MI	0.28	2.07E-05	WA	5.67E-16	3.83E-29
MN	2.40E-04	8.89E-11	WI	2.21E-10	4.77E-29
MO	1.73E-11	7.85E-20	WV	9.61E-15	7.81E-22
MS	2.94E-08	1.49E-19	WY	7.93E-26	2.39E-28

and death potential causes of six Google mobility indexes across 50 states in the US in **Supplementary Tables S5–7** and DS8, respectively. We did observe the opposite causal direction. **Table 6** presented the smallest *p*-value across 50 states in the US for testing two causal directions: from Google mobility indexes to the number of new cases of COVID-19 and from the number of new cases of COVID-19 to the Google mobility indexes. We observed significant causation from the number of COVID-19to the Google mobility indexes in some states. However, the *p*-values for testing causation from the number of new cases of COVID-19 to the Google mobility indexes were much larger than that from the Google mobility indexes to the number of new cases of COVID-19. The causal pattern for the COVID-19 deaths was similar to the number of new cases of COVID-19.

To illustrate the causal relationships between the risk factors and the number of new cases and deaths from COVID-19, we plotted **Figures 1** and **2**. **Figure 1** plotted the social distance index curves as a function of time from March 5, 2020 to August 25, 2020 in Florida (FL) and Rhode Island (RI). **Figure 1** showed that the social distance index in FL was much higher than that in RI state, which resulted in the larger number of new cases of COVID-19 in FL than that in RI. **Figure 2** showed the number of imported COVID-19 cases as a function of time from March 5, 2020 to August 25, 2020 in Maryland (MD) and Wyoming (WY). We observed a huge difference in the number of imported COVID-19 cases between MD and RI. The very low number of imported cases of COVID-19 in WY resulted in the very low number of deaths from COVID-19 in WY, while the high number of imported cases in MD state led to the increased deaths from COVID-19 in MD. These results were consistent with the finding in the literature. It was reported that strong

TABLE 4 | The most significant risk factor for the new cases of COVID-19 in each of 50 states in the US.

State	Risk factor	*p*-value	State	Risk factor	*p*-value
AK	Tests done/1,000 people	6.27E-09	MT	Tests done/1,000 people	1.87E-11
AL	Active cases/1,000 people	5.40E-08	NC	Active cases/1,000 people	7.49E-14
AR	Active cases/1,000 people	4.49E-14	ND	Active cases/1,000 people	3.35E-13
AZ	Active cases/1,000 people	7.12E-07	NE	Unemployment rate	6.10E-10
CA	Active cases/1,000 people	2.29E-18	NH	Imported COVID cases	9.04E-32
CO	Work places	1.48E-17	NJ	Unemployment claims/1,000 people	2.15E-07
CT	Test capacity	5.46E-25	NM	Active cases/1,000 people	5.68E-13
DE	Residential	3.64E-15	NV	Imported COVD cases	2.36E-09
FL	Work places	9.60E-03	NY	Residential	9.57E-05
GA	Active cases/1,000 people	2.76E-10	OH	Active cases/1,000 people	6.42E-10
HI	Active cases/1,000 people	1.31E-08	OK	Tests done/1,000 people	4.14E-10
IA	Active cases/1,000 people	5.07E-14	OR	Active cases/1,000 people	3.87E-12
ID	Imported COVD cases	4.53E-10	PA	Work places	4.21E-09
IL	Work places	5.82E-08	RI	Imported COVD cases	1.02E-27
IN	Active cases/1,000 people	1.12E-09	SC	Active cases/1,000 people	1.58E-04
KS	Tests done/1,000 people	1.21E-32	SD	Work places	3.89E-12
KY	Imported COVID cases	4.30E-26	TN	Active cases/1,000 people	3.45E-12
LA	% Working from Home	1.58E-19	TX	Active cases/1,000 people	3.05E-06
MA	Retail	3.20E-10	UT	Active cases/1,000 people	2.04E-07
MD	Work places	6.35W-09	VA	Active cases/1,000 people	2.08E-08
ME	Unemployment rate	7.15E-18	VT	Unemployment claims/1,000 people	2.07E-32
MI	Work places	1.16E-10	WA	Active cases/1,000 people	2.65E-21
MN	Active cases/1,000 people	9.01E-07	WI	Active cases/1,000 people	5.19E-12
MO	Tests done/1,000 people	2.42E-09	WV	Active cases/1,000 people	1.53E-16
MS	Tests done/1,000 people	2.16E-12	WY	Active cases/1,000 people	2.13E-21

TABLE 5 | The most significant risk factor for deaths of COVID-19 in each of 50 states in the US.

State	Risk factor	*p*-value	State	Risk factor	*p*-value
AK	Active cases/1,000 people	3.95E-24	MT	Active cases/1,000 people	4.41E-25
AL	Active cases/1,000 people	1.10E-18	NC	Work places	8.95E-18
AR	Tests Done/1,000 people	1.54E-30	ND	Unemployment rate	1.04E-20
AZ	Active cases/1,000 people	1.29E-27	NE	Work places	2.62E-23
CA	Work places	1.54E-13	NH	Imported COVID cases	3.41E-28
CO	Residential	6.44E-26	NJ	Residential	1.96E-09
CT	Active cases/1,000 people	3.12E-16	NM	Unemployment rate	1.15E-16
DE	Retail	1.35E-06	NV	Active cases/1,000 people	1.45E-17
FL	Active cases/1,000 people	1.47E-15	NY	% Change in Consumption	1.52E-05
GA	Work places	5.81E-17	OH	Unemployment rate	2.67E-17
HI	Active cases/1,000 people	1.14E-20	OK	Work places	1.89E-19
IA	Unemployment rate	8.06E-15	OR	% Working from Home	4.50E-26
ID	Active cases/1,000 people	7.11E-18	PA	Imported COVID cases	1.11E-23
IL	Residential	1.34E-12	RI	Imported COVID cases	2.16E-22
IN	Work places	3.39E-17	SC	Active cases/1,000 people	6.75E-25
KS	Grocery	1.42E-40	SD	Active cases/1,000 people	6.01E-24
KY	Work places	1.40E-24	TN	Test Done/1,000 people	2.06E-24
LA	Transit	2.38E-20	TX	Active cases/1,000 people	2.59E-23
MA	Active cases/1,000 people	3.92E-09	UT	Active cases/1,000 people	8.12E-29
MD	Grocery	7.60E-06	VA	Work places	2.61E-16
ME	Residential	4.29E-25	VT	Unemployment claims/1,000 people	1.23E-29
MI	Unemployment claims/1,000 people	2.07E-05	WA	Transit	3.83E-29
MN	Testing capacity	8.89E-11	WI	Work places	4.77E-29
MD	Work places	7.85E-20	WV	Active cases/1,000 people	7.81E-22
MS	Active cases/1,000 people	1.49E-19	WY	Active cases/1,000 people	2.39E-28

TABLE 6 | Lower bound of *p*-values for testing causation from Google mobility indexes to COVID-19 case and from COVID-19 case to Google mobility indexes.

Cause	Effect	Smallest *p*-value	Cause	Effect	Smallest *p*-value
Retail	COVID-19 case	7.67E-26	COVID-19 case	Retail	1.29E-04
Grocery		3.70E-25		Grocery	4.85E-11
Parks		3.87E-23		Parks	1.80E-12
Transit		1.90E-27		Transit	1.29E-04
Workplaces		1.07E-25		Workplaces	3.99E-06
Residential		1.09E-25		Residential	2.99E-07

interventions would substantially decrease the number of deaths (Davies et al. 2020; Gagnon et al. 2020).

The COVID-19 case curves as a function of time from March 5, 2020 to August 25, 2020 in FL, RI, MD and WY were plotted in **Figure 3**, respectively. The response of COVID-19 to public health intervention which were measured by Google mobility indexes were usually delayed.

DISCUSSION

Causal inference for COVID-19 is essential for selecting and implementing public intervention measures and understanding the role of the demographics in curbing the spread and reducing the deaths from COVID-19. In this paper, we systematically addressed the issues in identifying causal risk factors and evaluating the causal effects of risk factors and intervention measures on the spread and deaths from COVID-19 in the US. Risk factors and intervention measures included scalar variables and temporal variables. The ANMs were used to test for causal relationships between scalar risk factors and the average number of new cases or deaths from COVID-19 in the US. Transmission of COVID-19 is a dynamic system. Many risk factors and intervention measures are temporal variables. The Granger Causality Test was used to reveal the causal relationships between the temporal risk factors and intervention measures, and the number of new cases or new deaths from COVID-19 across the 50 states in the US.

The demographic risk factors were the major part of the scalar risk factors in the causal analysis of COVID-19. We found that population density was the most significant causal factor of both new cases and death from COVID-19. Population density measured the average number of people per square kilometer living in a built-up area. Densely populated states generated conditions where COVID-19 can spread quickly and undetected in the densely populated areas and created high levels of vulnerability. The second significant demographic factor was percentage of males. Our data

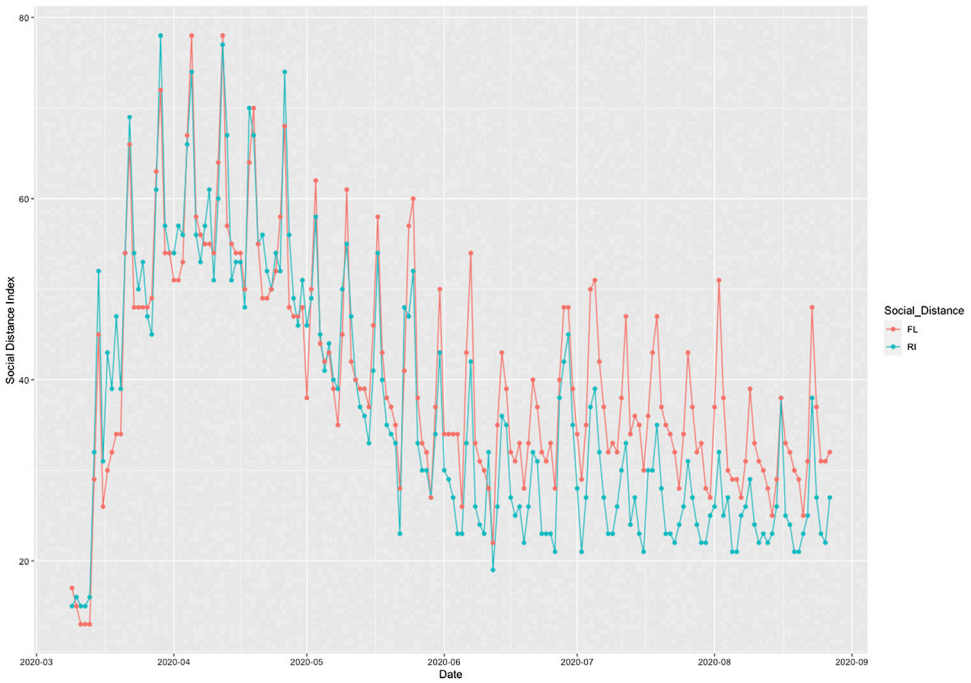


FIGURE 1 | Social distance index curves as a function of time from March 5, 2020 to August 25, 2020 in Florida (FL) (red color) and Rhode Island (RI) (blue color).

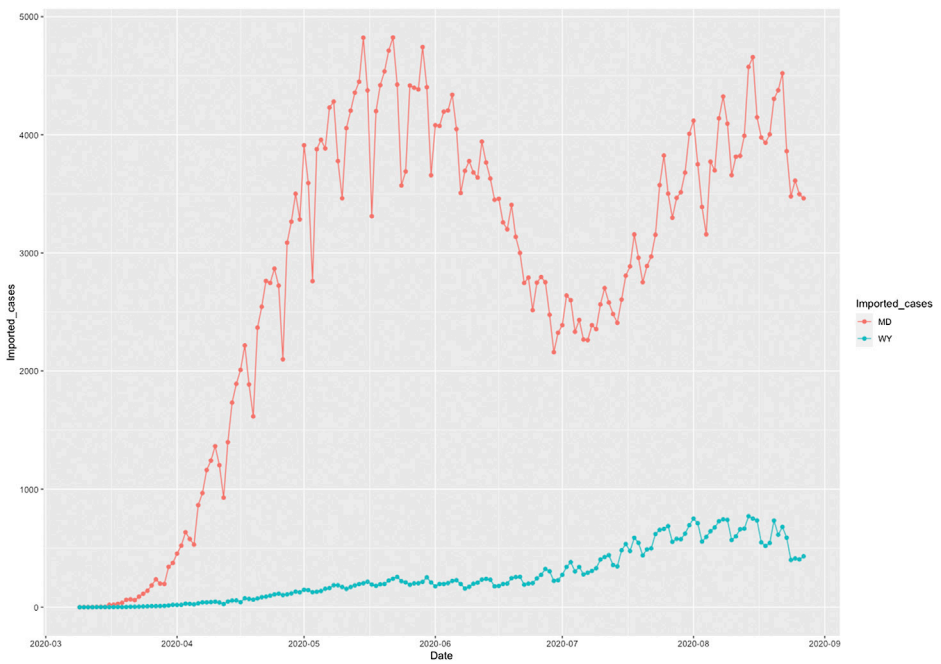


FIGURE 2 | Number of imported COVID-19 cases as a function of time from March 5, 2020 to August 25, 2020 in Maryland (MD) (red color) and Wyoming (WY) (blue color).

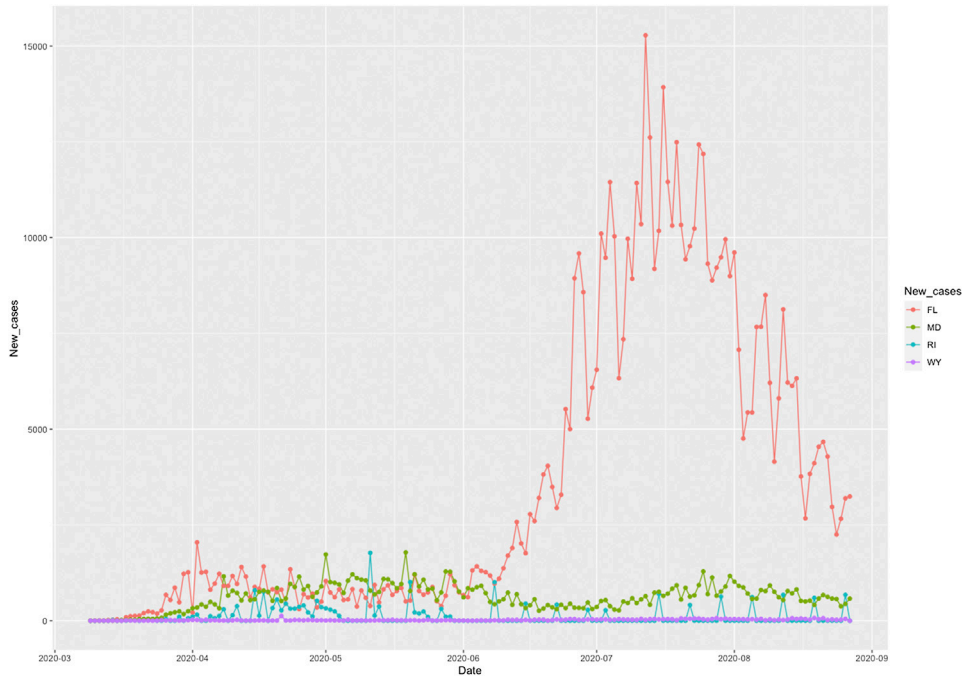


FIGURE 3 | Number of new cases of COVID-19 as a function of time from March 5, 2020 to August 25, 2020 in Florida (FL), Rhode Island (RI), Maryland (MD), and Wyoming (WY).

suggested that men were more vulnerable to COVID-19 than women. However, our analysis did not conclude that more men than women were dying from COVID-19.

We also discovered that more Black Americans were dying from COVID-19. The reasons for this were complex. Black Americans had higher rates of chronic disease conditions, including diabetes, heart disease, and lung disease, were poor and more easily exposed to the COVID-19, and lived in the cramped housing. Inequities in the social determinants of health affected mortality and morbidity of COVID-19 for Hispanic Americans with much milder significance.

We studied the causal effect of major public health interventions across the 50 states in the US. In the absence of centralized intervention measures and implementation of a timeline and presence of the complex dynamics of human mobility and the variable intensity of local outbreaks of COVID-19, evaluating the causal effect of public health intervention measures on COVID-19 transmission and deaths in the USA posed a great challenge. We used 6 Google mobility indexes and 12 daily metrics to measure the effects of COVID-19 spread and public health interventions on mobility and social distancing, derived from mobile device location data and COVID-19 case data, provided by the University of Maryland COVID-19 Impact Analysis Platform. These real time metrics capture the dynamics of social distancing. Granger causality tests were used to identify the causality relationships between time series metrics and time varying in the number of new cases or deaths from COVID-19. Although the risk factors differed by location, Active Cases/1,000 people were a significant risk factor for both number of new cases and deaths from COVID-19 in most states. The most popular intervention measure in the US was workplaces (mobility trends for places of work). Workplaces were the significant cause of the number new cases of COVID-19 in 44 states and significant cause of death in 49 states. Therefore, workplaces should be considered as a very important risk mitigation measure to reduce the number of new cases and deaths from COVID-19. Tests done/1,000 people was the second population intervention in the US. It was the significant cause of the new cases of COVID-19 in 46 states and significant cause of death in 47 states. Virus test results in quick case identification and isolation to contain COVID-19, and rapid treatment to reduce the number of deaths. Imported COVID cases were also a top significant risk factor for speeding the spread and increasing the deaths from COVID-19. Our results showed that the imported COVID case metric was the significant causal factor for the new cases in 46 states and the significant causal factor for the deaths in 47 states.

Our results showed that the high numbers of cases and deaths from COVID-19 were due to lacking strong interventions and high population density. We observed that no metrics showed significant evidence in mitigating the COVID-19 epidemic in FL and only a few metrics showed evidence in reducing the number of new cases of COVID-19 in AZ, NY and TX. Our results showed strong interventions were needed to contain COVID-19.

Although we tried to systematically and comprehensively analyze the data, this study has multiple limitations. First, we only analyzed the causal relationship between mobility patterns and the number of new cases or deaths and ignored the role of

other potential mitigating factors (e.g., wearing face masks) that could also have contributed to the reduction of new cases or deaths from COVID-19. When data are available, more metrics should be included in the analysis.

Second, we have not addressed the confounding bias issue. When confounding is unknown, adjusting for confounding methods cannot be applied to eliminate confounding bias from the causal analysis. Unadjusted confounding bias will distort the inferred (true) causal relationship between the number of new cases or deaths from COVID-19, and metrics for social distancing when these two variables share common causes. This will have substantive implications for developing interventions to mitigate the spread of COVID-19 and reduce the deaths from COVID-19. However, removing confounding from causal analysis for COVID-19 is complicated and will be investigated in the future.

In summary, our analysis has provided information for both individuals and governments to plan future interventions on containing COVID-19 and reduction of deaths from COVID-19.

DATA AVAILABLITY 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

ZL, data analysis; TX, data analysis; KZ, data acquisition and interpretation; H-WD, interpretation; EB, biological interpretation; MX, concept design, method development and write a manuscript.

FUNDING

H-WD was partially supported by NIH Grants Nos. U19AG05537301 and R01AR069055. MX was partially supported by NIH Grants No. U19AG05537301.

ACKNOWLEDGMENTS

The authors thank Sara Barton for editing the manuscript. The authors also thank the reviewers for their helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2020.611805/full#supplementary-material.

REFERENCES

1. Callaway E. Time to use the p-word? Coronavirus enters dangerous new phase. *Nature* (2020) doi:10.1038/d41586-020-00551-1
2. Priyadarsini SL, Suresh M. Factors influencing the epidemiological characteristics of pandemic COVID 19: a TISM approach. *Int J Healthc Manag* (2020) 13:89–98. doi:10.1080/20479700.2020.1755804
3. Irfan U. The math behind why we need social distancing, starting right now. (2020) Available from: https://www.vox.com/2020/3/15/21180342/coronavirus-covid-19-us-social-distancing (Accessed March 15, 2020)
4. Farseev A, Chu-Farseevac YY, Yanga Q, Loo DB. Understanding economic and health factors impacting the spread of COVID-19 disease. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.04.10.20058222 (Accessed June 9, 2020)
5. Tantrakarnapa K, Bhopdhornangkul B, Nakhaapakorn K. Influencing factors of COVID-19 spreading: a case study of Thailand. *J Public Health* (2020) [Epub ahead of print]. doi:10.1007/s10389-020-01329-5
6. Nakada LYK, Urban RC. COVID-19 pandemic: environmental and social factors influencing the spread of SARS-CoV-2 in the expanded metropolitan area of São Paulo, Brazil. *Environ Sci Pollut Res Int* (2020) 28:1–7. doi:10.1007/s11356-020-10930-w
7. Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *WClinical Medicine* (2020) 25:100464. doi:10.1016/j.eclinm.2020.100464
8. Baum CF, Henry M. Socioeconomic factors influencing the spatial spread of COVID-19 in the United States. In: *Boston college working papers in economics*, 1009. Boston College Department of Economics (2020)
9. Coccia M. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci Total Environ* (2020) 729:138474. doi:10.1016/j.scitotenv.2020.138474
10. Livadiotis G. Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19. *PloS One* (2020) 15(5):e0233875. doi:10.1371/journal.pone.0233875
11. 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* (2020) 395:1054–1062. doi:10.1016/S0140-6736(20)30566-3
12. Raghupathi V. An empirical investigation of chronic diseases: a visualization approach to Medicare in the United States. *Int J Healthc Manag* (2019) 12(4):327–39. doi:10.1080/20479700.2018.1472849
13. Anderson RM, Hollingsworth TD, Baggaley RF, Maddren R, Vegvari C. COVID-19 spread in the UK: the end of the beginning? *Lancet* (2020) 396(10251):587–90. doi:10.1016/S0140-6736(20)31689-5
14. Saadat S, Rawtani D, Hussain CM. Environmental perspective of COVID-19. *Sci Total Environ* (2020) 728(2020):138870. doi:10.1016/j.scitotenv.2020.138870
15. 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
16. MonodUnwin 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 Health* (2020) 4:397–404. doi:10.1016/S2352-4642(20)30095-X
17. BooyMytton J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. *Nat Med* (2020) 26:1183–92. doi:10.1038/s41591-020-1011-4
18. McKendryRees 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
19. Gumel SM, Radovic A. Digital approaches to remote pediatric health care delivery during the COVID-19 pandemic: existing evidence and a call for further research. *JMIR Pediatr Parent* (2020) 3(1):e20049. doi:10.2196/20049
20. Altman N, Krzywinski M Association, correlation and causation. *Nat Methods* (2015) 12(10):899–900. doi:10.1038/nmeth.3587
21. Sharkey P, Wood G. The causal effect of social distancing on the spread of SARS-CoV-2 (2020) Available from: https://doi.org/10.31235/osf.io/hzj7a (Accessed May 19, 2020)

22. Steigera E, Mußgnuga T, Kroll LE. Causal analysis of COVID-19 observational data in German districts reveals effects of mobility, awareness, and temperature. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.07.15.20154476 (Accessed July 23, 2020)
23. Li AY, Hannah TC, Durbin J, Dreher N, McAuley FM, Marayati NF, et al. Multivariate analysis of factors affecting COVID-19 case and death rate in U.S counties: the significant effects of Black race and temperature. *Am Med J Sci* (2020) 360(4):348–56. doi:10.1016/j.amjms.2020.06.015
24. Zenil H, Kiani NA, Zea AA, Tegnér J. Causal deconvolution by algorithmic generative models. *Nat Mach Intell* (2019) 1:58–66. doi:10.18148/srm/2020.v14i2.7723
25. Ramachandra V, Sun H. Causal inference for COVID-19 interventions. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.09.29.20203505 (Accessed September 29, 2020)
26. Friston KJ, Parr T, Zeidman P, Razi A, Flandin G, Daunizeau J, et al. Dynamic causal modelling of COVID-19. *Wellcome Open Res* (2020) 5:89. doi:10.12688/wellcomeopenres.15881.2
27. LambertHulme V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early COVID-19 pandemic in the U.S. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.05.27.20115139 (Accessed September 12, 2020)
28. Regis Annea W, Carolin Jeeva S. ARIMA modelling of predicting COVID-19 infections. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.04.18.20070631 (Accessed April 23, 2020)
29. Peters J, Mooij JM, Janzing D, Schölkopf B. Causal discovery with continuous additive noise models. *J Mach Learn Res* (2014) 15:2009–53. 10.15496/publikation-1672
30. Johansen S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* (1991) 59(6):1551–80. doi:10.2307/2938278
31. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* (1969) 37:424–38. doi:10.2307/1912791
32. Eichler M. Causal inference with multiple time series: principles and problems. *Philos Trans A Math Phys Eng Sci* (2013) 371:20110613. doi:10.1098/rsta.2011.0613
33. Bai ZD, Wong WK, Zhang BZ. Multivariate linear and nonlinear causality tests. *Math Comput Simulat* (2010) 81:5–17. doi:10.1371/journal.pone.0185155
34. Jiao R, Lin N, Hu Z, Bennett DA, Jin L, Xiong M. Bivariate causal discovery and its applications to gene expression and imaging data analysis. *Front Genet* (2018) 9:347. doi:10.3389/fgene.2018.00347
35. Mooij JM, Peters J, Janzing D, Zscheischler J, Schölkopf B. Distinguishing cause from effect using observational data: methods and benchmarks. *J Mach Learn Res* (2016) 17(32):1–102.
36. Heydari MR, Salehkaleybar S, Zhang K. Adversarial orthogonal regression: two non-linear regressions for causal inference (2019) Available at: https://arxiv.org/abs/1909.04454 (Accessed September 10, 2019)
37. Wang Y. *Smoohing splines: methods and applications*. 1st ed. Chapman and Hall/CRC (2011)
38. Lu S, Shen Y, Wang Y. Generalized high-precision simulation for TT&C channels using B-spline signal processing. *IEEE Signal Process Lett* (2017) 24(9):1383–7. doi:10.1109/LSP.2017.2727524
39. Cleveland WS. Robust locally weighted regression and smoothing scatterplot. *J Am Stat Assoc* (1979) 74(368):829–36. 10.1080/01621459.1979.10481038
40. Gretton A, Bousquet O, Smola A, Schölkopf B. Measuring statistical dependence with Hilbert–Schmidt norms. Proceedings of the International Conference on Algorithmic Learning Theory; 2005 October 8–11; Singapore (2005) 63–77.
41. Vakhania NN, Tarieladze VI, Chobanyan SA. Covariance operators. In: *Probability distributions on banach spaces. Mathematics and its application (soviet series)*, Vol. 14. Dordrecht: Springer (1987)
42. Abdalla I, Murinde V. Exchange rate and stock price interactions in emerging financial markets: evidence on India, Korea, Pakistan and the Philippines. *Appl Financ Econ* (1997) 7:25–35. doi:10.1080/096031097333826
43. Maryland Transportation Institute. *University of Maryland COVID-19 impact analysis platform*. College Park, USA: University of Maryland (2020) Available from: https://data.covid.umd.edu (Accessed 2020)
44. Zhang L, Ghader S, Pack M, Darzi A, Xiong C, Yang M, et al. *An interactive COVID-19 mobility impact and social distancing analysis platform*. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.04.29.20085472 (Accessed May 5, 2020)
45. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. *J Trav Med* (2020) 27(3):taaa038. doi:10.1093/jtm/taaa038
46. Pequeno P, Mendel B, Rosa C, Bosholn M, Souza JL, Baccaro F, et al. Air transportation, population density and temperature predict the spread of COVID-19 in Brazil. *PeerJ* (2020) 8:e9322. doi:10.7717/peerj.9322
47. MagnussonBarbosa E. High population density in India associated with spread of COVID-19. Medical Research News (2020) Available from: https://www.news-medical.net/news/20200703/High-population-density-in-India-associated-with-spread-of-COVID-19.aspx (Accessed July 3, 2020)
48. Rajan K, Dhana K, Barnes LL, Aggarwal NT, Evans L, Wilson RS, et al. Strong effects of population density and social characteristics on distribution of COVID-19 infections in the United States. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.05.08.20073239 (Accessed May 19, 2020)
49. Hamidi S, Sabouri S, Ewing R. Does density aggravate the COVID-19 pandemic? *J Am Plann Assoc* (2020) 86:495–509. doi:10.1080/01944363.2020.1777891
50. Golestaneh L, Neugarten J, Fisher M, Billett HH, Gil MR, Johns T, et al. The association of race and COVID-19 mortality. *EClinicalMedicine* (2020) 25:100455. doi:10.1016/j.eclinm.2020.100455
51. BellinYunes UV, Larkins-Pettigrew M. Racial demographics and COVID-19 confirmed cases and deaths: a correlational analysis of 2886 US counties. *J Public Health* (2020) 42(3):445–7. doi:10.1093/pubmed/fdaa070
52. Calo WA, Murray A, Francis E, Bermudez M, Kraschnewski J. Reaching the hispanic community about COVID-19 through existing chronic disease prevention programs. *Prev Chronic Dis* (2020) 17:200165. doi:10.5888/pcd17.200165
53. Bwire GM. Coronavirus: why men are more vulnerable to covid-19 than women?. *SN Compr Clin Med* (2020) 4:1–3. doi:10.1007/s42399-020-00341-w
54. *Google community mobility reports*. (2020) Available from: https://www.google.com/covid19/mobility/ (Accessed August 17, 2020)
55. Althouse BM, Wallace B, Case B, Scarpino SV, Berdhal A, White ER, et al. The unintended consequences of inconsistent pandemic control policies. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.08.21.20179473 (Accessed October 28, 2020)
56. Hebert-Dufresne S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. medRxiv [Preprint] (2020) Available from: https://doi.org/10.1101/2020.03.03.20029843 (Accessed March 13, 2020)
57. Masrur A, Yu M, Luo W, Dewan A. Space-time patterns, change, and propagation of COVID-19 risk relative to the intervention scenarios in Bangladesh. *Int J Environ Res Publ Health* (2020) 17(16):E5911. doi:10.3390/ijerph17165911
58. Goldschmidt-Clermont PJ. COVID-19 real-world data for the US and lessons to reopen business. *PLoS Pathog* (2020) 16(8):e1008756. doi:10.1371/journal.ppat.1008756

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 26 January 2021
doi: 10.3389/fphy.2020.609224

Social Heterogeneity Drives Complex Patterns of the COVID-19 Pandemic: Insights From a Novel Stochastic Heterogeneous Epidemic Model (SHEM)

Alexander V. Maltsev[1] *and Michael D. Stern*[2]*

[1]Translational Gerontology Branch, National Institute on Aging/NIH, Baltimore, MD, United States, [2]Laboratory of Cardiovascular Science, National Institute on Aging/NIH, Baltimore, MD, United States

In addition to vaccine and impactful treatments, mitigation strategies represent an effective way to combat the COVID-19 virus and an invaluable resource in this task is numerical modeling that can reveal key factors in COVID-19 pandemic development. On the other hand, it has become evident that regional infection curves of COVID-19 exhibit complex patterns which often differ from curves predicted by forecasting models. The wide variations in attack rate observed among different social strata suggest that this may be due to social heterogeneity not accounted for by regional models. We investigated this hypothesis by developing and using a new Stochastic Heterogeneous Epidemic Model that focuses on subpopulations that are vulnerable in the sense of having an increased likelihood of spreading infection among themselves. We found that the isolation or embedding of vulnerable sub-clusters in a major population hub generated complex stochastic infection patterns which included multiple peaks and growth periods, an extended plateau, a prolonged tail, or a delayed second wave of infection. Embedded vulnerable groups became hotspots that drove infection despite efforts of the main population to socially distance, while isolated groups suffered delayed but intense infection. Amplification of infection by these hotspots facilitated transmission from one urban area to another, causing the epidemic to hopscotch in a stochastic manner to places it would not otherwise reach; whereas vaccination only in hotspot populations stopped geographic spread of infection. Our results suggest that social heterogeneity is a key factor in the formation of complex infection propagation patterns. Thus, the mitigation and vaccination of vulnerable groups is essential to control the COVID-19 pandemic worldwide. The design of our new model allows it to be applied in future studies of real-world scenarios on any scale, limited only by computing memory and the ability to determine the underlying topology and parameters.

Keywords: social interactions, second wave, hotspots, stochastic model, infection, SARS-CoV2, COVID-19, superspreader

OPEN ACCESS

Edited by:
Sen Pei,
Columbia University, United States

Reviewed by:
Srinivasan Venkatramanan,
University of Virginia, United States
Weihua Lidslk,
Max Planck Institute of Animal Behaviour, Germany

***Correspondence:**
Michael D. Stern
SternMi@mail.nih.gov

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *22 September 2020*
Accepted: *30 November 2020*
Published: *26 January 2021*

Citation:
Maltsev AV and Stern MD (2021) Social Heterogeneity Drives Complex Patterns of the COVID-19 Pandemic: Insights From a Novel Stochastic Heterogeneous Epidemic Model (SHEM). Front. Phys. 8:609224. doi: 10.3389/fphy.2020.609224

INTRODUCTION

Coronaviruses represent one of the major pathogens that primarily target the human respiratory system. Previous outbreaks of coronaviruses (CoVs) that affected humans include the severe acute respiratory syndrome (SARS)-CoV and the Middle East respiratory syndrome (MERS)-CoV [1]. COVID-19 is a disease caused by the novel coronavirus SARS-CoV-2 virus that is both fatal and has a high transmission rate (R_0), almost twice that of the 2017–2018 common influenza [2, 3]. The World Health Organization stated that this combination of high health risk and susceptibility is of great global public health concern, and efforts must be directed to prevent further infection while vaccines are still being developed [4]. As of November 2020, there are almost sixty million confirmed COVID-19 cases worldwide and close to confirmed one and a half million deaths. Older adults seem to be at higher risk for developing more serious complications from COVID-19 illness [5, 6]. In addition to vaccines and treatments, an effective way to combat the virus is to find and implement mitigation strategies. An invaluable resource in this difficult task is numerical modeling studies that can reveal key factors in pandemic development.

What models could be useful? Direct study of the available data of COVID-19 is complicated because many cases and deaths are underrepresented. However, a simple model that correctly captures large-scale behaviors, but gets some details wrong, is useful, whereas a complicated model that gets some details correct but mischaracterizes the large-scale behaviors is misleading [7]. Previously, during the H1N1 pandemic, generic (i.e., non-specific) stochastic influenza models were important to understand and quantify the full effects of the virus in simulations of important scenarios [8]. Open source stochastic models such as FluTE (2010) or GLEaM (2011) [9, 10] were developed to simulate the spatial interaction and clusterization of millions of people to discover epidemic patterns.

Now, with respect to COVID-19, the FluTE model has recently been used to offer interventions to mitigate early spread of SARS-CoV-2 in Singapore [11], and GLEaM was adopted by Chinazzi et al. [12] to model the international propagation of COVID-19 to gain insight into the effect of travel restrictions on virus spread. Detailed statistical information about the social interactions and grouping of individuals is difficult to gather, but ultimately can be used to calibrate the parameters of agent-based models. Such calibrated agent-based models have been applied to model high-density housing in Brazil and their effect on viral spread to the rest of the population [13].

Despite extensive efforts to understand and predict the COVID-19 spread, the key factors that determine the multimodal rise patterns, the asymmetry of the recovery phase, and the emergence of a distinct second wave remain unclear. Therefore, instead of another data-based forecasting model, we chose to develop a scenario model to study the consequences of a set of hypothesis-driven conditions in a network of populations. One underexplored but important factor of pandemic spread is social heterogeneity which defines the degree of dissimilarity in the behaviors of embedded subpopulations. With regard to virus spread, the important characteristics of social heterogeneity to consider are levels of clusterization, societal interaction, and disease mitigation strategies. Our hypothesis is that complex infection curves that consist of multiple infection peaks and growth periods are the consequence of asynchronous propagation of infection among groups with widely varying degrees of intra-group interaction and isolation from main hubs (a metapopulation of infections).

To approach this problem, we developed a novel Stochastic Heterogeneous Epidemic Model (dubbed SHEM) which incorporates heterogeneous aspects of society. We also take into account over-dispersed stochasticity (super-spreading) [14], which is usually not incorporated into compartmental models but can be critical in small or virgin populations. The model design was inspired by our stochastic models of local calcium release dynamics inside heart cells, driven by explosive calcium-induced-calcium-release [15, 16]. We examine several key scenarios of heterogeneity where separate communities of various clusterization and transmission capabilities are linked to a large population hub. The basic reproduction number of infection (R_0) of the bulk of our population was assigned to $R_0 = 2.5$ which is within the range of SARS-CoV-2 basic reproduction number based on the early phase of COVID-19 outbreak in Italy [17]. Interplay of various degrees of heterogeneity and isolation periods in our model generated various dynamic patterns of infection, including a multi-modal growth periods, an extended plateau, prolonged tail, or a delayed second wave of infection. Most importantly, we found that vulnerable social subgroups play a key role in the propagation and unpredictability of the epidemic, and can defeat efforts at social distancing.

METHODS

Model Purpose

In view of the constantly changing behavioral environment for COVID-19 in the United States and worldwide, data-based predictive modeling of the future of the epidemic is difficult. Our model is specifically intended to examine the effect of heterogeneity, including not only geographic but also social heterogeneity, i.e. the existence of groups within one geographic location that have different social interaction patterns and may be partially isolated from neighboring groups, *e.g.*, nursing homes, prisons, campuses. Alternatively, subgroups can be partially embedded in the main population, e.g., meat processing plants or warehouse employees who are unable to socially distance at work, but spend part of their day in the main community where they can acquire and amplify infection. The model is fully stochastic and, unlike most compartmental models, incorporates the effect of over dispersion of secondary infections (super spreading).

Structure of the Model

The general model consists of a number of subpopulations ("villages") whose number is limited only by computing memory. The simulation is based on a generalization of the

SEIRD representation. The state of each village is represented by the numbers of individuals in each of five states: *Susceptible*, *Exposed* (destined to become infected), *Infected*, *Recovered* (immune) and *Dead* (however, see below under Super-spreading for additional state-dependence). Each village is, by definition, homogeneous and mixed. Villages could represent actual geographic units, but could also be groups or sub-regions that have different social interactions or behavior. The mean duration of infection (infectious period) was taken to be 7 days and the incubation period 5.5 days.

Each village J is characterized by its population, the expected mortality of virus infections, and its local value RINN(J) of the basic reproduction number R_0. R_0 is defined as the mean expected number of secondary infections spawned by one infected individual over the duration of their illness, *if the population were totally susceptible*. It is a property of both the virus and the behavior of individuals in the population, but is distinct from *R(t)*, the realized, time dependent, reproduction number that depends also on the fraction of susceptible individuals remaining during the epidemic.

Villages are connected by a user-specified network of formally unidirectional links along which infection or individuals can travel at user-specified rates, including links from each village to itself to represent internal infection/recovery processes. Infection can spread by two processes: transient contact between groups (*alpha* process) *e.g.,* nursing home staff coming from the city; or actual migration of individuals from one village to another (*beta* process). Each non-self link is characterized by four user-supplied parameters: *alphain* and *alphaout* describe the degree of transient contact (see below) along or against the direction of the link respectively; *betain* and *betaout* are rates of migration of individuals (time^{-1}).

Transient Contact (Alpha) Process

Infection transmitted by transient contact is modeled as though members of one village spend some (small) fraction *alpha(in/out)* of their time (*i.e.,* of their inter-personal contacts) "visiting" the opposite village at the other end of the link, adjusted for any mitigations (an example would be staff working at a nursing home, or meat-packing plant employees, treated as a separate, high-risk population but living in the surrounding county). The spread of infection in each direction of the link has two components: 1) exposure of susceptibles by visiting infectious individuals and 2) exposure of visiting susceptibles in the visited village, who then carry the infection back to their village. This formulation allows for the possibility that transmission is asymmetric. The generation of exposure by these "visitors" at home and abroad is scaled so that each infected individual, generates (in an otherwise susceptible population) his destined number of secondary cases (see below under super-spreading).

This arrangement allows for the possibility that "visitors" from different villages could cross-infect while visiting a common hub (picture UPS and FEDEX drivers) even if there is no direct link between them. To represent this process, "virtual links" are generated between pairs of physical links that meet in a hub (in graph-theory terms these are links of the adjoint graph of the network). Infection by this indirect process is second-order in the alpha's so it makes very little contribution in the case of highly isolated sub-populations (*e.g.,* nursing homes, prisons) but could be important for embedded sub-populations with high contact with the hub. Although each village is considered homogeneous by definition, further heterogeneity within a village could be represented by subdividing the population into several "villages" in close mutual contact via the alpha process (*e.g.,* students in a college split into those who go to bars and those who study alone).

Simulation Method

The entire collection of populations is simulated as a single, continuous-time Markov chain (birth-death process). There are 16 types of possible events associated with each link:

- Infection from source to target by transient contact
- Infection from target to source by transient contact
- Infected individual moves from source to target
- Exposed individual moves from source to target
- Susceptible individual moves from source to target
- Infected individual moves from target to source
- Exposed individual moves from target to source
- Susceptible individual moves from target to source
- Susceptible gets exposed inside village (self-link only)
- Exposed converts to infected inside village (self-link only)
- Infected recovers inside village (self-link only)
- Infected dies inside village (self-link only)
- Recovered moves from source to target
- Recovered moves from target to source
- Susceptible gets vaccinated
- Recovered loses immunity

The objective of the simulation is to generate a continuous-time sequence of Markov states, with transition rates determined by the SEIRD equations, modified as described below under Super Spreading. The algorithm consist of a front-end program that sets up the network of villages and the rates of spread of infections by the alpha and beta processes, and an engine module that is called repeatedly by the front-end to walk the Markov scheme under a sequence of imposed conditions, *e.g.,* open, lockdown etc. The operation of the program is described by the following simplified pseudocode:

```
PROGRAM FRONT_END
use module simulator
read parameter file !nh = number of villages
            do ih = 1,nh
            initialize village population sizes
            and states
            lrlinks(ilink,1:2) = ih !create
            self-links
            set r0's for first time period
            end do
!create network
        lrlinks(ilink,1) = source
        lrlinks(ilink,2) = target
        set   alphain,   alphaout,   betain,
        betaout(ilink)
        ilink++
```

```
call episim(...lrlinks ... tswitch, yflag) !
  invoke the engine in simulator module
if yflag = false on return then ! t reached a
  breakpoint
        change r0's,alphas, betas
        advance tswitch
        call episim again
else
        reached tmax
        write output history
end program front-end



MODULE SIMULATOR
contains
subroutine episim ! main engine
create bidirectional linked infectivity lists
!generate virtual links by extending link array
do ll1 = 1,nrlinks
        do ll2 = ll1, nrlinks ! triangular
        search for common hubs j3
        ilink++
        links(ilink,3) = j3
        alphav(ilink)    =    alphain/out(j1)
        *alphaout/in(j2)
        end do
end do
t = 0
! main loop
do while t < tswitch
        do over all links
              do event = 1,16 !generate
              cumulative rates of possible
              events
              rtot = rtot + rate(event, link)
              rtt(jtt) = rtot
              jtt++
              end do
        end do
! rtot is total rate of available markov transitions
! time of next event in Poissant point process
        time of next event = t-log(random)/rtot
        !exponential distribution
! choose the actual event link and type:
        find rtot*random2 in the cumulative
        array rtt at index jbin
        jl=(jbin-1)/16 + 1 ! find which link fired
        links(jl,1:3) gives the villages at the
        link ends and/or hub
        jp = jbin-16*(jl-1) ! remainder points
        to the event type
! carry out the event
        if the event creates a new infectious
        person then
              k = kranbin(random3,rinn(j),reff)
              ! personal infectivity
              push k on the top of infection list
              of village j
              inf(j) = inf(j)+k
! inf is the collective infectiousness of
village j, plays role of
! (numberinfected)*r0 in SEIRD equations
        end if
        if the event removes infectious person
        by recovery or death then
              pull k off bottom of infection list
              inf(j) = inf(j)-k
        end if
        if the event is migration of infectious
        move between tops(most recent) of
        infection list
        end if
        if t > tmax then
              return with yflag = true
        end if
        if t > tout then
              record state in kout array
              increment tout
        end if
        end do over links
end do while ! continue with time steps until t >
tswitch
        return with yflag = false ! continue to
        the next simulation period.
end subroutine episim

FUNCTION kranbin ! draws random negative
binomial integer with mean r0 and !
dispersion reff.

end module simulator
```

Super-Spreading

It is known that the distribution of secondary COVID-19 infections generated by a single, infected individual is over-dispersed (*i.e.* has a long tail compared to the Poisson distribution of infections expected if transmission were random). Although the average R_0 is estimated to be 2.5-4 in the absence of social distancing mitigations, contact tracing has shown that single individuals have infected up to a hundred others. This is known as super-spreading events, and can occur by several possible mechanisms, involving either a predilection of an individual (*e.g.,* a celebrity who travels widely and contacts many other people) or a situation in which individuals were placed in unusually close contact (*e.g.,* a church choir in an indoor location). On the other hand, the majority of infected individuals do not appear to spread the infection to anyone. It has been shown [14] that this over-dispersed distribution can be approximated by a negative binomial distribution, with mean R_0 (by definition) and dispersion parameter $r<<1$, for example 3 and 0.16. By iterating this distribution for several generations of viral spread, it is found that the eventual distribution of epidemic size is predicted to be quite different than found for a hypothetical stochastic transmission by Poison-distributed secondary infections with the same R_0. A recent model

of contact tracing assumed, based on data from the Netherlands, that the distribution of number of personal contacts outside the family is distributed as a negative binomial and used this to generate random changes to infection levels at 1-day intervals [18].

Unfortunately, viral generations do not remain synchronous in time, so it is not straightforward to incorporate super-spreading in a time-dependent epidemic evolution model except by following the interactions and infections of each individual in the population, as done for example in the FLuTE simulation for influenza [8]. This is very compute-intensive, but a more significant objection from our point of view is that it depends on knowing (statistically) the social interaction groups and travel behavior of the population at a fine-grained scale, and these have been severely disrupted by mitigation efforts during the current pandemic. It is possible to try to adjust for these mitigations by calibration against the evolving case data, but this is difficult. Rather than speculate on these variables, we have developed a modified Markov scheme that tries to reproduce the observed distribution of secondary infections by replacing R_0 in the event-rate calculations by an infectivity that is itself stochastic. This requires storing a partial history of individual infections, which makes the actual state-space, considered as a Markov process, much larger than that in a classic SEIRD model.

The stochastic process of infection generation by one infected case is in competition with the independent stochastic recovery process. In the model, recovery is a Poisson point process with a rate proportional to the number of infections. If we don't identify individuals, a super-spreader is likely to be "recovered" before (or after) generating his destined number of infections. To avoid this, we have adopted the following scheme:

- In each village *j*, at each event, an infectivity *inf(j)* is maintained that takes the place of $k_i{}^*R_0$ in the SEIRD rate equations.
- Whenever a new infection is created (by conversion of an exposed individual), a random number *K* is drawn from a negative binomial distribution of mean R_0 and dispersion r_{eff}, the latter to be determined. *Inf* is incremented by *K* and the individual infectivity *K* is placed on the top of a linked list.
- Whenever a random recovery event is generated at the above-mentioned rate, the oldest individual infectivity is removed from the bottom of the list and subtracted from *inf*.

The number of secondary infections actually realized by one infected individual is proportional to the actual length of time he remains infectious. Since infections recover in the order in which they were created, if there are *n* infections active, that lifetime will be the *n*th waiting time of the Poisson point process whose rate is *n* times the mean recovery rate (*i.e.,* the reciprocal of the mean infection duration). The secondary infections generated by individual *K* are a Poisson point process, which is then convolved with the recovery process to give the realized distribution of secondary infections generated by that individual. Further convolving that with the negative binomial distribution of *K* with mean r_0 and dispersion *r* we find:

$$p(j,n) = \frac{n^n \Gamma(n+j) r^r \sum_{k=0}^{\infty} \frac{r_0^k k^j (n+k)^{-n-j} (r+r_0)^{-r-k} \Gamma(r+k)}{\Gamma(k+1)}}{j\Gamma(j)\Gamma(n)\Gamma(r)} \quad (1)$$

As the distribution of the actual, realized number of secondary infections. This is a long-tailed probability distribution that can be fit, by an appropriate choice r_{eff} for the dispersion parameter *r* so as to approximate the empirical negative binomial distribution with $r = 0.16$ over the relevant range. With more than a few active infections present, the distribution converges to the limit:

$$p(j,\infty) = \sum_{k=0}^{\infty} \frac{r_0^k k^j e^{-k} r^r (r+r_0)^{-r-k} \Gamma(r+k)}{j\Gamma(j)\Gamma(k+1)\Gamma(r)} \quad (2)$$

We choose r_{eff} to give the best least-squares fit on a linear scale for the case $n = 1$, which is the most important stochastic case since it governs the chance that a single infected individual can start an outbreak, and gives the chance that an infected individual causes no secondary infections, $p(0,1) = 0.62$ similar to the empirical distribution. These distributions are all normalized and have mean R_0 and differ dramatically from the Poisson distribution (**Supplementary Figure S1**, dashed line) assumed in the classic SEIR model. Larger values of *n* are decreasingly important because the aggregate distribution of the actual infection rate controlled by the sum *inf* behaves similarly to *negbinomial* (R_0 ,n^*r) which converges to Poisson, so stochastic effects become less important once there are many active cases.

Super-Spreaders Vs Super-Spreading Events

Super-spreading can be a property of the individual or of the circumstances. What happens when an individual infected patient migrates to a new village? Does he keep his identity or does he assume the infectiousness typical of the local R_0 of his new environment? In the model we can make the choice, determined by a logical variable SPREADR (default TRUE. controlled in the demos by the input parameter *spreads*). If SPREADR is true, a migrant keeps his prior *K* value which simply migrates from the top (newest) link to be added to the top of the infection list in the new village, thereby preserving his infectious lifetime in his new home. If *SPREADR* is false then the *K* value of migrants is re-randomized using the local R_0 and r_{eff} and the infectivity of transient visitors in the alpha process is re-scaled to the local value of R_0. In the current version of the program, SPREADR is a single variable governing all events, but it could easily be made specific to individual links to distinguish groups that are vulnerable due to high density in their home village (e.g., factory or warehouse) vs. groups that are intrinsically super-spreaders due to their individual behavior (celebrities, bar hoppers).

Software Considerations

The model software is written in Fortran 77/95. The main simulation engine, described above, is in the form of a single Fortran module SIMULATOR. It is intended to be driven by a front-end program that sets up the network and scenario. For purpose of these demonstrations, we hand-coded a front end (epichainF) describing a chain of urban clusters (or a single cluster) connected by bidirectional travel, each linked to a large set of small subpopulations whose characteristics differ from the urban cluster. The single Markov-chain structure of the model is intrinsically serial, and is implemented in a single processor thread.

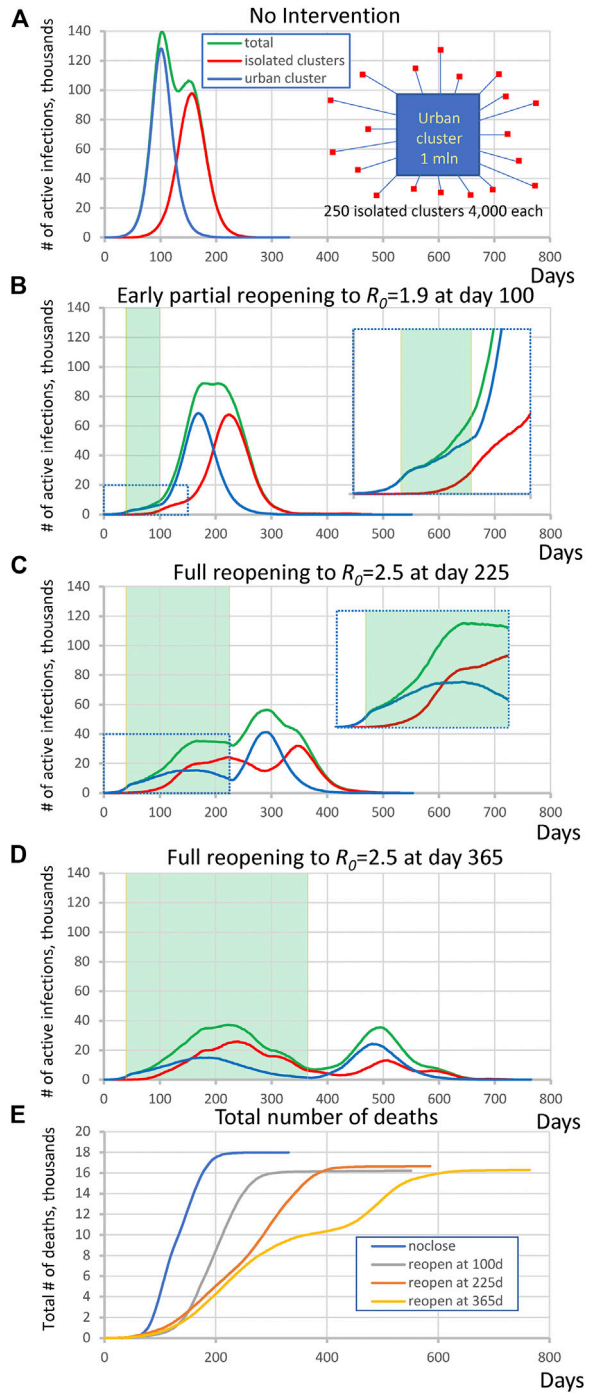


FIGURE 1 | Complex dynamic patterns of SARS-CoV2 infection in simulations in a heterogeneous society when infection in isolated clusters are ignited by an urban cluster implementing various lockdown strategies. **(A)** isolated clusters generate a second delayed peak when no intervention is implemented. Inset schematically illustrates the society structure in this scenario. Contributions are shown by different colors. **(B)** an apparent plateau after early reopening and complex rise pattern during close period (inset). Green shade shows the lockdown periods. **(C)** A multimodal rise (inset) with additional peak generated by rural cluster after full reopening at day 225. **(D)** A delayed second wave emerged after full reopening at day 365. **(E)** The dynamics of total number of deaths in each scenario.

For networks with many nodes and dense links this can be speeded up about 5-fold with 32 processors by parallelizing an inner loop.

RESULTS

Simulations of Infection in Isolated Clusters Driven by An Urban Cluster

In the first set of simulations we examined the virus spread in simple hypothetical scenarios with equal numbers of individuals in urban and isolated populations (**Figure 1A**, insert). The large urban cluster was composed of one million individuals set to $R_0 = 2.5$ (open level, but changing throughout the simulation). The isolated population consisted of 250 clusters, each with 4,000 ± 500 people and with the same internal $R_0 = 2.5$ that remained constant throughout all simulation stages. The urban cluster was weakly connected with 0.001% transient contact into the isolated clusters (*alphainpop*) while isolated clusters had 0.1% contact into the urban cluster (*alphaoutpop*), see Methods for the definition of transient contact. This can be visualized as a collection of small suburban neighborhoods or nursing homes that are attempting to isolate themselves from the city. We investigated four scenarios, specified below. In each scenario except #1, the urban cluster closed to $R_0 = 1.25$ at t = 40 days (closed level, e.g., this was New York City under lockdown, based on 21% antibody positive tests at the peak [19]).

(1) No mitigation, i.e., freely expending pandemic: The large cluster of individuals stays always open.
(2) Premature, partial reopening to $R_0 = 1.9$ at 100 days.
(3) Moderate lockdown period with full reopening at 225 days to $R_0 = 2.5$.
(4) Long lockdown period with full reopening at 365 days to $R_0 = 2.5$.

A general tendency throughout all four scenarios was that as the lockdown period increased, the magnitude of the infection decreased but its duration increased. At the same time, the interplay of the urban cluster and the isolated clusters generated a variety of specific patterns in virus spread dynamics. In the first "no mitigation" scenario (**Figure 1A**) the isolated areas generated a strong second peak at the time when infection in the urban cluster had gone through its peak and was decaying. On the other hand, the infection rise in the "premature reopening" scenario (**Figure 1B**) was multi-modal, and the cumulative peak in isolated clusters happened later than the urban cluster, creating an apparent plateau in active infection cases from day 175–225. The infection dynamics in the "moderate lockdown" scenario (**Figure 1C**) was more complex. During the closed stage (of the urban center), the infection in the urban cluster declined, but the delayed infection in isolated clusters continued to rise forming an additional peak in total infections (**Figure 1E**, inset). Then another peak in total infections emerged in the reopen stage that was generated mainly by the urban cluster, and then was echoed by the isolated subpopulations. Finally, in the "late reopening" (**Figure 1D**) scenario, infection decreased during the first wave in both urban and isolated clusters but a distinct delayed second wave of infection occurred.

We also performed a control simulation to validate that heterogeneity of isolated clusters is indeed important for the infection pattern. In the most complex scenario of "moderate lockdown" shown in **Figure 1C** we substitute 250 clusters by one big cluster with the same population of one million people keeping all other parameters the same. The simulations showed a different pattern in which the second big cluster always generated a peak of substantially larger amplitude (**Supplementary Figure S2**).

Simulations of Integrated Clusters Driving Infection in an Urban Cluster

By altering parameters in the same topology as **Figure 1A**, we found that the outlying clusters, if they are unable to socially distance, can become potential "hotspots" that can drive the infection in the urban population even against efforts of the latter to lock down. In this scenario the large urban cluster was composed of 1 million individuals with R_0 = 1.25 throughout all simulation stages while the highly susceptible population consists of 250 clusters each with 1,200 ± 500 people and internal R_0 = 3.0 that are partially embedded in the urban cluster. This R_0 value is based on data from four districts in Germany when essential manufacturing sectors were open–95%-prediction interval: 2.16–3.73 [20]. The potential hotspot clusters were connected with 20% out-coupling into the urban cluster (*alphaoutpop* = *0.20*, see Methods). This mechanism of transient contact implements short-term movement of the same people in and out regularly, which does not dilute the effect of the conditions in hotspots the way that random bidirectional migration would. In other words, the same people "virtually" move back and forth but spend most of their time in the high- R_0 locations where the infection regenerates. In this scenario, the small number of infections in the urban area are picked up by hotspots, amplified, and then drive a wave of infection among the urban population despite their efforts to keep their internal R_0 at 1.25 by social distancing.

We performed 10 runs of these simulations which demonstrated that the integrated clusters drove infection in the urban cluster, leading the late appearance of the epidemic in places that had seen few cases in a microcosm of the pattern (**Figure 2**). In the second "chain" topology multiple small urban areas (population 100 K each) are sequentially connected and 30 potential hotspots with R_0 = 2.0 drive infection within each urban cluster and facilitate propagation from cluster to cluster (**Figure 3**, **Supplementarys Figure S3 and Video S2** show the stochastic dynamics of individual hotspots). In this model, the first urban cluster began with R_0 = 2.5, then locked down to 1.25 at day 40, while the unsuspecting urban clusters connected through the chain kept R_0 = 1.25 throughout, signifying efforts at social distancing. Ultimately these efforts were defeated by the hotspots picking up the small number of arriving infections and amplifying them. These results demonstrate that subgroups who cannot or will not socially distance can drive the propagation of the epidemic to new regions against the best efforts of the majority of the populations. It follows that it is possible to control the spread of the epidemic through the mitigation of hotspot amplification. To validate this finding, we simulate the application of vaccine treatments to *just* the hotspot members, who constitute only about 30% of the population. The vaccine treatment is applied to individuals in hotspots at the rate of 5% per day, and, as a result, the geographic spread of infection is sufficiently stopped and the entire downstream region is protected from infection and deaths (**Figure 4**).

Reopening Urban Cluster After Hotspots Drive First Wave of Infection

We extended the single urban cluster hotspot scenario to reopen when infection numbers substantially drop. Here, the main cluster was composed of 1 million individuals which starts off closed with R_0 = 1.05 and reopens to R_0 = 2.50 at day 360. The cluster was connected to 30 potential hotspots each with 1,200 ± 500 people with R_0 = 3.0 which remained constant throughout all simulation stages. The urban cluster was connected with 0.1% transient contact into the isolated clusters (*alphainpop*) while isolated clusters had 1% contact into the urban cluster (*alphaoutpop*). The results show two distinct waves of infection (**Figure 5**). The hotspots drove the first wave of infection, whereas the second wave was almost entirely composed of infection from the urban area, demonstrating that the hotspots acquired immunity and did not participate at all in the second wave. The ending of the first wave, dominated by the vulnerable groups, created the illusion that the epidemic was nearly over, while a large fraction of the surrounding populations was in fact still susceptible when reopening occurred.

DISCUSSION

Interpretations and Implications

Since Summer of 2020, the infection curves of the COVID-19 pandemic in various locations have been very different from standard smooth bell curves. Here we tested the hypothesis that multiple, asynchronous waves and plateaus are in part due to stochasticity and heterogeneity, as well as due to changing efforts at mitigation. Geographic heterogeneity is included in forecasting models [12, 21, 22] which use extensive, public databases of population characteristics and travel patterns, but these do not fully account for the stratification of social behaviors that controls the spread of the virus. Therefore, instead of building another data-based forecasting and estimation model, we developed a numerical scenario model that we used to explore mechanisms of infection dynamics with regards to social stratification. The model was built as a network of "populations" which represent social and behavioral strata of geographic populations. Our model can be considered a metapopulation of SARS-CoV2, when a single species is spread among different environments that determine its local survival or extinction.

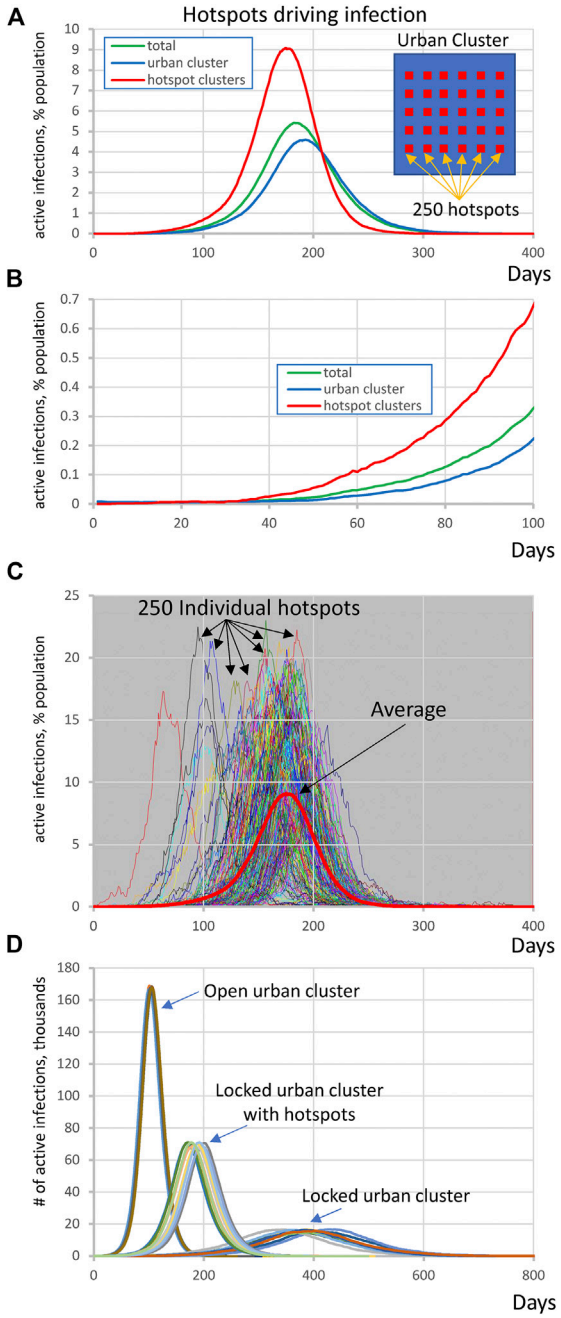


FIGURE 2 | Highly susceptible integrated clusters (hotspots) drive SARS-CoV2 infection in an urban cluster. **(A,B)** Initial rise of infection in hotspot clusters is followed by the infection in urban cluster with a delay of about 30 days. *Y*-axis represents active infections in % population reflecting for hotspots (red line) the ratio of all active cases in all hotspots to entire population of all 250 hotspots. Inset shows schematically the society structure in this scenario. **(C)** Infection in individual hotspots (multiple colors) substantially fluctuates in terms of time of ignition and magnitude from the mean (red bold curve). See also **Supplementary Video S1**. **(D)** Explosive infection in hotspots within locked urban cluster substantially increased the peak of infection in the entire society and shifted it toward much earlier occurrence from about 400 to 200 days. Shown are 10 simulation runs for each scenario.

We examined several scenarios which included one or more large urban populations connected to vulnerable subgroups that are unable/unwilling to socially distance and thus represent potential COVID-19 hotspots. Depending on the degree of interaction, these subgroups were either driven by infection from the main population, or acted as major drivers of the epidemic. Isolated subpopulations were infection-driven (*e.g.,* nursing homes, prisons, remote suburbs, clustered religious groups) and had a substantially delayed contribution to total infection cases, ultimately forming an infection curve which could include multi-modal growth periods, an extended plateau, a prolonged tail, or a delayed second wave of infection (**Figure 1**). These communities, due to their isolated nature, had low herd immunity that put them at risk for explosive scenarios when basic mitigation strategies were not implemented. Alternatively, partially integrated subpopulations were driving infection (*e.g.,* employees of factories, warehouses, meat packing plants, church groups, campuses, shelters, and other essential workers) in its connected urban population by picking up infection and amplifying it by (**Figure 2**, **Supplementary Video S1**). We found that these "hotspots" ignite infection even in a locked down population, ultimately propagating and igniting other isolated populations (**Figure 3**, **Supplementary Video S2**). The locked down population however does not acquire herd immunity, as opposed to the hotspots, and thus when lockdown is lifted, a second wave is generated by the main cluster (**Figure 5**).

There are several implications that arise from our results. We can expect social heterogeneity to form delayed local asynchronous epidemics, creating a variety of infection profiles in various regions over time, prolonging the pandemic time span, and spreading to new areas unpredictably due to the stochasticity of infection in small subgroups, as is becoming increasingly obvious in the United States in the Fall of 2020 Effective mitigation of the epidemic in the main population requires close attention to vulnerable subgroups in order to prevent the formation of COVID-19 infection hotspots. Otherwise vulnerable subgroups that cannot implement mitigation strategies spread infection to the socially distanced populations, defeating their efforts at mitigation. Despite hotspots possibly acquiring immunity, there still exists a threat of a second wave of infection in the socially distanced main population. Thus, an effective treatment or vaccination needs to be developed prior to full reopening. As vaccines become readily available, the selection and timing of their administration will be an important policy consideration. Our simulations in idealized scenarios (**Figure 4**) suggest that focusing vaccination on the small fraction of the population that is unable or unwilling to socially distance may be sufficient to interrupt regional spread and protect a much wider fraction of the public. Notably, achieving this effect requires vaccinating all hotspot groups, not merely medical personnel, and essential workers, but also uncooperative college students and those with an aversion to mitigations. This creates a kind of moral hazard–rewarding bad behavior–but the model suggests that it is the public interest.

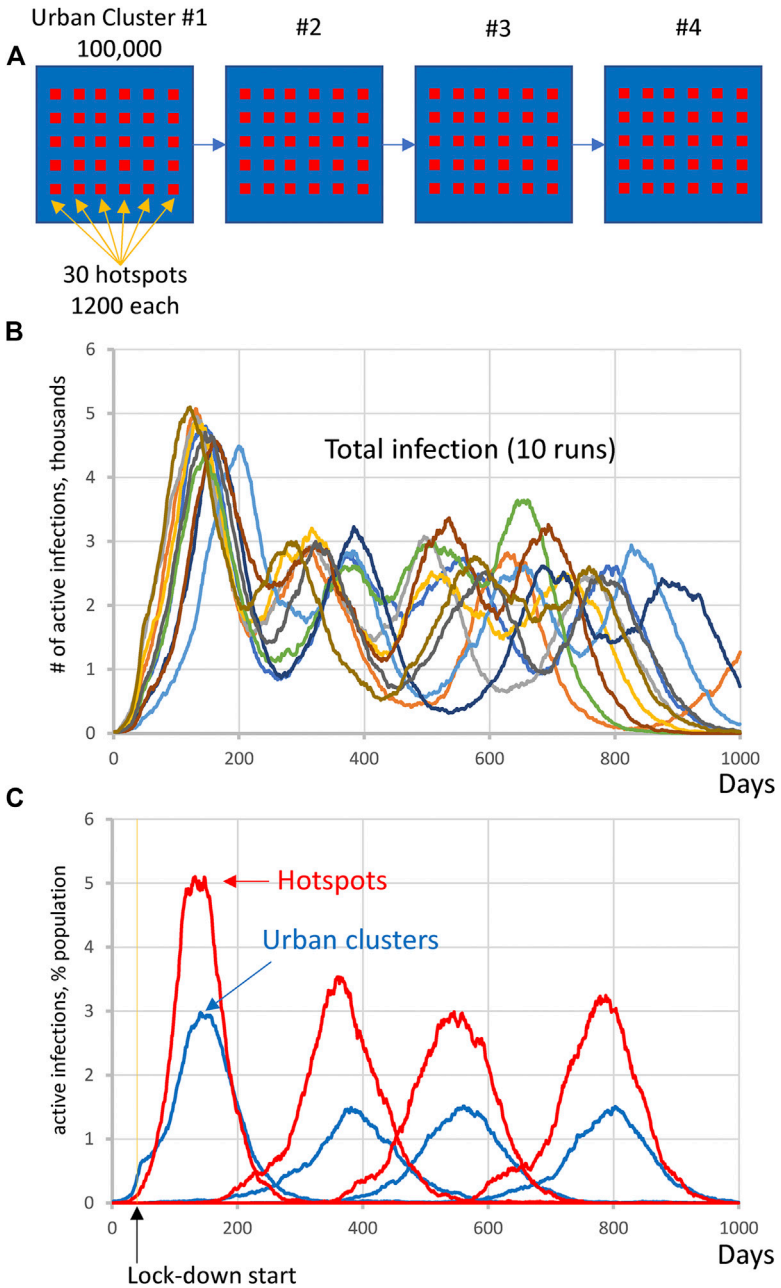


FIGURE 3 | Complex infection propagation patterns in multiple urban areas containing hotspots. **(A)** Schematic illustration of the heterogeneous society used in simulations. **(B)** Total infection count oscillates as infection propagates. While individual oscillations exhibit substantial variations in timing and amplitude, the patterns remain the same (i.e., four oscillations, reflecting infection surge in each urban cluster). **(C)** The infection in hotspots is delayed before the lockdown at day 40, but then is always in the lead (red curves), driving infection in each urban cluster (blue curves) and facilitating infection propagation among clusters (**Supplementary Video S2**).

Comparison With Other Studies

While our study is focused on vulnerable subpopulations in pandemic development, there are other important factors regarding social heterogeneity identified by previous studies.

The study by Dolbeault et al. [23], using their multi-group SEIR model, underlined the importance of mitigation measures on single individuals with a high level of social interactions. Indeed, their study showed that even a small group of individuals with high transmission rate can trigger an outbreak even if the R_0 of the majority is below 1. Althouse et al. [14] identified and explored in depth another important factor, explosive super-spreading events originating from long-term care facilities, prisons, meat-packing plants, fish factories, cruise ships, family gatherings, parties and night clubs. This study further

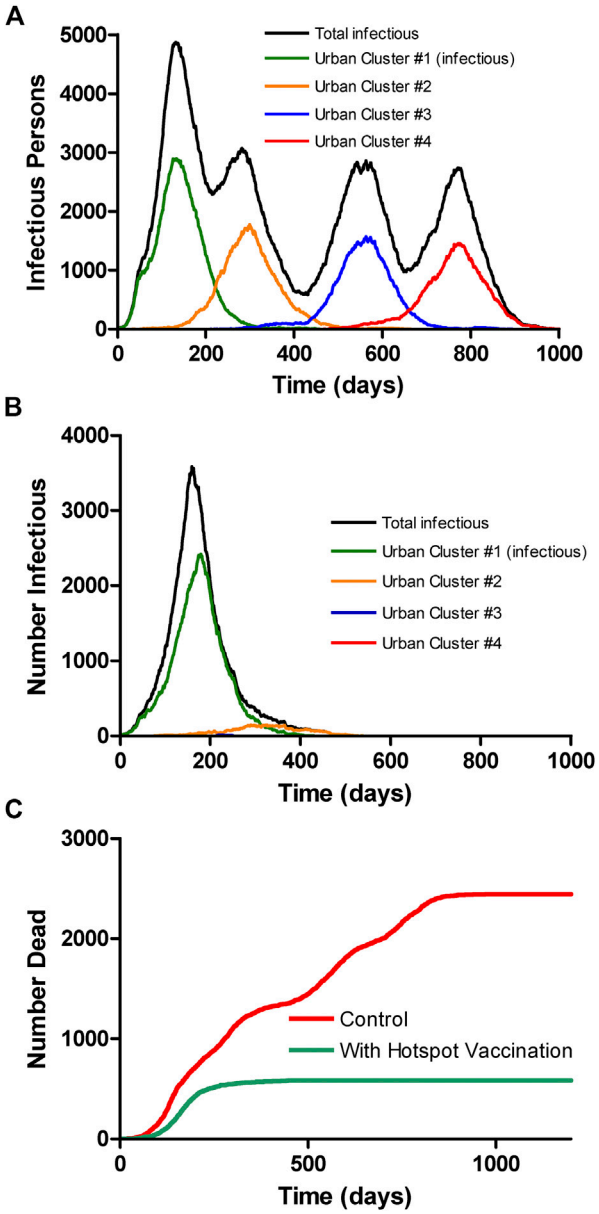


FIGURE 4 | Multicity model as in **Figure 3** without **(A)** and with **(B)** vaccination of only the hotspot populations at a rate of 5% of the population per day, starting at day 150. Vaccination of hotspot individuals prevents geographic spread of the virus even though they are only about 1/3 of the population, thereby protecting the general population. The colored curves show only the infections among the socially distanced majority of the city population. **(C)** Overall mortality with and without vaccine, assuming case mortality of 1% in all groups.

demonstrated the urgent need for targeted interventions as routes of effective virus transmission. Taking into account the importance of these super-spreading events and individuals, they were included in the design of our model (see Methods, Super-spreading) to generate more realistic outcomes of scenarios.

With regard to agent-based models, Chinazzi et al. [12] used GLEaM to demonstrate that travel restrictions introduced in Wuhan in January 2020 only delayed epidemic progression by 3–5 days within China, and international travel restrictions only helped slow infectious spread until mid-February. Our simulations of COVID-19 spread also show that ultimately, when enough time goes by, isolation does not prevent infection of vulnerable subpopulations (**Figure 1**). Chinazzi et. al. suggests that early detection, hand washing, self-isolation, and household quarantine are more effective than travel restrictions at mitigating this COVID-19 pandemic. Our recommendations are in accord, and we advocate for communities to take extra care of vulnerable subpopulations internally, as so to prevent a possible hotspot formation that may evolve into a regional epidemic.

Model Features, Limitations, and Future Studies

An epidemic can be likened to a forest fire, which spreads by diffusion along a front, but can also jump by embers that may or may not start a new blaze. Such spread to virgin areas, with a virus as with a fire, is intrinsically stochastic and such stochasticity, which is not explicitly included in mean-field models, may contribute to the remarkable patchiness of the COVID-19 epidemic. This has caused the epidemic to appear entirely different to observers in different locations, leading to politicization of the response, which is, itself, a form of social heterogeneity. For rare spread to small, isolated subgroups (embers) this stochasticity is crucial. Patchiness is aggravated by the over-dispersion (super-spreading) of secondary cases of COVID-19, where the majority of infected individuals do not spread the virus, but some can cause up to a hundred secondary infections [14]. Our model is explicitly stochastic, with a mechanism to account for over-dispersion, by keeping a partial history of individual infections. Furthermore, the design of our new model allows it to be applied in future studies of real-world scenarios on any scale, limited only by memory and the ability to determine the underlying topology and parameters.

However in our model, we make no attempt to distinguish between symptomatic and asymptomatic cases, despite recent findings by Chao et al. [24] in their agent-based model (dubbed Corvid) that demonstrated that most infections actually originate from pre-symptomatic people. Since the relative infectivity of symptomatic and non-symptomatic is uncertain, there is no direct way to accurately determine the number of asymptomatic infections at present. Such a distinction (included in a number of other models) could easily be added by subdividing the five compartments, at the risk of added complexity and more parameters needed in a scenario.

We did not take into account recent suggestions that infectivity is concentrated in a short time window just before and after symptom onset. Instead, we used the standard SEIRD assumption that infections are generated throughout the period of infection, using a mean clinical duration of 7 days. The model does not consider the physical mechanisms of transmission of COVID-19, or the

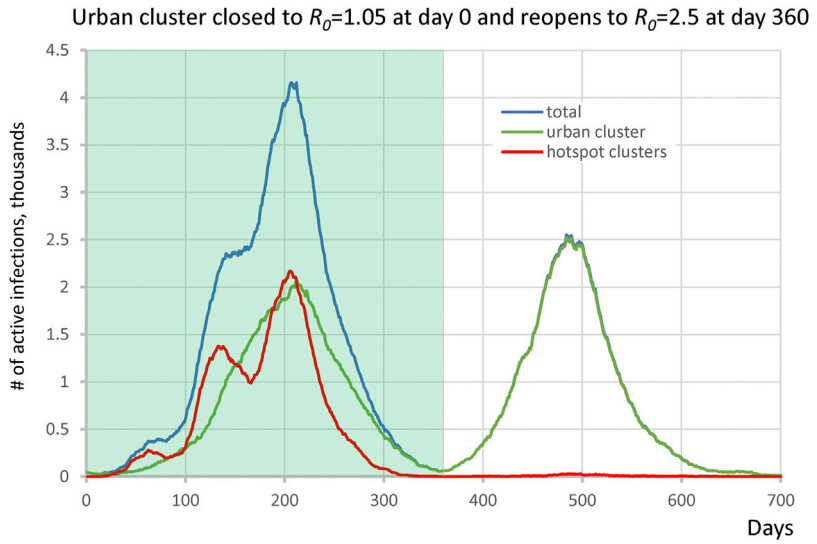


FIGURE 5 | Second wave in the hotspot scenario. Urban cluster generates a second wave of infection when it reopens from R_0 = 1.05 to R_0 = 2.50 on day 360 (green line), whereas hotspots with R_0 = 3.0 (red line) have acquired immunity in the first wave and do not participate in the second wave.

possibility that many recovered patients do not quickly re-enter their normal social circles, delaying herd immunity. An additional compartment, with a pipeline mechanism, could also be added to account for this.

We examined several simple scenarios as a demonstration of our model, which revealed the important role of embedded, non-distancing sub-populations in infection propagation. Further studies require consideration of the role of model network topology. Several studies have shown that epidemic propagation in large, scale-free networks can result in the establishment of an endemic state even with small infection rates, preventing random vaccination from effectively ending the epidemic [25, 26]. Strictly speaking this cannot happen in the scenarios we considered, which assumed that recovered individuals are permanently immune–a choice we made because of the extreme rarity of re-infections with SARS-CoV-2. A more important point is that prior theoretical analyses pertained to networks of individuals, each of whom can be either infected or susceptible. Within a single population cluster, over-dispersed link distributions such as in scale-free networks can enable persistence of infection because highly connected individuals can scavenge rare infections and widely redistribute them [27]. This is a major mechanism of super-spreading, which is incorporated in our model by heuristically handling super-spreading in each homogeneous cluster. However, stratification of the connectivity of individuals is not included in the model: Individual villages were taken to be homogeneous, characterized by their populations, R_0 and r_{eff} that determine the effective dispersion of secondary infections. Further stratification of individual connectivity could be handled by splitting social behavior into separate, mutually embedded clusters e.g., college students who study together vs. those who study alone.

It requires further studies to see if similar topological considerations pertain to networks of populations as in our model. With that in mind, the model includes the possibility that a recovered individual may revert to being susceptible, with a specified rate constant. How the topology of the larger-scale network of populations affects the propagation of the virus requires simulation studies too extensive to be considered in this paper. For example, whether physical transportation and communication networks are scale-free is controversial [28–30]. In our preliminary simulations (not shown), we found that a scale-free random network of 500 villages with populations proportional to the link numbers, and uniform behavior, had a significant probability of entering an endemic state even when the lifetime of immunity was as long as 500 days. However, the same was true of Erdös-Rényi random networks with a similar number of links. Interestingly, both types of random networks produced smooth single-peak epidemics resembling a single population suggesting that the increasingly complex patterns now being observed do depend on behavioral heterogeneity.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

AM and MS designed research, performed research, analyzed data, and wrote the paper.

FUNDING

The work was supported by the Intramural Research Program of the National Institute on Aging, National Institutes of Health.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at https://www.medrxiv.org/content/10.1101/2020.07.10.20150813v4 [31]. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (https://hpc.nih.gov/).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.609224/full#supplementary-material.

REFERENCES

1. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* (2020) 109:102433. doi:10.1016/j.jaut.2020.102433
2. Xiaofang F, Yuqing Z, Jie W, Xiaoxiao L, Cheng D, Chenyang H, et al. A severe seasonal influenza epidemic during 2017–2018 in china after the 2009 pandemic influenza: a modeling study. *Infect Microbes Dis* (2019) 1:20–6. doi:10.1097/IM9.0000000000000006
3. Liu Y, Gayle AA, Wilder-Smith A, Rocklov J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Trav Med* (2020) 27. doi:10.1093/jtm/taaa021
4. World-Health-Organization. Coronavirus disease 2019 (COVID-19) situation (2020). Report No. 72. Available from: https://apps.who.int/iris/bitstream/handle/10665/331685/nCoVsitrep01Apr2020-eng.pdf. (Accessed November 23, 2020).
5. Liu K, Chen Y, Lin R, Han K. Clinical features of COVID-19 in elderly patients: a comparison with young and middle-aged patients. *J Infect* (2020) 80:e14–e8. doi:10.1016/j.jinf.2020.03.005
6. Li P, Chen L, Liu Z, Pan J, Zhou D, Wang H, et al. Clinical features and short-term outcomes of elderly patients with COVID-19. *Int J. Infect. Dis* (2020) 97:245–50. doi:10.1016/j.ijid.2020.05.107
7. Siegenfeld AF, Taleb NN, Bar-Yam Y. Opinion: what models can and cannot tell us about COVID-19. *Proc Natl Acad Sci* (2020) 117 (28):16092–16095. doi:10.1073/pnas.2011542117
8. Chao DL, Halloran ME, Obenchain VJ, Longini IM, Jr. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput. Biol* (2010) 6:e1000656. doi:10.1371/journal.pcbi.1000656
9. Van den Broeck W, Gioannini C, Goncalves B, Quaggiotto M, Colizza V, Vespignani A. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infect Dis* (2011) 11:37. doi:10.1186/1471-2334-11-37
10. Balcan D, Goncalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. Modeling the spatial spread of infectious diseases: the GLobal epidemic and mobility computational model. *J Comput Sci* (2010) 1:132–45. doi:10.1016/j.jocs.2010.07.002
11. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis* (2020) 20:678–88. doi:10.1016/S1473-3099(20)30162-6
12. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020) 368:395–400. doi:10.1126/science.aba9757
13. Klôh V, Silva G, Ferro M, Araújo E, Melo C, Andrade Lima JRP, et al. The virus and socioeconomic inequality: an agent-based model to simulate and assess the impact of interventions to reduce the spread of COVID-19 in rio de janeiro, brazil. *Brazilian J Health Rev* (2020) 3:647–3673. doi:10.34119/bjhrv3n2-192
14. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2 (2020) *arxiv [Preprint]* https://arxiv.org/abs/2005.13689
15. Stern MD, Maltseva LA, Juhaszova M, Sollott SJ, Lakatta EG, Maltsev VA. Hierarchical clustering of ryanodine receptors enables emergence of a calcium clock in sinoatrial node cells. *J Gen Physiol* (2014) 143:577–604. doi:10.1085/jgp.201311123
16. Maltsev AV, Maltsev VA, Stern MD. Clusters of calcium release channels harness the Ising phase transition to confine their elementary intracellular signals. *Proc Natl Acad Sci U.S.A* (2017) 114:7525–30. doi:10.1073/pnas.1701409114
17. Rahman B, Sadraddin E, Porreca A. The basic reproduction number of SARS-CoV-2 in Wuhan is about to die out, how about the rest of the world?. *Rev Med Virol* (2020) e2111. doi:10.1002/rmv.2111
18. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert J, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* (2020) 5:e452–e9. doi:10.1016/S2468-2667(20)30157-2
19. Schultz C. *More than one in five people who were tested for virus antibodies in N.Y.C. had them* New York, NY: The New York Times (2020).
20. The-Robert-Koch-Institute. COVID-19 situation (2020). Report No.: 21/06/2020. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-06-21-en.pdf. (Accessed November 23, 2020).
21. Friston KJ, Parr T, Zeidman P, Razi A, Flandin G, Daunizeau J, et al. Second waves, social distancing,and the spread of COVID-19 across america. *arxiv [Preprint]* (2020). https://arxiv.org/abs/2004.13017
22. Pei S, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in the united states. *medRxiv [Preprint]* (2020). doi:10.1101/2020.05.15.20103655
23. Dolbeault J, Turinici G. Social heterogeneity and the COVID-19 lockdown in a multi-group SEIR model. *medRxiv [Preprint]* (2020). doi:10.1101/2020.05.15.20103010
24. Chao DL, Oron AP, Srikrishna D, Famulare M. Modeling layered non-pharmaceutical interventions against SARS-CoV-2 in the united states with corvid. *medRxiv [Preprint]* (2020). doi:10.1101/2020.04.08.20058487
25. Pastor-Satorras R, Vespignani A. Immunization of complex networks. *Phys Rev E Stat Nonlinear Soft Matter Phys* (2002) 65:036104. doi:10.1103/PhysRevE.65.036104
26. Pastor-Satorras R, Vespignani A. Epidemic dynamics in finite size scale-free networks. *Phys Rev E Stat Nonlinear Soft Matter Phys* (2002) 65:035108. doi:10.1103/PhysRevE.65.03510
27. Ding Q, Li W, Hu X, Zheng Z, Tang S. The SIS diffusion process in complex networks with independent spreaders. *Phys Stat Mech Appl* (2020) 546:122921. doi:10.1016/j.physa.2019.122921
28. Zhang L, Zeng G, Li D, Huang HJ, Stanley HE, Havlin S. Scale-free resilience of real traffic jams. *Proc Natl Acad Sci U.S.A* (2019) 116:8673–8. doi:10.1073/pnas.1814982116
29. Kocur-Bera K. Scale-free network theory in studying the structure of the road network in poland. *Promet Traffic Traffico* (2014) 26:235–42. doi:10.7307/ptt.v26i3.1316
30. Broido AD, Clauset A. Scale-free networks are rare. *Nat Commun* (2019) 10:1017. doi:10.1038/s41467-019-08746-5
31. Maltsev AV, Stern M. Social heterogeneity drives complex patterns of the COVID-19 pandemic: insights from a novel stochastic heterogeneous epidemic model (SHEM). *medRxiv [Preprint]* (2020). doi:10.1101/2020.07.10.20150813

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.

frontiers in Physics

PERSPECTIVE
published: 01 February 2021
doi: 10.3389/fphy.2020.631264

The Effect of An Emergency Evacuation on the Spread of COVID19

Sachit Butail[1]* *and Maurizio Porfiri*[2,3,4]*

[1]Department of Mechanical Engineering, Northern Illinois University, DeKalb, IL, United States, [2]Department of Mechanical and Aerospace Engineering, Tandon School of Engineering, New York University, Brooklyn, NY, United States, [3]Department of Biomedical Engineering, Tandon School of Engineering, New York University, Brooklyn, NY, United States, [4]Center for Urban Science and Progress, Tandon School of Engineering, New York University, Brooklyn, NY, United States

In an emergency evacuation, people almost always come in close proximity as they quickly leave a built environment under a potential threat. With COVID19, this situation presents yet another challenge: that of getting unintentionally exposed to an infected individual. To assess the epidemiological consequences of an emergency evacuation, we expanded a popular pedestrian dynamic model to enable social distancing during a normal exit and analyze the effect of possible transmission through respiratory droplets and aerosol. Computer simulations point to a troubling outcome, whereby the benefits of a quick exit could be outweighed by the risk of infection.

Keywords: COVID19, evacuation, aerosol, crowds, proximity

OPEN ACCESS

Edited by:
Lin Wang,
University of Cambridge,
United Kingdom

Reviewed by:
Jianbo Wang,
Southwest Petroleum University,
China
Hui-Jia Li,
Beijing University of Posts and
Telecommunications, China

***Correspondence:**
Sachit Butail
sbutail@niu.edu
Maurizio Porfiri
mporfiri@nyu.edu

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *19 November 2020*
Accepted: *30 December 2020*
Published: *01 February 2021*

Citation:
Butail S and Porfiri M (2021) The Effect of An Emergency Evacuation on the Spread of COVID19. Front. Phys. 8:631264. doi: 10.3389/fphy.2020.631264

INTRODUCTION

As schools and universities continue to evaluate various social distancing strategies to mitigate the spread of COVID19, a critical feature of human behavior is being overlooked—the response to a sudden alarm in a built environment that may trigger an emergency evacuation. The alarm may come from a fire in the building, the presence of an active shooter, or even a simple drill to prepare for true emergencies. Perhaps, in the current context, even someone blatantly unwilling to comply with social distancing regulations and use of masks could trigger an alarm. Whatever the source of the alarm is, during an evacuation, individuals will likely weigh the risks of being injured from the perceived threat heavily against the possibility of contracting an infection from a classmate or the instructor. Upon exiting however, people may wonder if they got too close to an infected person and if they breathed the same air for too long. These aspects represent an important discussion in relation to airborne transmission of COVID19 [1, 2].

The recommended separation distance of 2 m (six feet) is largely based on the transport of "respiratory droplets produced when an infected person coughs, sneezes, or talks." [2]. These droplets can be propelled through air for up to 2 m and "land in the mouths or noses of people who are nearby or possibly be inhaled into the lungs." [2]. More recently, results from fluid mechanics research have shown that aerosol could be the dominant driving mechanism for transmission between people in close proximity [3]. In comparison with respiratory droplets, aerosol includes much smaller particles that remain suspended in air for long periods of time to be inhaled by others. Irrespective of the driving mechanism, proximity to an infected individual is likely to increase the risk of infection, especially in the presence of screaming.

RISK OF INFECTION FROM AN INDIVIDUAL AS A FUNCTION OF THEIR PHYSICAL PROXIMITY

Under the premise that the risk of an infection increases with the proximity to an infected individual, we could quantify the epidemiological consequences of an evacuation by tracking the separation distance within the crowd. We assume that the risk of infection decays exponentially with distance [3], and that this risk accumulates over time. More specifically, we measure the risk of exposure in a crowd of N people from an infected individual I as the instantaneous exposure integrated over time for the duration of the evacuation: $E = \max_{j \neq I} \int_t e^{-\tau d_{Ij}(t)} dt$, where $d_{Ij}(t)$ is the instantaneous distance between individual I and any other individual in the crowd (**Figure 1A**), and τ is the spatial decay rate of the transmission. The maximization ensures that we select the individual who receives the highest exposure within the crowd and quantifies the risk in terms of a worst-case scenario. The higher the value of E, the more likely the infected individual will create a new infection in the crowd. This definition is agnostic to the specific mechanism of transmission, be it respiratory droplets or aerosol, and allows for a direct comparison among feasible scenarios.

An estimate of the value of τ can be obtained through a linear regression of the plots in logarithmic scale presented in Figure 7 of Chen et al. [3]. These plots include exposure from both talking and coughing for droplets of size more than 100 μM (respiratory droplets) and short-range airborne (aerosol) as functions of distance. Hence, we obtain the following estimates: talking/respiratory droplets: $\tau = 16.29\ \mathrm{m}^{-1}$; talking/aerosol: $\tau = 9.46\ \mathrm{m}^{-1}$; coughing/respiratory droplets: $\tau = 7.64\ \mathrm{m}^{-1}$; and coughing/aerosol: $\tau = 5.29\ \mathrm{m}^{-1}$. As a reference for values of E that could lead to an infection, we can follow guidelines of the Centers for disease Control and Prevention (CDC)[4] that define a "close contact" as one that may trigger an infection by being within 2 m of an infected individual for more than 15 min. By considering the most extreme case of aerosol transmission during coughing, the value of E that corresponds to close contact is 2.29×10^{-2} s. This value can be used as a simple threshold to assess a close contact in an evacuating crowd.

SIMULATING EMERGENCY EVACUATIONS AND NORMAL EXIT WITH SOCIAL DISTANCING

Emergency evacuations represent a dire situation where people exit a built environment as quickly as possible to escape the perceived danger. In an evacuation, the resulting crowd dynamics arise from a complex interplay between psychological, social, and physical factors. Individuals use social, cognitive, visual, and physical cues to stay with friends and family [5], look for the exit [6], and avoid collisions and injury [7]. Evacuation is therefore a cognitively demanding situation, which makes it inevitable for individuals to come close to each other—much less than the stipulated 2-m distance. Could this increase the risk of contracting COVID19?

Experiments on evacuation are impractical and potentially dangerous to conduct. A number of agent-based, mathematical models have been proposed over the years to predict human response and support hypothesis-driven experiments to clarify the mechanisms of the crowd dynamics. Among those, the social force model originally proposed by Helbing et al. [8] constitutes a viable compromise between model complexity and predictive power. The social force model is a physics-based model that captures interactions between finite-sized particles (agents) in the form of four kinds of forces: a social force that keeps agents apart; a goal force that makes them orient and move toward a goal location; a physical force in the event of friction and collision between agents; and a wall force, which is the same as the social force but captures interaction with walls and obstacles instead of other agents. Computer simulations can reproduce several real-world phenomena, including occurrences of bottlenecks near exits, injuries during an evacuation of a large crowd, and lane formation in corridors. The social force model has been validated in laboratory experiments [9], as well as real-world scenarios [7], thereby constituting a valid framework for exploring the potential epidemiological implications of an evacuation. By combining the classical evacuation model from Helbing et al. [8] with the proposed definition of risk of exposure, it is possible to provide a first assessment of the epidemiological consequences of an evacuation, compared to a normal exit where people can exercise social distancing.

The social force model [8] captures the motion of agent i as the combination of three effects, a desire to move toward the exit goal, maintain separation from others, and maintain distance from walls. This is mathematically written as $m\ddot{\boldsymbol{x}}_i = \boldsymbol{f}_g + \sum_j \boldsymbol{f}_{ij} + \sum_W \boldsymbol{f}_{iW}$, where m is the common mass of each agent (80 kg), $\boldsymbol{x}_i$ is the two-dimensional position vector of agent i, $\boldsymbol{f}_g$ is the goal force, $\boldsymbol{f}_{ij}$ includes the social force and the physical force between agents i and j, and $\boldsymbol{f}_{iW}$ is the wall interaction force for agent i with respect to the wall W. The goal force is modeled as $\boldsymbol{f}_g = \frac{v_0 \boldsymbol{e} - \boldsymbol{v}}{\alpha}$, where v_0 is the desired speed that encapsulates the urgency with which the agent must leave the built environment, $\boldsymbol{e}$ is the direction toward the exit, $\boldsymbol{v}$ is the instantaneous velocity, and $\alpha = 0.5$ s is the relaxation time. The interaction force is $\boldsymbol{f}_{ij} = \left\{ A e^{\frac{(r_{ij} - d_{ij})}{B}} + k g(r_{ij} - d_{ij}) \right\} \boldsymbol{n}_{ij} + \kappa g(r_{ij} - d_{ij}) \Delta v_{ji} \boldsymbol{t}_{ij}$, where r_{ij} is the sum of the radii of agents i and j (modeled as circles); d_{ij} is the distance between agents i and j ; $\boldsymbol{n}_{ij}$ identifies the direction from j to i, and $\boldsymbol{t}_{ij}$ denotes the direction that is perpendicular to $\boldsymbol{n}_{ij}$; A and B are constants that determine the strength of social interaction, with higher values leading to larger distances between agents; and $k = 1.2 \times 10^5\ \mathrm{kg\,s^{-2}}$ and $\kappa = 2.4 \times 10^5\ \mathrm{kg\,m^{-1}s^{-1}}$ determine the strength of physical interaction and friction effects, with the function g being equal to $r_{ij} - d_{ij}$ if $r_{ij} > d_{ij}$ and is zero otherwise. The wall interaction force $\boldsymbol{f}_{iW}$ has the same form of the social interaction force, so that an agent stays away from the wall and experiences physical force when in contact.

To quantify and compare the risk associated with an emergency evacuation, we simulated two scenarios: evacuation and normal exit with social distancing. To simulate these two scenarios, we varied the interaction range (parameter B in the

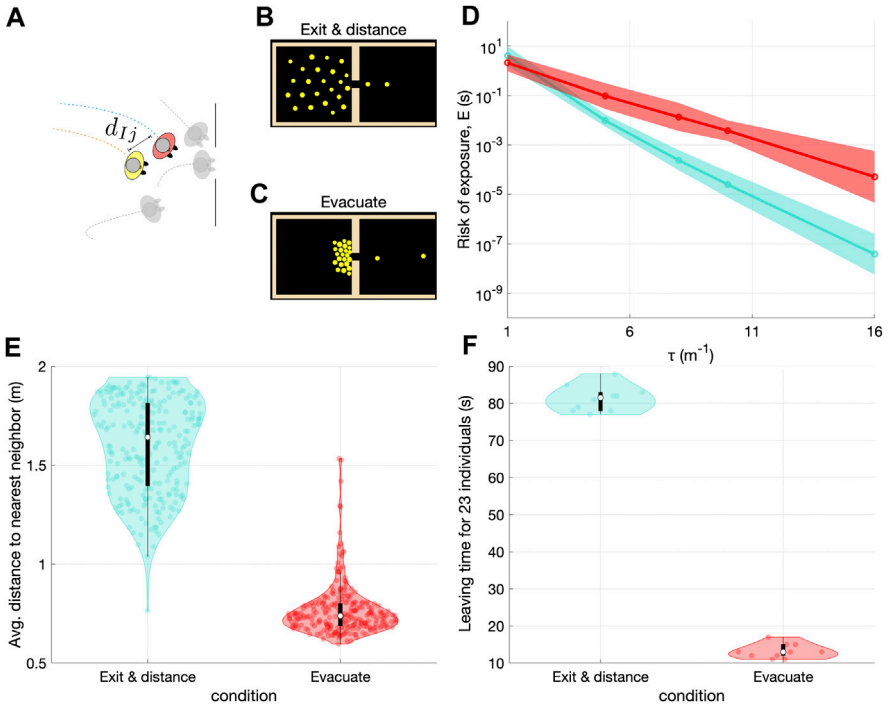


FIGURE 1 | The epidemiological risks from coming close to an individual during an evacuation could outweigh the benefits of being able to quickly leave a room under a potential threat **(A)** The framework used here combines a pedestrian dynamic model and an exposure model to quantify the risk of exposure during an emergency evacuation **(B)** and **(C)**, Snapshots from a simulation of the social force model [8] as 25 agents exit a 10 × 10 m room through a 1-m wide door while maintaining social distance or evacuate without maintaining social distance **(D)**, Risk of exposure of an individual in the crowd as a function of the decay rate of the transmission; a low value of τ indicates high transmission at larger separation distances (red denotes evacuating, and turquoize identifies exiting and social distancing) **(E)**, Average distance to the nearest neighbor during the two types of simulations performed (red denotes evacuation without distancing, and turquoize identifies exiting with distancing) **(F)**, Leaving time for more than ninety percent of the crowd for the two scenarios.

model), interaction repulsive force (parameter A in the model), and desired speed (parameter v_o in the model) within the social force model (**Figure 1**). For a normal exit where people exercise social distancing (**Figure 1B**), we set a large interaction range and a strong interaction repulsive force, along with a low desired speed of 1 m/s. On the other hand, for an evacuation, we utilize a low interaction range and a weak repulsive force (**Figure 1C**), accompanied by a high desired speed of 5 m/s. The selection of these desired speeds reflect walking and running speeds during normal and emergency situations [8].

Specifically, evacuation was simulated by setting $A = 20$ kN, $B = 0.08$ m, $v_0 = 5$ m s^{-1}, which were the default values proposed in Helbing et al. [8] to simulate an evacuation; the A and B parameter values for exit with social distancing were selected by simulating exit scenarios with a range of values $A \in \{20, 40, 60, \ldots, 200\}$ kN, $B = \{0.08, 0.16, 0.24, \ldots, 0.72\}$ m, for a normal walking speed of $v_0 = 1$ m s^{-1} and calculating the average distance to the nearest neighbor for all agents in the room for the first 10 s; we found that the average distance to nearest neighbor increased steadily with A and B before it plateaued at approximately 1.7 m due to the wall and room size constraints. The dependence of average distance to the nearest neighbor on parameters A and B is highly nonlinear and multiple parameter choices could yield the same value. Specifically, we record an average distance of 1.7 m for $A \times B \approx 5$ kNm as we vary A from 6 kN to 14 kN and B from 0.4 m to 0.72 m. We selected A = 10 kN, and $B = 0.48$ m as the centroid of this region. In the event that experimental data is available, the parameters of the pedestrian dynamics model can be set by minimizing the difference between simulated and experimental trajectories [9], or between distribution of individual speeds [10]. All other parameter values were kept the same as set in the open source code provided as part of Helbing et al [8]. Simulations were performed using the C source code provided as Supplement to the paper by Helbing et al. [8].

To prevent goal and interaction forces from balancing out to an equilibrium for the exit with social distancing scenario, the goal force was multiplied by a factor k_g that was a function of the distance to the exit d_e. This distance-dependent factor was set to an exponentially decaying value, namely, $k_g = 1 + C_1 \exp(-C_2 d_e)$, with $C_1 = 100$, and $C_2 = 1$ m^{-1} so that agents felt a stronger pull toward the exit as they got closer to it.

For each simulation, we randomly placed 25 agents (modeled as finite-sized circles) within a 10 × 10 m room with a single 1 m wide exit; this number of individuals is sufficiently low to allow for maintaining a separation distance of 2 m within the room. Randomness in the simulation was introduced through two means: first, ten simulations were performed in each scenario, where each simulation corresponded to a different initial condition and the distribution of agent size (circles with

diameters ranging uniformly between 0.5 and 0.7 m), and, second, by selecting a different agent as the single infected agent within the crowd. This amounted to hundred different realizations of each scenario. The risk of exposure, E, was computed for each scenario for different values of τ.

Figure 1D shows that the exposure for an agent within the evacuating crowd without social distancing is much larger than when the crowd leaves normally and maintains social distance, despite the evacuating crowd leaves the arena much sooner than a crowd that normally exits and maintains social distance (**Figure 1F**). **Figure 1E** confirms that the agents maintain larger distances as a result of the higher interaction range and repulsive force encoded into the model. Should the simulation be performed with a larger group of 100 agents within a room of 20×20 m dimensions, the results would be statistically indistinguishable from those presented herein. This evidence suggests that our claims should be robust with respect to size of the room and the evacuating crowd.

DISCUSSION

Our simulations show that despite the evacuating crowd taking only a sixth of the time to leave the room than when the crowd which is exiting normally, evacuation presents a far greater threat for possible transmission of COVID19 than a crowd that is exiting with social distancing. For example, in the case of aerosol transmission, evacuating in the presence of an infected individual who is coughing will yield a risk of exposure due to aerosol transmission of about 0.1 s (above the estimate of the threshold of close contact), while exit with social distancing will cause an average exposure ten time smaller (below the estimate of the threshold of close contact).

A vast community of researchers is focused on understanding how the flow of individuals during an emergency evacuation can be eased to avoid bottlenecks and high pressures that could lead to injuries and fatalities [11]. COVID19 presents yet another complication, where we must also weigh our compulsion to run away from a potential threat against the possible risks involved in being in proximity to an infected individual. Our results indicate that maintaining social distancing during an exit could increase the time required to leave the built environment by a factor of ten, which may be fatal in the case of a fire or a mass shooting. At the same time, evacuating without maintaining a social distance dramatically increases the risk of exposure, potentially leading to further infections. Face coverings can certainly help mitigate these risks, although more research is required to precisely evaluate the reduction in the decay rate associated with the proper use of masks, especially in the context of aerosol intake. Overall, this study points to a critical gap in the current guidelines for resuming in-presence learning, as well as opening up businesses during the coming fall.

Our analysis is not free of limitations, which should be investigated in further efforts, beyond the scope of this perspective. First of all, the pedestrian dynamics is described by one of the very first mathematical models in the field [8], which has seen several refinements throughout the last 20 years [12]. As such, one may attempt at more complex simulation of the evacuation process to capture perceptual and psychological factors that are missing in this model. Second, the simple evacuation scenario presented here is not able to fully capture the combination of exposure risks one may face when evacuating a building with multiple rooms, floors, stairways, and doorways. In such a complex scenario, it will also be important to trace the individuals who have experienced the strongest exposure; toward this aim it may be possible to borrow techniques from community detection in networks [13–15]. Third, the contact process examined herein does not account for individual orientation, which is likely to play an important role on droplet-based exposure, and, to a lesser extent, on airborne transmission [3]. This limitation for example could be overcome by following the approach proposed by Ronchi & Lovreglio [16] to combine risks of infections across a range of viable scenarios in a built environment. Overall, this study contributes to the general topic of safety-related issues during the current pandemic [17], by bringing forward preliminary evidence for the expected risks of infection during evacuations.

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

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

SB was supported by the National Science Foundation, under grant # CMMI-2027988, and MP was supported by the National Science Foundation, under grant # CMMI-2027990.

REFERENCES

1. Morawska L, Tang JW, Bahnfleth W, Bluyssen PM, Boerstra A, Buonanno G, et al. How can airborne transmission of COVID-19 indoors be minimised?. *Environ Int* (2020) 142:105832. doi:10.1016/j.envint.2020.105832
2. Seibert T, Allen DB, Eickhoff JC, Carrel AL. How to protect yourself and others. Centers for Disease Control and prevention (CDC) (2020). https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.
3. Chen W, Zhang N, Wei J, Yen H-L, Li Y. Short-range airborne route dominates exposure of respiratory infection during close contact. *Build Environ* (2020) 176: 106859. doi:10.1016/j.buildenv.2020.106859

4. Seibert T, Allen DB, Eickhoff JC, Carrel AL. When to quarantine. Centers for Disease Control and prevention (CDC) (2020). https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/quarantine.html.
5. Mawson AR. *Mass panic and social attachment: the dynamics of human behavior*. London: Routledge (2017).
6. Moussaïd M, Helbing D, Theraulaz G. How simple rules determine pedestrian behavior and crowd disasters. *Proc Natl Acad Sci USA* (2011) 108(17):6884–8. doi:10.1073/pnas.1016507108
7. Helbing D, Buzna L, Johansson A, Werner T. Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions. *Transport Sci* (2005) 39(1):1–24. doi:10.1287/trsc.1040.0108
8. Helbing D, Farkas I, Vicsek T. Simulating dynamical features of escape panic. *Nature* (2000) 407:487–90. doi:10.1038/35035023
9. Moussaïd M, Helbing D, Garnier S, Johansson A, Combe M, Theraulaz G. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proc Biol Sci* (2009) 276(1668):2755–62. doi:10.1098/rspb.2009.0405
10. Lee HRL, Bhatia A, Brynjarsdóttir J, Abaid N, Barbaro A, Butail S. Speed modulated social influence in evacuating pedestrian crowds. *Collect Dyn* (2020) 5:1–24. doi:10.17815/cd.2020.25
11. Schadschneider A, Klingsch W, Klüpfel H, Kretz T, Rogsch C, Seyfried A. Evacuation dynamics: Empirical results, modeling and applications In: R Meyers, editor *Encyclopedia of complexity and systems science*. New York: Springer (2009) p 3142–3176.
12. Martinez-Gil F, Lozano M, García-Fernández I, Fernández F. Modeling, evaluation, and scale on artificial pedestrians. *ACM Comput Surv* (2017) 50(5):1–35. doi:10.1145/3117808
13. Li HJ, Wang L, Zhang Y, Perc M. Optimization of identifiability for efficient community detection. *New J Phys* (2020) 22(6):063035. doi:10.1088/1367-2630/ab8e5e
14. Fortunato S, Hric D. Community detection in networks: a user guide. *Phys Rep* (2016) 659:1–44. doi:10.1016/j.physrep.2016.09.002
15. Cazabet R, Boudebza S, Rossetti G. Evaluating community detection algorithms for progressively evolving graphs. arXiv:2007.08635 (2020).
16. Ronchi E, Lovreglio R. EXPOSED: an occupant exposure model for confined spaces to retrofit crowd models during a pandemic. arXiv:2005.04007 (2020).
17. Haghani M, Bliemer MC, Goerlandt F, Li J. The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: a scientometric analysis and scoping review. *Saf Sci* (2020) 129:104806. doi:10.1016/j.ssci.2020.104806

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 February 2021
doi: 10.3389/fphy.2020.622485

Isolation and Contact Tracing Can Tip the Scale to Containment of COVID-19 in Populations With Social Distancing

Mirjam E. Kretzschmar[1]*, *Ganna Rozhnova*[1,2] *and Michiel van Boven*[1]

[1]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands,* [2]*BioISI—Biosystems & Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal*

SARS-CoV-2 has established itself in all parts of the world, and many countries have implemented social distancing as a measure to prevent overburdening of health care systems. Here we evaluate whether and under which conditions containment of SARS-CoV-2 is possible by isolation and contact tracing in settings with various levels of social distancing. To this end we use a branching process model in which every person generates novel infections according to a probability distribution that is affected by the incubation period distribution, distribution of the latent period, and infectivity. The model distinguishes between household and non-household contacts. Social distancing may affect the numbers of the two types of contacts differently, for example while work and school contacts are reduced, household contacts may remain unchanged. The model allows for an explicit calculation of the basic and effective reproduction numbers, and of exponential growth rates and doubling times. Our findings indicate that if the proportion of asymptomatic infections in the model is larger than 30%, contact tracing and isolation cannot achieve containment for a basic reproduction number ($\mathcal{R}_0$) of 2.5. Achieving containment by social distancing requires a reduction of numbers of non-household contacts by around 90%. If containment is not possible, at least a reduction of epidemic growth rate and an increase in doubling time may be possible. We show for various parameter combinations how growth rates can be reduced and doubling times increased by contact tracing. Depending on the realized level of contact reduction, tracing and isolation of only household contacts, or of household and non-household contacts are necessary to reduce the effective reproduction number to below 1. In a situation with social distancing, contact tracing can act synergistically to tip the scale toward containment. These measures can therefore be a tool for controlling COVID-19 epidemics as part of an exit strategy from lock-down measures or for preventing secondary waves of COVID-19.

Keywords: contact tracing, branching process model, effective reproduction number, exponential growth rate, doubling time, social distancing

OPEN ACCESS

Edited by:
Sen Pei,
Columbia University, United States

Reviewed by:
Hao Lei,
Zhejiang University, China
Wei Wang,
Sichuan University, China
Hongyu Miao,
University of Texas Health Science Center at Houston, United States

****Correspondence:***
Mirjam E. Kretzschmar
m.e.e.kretzschmar@umcutrecht.nl

Specialty section:
This article was submitted to Social Physics, a section of the journal Frontiers in Physics

Received: *28 October 2020*
Accepted: *31 December 2020*
Published: *10 February 2021*

Citation:
Kretzschmar ME, Rozhnova G and van Boven M (2021) Isolation and Contact Tracing Can Tip the Scale to Containment of COVID-19 in Populations With Social Distancing. Front. Phys. 8:622485. doi: 10.3389/fphy.2020.622485

1 INTRODUCTION

The novel coronavirus (SARS-CoV-2) has established itself in all parts of the world. There are still no registered vaccines and treatment options to COVID-19 disease remain mainly supportive. Control of virus transmission and associated disease thus depends on preventive measures such as social distancing combined with isolation of infected persons and those that have high likelihood of being

infected, for instance because they have been traced as contacts of infected persons [1, 2]. It has become clear that additional measures are needed to control epidemic transmission, for example by using active tracing of contacts in combination with isolation of infected contacts. Also, such measures are important in the context of exit strategies, i.e. once social distancing measures are reduced or lifted, as has been suggested recently [3]. It is unclear how effective such combinations of interventions can be in populations with social distancing in place [4].

To what extent local containment or local slowing down of an epidemic by isolation and contact tracing is successful depends on the fraction of infections that remain asymptomatic or have mild disease, on the infectiousness before the onset of symptoms [5, 6], and on testing rates. It is known that occurrence of asymptomatic infections, a high proportion of transmission occurring before the onset of symptoms, a long delay between case finding and isolation, and high overall transmissibility all factor in negatively in the likelihood that an outbreak can be contained [7–11]. For SARS-CoV-2, evidence indicates that a high fraction of infected persons is infectious before they show symptoms (up to 50%), that a substantial fraction of infections may be asymptomatic or show only mild symptoms (up to 80%), and that the epidemic doubling time in the absence of interventions may be one week or even less [6, 12–18]. On the other hand, it is also reported that with intensive contact tracing it could be possible to trace the majority (>80%) of secondary infections [11, 19].

Here we provide a model-based analysis of the impact of isolation and contact tracing in a setting with various levels of social distancing measures, using varying levels of the effectiveness and timeliness of contact tracing. It is important to consider the impact of each of these interventions in isolation but also in combination, as it is known that each intervention that reduces transmission is expected to increase the effectiveness of additional interventions in a synergistic manner [20]. The current analyses extend and complement our earlier study in which the focus was purely on the impact on delays in testing and tracing of contacts of infected individuals [11]. Here we report effective reproduction number, the (exponential) rate of increase, and the doubling time of the epidemic for scenarios with various combinations of interventions. Considering that the capacity of healthcare systems is limited, it is important to assess which interventions are most effective in slowing down the rate of increase of case numbers during an ongoing outbreak. As it is likely that, on the one hand, isolation and contact tracing will be more effective in close contact settings with well-defined contacts (household) than in the community (commuting, public spaces), while, on the other hand, the potential impact of household interventions on the epidemic could be smaller, we stratify the analyses by transmission setting (henceforth called household and non-household) [9].

2 METHODS

2.1 Overview

We use a stochastic transmission model based on a model that has been developed earlier [9], and which has been adapted to describe the biological characteristics of SARS-CoV-2 [11]. The model describes an epidemic while the proportion of immunes is low as a branching process. The model does not take into account clustering of infections, small world network effects, or other density dependent effects. Starting from a small set of initially infected individuals, the model calculates the numbers of latently infected persons, infectious persons, and persons that are diagnosed and isolated in time steps of one day. Latent infection, infectivity during the infectious period, and daily contact rates are quantified using distributions taken from the literature (**Table 1**). We distinguish between household contacts (e.g. housemates, but also other persons with whom contact is regular and close like care takers), and non-household contacts with whom frequency and duration of contact is lower. The two types of contacts differ in the risk of infection, and the delay and effectiveness of tracing and isolation may be different. Intervention effectiveness is determined by the daily probability of being diagnosed during the infectious period (**Table 2**). Furthermore, intervention effectiveness depends on the delays in tracing household and non-household contacts, respectively, and the proportions of contacts can be found and isolated or quarantined. Here isolation applies to stopping contacts of a person who is diagnosed with COVID-19, while quarantine means that a person who is not yet tested refrains from contacts. We assume that isolation and quarantine are perfect, i.e. that isolated and quarantined persons cannot transmit any longer. See **Figure 1** for a schematic description of the transmission and contact tracing process. The model is described by a set of difference equations, and allows for explicit computation of the basic reproduction number $\mathcal{R}_0$ and the effective reproduction number under interventions $\mathcal{R}_e$. Although the model is a dynamic stochastic model, here we only report results on expectations of effective reproduction numbers, exponential growth rates and doubling times. For more information about the time dependent version of the model and some results concerning the exponential growth phase of the COVID-19 epidemic, we refer the reader to [9, 21]. The model is coded in Mathematica 12.1 and is available in our GitHub repository. We give a summary of the model assumptions here, and provide a technical description in the **Supplementary Appendix**.

2.2 Natural History of Infection

We assume that the latent period lasts between 1 and 3 days with a given probability per day of moving to the infectious state. Individuals then become infectious for at most 10 days [22]. Infectivity is high at the beginning of the infectious period and decays to low levels during these 10 days (**Figure 2A**). The probability of symptom onset increases during the first 3 days of the infectious period, thereby influencing the daily probability of diagnosis during the infectious period (see **Section 2.4** and **Figure 2B**). Incubation period distribution and infectivity were fitted to recent estimates by Li et al [15], He et al [23], and Ashcroft et al [24]. The average incubation period in our model was 5.2 days with standard deviation of 3.9 days. An infectious individual makes contacts with household members and persons outside the household. We model the daily number of household

TABLE 1 | Disease and transmission parameters.

Parameter	Distribution/values	References
Latent period	1–3 days	Based on incubation period and infectivity distribution
Infectious period	10 days (variable infectivity)	[22]
Incubation period	*Lognormal* (1.434065, 0.6612)	[15]
Infectivity	*Gamma* (97.1875, 0.2689) shifted by 25.625	[23, 24]
Number of household contacts	*Poisson* (2.15)	Statistics Netherlands
Number of non-household contacts	*Negbin* (2.0, 0.15)	[25]
Relative transmissibility of non-household contacts	0.25	[43]
Scaling factor for infectivity	0.152	Calibrated such that $\mathcal{R}_0$ = 2.5

TABLE 2 | Parameters related to diagnosis and contact tracing.

Parameter	Value/range
Testing delay (delay between symptom onset and testing of index case)	Baseline: 0 days; varied from 0 to 7 days
Testing coverage (percentage symptomatic persons tested)	Baseline: 100%; alternative values considered are 60% and 80%
Tracing delay household contacts	Baseline: 0 days; varied from 0 to 4 days
Tracing delay non-household contacts	Baseline: 0 days; varied from 0 to 4 days
Tracing coverage household contacts	Baseline: 100%; varied from 0 to 100%
Tracing coverage non-household contacts	Baseline: 100%; varied from 0 to 100%
Percent symptomatic infections	Baseline: 80%; varied from 0 to 100%
Percent reduction of non-household contact rates	Baseline: 0%; varied from 0 to 95%

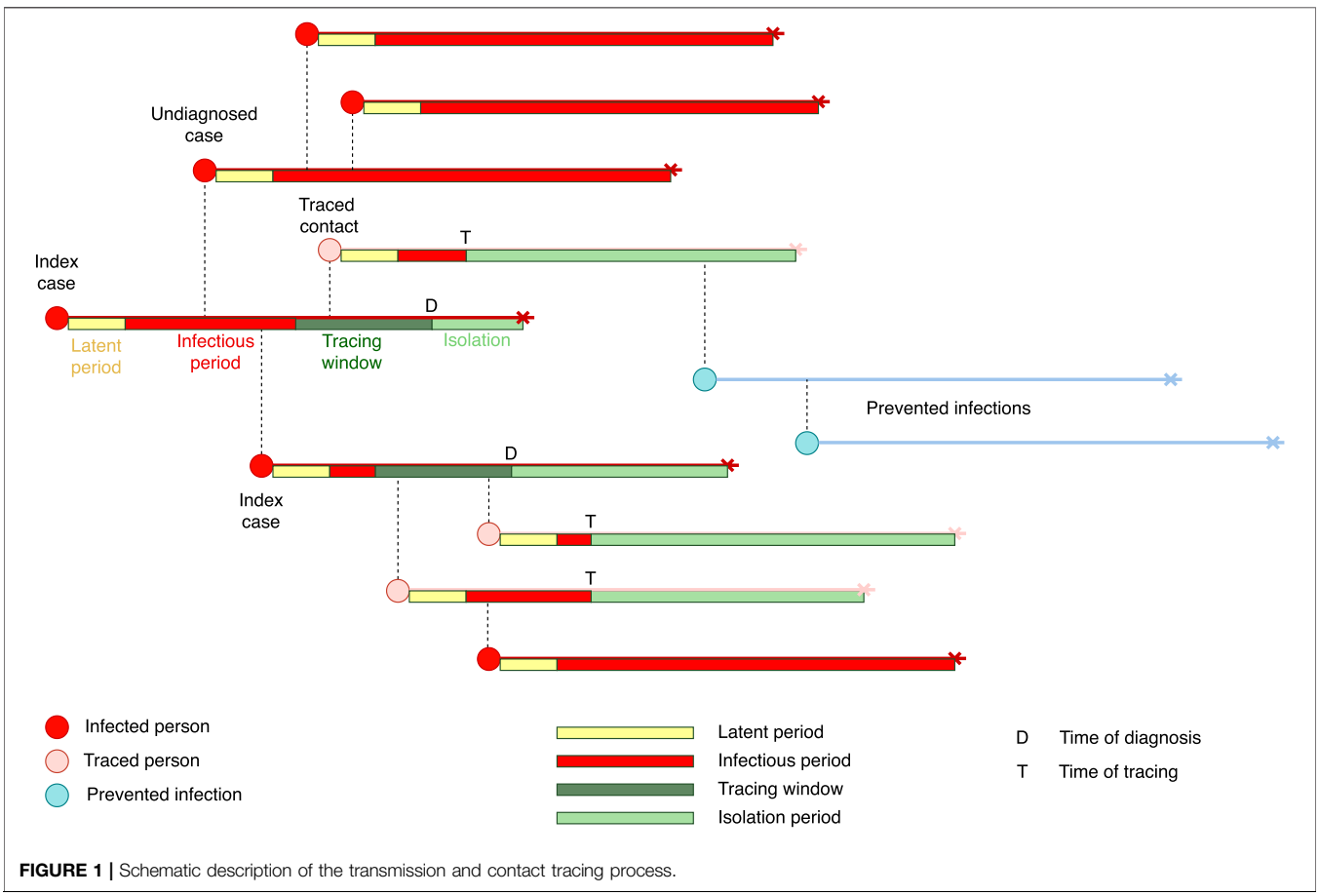


FIGURE 1 | Schematic description of the transmission and contact tracing process.

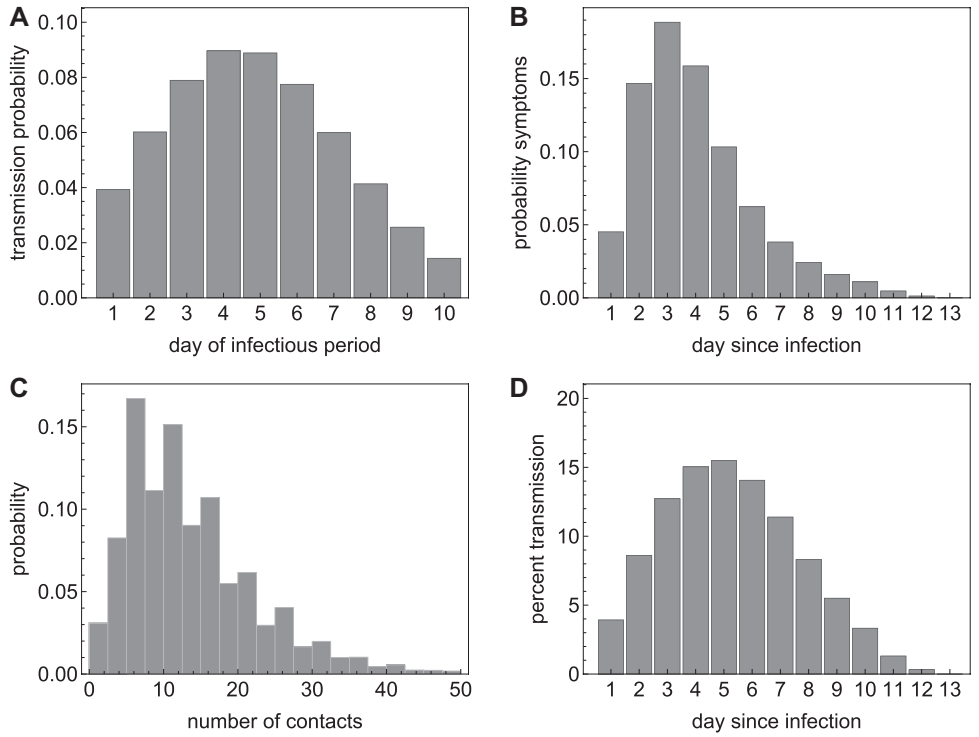


FIGURE 2 | (A) The transmission probability per day of the infectious period; **(B)** the probability of developing symptoms per day since infection; **(C)** the probability distribution of the number of contacts per day; and **(D)** percent onward transmission per day since infection.

contacts with a Poisson distribution, and the numbers of non-household contacts with a negative binomial distribution (**Table 1**), with parameters based on the average household size in the The Netherlands, and numbers of contacts observed in a contact study in the The Netherlands (**Figure 2C**) [25]. With the chosen parameters, the mean number of contacts per day is 13.2 with standard deviation of 8.5 days.

On each day of the infectious period, an individual makes a number of contacts according to the contact distribution. This number is reduced by a factor describing the probability that the contact person has already been infected during earlier contacts with the index person. **Figure 2A** shows the probability distribution of transmission upon contact with a susceptible household contact. As contacts with persons outside the household are often less close, and secondary attack rates in non-household contacts are observed to be lower than in household contacts [26], we assume that the transmission probability for these contacts is lower by factor 0.25. For this reduction factor, **Figure 2D** shows the percentage of onward transmissions per day since becoming infected, e.g. around 40% of transmission occurs in the first 4 days after acquisition of infection, i.e. before the average time of symptoms onset [24].

2.3 Social Distancing

Social distancing can be self-imposed, if people decide to reduce their social contacts during the outbreak, and it can be government-imposed by closing schools, workplaces, and other venues of social gatherings [27]. Here we assume that when social distancing is applied, household contacts remain unchanged, but the mean number of non-household contacts is reduced. This is implemented by a reduction factor in the mean of the negative binomial distribution describing non-household contact numbers. The reduction factor for social distancing was varied between 0 and 95%. In scenarios with social distancing we assumed that 80% of cases are symptomatic or can be ascertained [28]. In surveys during the lock-down in the first wave of SARS-CoV-2 in the The Netherlands, it was shown that the daily number of community contacts was reduced by 71% to around 3.7 per day [29] and after the partial lifting of the measures the number of contacts slowly increased again. Similar decreases in contact numbers during the lock-down in the United Kingdom were reported by [30].

2.4 Diagnosis, Contact Tracing, and Isolation

An infectious person becomes symptomatic with a given probability per day since infection (**Figure 2B**). For SARS-CoV-2 the probability of developing symptoms is high in the first few days of the infectious period and then declines. If an infected and infectious person has not developed symptoms 10 days after acquisition of the virus, the probability that he/she will still do so is very small. The probability of developing symptoms determines whether he/she will be diagnosed and isolated. The total probability of developing symptoms

determines the fraction that remains asymptomatic or otherwise undiagnosed, i.e. if the total probability of developing symptoms is smaller than 1, a proportion of the infected persons will remain undiagnosed and can transmit throughout their infectious period. With the assumed distributions, on average at least half of all potential onward transmissions will have occurred before an infected person is diagnosed and isolated. If diagnosis is delayed, because a person does not get tested immediately at symptom onset, and then it takes time until a test result is available, this proportion will be higher. This delay, i.e. the time between symptom onset until a symptomatic person gets a positive diagnosis, is denoted here as the testing delay. A testing delay is implemented in the model by setting the diagnosis probability to zero for the number of days of delay, and shifting the probability of diagnosis distribution to the right.

If an individual is diagnosed, contacts will be traced. Traced infected persons will be diagnosed and isolated. Tracing goes back in time for a given number of days to trace all contacts of the index case during this time window. There may be a delay before contacts are found and diagnosed, and only a fraction of all contacts may be found. These parameters, tracing delay and tracing coverage, may be different for household and non-household contacts. We assume that all traced infected persons are immediately isolated and cannot transmit any further. In reality, there might be a delay between tracing a contact and its effective isolation, but we interpret the tracing as an 'effective tracing delay' that encompasses the time from positive diagnosis of the index case until isolation of the contact. Therefore, the only individuals who will continue transmitting are those who are not found by tracing and are not yet diagnosed. **Table 2** shows the parameter values related to diagnosis and contact tracing.

2.5 Baseline Scenario

For assessing the effectiveness of contact tracing and isolation, we use a best case scenario, where all parameters are set to optimistic values. We assume that when a case is diagnosed, he/she will immediately be isolated and this will stop onward transmission completely. Furthermore, we assume that all contacts will be traced, and if found infected will be isolated immediately. We assume that it takes 0 days to find and isolate both household and non-household contacts. The rationale for using these optimistic assumptions as a baseline is that it enables investigation of the maximum contribution contact tracing can provide for achieving containment. We then investigated for various control parameters at which point of diverging from the baseline parameters control of the outbreak will be lost. We also considered more realistic parameter combinations with imperfect contact tracing, in particular including delays and reduced tracing coverages (see also [11]).

2.6 Output Variables

The model allows an explicit calculation of the basic reproduction number $\mathcal{R}_0$ and effective reproduction number $\mathcal{R}_e$ [9, 11]. $\mathcal{R}_0$ is defined as the number of secondary cases an index case generates on average in a susceptible population without any intervention. $\mathcal{R}_e$ is the number of secondary infections per case when an intervention is in place. $\mathcal{R}_0$ is determined by daily transmission probabilities and numbers of contacts, and $\mathcal{R}_e$ in addition by the level of social distancing, diagnosis probabilities, tracing delays, and tracing coverage per day of the infectious period. We can therefore investigate how $\mathcal{R}_e$ depends on $\mathcal{R}_0$, and on the intervention parameters. Details are given in the **Supplementary Appendix**.

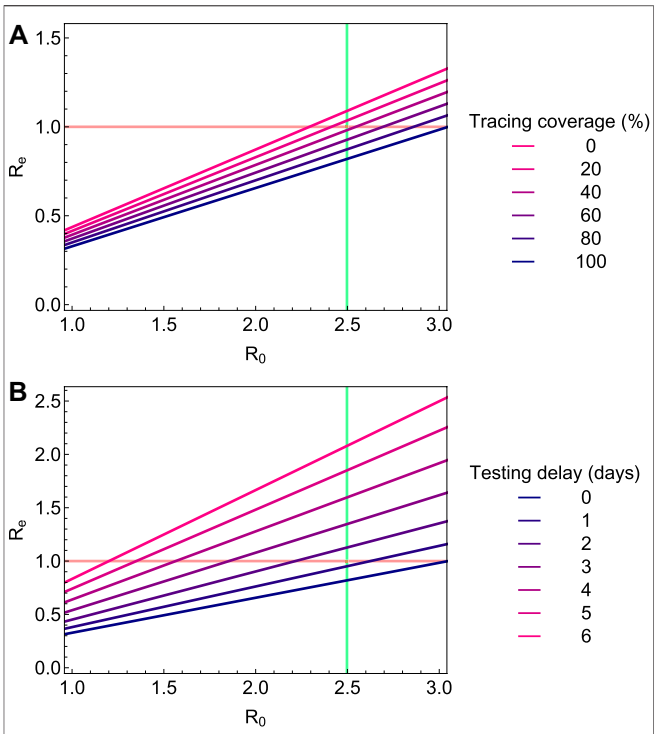


FIGURE 3 | The effective reproduction number $\mathcal{R}_e$ for varying values of the basic reproduction number $\mathcal{R}_0$ in the optimistic baseline scenario, and for various values of the tracing coverage for non-household contacts **(A)** or testing delay **(B)**. Tracing coverage of household contacts is assumed to be 100%.

We are interested in the critical tracing coverage, i.e. what proportion of non-household contacts needs to be found and isolated to control the outbreak, for populations with various levels of social distancing. Furthermore, we study the epidemic growth rate (or epidemic doubling time) without and with contact tracing and isolation and various levels of social distancing. In sensitivity analyses, we study how these quantities depend on the testing delay of the index case and on the tracing delay in contact tracing. For example, we assume that household contacts can be traced with a high coverage without delay, but that tracing of non-household contact may take longer and be less complete.

Based on the distributions of the latent and infectious periods and infectivity, we calculate the exponential growth rates and doubling times under various assumptions on the intervention parameters. This gives additional information for situations where the outbreak is not controllable, because intervention measures will lower the growth rate and increase the epidemic doubling time.

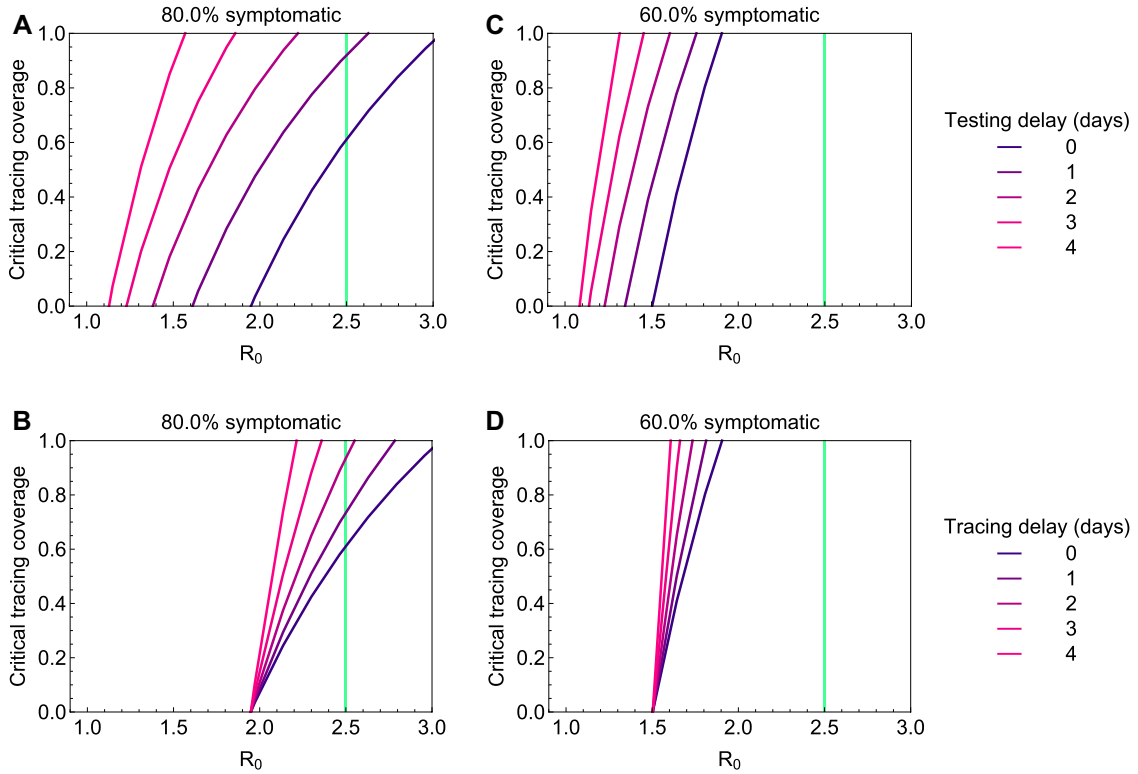


FIGURE 4 | The critical tracing coverage for varying values of the basic reproduction number $\mathcal{R}_0$; **(A)** 80% symptomatic and diagnosed index cases and varying testing delay, **(B)** 80% symptomatic and diagnosed index cases and varying tracing delay for both household and non-household contacts; **(C)** 60% symptomatic and diagnosed index cases and varying testing delay, **(D)** 60% symptomatic and diagnosed index cases and varying tracing delay for both household and non-household contacts.

We investigated how controllability of the outbreak depends on the fraction of infections that develop symptoms and therefore vary this percentage between 0 and 100%. We then considered combinations of interventions and their impact on the effective reproduction number, growth rate, and doubling time of the epidemic. We varied levels of social distancing, and coverage of tracing of household and non-household contacts. In our analysis for different levels of social distancing we assumed that 80% of infected persons develop symptoms [28, 31].

3 RESULTS

3.1 Basic and Effective Reproduction Numbers

In the baseline scenario without interventions we calibrate the transmission probability such that $\mathcal{R}_0 = 2.5$. In this case, 39% of transmission events take place in the household. The basic reproduction number of household contacts is 0.97, and that of non-household contacts 1.53. Hence, if all non-household transmissions could be prevented, the outbreak would be just under the control limit. In the baseline scenario without interventions the exponential growth rate is 0.16 per day and the doubling time is 4.4 days, which agrees with published estimates [32, 33]. **Figure 3** shows the relation between $\mathcal{R}_0$ and $\mathcal{R}_e$ for varying levels of the tracing coverage and testing delay. In **Figure 3A**, where testing delay is kept at 0 days, we find that for a tracing coverage of 40% and higher, $\mathcal{R}_e < 1$, i.e. the epidemic can be controlled by contact tracing and isolation. Similarly, if tracing coverage is 100%, as shown in **Figure 3B** testing delay can be at most 1 day to keep $\mathcal{R}_e < 1$. For lower values of $\mathcal{R}_0$, for example if reproduction numbers are reduced by social distancing, control is possible at longer delays and lower tracing coverages. However, this is only possible if all other parameters are at optimal values.

3.2 Fraction of Non-household Contacts Needed to Be Traced and Isolated

The question arises how effective contact tracing has to be to keep the outbreak under control if there is a testing delay. We therefore compute the minimum fraction of non-household contacts that need to be traced and isolated (henceforth termed "critical tracing coverage") to bring $\mathcal{R}_e$ below 1 (**Figure 4**). If 80% of infected persons develop symptoms [28, 31] and subsequently get tested, there is a chance of controlling the outbreak if the coverage of tracing non-household contacts is above the critical tracing fraction for a testing delay of at most a single day (**Figure 4A**). If 60% of infected persons develop symptoms and are tested even perfect contact tracing cannot

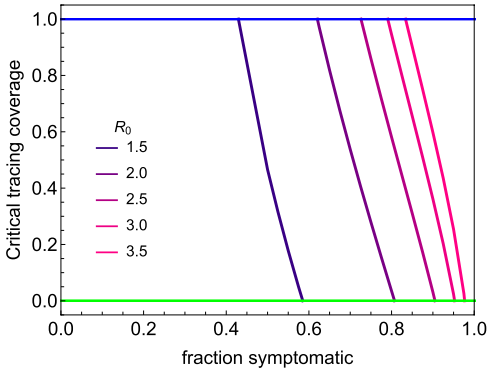


FIGURE 5 | The critical tracing coverage for varying percentages of symptomatic infections and values of the basic reproduction number $\mathcal{R}_0$ between 1.5 and 3.5. If the critical tracing coverage is above the line, the outbreak can be controlled. However, if the coverage reaches the blue line, control is lost. If more than 20% of cases escape diagnosis, because they are asymptomatic or have only mild symptoms, for $\mathcal{R}_0 = 2.5$ the outbreak is not controllable even with our optimistic baseline values for the intervention parameters.

control the outbreak. Further, in **Figures 4C,D** we vary the tracing delay of non-household contacts from 0 to 4 days, assuming no testing delays. If 80% of infection are symptomatic and diagnosed, the tracing delay should not be more than 2 days, while control is not possible if only 60% of infections are symptomatic and diagnosed.

3.3 Impact of Asymptomatic Cases

Not being diagnosed can be a consequence of not developing symptoms, having only mild symptoms, or any other reason why infected persons might not be identified by healthcare system. We subsume these possible reasons for cases not being ascertained under the term "asymptomatic". With increasing proportion of asymptomatic cases, the possibility of controlling the outbreak with contact tracing and isolation quickly fades. This is illustrated in **Figure 5**, in which we plot the critical tracing coverage for non-household contacts for several values of $\mathcal{R}_0$ as a function of the fraction of symptomatic cases (i.e. the fraction of those who will eventually develop symptoms during their entire infectious period). Household contacts are assumed to be always traced and isolated. The figure shows that for $\mathcal{R}_0 = 2.5$ control is not possible with isolation and contact tracing, if less than 80% of all infected persons develop symptoms or are otherwise not detected by the healthcare system. This is true even if all other parameters are at their most optimistic values. Other control measures such as social distancing are then needed for containment.

3.4 Exponential Growth Rates and Doubling Times

If epidemic control is not possible with isolation and contact tracing only, it might still be possible to slow down the epidemic and thereby lower demand for the healthcare system. We find that contact tracing has a significant impact on the epidemic growth rate for short testing delays (**Figure 6A**) and high coverage of tracing non-household contacts (**Figure 6C**). **Figures 6B,D** show the associated epidemic doubling times (see the **Supplementary Appendix** for details). If 60% of cases are symptomatic and diagnosed, while contact tracing is efficient (i.e. short testing delay and high tracing coverage) the doubling time can be increased to about 15 days. If less than 60% of infections are symptomatic and ascertained, however, the impact of contact tracing on the doubling time is small.

3.5 Social Distancing and Contact Tracing

Social distancing in theory could reduce the effective reproduction number to below 1, but only if the number of non-household contacts is reduced to near zero. In practice this will be hard to achieve. Additional effort into tracing and isolation of household contacts are then needed to achieve containment. In **Figure 7A** we consider a scenario in which 80% of infected persons who develop symptoms are tested and isolated, and social distancing is implemented. The figure shows the critical tracing coverage as a function of the reduction of non-household contacts. In the brown area, it is sufficient to trace and isolate household contacts with a coverage above the critical coverage. In the purple area, also non-household contacts need to be traced with a coverage above the critical coverage in addition to tracing and isolating 100% of household contacts. We find that, if social contacts outside the household are reduced by at least 30%, isolating all household contacts is sufficient for control. If non-household contacts are reduced by more than 50%, testing and isolating of cases without tracing is sufficient to bring $\mathcal{R}_e$ below 1. Similarly, **Figure 7B** shows the critical tracing coverage if testing coverage is only 60%. Here, there is less testing and the tracing coverage needs to be higher, or, alternatively, there needs to be a larger reduction in non-household contacts. Note that all diagnosed cases must be isolated immediately and isolation needs to be perfect.

In **Figure 8**, we explore the impact of social distancing on the exponential growth rate and doubling time, again for the scenario with a testing coverage of 80%. The curves show how the exponential growth rate and doubling time are affected by the reduction of non-household contacts, for various coverage levels of tracing household and non-household contacts. **Figure 8A** shows how the exponential growth rate decreases with increasing level of social distancing, and also how increasing coverage of tracing household and non-household contacts lowers the exponential growth rate. **Figure 8B** shows the associated doubling times. We consider scenarios of with increasing coverage of contact tracing. First, only household contact are traced with coverage increasing from 0 to 100% in increments of 20% (green to yellow curves in **Figure 7**); then in addition to tracing 100% of household contacts, an increasing fraction varying from 0 to 100% in increments of 20% of non-household contacts are traced (blue to magenta curves in **Figure 7**). We find that in situations where control of the epidemic is not possible, i.e. when the reduction of non-household contacts remains lower than about 50%, an effective tracing may help to greatly increase epidemic doubling times.

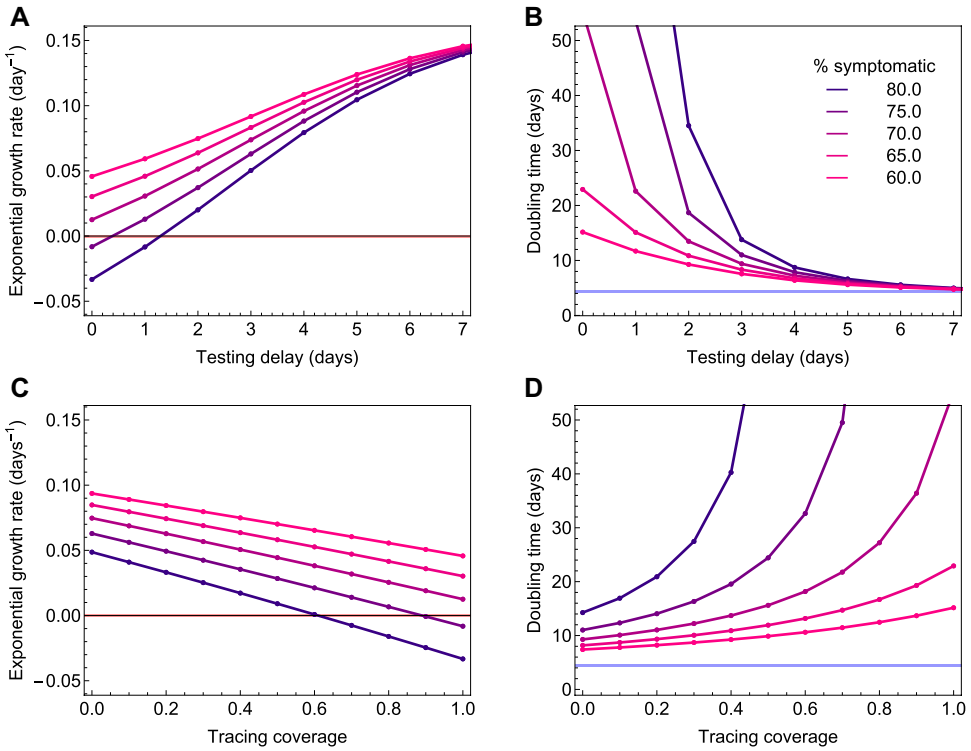


FIGURE 6 | The exponential growth rate and associated epidemic doubling time as a function of the testing delay **(A and C)** and coverage of contact tracing **(B and D)**. The colors depict the fraction of infections that is symptomatic, and which is varied from 60% (pink) to 80% (blue). The outbreak is controlled when the exponential growth rate is smaller than zero (red horizontal line). If 60% of infections are symptomatic, contact tracing and isolation can slow down the epidemic to around 15 days, but they cannot control it, even if there is no diagnosis delay and all non-household contacts are traced.

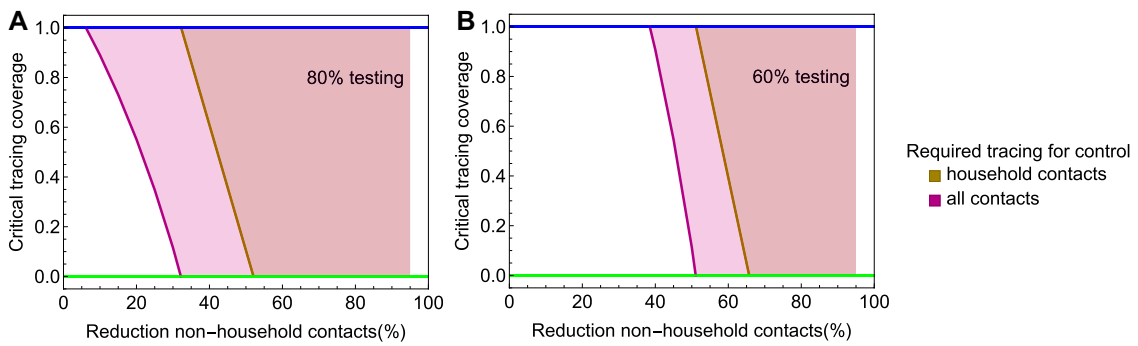


FIGURE 7 | The critical tracing coverage in a population with social distancing and $\mathcal{R}_0 = 2.5$. In the brown shaded area tracing of household contacts is sufficient, in the purple area a certain level of non-household contacts have to be traced in addition to 100% of household contacts. **(A)** 80% testing coverage; **(B)** 60% testing coverage.

4 DISCUSSION

Our analyses show that rapid diagnosis and isolation of infections based on COVID-19 disease alone cannot control outbreaks of SARS-CoV-2, but that the addition of tracing and isolation of traced cases could in theory be successful (**Figure 3**) [3, 11, 19, 34–37]. In practice, however, the potential for containment will be seriously jeopardized by delays and imperfections in the tracing process. Especially delays in diagnosis and isolation, and the existence of asymptomatic and mild infections that contribute to onward transmission could make control difficult. As evidence is mounting that the proportion of asymptomatic and mild cases is large and leads to substantial numbers of unascertained cases, most countries have

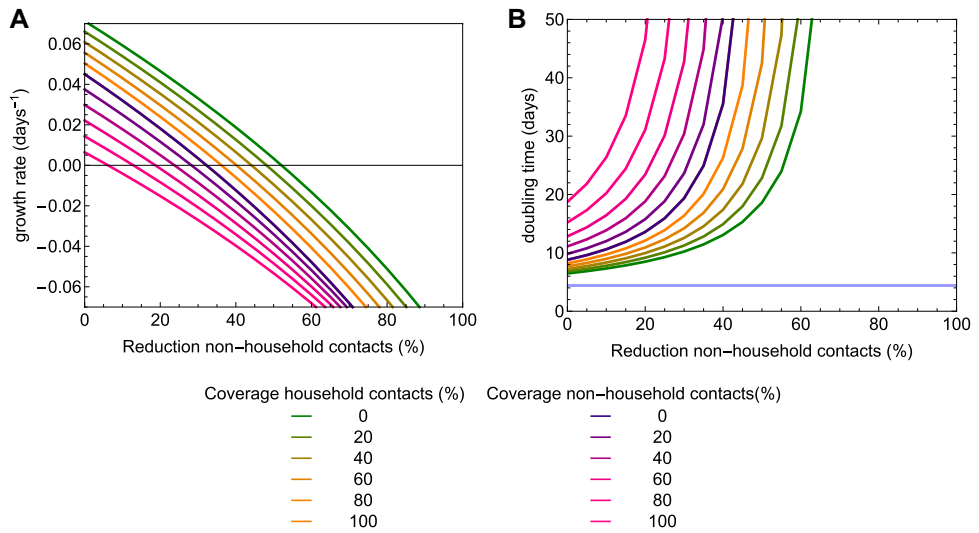


FIGURE 8 | The exponential growth rate **(A)** and epidemic doubling time **(B)** as a function of the level of social distancing and coverage of contact tracing. Testing coverage is set to 80%. The colors represent different coverage levels of contact tracing. For the lines colored from green to yellow only household contacts are traced and isolated; for the lines colored from blue to magenta, 100% of household contacts are traced and an additional fraction of non-household contacts are traced and isolated.

implemented strategies of social distancing or full lock-downs. Such measures have proven effective earlier during the 2009 influenza pandemic [38, 39]. However, social distancing can never be complete, as healthcare workers and doctors have to continue their work, but also personnel of supermarkets, public transport employees, and others will have contact outside their households. We find that in a situation where 60% of cases are ascertained, social distancing of non-household contacts fails to contain the epidemic even if contacts outside the household are reduced by 80%. In this case, combining the social distancing with tracing and isolation of household contacts may suffice to bring the balance toward containment. If social distancing is less severe, more intensive contact tracing and also tracing of non-household contacts is needed (**Figure 7**). If social distancing reduces non-household contacts only by 50%, tracing and isolation also of non-household contacts is needed for containment. If this is not possible, for example due to constraints of the public health system, tracing and isolation of household contacts can at least substantially increase the doubling time of the epidemic (**Figure 8**).

Even though the SARS-CoV-2 pandemic cannot be contained by contact tracing and rapid isolation alone, this does not render contact tracing useless. On the contrary, contact tracing and isolation when used in addition to social distancing, may be the tool needed to make this mix of strategies successful. Our analyses show that isolation and contact tracing when combined with social distancing can contribute to reducing the growth rate and increasing the doubling time of epidemics, thereby buying time, spreading the number of severe cases out over a longer period of time, and potentially also reducing the total number of infections [40]. This will lower peak healthcare demand, alleviate the stress on healthcare systems, and contribute to reducing the burden of disease.

Our analyses of contact tracing add to an earlier study by a more systematic analysis of the relation between key parameters (transmissibility, fraction asymptomatic, fraction of contacts traced, diagnosis delays), and by incorporating household vs. non-household contacts [34]. Household contacts are at a higher risk of becoming infected than non-household contacts as persons in a household will usually have repeated contacts. On the other hand, our analyses show that household infections contribute less to onward transmission than non-household infections simply because the numbers of household contacts are much lower than numbers of other contacts. As a consequence, the effectiveness of isolating non-household contacts is key for a successful contact tracing strategy. Our assumption that asymptomatic cases are as infectious as symptomatic cases may result in an overestimation of the contribution of asymptomatic cases to transmission. This might mean that effectiveness of contact tracing is more favourable than found in our analyses.

A strength of our model is that quantitative information about distributions of the latent and infectious periods, and the infectivity per day of the infectious period can be incorporated easily and detailed, such that if new and better data become available, the analyses can be updated quickly. In particular, the model can incorporate non-standard distributions based on empirical data (e.g. viral load measurements to quantify infectiousness per day).

A limitation of the analyses presented here is that they apply to a situation in which the epidemic is described by a branching process and is growing exponentially. This also applies to

another modeling using a (one-type) branching process [34]. Ultimately, as the number of persons who are or have been infected increases, the number of persons that are still susceptible will start to dwindle, and epidemic growth will ultimately come to a halt. Hence, strictly speaking our results apply to the early stages of an epidemic. However, in the present situation, the proportion of the population who is immune for COVID-19 does in general not exceed 10% in most places [41]. In fact, even when the number of infected persons is still relatively small in the early stage of an epidemic it is possible that exponential growth is not observed, for instance due to local depletion of susceptible persons in combination with clustering in contact patterns, spatial effects, and inhomogeneous mixing [42]. However, estimates of the effective reproduction number are independent of the dynamics and give information about the ability of an intervention to slow down epidemic spread. Also, at present it is only in few places in the world where cumulative infection attack rates may have surpassed the 10% level.

In conclusion, our results show that in populations where social distancing is implemented, isolation and contact tracing can play an essential role in gaining control of the COVID-19 epidemic. On their own, none of these strategies are able to contain COVID-19 for realistic parameter settings, but in a combined strategy they can just tip the balance toward containment. These insights provide guidance for policy makers, who will have to decide when and how to release severe lock-down or social distancing measures, and whether additional contact tracing and isolation is then a feasible alternative to keep a resurging epidemic at bay.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/mirjamkretzschmar/ContacttracingModel.

AUTHOR CONTRIBUTIONS

MK, MB, and GR conceived the study. MK designed and programmed the model, and produced output. All authors interpreted the results, contributed to writing the manuscript, and approved the final version for submission.

ACKNOWLEDGMENTS

We thank Martin Bootsma for help with the estimation procedure. MK acknowledges support from the Netherlands Organization for Health Research and Development (ZonMw; Grant no. 91216062 and Grant no. 10430022010001). GR acknowledges support from the Portuguese Foundation for Science and Technology (FCT; Grant no. 131_596787873).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.622485/full#supplementary-material.

REFERENCES

1. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD How will country-based mitigation measures influence the course of the covid-19 epidemic?. *The Lancet* (2020) 395(10228):931–4. doi:10.1016/S0140-6736(20)30567-5
2. Bernard Stoecklin S, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, et al. First cases of coronavirus disease 2019 (covid-19) in France: surveillance, investigations and control measures. *Euro Surveill* (2020) 25(6), 94. doi:10.2807/1560-7917.ES.2020.25.6.2000094
3. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science* (2020) 368(6491):6936. doi:10.1126/science.abb6936
4. Thompson RN, Hollingsworth TD, Isham V, Arribas-Bel D, Ashby B, Britton T, et al. Key questions for modelling covid-19 exit strategies. *Proc R Soc B* (2020) 287:20201405. doi:10.1098/rspb.2020.1405
5. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* (2020) 172(9):577–582. doi:10.7326/M20-0504
6. Cheng HY, Jian SW, Liu DP, Ng TC, Huang WT, Lin HH. Taiwan COVID-9 Outbreak Investigation Team. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Intern Med* (2020). 180(9): 1156–1163. doi:10.1001/jamainternmed.2020.2020
7. Fraser C, Riley S, Anderson RM, Ferguson NM Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci USA* (2004) 101(16):6146–51. doi:10.1073/pnas.0307506101
8. Klinkenberg D, Fraser C, Heesterbeek H The effectiveness of contact tracing in emerging epidemics. *PLoS One* (2006) 1(1):12. doi:10.1371/journal.pone.0000012
9. Kretzschmar M, van den Hof S, Wallinga J, van Wijngaarden J Ring vaccination and smallpox control. *Emerg Infect Dis* (2004) 10(5):832–41. doi:10.3201/eid1005.030419
10. Müller J, Kretzschmar M, Dietz K Contact tracing in stochastic and deterministic epidemic models. *Math Biosci* (2000) 164(1):39–64. doi:10.1016/s0025-5564(99)00061-9
11. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHM, Bonten MJM Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* (2020) 5(8):e452–e459. doi:10.1016/S2468-2667(20)30157-2
12. Backer JA, Klinkenberg D, Wallinga J Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from Wuhan, China, 20–28 January 2020. *Euro Surveill* (2020) 25(5):62. doi:10.2807/1560-7917.ES.2020.25.5.2000062
13. Chen J Pathogenicity and transmissibility of 2019-ncov-a quick overview and comparison with other emerging viruses. *Microb Infect* (2020) 22(2):69–71. doi:10.1016/j.micinf.2020.01.004
14. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* (2020) 395: 497–506. doi:10.1016/S0140-6736(20)30183-5
15. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* (2020) 382:1199. doi:10.1056/NEJMoa2001316
16. Riou J, Althaus CL Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-ncov), december 2019 to january 2020. *Euro Surveill* (2020) 25(4). doi:10.2807/1560-7917.ES.2020.25.4.2000058

17. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-ncov infection from an asymptomatic contact in Germany. *N Engl J Med* (2020) 382(10):970–971. doi:10.1056/NEJMc2001468
18. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *JCM* (2020) 9(2):462. doi:10.3390/jcm9020462
19. Keeling MJ, Hollingsworth TD, Read JM The efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19). *J Epidemiol Community Health* (2020) 74(10). 861–866. doi:10.1136/jech-2020-214051
20. Dodd PJ, White PJ, Garnett GP Notions of synergy for combinations of interventions against infectious diseases in heterogeneously mixing populations. *Math Biosci* (2010) 227(2):94–104. doi:10.1016/j.mbs.2010.06.004
21. Kretzschmar ME, Rozhnova G, van Boven M Isolation and contact tracing can tip the scale to containment of covid-19 in populations with social distancing. medRxiv (2020). Available from: https://www.medrxiv.org/content/ doi:10.1101/2020.03.10.20033738
22. Singanayagam A, Patel M, Charlett A, Bernal JL, Saliba V, Ellis J, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of covid-19, England. *Euro Surveill* (2020) 25(32):2001483. doi:10.2807/1560-7917.es.2020.25.32.2001483
23. 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(9):672–675. doi:10.1038/s41591-020-0869-5
24. Ashcroft P, Huisman JS, Lehtinen S, Bouman JA, Althaus CL, Regoes RR, et al. Covid-19 infectivity profile correction. *Swiss Med Wkly* (2020) 150:20336. doi:10.4414/smw.2020.20336
25. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* (2008) 5(3):e74. doi:10.1371/journal.pmed.0050074
26. Lei H, Xu X, Xiao S, Wu X, Shu Y Household transmission of covid-19-a systematic review and meta-analysis. *J Infect* (2020) 81(6):979–997. doi:10.1016/j.jinf.2020.08.033
27. Teslya A, Pham TM, Godijk NG, Kretzschmar ME, Bootsma MCJ, Rozhnova G Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a covid-19 epidemic: a modelling study. *PLoS Med* (2020) 17(12):1003166. doi:10.1371/journal.pmed.1003166
28. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic sars-cov-2 infections: a living systematic review and meta-analysis. *PLoS Med* (2020) 17(9):1003346. doi:10.1371/journal.pmed.1003346
29. Backer JA, Mollema L, Klinkenberg D, van der Klis FR, de Melker HE, van den Hof S, et al. The impact of physical distancing measures against covid-19 transmission on contacts and mixing patterns in The Netherlands: repeated cross-sectional surveys. medRxiv (2020). doi:10.1101/2020.06.30.20143560
30. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of covid-19 in the UK. *BMC Med* (2020) 18(1):1–10. doi:10.1186/s12916-020-01597-8
31. Mizumoto K, Kagaya K, Zarebski A, Chowell G Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill* (2020) 25(10):180. doi:10.2807/1560-7917.ES.2020.25.10.2000180
32. Roser M, Ritchie H, Hasell J. *Coronavirus pandemic (covid-19). Our World in Data* (2020). https://ourworldindata.org/coronavirus.
33. Lurie MN, Silva J, Yorlets RR, Tao J, Chan PA Coronavirus disease 2019 epidemic doubling time in the United States before and during stay-at-home restrictions. *J Infect Dis* (2020) 222(10):1601–1606. doi:10.1093/infdis/jiaa491
34. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* (2020) 74:109. doi:10.1016/S2214-109X(20)30074-7
35. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. Covid-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss Med Wkly* (2020) 150:20225. doi:10.4414/smw.2020.20225
36. Steinbrook R Contact tracing, testing, and control of covid-19-learning from Taiwan. *JAMA Intern Med* (2020) 180:1163. doi:10.1001/jamainternmed.2020.2072
37. Sun K, Viboud C Impact of contact tracing on sars-cov-2 transmission. *Lancet Infect Dis* (2020) 20(8):876–877. doi:10.1016/S1473-3099(20)30357-1
38. 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(8):473–481. doi:10.1016/S1473-3099(09)70176-8
39. te Beest DE, Birrell PJ, Wallinga J, De Angelis D, van Boven M Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza a pandemic of 2009 in The Netherlands. *J R Soc Interface* (2015) 12(103):20141244. doi:10.1098/rsif.2014.1244
40. Diekmann O, Heesterbeek H, Britton T. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton, NJ: Princeton University Press (2013). doi:10.2307/j.cttq9530
41. Vos ER, den Hartog G, Schepp RM, Kaaijk P, van Vliet J, Helm K, et al. Nationwide seroprevalence of sars-cov-2 and identification of risk factors in the general population of The Netherlands during the first epidemic wave. *J Epidemiol Commun Health* (2020) 23:9530. doi:10.1136/jech-2020-215678
42. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of covid-19 in shenzhen China: analysis of 391 cases and 1,286 of their close contacts. *Lancet Infect Dis* (2020) 20(8). 911–919. doi:10.1016/S1473-3099(20)30287-5
43. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* (2020) 395(10223):514–23. doi:10.1016/S0140-6736(20)30154-9

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership