Fields Institute Communications 85

The Fields Institute for Research in Mathematical Sciences

V. Kumar Murty
Jianhong Wu
Editors

Mathematics of Public Health

Proceedings of the Seminar on the Mathematical Modelling of COVID-19

FIELDS

MOREMEDIA

Springer

Fields Institute Communications

Volume 85

The Communications series features conference proceedings, surveys, and lecture notes generated from the activities at the Fields Institute for Research in the Mathematical Sciences. The publications evolve from each year's main program and conferences. Many volumes are interdisciplinary in nature, covering applications of mathematics in science, engineering, medicine, industry, and finance.

More information about this series at https://link.springer.com/bookseries/10503

V. Kumar Murty • Jianhong Wu
Editors

Mathematics of Public Health

Proceedings of the Seminar on the Mathematical Modelling of COVID-19

Springer

FIELDS

Editors
V. Kumar Murty
Fields Institute for Research
in Mathematical Sciences
Toronto, ON, Canada

Jianhong Wu
Department of Mathematics and Statistics
York University
Toronto, ON, Canada

ISSN 1069-5265 ISSN 2194-1564 (electronic)
Fields Institute Communications
ISBN 978-3-030-85052-4 ISBN 978-3-030-85053-1 (eBook)
https://doi.org/10.1007/978-3-030-85053-1

Mathematics Subject Classification: 92-10, 68Txx, 65Y10, 68W20

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

At noon on March 17, 2020, the doors of the Fields Institute closed as the coronavirus pandemic reached Toronto, and the Province of Ontario declared a state of emergency.

The Institute building on College Street may be considered the physical embodiment of the legacy of John Charles Fields. It is a wonderful sunlit space, lined with blackboards and covered in a fine film of chalk dust. It is the place where mathematical scientists from around the world congregate to exchange ideas. While there are many structured events, the most precious aspect of the Institute is in its ability to provide a physical environment to stimulate serendipitous meetings of minds in which creative thoughts, embedded in a string of symbols, diagrams, and formulae, are bounced around and collectively shaped into new discoveries.

About a month earlier, in mid-February, the Institute recognized the coming storm and started planning the role it might play in dealing with the virus. With barely a week's notice, the Institute summoned mathematical modelers from across Canada to convene in Toronto for a 2-day seminar and brainstorming session on February 14–15, 2020. During that session, the core group was formed, and soon afterwards, a proposal was submitted to the Canadian Institutes of Health Research (CIHR) for funding the work of the Task Force.

That proposal was successful, and the research activity of the group began in earnest. During the period of March 2020 to June 2021, the Task Force produced 46 seminars, publishing dozens of papers, participating in and running knowledge-sharing events for mathematicians and nonmathematicians alike, and supporting decision makers across the country to reduce the spread of COVID-19 by applying mathematical modeling to a host of public health problems.

Particularly noteworthy about this Task Force is that it had members from across Canada as well as the support of the mathematical science institutes, namely AARMS (the Atlantic Association for Research in Mathematical Sciences), CRM (Centre de Recherches Mathematiques), and PIMS (Pacific Institute for Mathematical Sciences) in addition to the Fields Institute itself. In addition, we also had the support of the Public Health Agency of Canada (PHAC), the Vaccine

and Infectious Disease Organization (VIDO-Intervac), and the National Research Council (NRC).

A central feature of the research activity of the Task Force was the weekly COVID-19 Mathematical Modelling Seminar. The seminar brought together experts in mathematical modeling from across Canada and the world, presenting modeling methods as they related to the COVID-19 pandemic. While the primary aim of the seminar was to advance the work of the Canadian Mathematical Modelling of COVID-19 Task Force, we hoped that sharing the methods and mathematics that have proved useful in a particular geography might provide useful insights that could be applied elsewhere. This volume largely represents talks from that seminar.

The mathematics in this book can be used to support decision makers on critical issues such as projecting outbreak trajectories, evaluating public health interventions for infection prevention and control, and developing vaccines and decisions around vaccine optimization. Readers of this book will find chapters on

- Compartment modeling involving categories of susceptible, exposed, infected, and recovered (SEIR) as well as versions of compartment models which are more nuanced to include age stratification and other subdivisions
- Forecasting for personal protective equipment (PPE)
- Predicting COVID-19 deaths
- The impact of delays of contact tracing
- Heterogeneity in social distancing
- The challenge of providing daily COVID-19 forecasts
- Analytics of contagion
- Modeling point-of-care diagnostics of COVID-19
- Understanding unreported cases

Readers will have the opportunity to learn about current modeling methodologies for infectious diseases, and the mathematics behind them, and understand the important role that mathematics has to play during this crisis, in supporting governments and public health agencies.

Most researchers who have been grappling with issues related to the pandemic understand that the challenges that had to be dealt with evolved quite rapidly over the course of the past year, and in some sense, they continue to evolve. In the early months, data was in short supply, but that changed very quickly, and there is now an abundance of data from around the world. At the same time, specific regions had, and still have, data silos, and accessing that data was a nontrivial matter which was addressed in some specific cases only through the leadership shown by policy makers and decision makers. Data access remains an issue for those who are thinking about the future of public health preparedness and resilience. Mathematical scientists have shown that access to good data can provide a formidable tool in dealing with public health crises.

The availability of data at a speed and a scale that we did not have in the past naturally led to a keen interest in turning this data to knowledge and to actionable science. With polymerase chain reaction (PCR) tests giving rapid results for infections and cell phones providing data on people's mobility (more accurate

than human recall), financial data on spending, and digital government data that records information like employment status, we are challenged to quickly and accurately calculate the spread of the disease and observe its impact on people's lives. What will happen to rates of infection if we open the border? What will happen if we reopen schools? How will we cope with flu and COVID-19? How do we optimize testing or make sure that testing is equitable? Who will get the vaccine? These are complex questions that require an assortment of tools including mathematical modeling.

While the pandemic forced us to temporarily close our physical space, the problems we were trying to solve gave birth to new virtual collaborations. As everyone worked on these complex problems presented by COVID-19, mathematicians found themselves alongside epidemiologists and public health experts applying mathematical tools in novel ways, where clear gaps between knowledge creation and translation existed. The immediacy of these problems did not allow the luxury of time to polish and discuss as in pre-COVID times.

Virtual environments have made it possible for us to work quickly, to all be in the same room without the obstacles of travel and room bookings, and to include those who could not come to the meeting because of other commitments, career responsibilities, and mobility or health issues. While we definitely want to return to our beautiful physical Institute, this new, virtual world is also quite beautiful and a "new frontier of math," one which also is "strengthening collaboration, innovation, and learning in mathematics and across a broad range of disciplines" beyond what is possible at 222 College Street. If meeting in physical space enables the possibility of serendipitous collaboration, meeting in virtual space enables introduction to an expanded world of researchers with new ideas and new approaches. Effectively utilizing both kinds of space will be crucial for the successful functioning of the Institute in the future.

And we bring these prefatory remarks to a close on this optimistic note. While there has been much suffering around the world as a result of the pandemic, perhaps it has also opened our minds to the possibility of broader and deeper collaboration through the combined use of virtual and in-person interactions.

Toronto, ON, Canada V. Kumar Murty

Toronto, ON, Canada Jianhong Wu

Contents

Diverse Local Epidemics Reveal the Distinct Effects of Population Density, Demographics, Climate, Depletion of Susceptibles, and Intervention in the First Wave of COVID-19 in the United States

Niayesh Afshordi, Benjamin P. Holder, Mohammad Bahrami, and Daniel Lichtblau

1 Introduction

The new human coronavirus SARS-CoV-2 first reported in Wuhan Province, China in December 2019 [1, 2], reached 10,000 confirmed cases and 200 deaths due to the disease (known as COVID-19) by the end of January 2020. Although travel from China was halted by late-January, dozens of known introductions of the virus to North America occurred prior to that [3, 4], and dozens more known cases were imported to the US and Canada during February from Europe, the Middle East, and elsewhere. Community transmission of unknown origin was first detected in California on February 26, followed quickly by Washington State [5], Illinois and Florida, but only on March 7 in New York City. Retrospective genomic analyses have demonstrated that case-tracing and self-quarantine efforts were effective in preventing most known imported cases from propagating [6–8], but that the eventual outbreaks on the West Coast [5, 8, 9] and New York [7] were likely seeded by

N. Afshordi (✉)
Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada

Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON, Canada

Perimeter Institute of Theoretical Physics, Waterloo, ON, Canada
e-mail: nafshordi@pitp.ca; nafshordi@perimeterinstitute.ca

B. P. Holder
Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, Canada

Department of Physics, Grand Valley State University, Grand Rapids, MI, USA

M. Bahrami · D. Lichtblau
Wolfram Research Inc., Champaign, IL, USA

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_1

unknown imports in mid-February. By early March, cross-country spread was primarily due to interstate travel rather than international imports [10].

In mid-March 2020, nearly every region of the United States of America saw a period of uniform exponential growth in daily confirmed cases—signifying robust community transmission—followed by a plateau in late March, likely due to social mobility reduction. The same qualitative dynamics were seen in COVID-19 mortality counts, delayed by approximately 1 week. Although the qualitative picture was similar across locales, the quantitative aspects of localized epidemics—including initial rate of growth, infections/deaths per capita, duration of plateau, and rapidity of resolution—were quite diverse across the country. Understanding the origins of this diversity will be key to predicting how the relaxation of social distancing, annual changes in weather, and static local demographic/population characteristics will affect the resolution of the first wave of cases, and will drive coming waves, prior to the availability of a vaccine.

The exponential growth rate of a spreading epidemic is dependent on the biological features of the virus-host ecosystem—including the incubation time, susceptibility of target cells to infection, and persistence of the virus particle outside of the host—but, through its dependence on the transmission rate between hosts, it is also a function of external factors such as population density, air humidity, and the fraction of hosts that are susceptible. Initial studies have shown that SARS-CoV-2 has a larger rate of exponential growth (or, alternatively, a lower doubling time of cases)[1] than many other circulating human viruses [11]. For comparison, the pandemic influenza of 2009, which also met a largely immunologically-naive population, had a doubling time of 5–10 d [12, 13], while that of SARS-CoV-2 has been estimated at 2–5 d [14, 15] (growth rates of $\sim 0.10\ \mathrm{d}^{-1}$ vs. $\sim 0.25\ \mathrm{d}^{-1}$). It is not yet understood which factors contribute to this high level of infectiousness.

While the dynamics of an epidemic (e.g., cases over time) must be described by numerical solutions to nonlinear models, the exponential growth rate, λ, usually has a simpler dependence on external factors. Unlike case or mortality incidence numbers, the growth rate does not scale with population size. It is a directly measurable quantity from the available incidence data, unlike, e.g., the reproduction number, which requires knowledge of the serial interval distribution [16–18], something that is difficult to determine empirically [19, 20]. Yet, the growth rate contains the same threshold as the reproduction number ($\lambda = 0$ vs. $R_0 = 1$), between a spreading epidemic (or an unstable uninfected equilibrium) and a contracting one (or an equilibrium that is resistant to flare-ups). Thus, the growth rate is an informative direct measure on that space of underlying parameters.

In this work, we leverage the enormous data set of epidemics across the United States to evaluate the impact of demographics, population density and structure, weather, and non-pharmaceutical interventions (i.e., mobility restrictions) on the exponential rate of growth of COVID-19. Following a brief analysis of the initial spread in metropolitan regions, we expand the meaning of the exponential rate

[1] The doubling time is $\ln 2$ divided by the exponential growth rate.

to encompass all aspects of a local epidemic—including growth, plateau and decline—and use it as a tracer of the dynamics, where its time dependence and geographic variation are dictated solely by these external variables and per capita cumulative mortality. Finally, we use the results of that linear analysis to calibrate a new nonlinear model—a renewal equation that utilizes the excursion probability of a random walk to determine the incubation period—from which we develop local predictions about the impact of social mobility relaxation, the level of herd immunity, and the potential of rebound epidemics in the Summer and Fall of 2020. The methodology can be modified to make local predictions as the pandemic evolves.

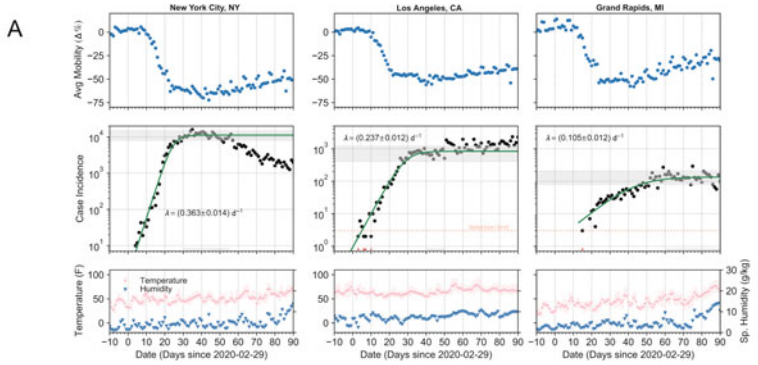
2 Results

2.1 Initial Growth of Cases in Metropolitan Regions Is Exponential with Rate Depending on Mobility, Population, Demographics, and Humidity

As an initial look at COVID-19's arrival in the United States, we considered the ∼100 most populous metropolitan regions—using maps of population density to select compact sets of counties representing each region (see [21])—and estimated the initial exponential growth rate of cases in each region. We performed a linear regression to a large set of demographic (sex, age, race) and population variables, along with weather and social mobility [22] preceding the period of growth (Fig. 1). In the best fit model ($R^2 = 0.75$, BIC $= -183$), the baseline value of the initial growth rate was $\lambda = 0.21\ \mathrm{d}^{-1}$ (doubling time of 3.3 d), with average mobility 2 weeks prior to growth being the most significant factor (Fig. 1b). Of all variables considered, only four others were significant: population density (including both *population-weighted density* (PWD)—also called the "lived population density" because it estimates the density for the average individual [23]—and *population sparsity*, γ, a measure of the difference between PWD and standard population density, see Methods), $p < 0.001$ and $p = 0.006$; specific humidity 2 weeks prior to growth, $p = 0.001$; and median age, $p = 0.04$.

While mobility reduction certainly caused the "flattening" of case incidence in every region by late-March, our results show (Fig. 1c) that it likely played a key role in reducing the *rate* of growth in Boston, Washington, DC, and Los Angeles, but was too late, with respect to the sudden appearance of the epidemic, to have such an effect in, e.g., Detroit and Cleveland. In the most extreme example, Grand Rapids, MI, seems to have benefited from a late arriving epidemic, such that its growth (with a long doubling time of 7 d) occurred almost entirely post-lockdown.

Specific humidity, a measure of absolute humidity, has been previously shown to be inversely correlated with respiratory virus transmission [24–27]. Here, we found it to be a significant factor, but weaker than population density and mobility (Fig. 1c). It could be argued that Dallas, Los Angeles, and Atlanta saw a small



B

Variable		Mean value	Coef.	Std. Err.	t	95% CI
average growth rate, λ_{avg}			0.232	0.005	43.1	[0.220, 0.242]
avg mobility, 2wk prior	–	+2.2%	0.0063	0.0006	10.8	[0.0051,0.0074]
ln[PWD]	–	8.1	0.048	0.010	4.7	[0.028,0.069]
sp. humidity, 2wk prior	–	4.9 g/kg	-0.010	0.0027	-3.7	[-0.016, -0.005]
pop. sparsity, γ	–	0.21	-0.33	0.11	-3.0	[-0.55, -0.11]
median age	–	35.7	0.0049	0.0024	2.1	[0.0002,0.0096]

C

	med. age	mob. 2wk pr.	pop. sparsity	hum. 2wk pr.	ln[pwpd]	λ_{exp}	λ_{exp}	λ_{exp} err
ny_newyorkcity	0.004	0.011	0.007	0.021	0.026	0.343	0.363	0.014
il_chicago	-0.002	0.022	0.02	0.018	0.017	0.307	0.316	0.017
mi_detroit	0.008	0.044	0.019	0.02	-0.023	0.301	0.412	0.035
oh_cleveland	0.014	0.035	0.02	0.018	-0.024	0.294	0.298	0.027
pa_philadelphia	0.004	0.01	0.015	0.014	0.012	0.286	0.291	0.009
ma_boston	0.003	-0.03	0.015	0.021	0.017	0.258	0.222	0.013
ct_hartford	0.012	-0.003	0.01	0.019	-0.014	0.255	0.287	0.019
co_denver	-0.009	0.022	-0.018	0.027	-0.002	0.252	0.24	0.014
wa_seattle	-0.003	0.02	0.002	0.007	-0.006	0.252	0.223	0.019
ma_worcester	0.009	0.003	0.001	0.024	-0.02	0.248	0.237	0.026
tx_dallas	-0.015	0.027	0.011	-0.016	0.021	0.24	0.252	0.021
ga_atlanta	-0.008	0.022	0.025	-0.006	-0.023	0.24	0.242	0.019
ca_losangeles	-0.007	0.0	-0.006	-0.016	0.027	0.229	0.237	0.012
dc_washington	-0.002	-0.047	0.008	0.003	0.012	0.206	0.213	0.009
mi_grandrapids	-0.006	-0.12	-0.001	0.015	-0.038	0.083	0.105	0.012

Fig. 1 Mobility and COVID-19 incidence data examples, and the results of linear regression to extracted initial exponential growth rates, λ_{exp}, in the top 100 metropolitan regions. (**a**) Three example cities with different initial growth rates. Data for Google mobility (blue points), daily reported cases (black points), and weather (red and blue points, bottom) are shown with a logistic fit to cases (green line). Data at or below detection limit were excluded from fits (dates marked by red points). Thin grey bars at base of cases graphs indicate region considered "flat", with right end indicating the last point used for logistic fitting; averaging over "flat" values generates the thick grey bars to guide the eye. [See Supp. Matt. in [21] for additional information and for complete data sets for all metropolitan regions.] (**b**) Weighted linear regression results in fit to λ_{exp} for all metropolitan regions. (**c**) Effect of each variable on growth rate (i.e., $\Delta\lambda$ values) for those regions with well-estimated case and death rates; white/yellow indicates a negative effect on λ, red indicates positive

benefit from higher humidity at the time of the epidemic's arrival, while the dry late-winter conditions in the Midwest and Northeast were more favorable to rapid transmission of SARS-CoV-2.

2.2 Exponential Growth Rate of Mortality as a Dynamical, Pan-Epidemic, Measure

In the remainder of this report, we consider the exponential rate of growth (or decay) in local confirmed deaths due to COVID-19. The statistics of mortality is poorer compared to reported cases, but it is much less dependent on unknown factors such as the criteria for testing, local policies, test kit availability, and asymptomatic individuals [28]. Although there is clear evidence that a large fraction of COVID-19 mortality is missed in the official counts (e.g., [29, 30]), mortality is likely less

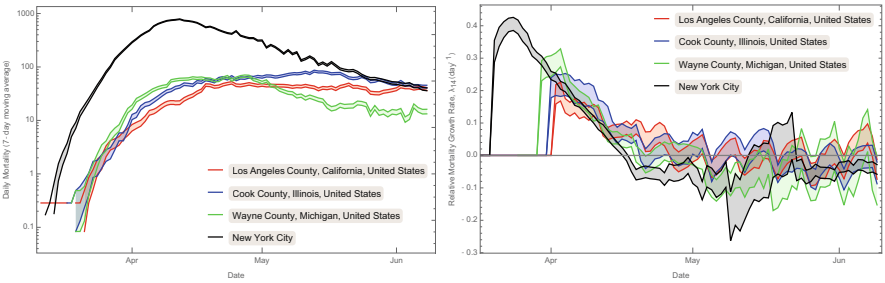


Fig. 2 COVID-19 mortality incidence (7-day rolling average, left) and exponential growth rate (λ_{14}, determined by regression of the logged mortality data over 14-day windows, right) for the four US counties with >2400 confirmed COVID-19 reported deaths (as of 8th June, 2020)

susceptible to rapid changes in reporting, and, as long as the number of reported deaths is a monotonic function of the actual number of deaths (e.g., a constant fraction, say 50%), the sign of the exponential growth rate will be unchanged, which is the crucial measure of the success in pandemic management.

To minimize the impact of weekly changes, such as weekend reporting lulls, data dumps, and mobility changes from working days to weekends, we calculate the regression of $\ln\left[\text{Mortality}\right]$ over a 14-day interval, and assign this value, $\lambda_{14}(t)$, and its standard error to the last day of the interval. Since only the data for distinct 2-week periods are independent, we multiply the regression errors by $\sqrt{14}$ to account for correlations between the daily estimates. Together with a "rolling average" of the mortality, this time-dependent measure of the exponential growth rate provides, at any day, the most up-to-date information on the progression of the epidemic (Fig. 2).

In the following section, we consider a linear fit to λ_{14}, to determine the statistically-significant external (non-biological) factors influencing the dynamics of local exponential growth and decline of the epidemic. We then develop a first-principles model for λ_{14} that allows for extrapolation of these dependencies to predict the impact of future changes in social mobility and climate.

2.3 Epidemic Mortality Data Explained by Mobility, Population, Demographics, Depletion of Susceptible Population and Weather, Throughout the First Wave of COVID-19

We considered a spatio-temporal dataset containing 3933 estimates of the exponential growth measure, λ_{14}, covering the 3 month period of 8 March 2020–8 June 2020 in the 187 US counties for which information on COVID-19 mortality and all potential driving factors, below, were available (the main barrier was social mobility information, which limited us to a set of counties that included 69% of

Table 1 Joint linear fit to $\lambda_{14}(t)$ data (Top). Any dependence with t-statistic below 2.5σ is considered not statistically significant. Joint linear fit to $\lambda_{14}(t)$, including only statistically significant dependencies (Bottom). For all coefficients, the population-weighted baseline is subtracted from the linear variable

	Estimate	Std. err	t-Statistic
Joint fit to all potential drivers			
Baseline mortality growth rate λ_{14}	0.195	0.011	17.2
COVID death fraction	−59.4	6.1	−9.7
Social mobility (2wks prior)	0.00238	0.00028	8.5
ln(population weighted density) 8.24	0.0412	0.0058	7.1
Social mobility (4wks prior)	0.00122	0.00019	6.6
Population sparsity 0.188	−0.249	0.063	−3.9
log(annual death) 4.04	−0.0301	0.0091	−3.3
Median age 37.47	0.0038	0.0012	3.0
People per household 2.76	0.023	0.014	1.6
Specific humidity (2wks prior) 5.92 g/kg	−0.0033	0.0031	−1.1
Temperature (2wks prior) 13.11 C	−0.00083	0.0013	−0.6
Temperature (4wks prior) 11.60 C	−0.00060	0.0014	−0.4
Specific humidity (4wks prior) 5.53 g/kg	0.00058	0.0032	0.2
Joint fit to statistically significant drivers			
Baseline mortality growth rate λ_{14}	0.198	0.011	18.7
COVID death fraction	−56.7	5.9	−9.7
Social mobility (2wks prior)	0.00236	0.00027	8.8
Social mobility (4wks prior)	0.00131	0.00017	7.6
ln(population weighted density) 8.24	0.0413	0.0058	7.2
Population sparsity 0.188	−0.260	0.061	−4.3
Specific humidity (2wks prior) 5.92 g/kg	−0.0047	0.0011	−4.1
log(annual death) 4.04	−0.0324	0.0088	−3.7
Median age 37.48	0.0040	0.0012	3.3

US mortality). A joint, simultaneous, linear fit of these data to 12 potential driving factors (Table 1) revealed only 7 factors with *independent* statistical significance. Re-fitting only to these variables returned the optimal fit for the considered factors (BIC $= -5951$; $R^2 = 0.674$).

We found, not surprisingly, that higher population density, median age, and social mobility correlated with positive exponential growth, while population sparsity, specific humidity, and susceptible depletion correlated with exponentially declining mortality. Notably the coefficients for each of these quantities was in the 95% confidence intervals of those found in the analysis of metropolitan regions (and vice versa). Possibly the most surprising dependency was the negative correlation, at $\simeq -3.7\sigma$ between λ_{14} and the *total* number of annual deaths in the county. In fact, this correlation was marginally more significant than a correlation with log(population), which was -3.3σ. One possible interpretation of this negative correlation is that the number of annual death is a proxy for the number of potential

outbreak clusters. The larger the number of clusters, the longer it might take for the epidemic to spread across their network, which would (at least initially) slow down the onset of the epidemic.

2.4 Nonlinear Model

To obtain more predictive results, we developed a mechanistic nonlinear model for infection (see [21] for details). We followed the standard analogy to chemical reaction kinetics (infection rate is proportional to the product of susceptible and infectious densities), but defined the generation interval (approximately the incubation period) through the excursion probability in a 1D random walk, modulated by an exponential rate of exit from the infected class. This approach resulted in a *renewal equation* [19, 31, 32], with a distribution of generation intervals that is more realistic than that of standard SIR/SEIR models, and which could be solved formally (in terms of the Lambert W function) for the growth rate in terms of the infection parameters:

$$\lambda = \frac{1}{2\tau}\left[\mathrm{W}\left(\sqrt{\frac{\beta\mathcal{S}\tau}{2}}\right)\right]^2 - d \tag{1}$$

The model has four key dependencies, which we describe here, along with our assumptions about their own dependence on population, demographic, and climate variables. As mortality (on which our estimate of growth rate is based) lags infection (on which the renewal equation is based), we imposed a fixed time shift of Δt for time-dependent variables:

1. We assumed that the susceptible population, which feeds new infections and drives the growth, is actually a sub-population of the community, consisting of highly-mobile and frequently interacting individuals, and that most deaths occurred in separate sub-population of largely immobile non-interacting individuals. Under these assumptions, we found (see Supp. Mat. in [21]) that the susceptible density, $S(t)$, could be estimated from the cumulative per capita death fraction, f_D, as:

$$S(t - \Delta t) = S(0)\exp\left[-C_D\, f_D(t)\right] \quad (f_D = D_{\mathrm{tot}}/N)\ ,$$

   where D_{tot} is the cumulative mortality count, N is the initial population, and the initial density is $S(0) = k$ PWD.
2. We assumed that the logarithm of the "rate constant" for infection, β, depended linearly on social mobility, m, specific humidity, h, population sparsity, γ, and

total annual death, A_D, as:

$$\begin{aligned}\ln\left[\beta\left(\mathcal{M},\mathcal{H},\gamma,A_D\right)\right] &= \ln\left[\beta_0\right] \\ &+ C_{\mathcal{M}}\left(\mathcal{M}-\bar{\mathcal{M}}\right) + C_{\mathcal{H}}\left(\mathcal{H}-\bar{\mathcal{H}}\right) \\ &+ C_{\gamma}\left(\gamma-\bar{\gamma}\right) + C_{A_D}\left(A_D-\bar{A}_D\right)\end{aligned} \tag{2}$$

where a barred variable represents the (population-weighted) average value over all US counties, and where the mobility and humidity factors were time-shifted with respect to the growth rate estimation window: $\mathcal{M} = m\left(t-\Delta t\right)$ and $\mathcal{H} = h\left(t-\Delta t\right)$.

3. The characteristic time scale to infectiousness, τ, is intrinsic to the biology and therefore we assumed it would depend only on the median age of the population, A. We assumed a power law dependence:

$$\tau = \tau_0\left(\frac{A}{A_0}\right)^{C_A} \tag{3}$$

where we fixed the pivot age, A_0, to minimize the error in τ_0.

4. The exponential rate of exit from the infected class, d, was assumed constant, since we found no significant dependence for it on other factors in our analysis of US mortality. From the properties of the Lambert W function, when the infection rate or susceptibility density approach zero—through mobility restrictions or susceptible depletion—the growth rate will tend to $\lambda \approx -d$, its minimum value.

With these parameterizations, we performed a nonlinear regression to $\lambda_{14}(t)$ using the entire set of US county mortality incidence time series (Table 2).

Table 2 Best-fit parameters for the nonlinear model using parametrization defined in the text

Parameter	Best-fit ± Std. err	Description
$\tau = \tau_0(\text{Median Age}/26.2\text{ years})^{C_A}$		Time from exposure to contagiousness
τ_0(day)	160 ± 58	Normalization
C_A	-2.26 ± 0.95	Age dependence
d^{-1}(day)	17.6 ± 2.2	Time from exposure to quarantine/recovery
C_D	3460 ± 610	Conversion constant, $f_D \to f_I$
β: Eq. (2)		Rate constant for infection
$\ln\left[k\beta_0\tau_0^{-2}(\text{m}^2/\text{day}^3)\right]$	0.37 ± 1.25	Normalization
$100C_{\mathcal{M}}$	8.08 ± 1.76	Dependence on social mobility
$C_{\mathcal{H}}$	-0.154 ± 0.055	Dependence on specific humidity
C_{γ}	-5.52 ± 2.35	Dependence on population sparsity
C_{A_D}	-1.05 ± 0.25	Dependence on total annual deaths

Compared to the linear model of the previous section (Table 1b), the fit improved by 7.6σ (BIC $= -6008$; $R^2 = 0.724$), despite both having 9 free parameters. Through the estimated parameter values, the model makes predictions for an individual's probability of becoming infectious, and the distributions of incubation period and generation interval, all as a function of the median age of the population (see Supplementary Material in [21]).

The model was very well fit to the mortality growth rate measurements for counties with a high mortality (Fig. 3). More quantitatively, the scatter of measured

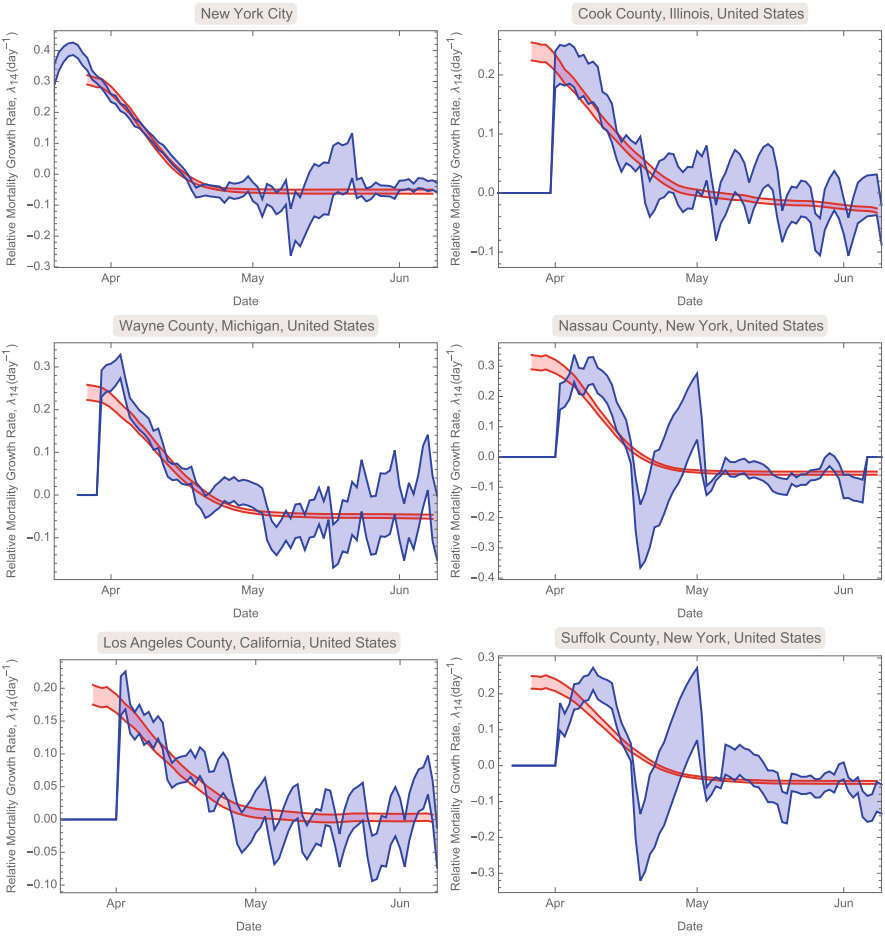


Fig. 3 Nonlinear model prediction (Eq. 1, red) for the actual (blue) mortality growth rate, in the six counties with highest reported death. Bands show 1-σ confidence region for both the model mean and the λ_{14} value

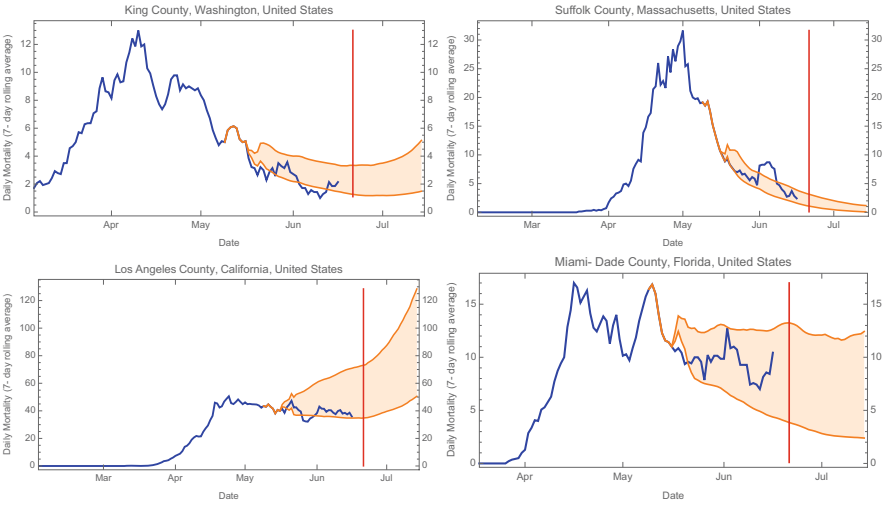


Fig. 4 Forecasts of COVID-19 mortality (orange)—based on the best-fit nonlinear model to data prior to May 16th, 2020—versus actual reported mortality (blue) for 4 large US counties. The 68% confidence range (orange regions) were determined from 100 random 60-day long simulations (see Supplementary Methods in [21]). The vertical red lines indicate June 21st. Forecasts for most US counties can be found at our online dashboard: https://wolfr.am/COVID19Dash

growth rates around the best-fit model predictions was (on average) only 13% larger than the measurement errors, independent of the population of the county.[2]

Importantly, when the model was calibrated on only a subset of the data—e.g., all but the final month for which mobility data is available—its 68% confidence prediction for the remaining data was accurate (Fig. 4) given the known mobility and weather data for that final month. This suggests that the model, once calibrated on the first wave of COVID-19 infections, can make reliable predictions about the ongoing epidemic, and future waves, in the United States.

2.5 Predictions for Relaxed Mobility Restrictions, the Onset of Summer, and the Potential Second Wave

Possibly the most pressing question for the management of COVID-19 in a particular community is the combination of circumstances at which the virus fails to propagate, i.e., at which the growth rate, estimated here by λ_{14}, becomes negative (or, equivalently, the reproduction number R_t falls below one). In the absence of mobility restrictions this is informally called the threshold for "herd immunity,"

[2] See [21] for more detailed discussion of Error Diagnostics.

which is usually achieved by mass vaccination (e.g., [33, 34]). Without a vaccine, however, ongoing infections and death will deplete the susceptible population and thus decrease transmission. Varying the parameters of the nonlinear model individually about their Spring 2020 population-weighted mean values (Fig. 5) suggests that this threshold will be very much dependent on the specific demographics, geography, and weather in the community, but it also shows that reductions in social mobility can significantly reduce transmission prior the onset of herd immunity.

To determine the threshold for herd immunity in the absence or presence of social mobility restrictions, we considered the "average US county" (i.e., a region with population-weighted average characteristics), and examined the dependence of the growth rate on the cumulative mortality. We found that in the absence of social distancing, a COVID-19 mortality rate of 0.13% (or 1300 per million population) would bring the growth rate to zero. However, changing the population density of this average county shows that the threshold can vary widely (Fig. 5).

Examination of specific counties showed that the mortality level corresponding to herd immunity varies from 10 to 2500 per million people (Fig. 6). At the current levels of reported COVID-19 mortality, we found that, as of June 22nd, 2020, only 128 ± 55 out of 3142 counties (inhabiting $9.4 \pm 2.1\%$ of US population) have surpassed this threshold at 68% confidence level (Fig. 7). Notably, New York City, with the highest reported per capita mortality (2700 per million) has achieved mobility-independent herd immunity at the 10σ confidence level, according to the model (Fig. 8). A few other large-population counties in New England, New Jersey, Michigan, Louisiana, Georgia and Mississippi that have been hard hit by the pandemic also appear to be at or close to the herd immunity threshold. This is not the case for most of the United States, however (Fig. 7). Nationwide, we predict that COVID-19 herd immunity would only occur after a death toll of $340,000 \pm 61,000$, or 1058 ± 190 per million of population.

We found that the approach to the herd immunity threshold is not direct, and that social mobility restrictions and other non-pharmaceutical interventions must be applied carefully to avoid excess mortality beyond the threshold. In the absence of social distancing interventions, a typical epidemic will "overshoot" the herd immunity limit (e.g., [35, 36]) by up to 300%, due to ongoing infections (Fig. 8). At the other extreme, a very strict "shelter in place" order would simply delay the onset of the epidemic; but if lifted (see Figs. 8 and 9), the epidemic would again overshoot the herd immunity threshold. A modest level of social distancing, however—e.g., a 33% mobility reduction for the average US county—could lead to fatalities "only" at the level of herd immunity. Naturally, communities with higher population density or other risk factors (see Fig. 5), would require more extreme measures to achieve the same.

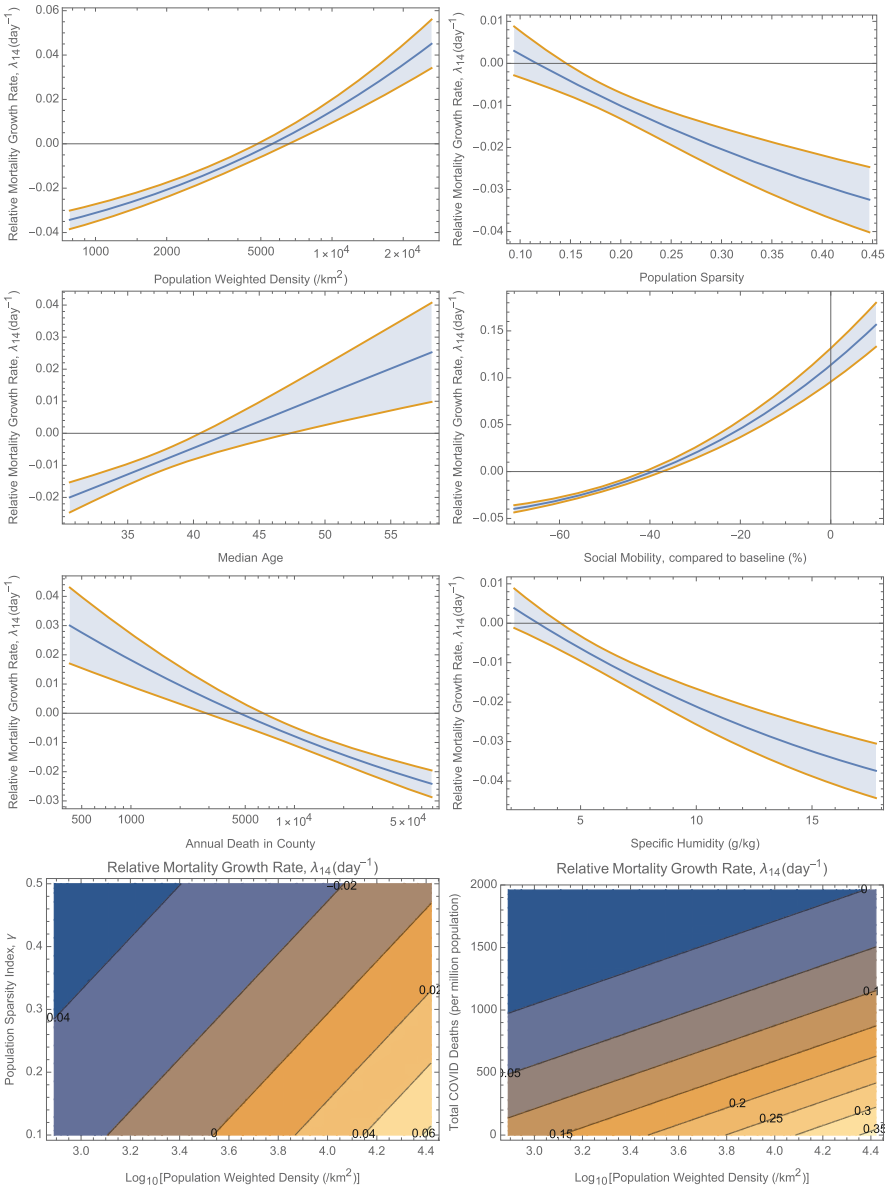


Fig. 5 Dependence and 68% confidence bands of the mortality growth rate—as specified by the nonlinear model (Eq. 1)—on various parameters for an "average county." All parameters not being varied are fixed at their population-weighted mean values (as of 8th June, 2020): $\log_{10}[\mathrm{PWD/km}^{-2})] = 3.58$, population sparsity = 0.188, COVID death fraction $= 5.1 \times 10^{-4}$ (510 deaths/million population), Median age = 37.5 yr, log(annual death) = 4.04, social mobility $\bar{\mathcal{M}} = -44\%$, and specific humidity $\bar{\mathcal{H}} = 5.7$ g/kg

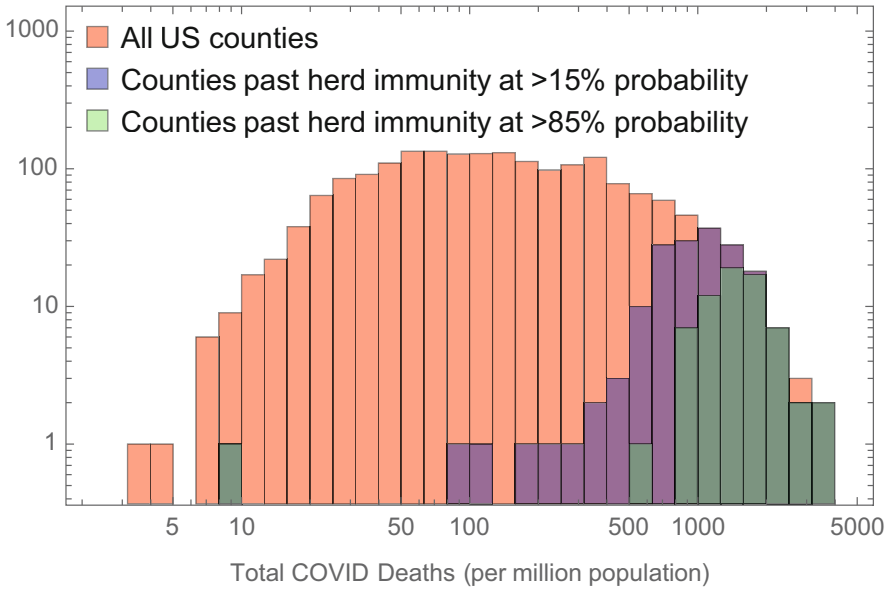


Fig. 6 Histogram of reported COVID-19 deaths per million for all US counties, showing the proportion that have passed "herd immunity" threshold, according to fit of the nonlinear model

Avoiding the level of mortality required for herd immunity will require long-lasting and effective non-pharmaceutical options, until a vaccine is available. The universal use of face masks has been suggested for reducing the transmission of SARS-CoV-2, with a recent meta-analysis [37] suggesting that masks can suppress the rate of infection by a factor of 0.07–0.34 (95% CI), or equivalently $\Delta \ln(\text{transmission}) = -1.9 \pm 0.4$ (at 1σ). Using our model's dependence of the infection rate constant on mobility, this would correspond to an equivalent social mobility reduction of $\Delta\bar{\mathcal{M}}_{\text{mask}} \simeq -24\% \pm 9\%$. Warmer, more humid weather has also considered a factor that could slow the epidemic (e.g., [38–40]). Annual changes in specific humidity are $\Delta\bar{\mathcal{H}} \simeq 6$ g/kg (Figure 10b in [21]), which can be translated in our model to an effective mobility decrease of $\Delta\bar{\mathcal{M}}_{\text{summer}} \simeq -12\% \pm 5\%$. Combining these two effects could, in this simple analysis, yield a modestly

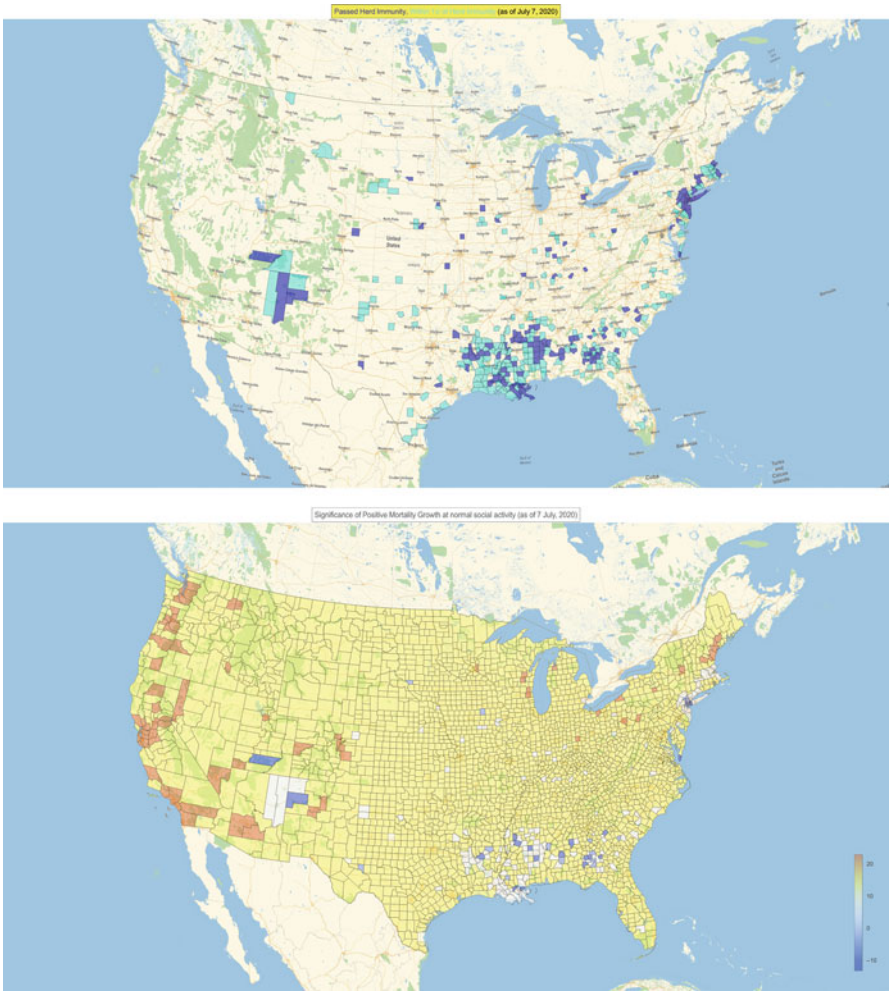


Fig. 7 Top: United States counties that have passed (blue), or are within (cyan), the threshold for "herd immunity" at the 1-σ level, as predicted by the nonlinear model. Bottom: Predicted confidence in the growth of COVID-19 outbreak (defined as predicted daily growth rate divided by its uncertainty), for all counties should they return today to their baseline (pre-COVID) social mobility. Counties that have approached the threshold of herd immunity have lower growth rates due to the depletion of susceptible individuals

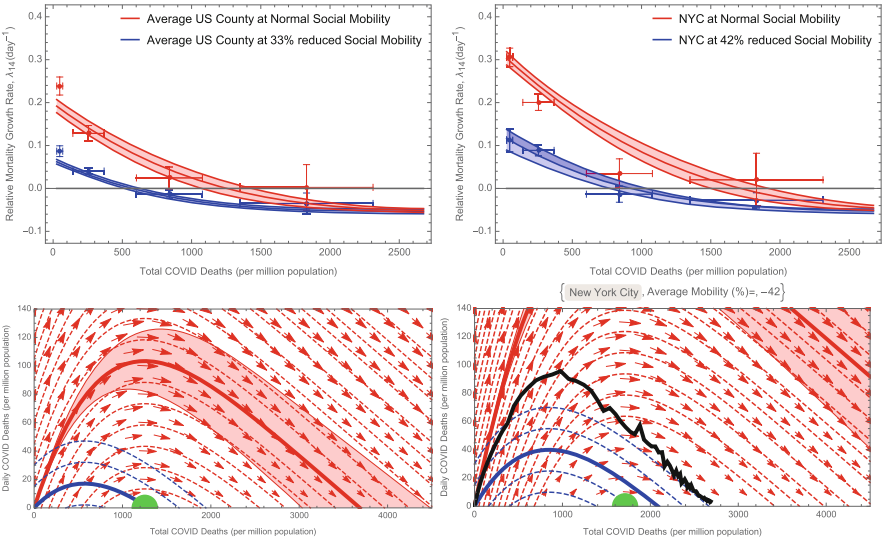


Fig. 8 Nonlinear model prediction of the exponential growth rate, λ_{14}, vs. cumulative COVID-19 mortality (top panels), assuming baseline social mobility, $\bar{\mathcal{M}} = 0$, in the "average US county" (see caption of Fig. 5) on the left, and New York City, on the right. The curves show 68% predictions for the nonlinear model (Table 2), while the points with errorbars are linear fits to all the data within bins of death fraction. The threshold for "herd immunity" ($\lambda_{14} = 0$) is reached at a mortality of approximately 1300 (1700) per million for an average county (NYC), but this would be higher in counties with more unfavorable values of the drivers. The eventual mortality burden of the average county will be determined by its path through a "phase space" of Daily vs. Total Mortality (bottom panel). An epidemic without intervention (red curves, with the particular trajectory starting at zero death shown in bold) will pass the threshold for herd immunity (1300 deaths per million; note that at zero daily deaths this is a fixed point) and continue to three times that value due to ongoing infections. A modest 33% reduction in social mobility (blue curves), however, leads to mortality at "only" the herd immunity level (the green disk). The black curve on the bottom right panel shows the 7-day rolling average of reported mortality for NYC, which appears to have "overshot" the "herd immunity threshold"

effective defense for the summer months: $\Delta\bar{\mathcal{M}}_{\text{mask+ summer}} \simeq -37\% \pm 10\%$. Therefore, this could be a reasonable strategy for most communities to manage the COVID-19 epidemic at the aforementioned -33% level of mobility needed to arrive at herd immunity with the least excess death. More stringent measures would be required to keep mortality below that level. Of course, this general prescription would need to be fine-tuned for the specific conditions of each community.

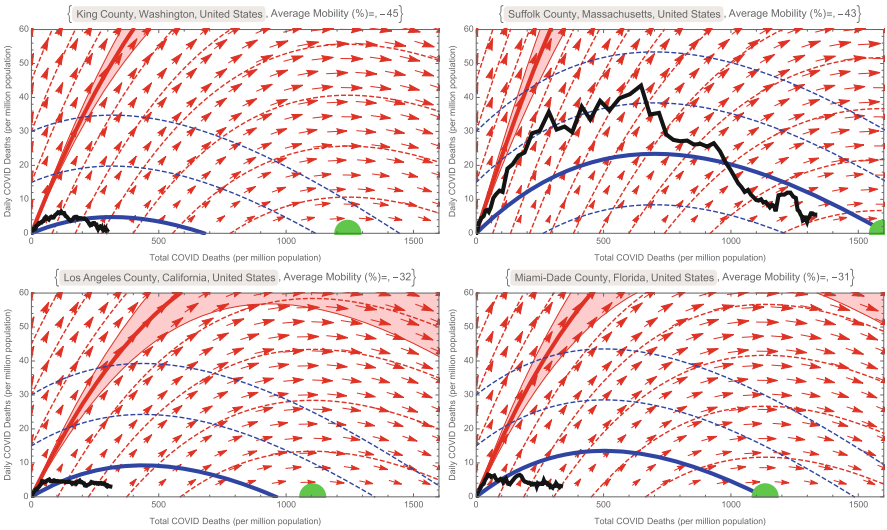


Fig. 9 Epidemic Phase Portraits for the same four counties as in Fig. 4, similar to the Phase portrait in Fig. 8. The blue curves are for the county's average Social Mobility during Feb. 15 through June 12, 2020, while red curves/arrows are at normal (pre-covid) social mobility. The thick black curve is the 7-day rolling average of the official reported mortality, while the green disk shows the threshold for "herd immunity"

3 Discussion and Conclusions

By simultaneously considering the time series of mortality incidence in every US county, and controlling for the time-varying effects of local social distancing interventions, we demonstrated for the first time a dependence of the epidemic growth of COVID-19 on population density, as well as other climate, demographic, and population factors. We further constructed a realistic, but simple, first-principles model of infection transmission that allowed us to extend our heuristic linear model of the dataset into a predictive nonlinear model, which provided a better fit to the data (with the same number of parameters), and which also accurately predicted late-time data after training on only an earlier portion of the data set. This suggests that the model is well-calibrated to predict future incidence of COVID-19, given realistic predictions/assumptions of future intervention and climate factors. We summarized some of these predictions in the final section of Results, notably that only a small fraction of US counties (with less than 10% of the population) seem to have reached the level of herd immunity, and that relaxation of mobility restrictions without counter-measures (e.g., universal mask usage) will likely lead to increased daily mortality rates, beyond that seen in the Spring of 2020.

In any epidemiological model, the infection rate of a disease is assumed proportional to population density [41], but, to our knowledge, its explicit effect in a real-world respiratory virus epidemic has not been demonstrated. The universal

reach of the COVID-19 pandemic, and the diversity of communities affected have provided an opportunity to verify this dependence. Indeed, as we show here, it must be accounted for to see the effects of weaker drivers, such as weather and demographics. A recent study of COVID-19 in the United States, working with a similar dataset, saw no significant effect due to population density [42], but our analysis differs in a number of important ways. First, we have taken a dynamic approach, evaluating the time-dependence of the growth rate of mortality incidence, rather than a single static measure for each county, which allowed us to account for the changing effects of weather, mobility, and the density of susceptible individuals. Second, we have included an explicit and real-time measurement of social mobility, i.e., cell phone mobility data provided by Google [22], allowing us to control for the dominant effect of intervention. Finally, and perhaps most importantly, we calculate for each county an estimate of the "lived" population density, called the population-weighted population density (PWD) [23], which is more meaningful than the standard population per political area. As with any population-scale measure, this serves as a proxy—here, for estimating the average rate of encounters between infectious and susceptible people—but we believe that PWD is a better proxy than standard population density, and it is becoming more prevalent, e.g., in census work [43, 44].

We also found a significant dependence of the mortality growth rate on specific humidity (although since temperature and humidity were highly correlated, a replacement with temperature was approximately equivalent), indicating that the disease spread more rapidly in drier (cooler) regions. There is a large body of research on the effects of temperature and humidity on the transmission of other respiratory viruses [27, 45], specifically influenza [46]. Influenza was found to transmit more efficiently between guinea pigs in low relative-humidity and temperature conditions [24], although re-analysis of this work pointed to absolute humidity (e.g., specific humidity) as the ultimate controller of transmission [25]. Although the mechanistic origin of humidity's role has not been completely clarified, theory and experiments have suggested a snowballing effect on small respiratory droplets that cause them to drop more quickly in high-humidity conditions [47–49], along with a role for evaporation and the environmental stability of virus particles [49, 50]. It has also been shown that the onset of the influenza season [26, 51]—which generally occurs between late-Fall and early-Spring, but is usually quite sharply peaked for a given strain (H1N1, H3N2, or Influenza B)—and its mortality [46] are linked to drops in absolute humidity. It is thought that humidity or temperature could be the annual periodic driver in the resonance effect causing these acute seasonal outbreaks of influenza [52, 53], although other influences, such as school openings/closings have also been implicated [54]. While little is yet known about the transmission of SARS-CoV-2 specifically, other coronaviruses are known to be seasonal [45, 55], and there have been some preliminary reports of a dependence on weather factors [56, 57]. We believe that our results represent the most definitive evidence yet for the role of weather, but emphasize that it is a weak, secondary driver, especially in the early stages of this pandemic where the susceptible fraction of the population remains large [58]. Indeed, the current early-summer rebound of COVID-19 in the

relatively dry and hot regions of the Southwest suggests that the disease spread will not soon be controlled by seasonality.

We developed a new model of infection in the framework of a renewal equation (see, e.g., [32] and references therein), which we could formally solve for the exponential growth rate. The incubation period in the model was determined by a random walk through the stages of infection, yielding a non-exponential distribution of the generation interval, thus imposing more realistic delays to infectiousness than, e.g., the standard SEIR model. In this formulation, we did not make the standard compartmental model assumption that the infection of an individual induces an autonomous, sequential passage from exposure, to infectiousness, to recovery or death; indeed, the model does not explicitly account for recovered or dead individuals. This freedom allows for, e.g., a back passage from infectious to noninfectious (via the underlying random walk) and a variable rate of recovery or death. We assumed only that the exponential growth in mortality incidence matched (with delay) that of the infected incidence—the primary dynamical quantity in the renewal approach—and we let the cumulative dead count predict susceptible density—the second dynamical variable in the renewal approach—under the assumption that deaths arise from a distinct subset of the population, with lower mobility behavior than those that drive infection (see [21]). Therefore, we fitted the model to the (rolling 2-week estimates of the) COVID-19 mortality incidence growth rate values, λ_{14}, for all counties and all times, and used the per capita mortality averaged over that period, f_D, to determine susceptible density. Regression to this nonlinear model was much improved over linear regression, and, once calibrated on an early portion of the county mortality incidence time series, the model accurately predicted the remaining incidence.

Because we accounted for the precise effects of social mobility in fitting our model to the actual epidemic growth and decline, we were then able to, on a county-by-county basis, "turn off" mobility restrictions and estimate the level of cumulative mortality at which SARS-CoV-2 would fail to spread even without social distancing measures, i.e., we estimated the threshold for "herd immunity." Meeting this threshold prior to the distribution of a vaccine should not be a goal of any community, because it implies substantial mortality, but the threshold is a useful benchmark to evaluate the potential for local outbreaks following the first wave of COVID-19 in Spring 2020. We found that a few counties in the United States have indeed reached herd immunity in this estimation—i.e., their predicted mortality growth rate, assuming baseline mobility, was negative—including counties in the immediate vicinity of New York City, Detroit, New Orleans, and Albany, Georgia. A number of other counties were found to be at or close to the threshold, including much of the greater New York City and Boston areas, and the Four Corners, Navajo Nation, region in the Southwest. All other regions were found to be far from the threshold for herd immunity, and therefore are susceptible to ongoing or restarted outbreaks. These determinations should be taken with caution, however. In this analysis, we estimated that the remaining fraction of susceptible individuals in the counties at or near the herd immunity threshold was in the range of 0.001% to 5% (see [21]). This is in strong tension with initial seroprevalence studies [59, 60]

which placed the fraction of immune individuals in New York City at 7% in late March and 20% in late April, implying that perhaps 75% of that population remains susceptible today. We hypothesize that the pool of susceptible individuals driving the epidemic in our model is a subset of the total population—likely those with the highest mobility and geographic reach—while a different subset, with very low baseline mobility, contributes most of the mortality (see [21]). Thus, the near total depletion of the susceptible pool we see associated with herd immunity corresponds to the highly-mobile subset, while the low-mobility subset could remain largely susceptible. One could explicitly consider such factors of population heterogeneity in a model—e.g., implementing a saturation of infectivity as a proxy for a clustering effect [61–64]—but we found (in results not shown) that the introduction of additional of parameters left portions of the model unidentifiable. Despite these cautions, it is interesting to note that the epidemic curves (mortality incidence over time) for those counties that we have predicted an approach to herd immunity are qualitatively different than those we have not. Specifically, the exponential rise in these counties is followed by a peak and a sharp decline—rather than the flattening seen in most regions—which is a typical feature of epidemic resolution by susceptible depletion.

At the time of this writing, in early Summer 2020, confirmed cases are again rising sharply in many locations across the United States—particularly in areas of the South and West that were spared significant mortality in the Spring wave. The horizon for an effective and fully-deployed vaccine still appears to be at least a year away. Initial studies of neutralizing antibodies in recovered COVID-19 patients, however, suggest a waning immune response after only 2–3 months, with 40% of those that were asymptomatic becoming seronegative in that time period [65]. Although the antiviral remdesivir [66–68] and the steroid Dexamethasone [69] have shown some promise in treating COVID-19 patients, the action of remdesivir is quite weak, and high-dose steroids can only be utilized for the most critical cases. Therefore, the management of this pandemic will likely require non-pharmaceutical intervention—including universal social distancing and mask-wearing, along with targeted closures of businesses and community gathering places—for years in the future. The analysis and prescriptive guidance we have presented here should help to target these approaches to local communities, based on their particular demographic, geographic, and climate characteristics, and can be facilitated through our http://mylocalcovid.uwaterloo.ca/ online simulator dashboard. Finally, although we have focused our analysis on the United States, due to the convenience of a diverse and voluminous data set, the method and results should be applicable to any community worldwide, and we intend to extend our analysis in forthcoming work.

Acknowledgments We are indebted to helpful comments and discussions by our colleagues, in particular Bruce Bassett, Ghazal Geshnizjani, David Spergel, and Lee Smolin. NA is partially supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. BPH acknowledges sabbatical support from Grand Valley State University, and is grateful to the hospitality of the University of Waterloo during his stay.

References

1. N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, et al., The Lancet **395**(10223), 507 (2020)
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, et al. 2020 Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine 382, 1199–1207, doi: https://doi.org/10.1056/NEJMoa2001316.
3. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A, et al. 2020 First case of 2019 novel coronavirus in the United States. New England Journal of Medicine 382, 929–936, doi: https://doi.org/10.1056/NEJMoa2001191.
4. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, Davies N, Gimma A, van Zandvoort K, Gibbs H, Hellewell J, Jarvis CI, Cliord S, Quilty BJ, Bosse NI, Abbott S, Klepac P, Flasche S. 2020 Early dynamics of transmission and control of COVID-19: A mathematical modelling study. The Lancet Infectious Diseases 20, 553–558, doi: https://doi.org/10.1016/S1473-3099(20)30144-4.
5. Chu HY, Englund JA, Starita LM, Famulare M, Brandstetter E, Nickerson DA, Rieder MJ, Adler A, Lacombe K, Kim AE, et al. 2020 Early detection of COVID-19 through a citywide pandemic surveillance platform. New England Journal of Medicine 383, 185–187.
6. Ladner JT, Larsen BB, Bowers JR, Hepp CM, Bolyen E, Folkerts M, Sheridan K, Pfeier A, Yaglom H, Lemmer D, Sahl JW, Kaelin EA, Maqsood R, Bokulich NA, Quirk G, Watts TD, Komatsu KK, Waddell V, Lim ES, Caporaso JG, Engelthaler DM, Worobey M, Keim P, Fraser CM. 2020 An early pandemic analysis of SARS-CoV-2 population structure and dynamics in Arizona. mBio 11, e02107-20, doi: https://doi.org/10.1128/mBio.02107-20.
7. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, Fabre S, Kleiner G, Polanco J, Khan Z, et al. 2020 Introductions and early spread of SARS-CoV-2 in the New York City area. Science 369, 297–301.
8. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, Lemey P. 2020 The emergence of SARS-CoV-2 in Europe and North America. Science 370, 564–570.
9. X. Deng, W. Gu, S. Federman, L. Du Plessis, O. Pybus, N. Faria, C. Wang, G. Yu, C.Y. Pan, H. Guevara, et al., Science (2020). https://doi.org/10.1126/science.abb9263
10. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, Vogels CB, Brito AF, Alpert T, Muyombwe A, et al. 2020 Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. Cell 181, 990–996.
11. Park SW, Bolker BM, Champredon D, Earn DJ, Li M, Weitz JS, Grenfell BT, Dusho_J. 2020 Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: Framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. Journal of the Royal Society Interface 17, 20200144.
12. H. Yu, S. Cauchemez, C.A. Donnelly, L. Zhou, L. Feng, N. Xiang, J. Zheng, M. Ye, Y. Huai, Q. Liao, et al., Emerging infectious diseases **18**(5), 758 (2012)
13. A.D. Storms, M.D. Van Kerkhove, E. Azziz-Baumgartner, W.K. Lee, M.A. Widdowson, N.M. Ferguson, A.W. Mounts, Influenza and other respiratory viruses **7**(6), 1328 (2013)
14. S. Sanche, Y. Lin, C. Xu, E. Romero-Severson, N. Hengartner, R. Ke, Emerging infectious diseases **26**(7) (2020)
15. B. Oliveiros, L. Caramelo, N.C. Ferreira, F. Caramelo, medRxiv (2020)
16. J. Wallinga, M. Lipsitch, Proceedings of the Royal Society B: Biological Sciences **274**(1609), 599 (2007)
17. M. Roberts, J. Heesterbeek, Journal of mathematical biology **55**(5–6), 803 (2007)
18. Dusho_J, Park SW. 2021 Speed and strength of an epidemic intervention. Proceedings of the Royal Society B 288, 20201556.
19. D. Champredon, J. Dushoff, Proceedings of the Royal Society B: Biological Sciences **282**(1821), 20152026 (2015)
20. H. Nishiura, Mathematical Biosciences & Engineering **7**(4), 851 (2010)

21. N. Afshordi, B. P. Holder, M. Bahrami, D. Lichtblau, arXiv preprint arXiv:2007.00159 (2020)
22. J. Fitzpatrick, D. Karen, COVID-19 community mobility reports. Tech. rep., Google (2020). Available at: https://www.google.com/covid19/mobility/
23. J. Craig, Population Trends **39**, 16 (1985)
24. A. Lowen, S. Mubareka, J. Steel, P. Palese, **3**(10), e151 (2007). https://doi.org/10.1371/journal.ppat.0030151
25. J. Shaman, M. Kohn, Proceedings of the National Academy of Sciences **106**(9), 3243 (2009)
26. J. Shaman, E. Goldstein, M. Lipsitch, American journal of epidemiology **173**(2), 127 (2011)
27. E. Kudo, E. Song, L.J. Yockey, T. Rakib, P.W. Wong, R.J. Homer, A. Iwasaki, Proceedings of the National Academy of Sciences **116**(22), 10905 (2019)
28. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. 2020 Accurate statistics on COVID-19 are essential for policy guidance and decisions. American Journal of Public Health 110, 949–951, doi: https://doi.org/10.2105/AJPH.2020.305708.
29. D.A. Leon, V.M. Shkolnikov, L. Smeeth, P. Magnus, M. Pechholdová, C.I. Jarvis, The Lancet **395**(10234), e81 (2020)
30. C. Modi, V. Boehm, S. Ferraro, G. Stein, U. Seljak, medRxiv (2020). https://doi.org/10.1101/2020.04.15.20067074. URL https://www.medrxiv.org/content/early/2020/05/14/2020.04.15.20067074
31. J. Heesterbeek, K. Dietz, Statistica Neerlandica **50**(1), 89 (1996)
32. D. Champredon, J. Dushoff, D.J. Earn, SIAM Journal on Applied Mathematics **78**(6), 3258 (2018)
33. T.J. John, R. Samuel, European Journal of Epidemiology **16**(7), 601 (2000). URL https://doi.org/10.1023/A:1007626510002
34. P. Fine, K. Eames, D.L. Heymann, Clinical Infectious Diseases **52**(7), 911 (2011). URL https://doi.org/10.1093/cid/cir007
35. A. Handel, I.M. Longini Jr, R. Antia, Proceedings of the Royal Society B: Biological Sciences **274**(1611), 833 (2007)
36. I.C.H. Fung, R. Antia, A. Handel, PLoS One **7**(6), e36573 (2012)
37. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schunemann HJ, El-harakeh A, Bognanni A, Lot T, Loeb M, et al. 2020 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. The Lancet 395, 1973–1987.
38. Wang J, Tang K, Feng K, Lin X, Lv W, Chen K, Wang F. 2021 Impact of temperature and relative humidity on the transmission of COVID-19: A modelling study in China and the United States. BMJ Open 11, e043863.
39. A. Notari, arXiv e-prints arXiv:2003.12417 (2020)
40. R. Xu, H. Rahmandad, M. Gupta, C. DiGennaro, N. Ghaffarzadegan, H. Amini, M.S. Jalali, medRxiv (2020). https://doi.org/10.1101/2020.05.05.20092627. URL https://www.medrxiv.org/content/early/2020/05/24/2020.05.05.20092627
41. M.C. de Jong, O. Diekmann, H. Heesterbeek, in *Epidemic Models: Their structure and relation to data*, ed. by D. Mollison (Cambridge University Press, Cambridge, 1995), pp. 84–94
42. S. Hamidi, S. Sabouri, R. Ewing, Journal of the American Planning Association **0**(0), 1 (2020). URL https://doi.org/10.1080/01944363.2020.1777891
43. Dorling D, Atkins D. 1995 Population density, change and concentration in Great Britain 1971, 1981 and 1991, number 58 in Studies on Medical and Population Subjects, London: HMSO.
44. S.G. Wilson, *Patterns of metropolitan and micropolitan population change: 2000 to 2010* (US Department of Commerce, Economics and Statistics Administration, 2012)
45. M. Moriyama, W.J. Hugentobler, A. Iwasaki, Annual Review of Virology **7**(1), null (2020). URL https://doi.org/10.1146/annurev-virology-012420-022445. PMID: 32196426
46. A.I. Barreca, J.P. Shimshack, American Journal of Epidemiology **176**(suppl_7), S114 (2012). URL https://doi.org/10.1093/aje/kws259
47. R. Tellier, Journal of the Royal Society Interface **6**(suppl_6), S783 (2009)
48. J.D. Noti, F.M. Blachere, C.M. McMillen, W.G. Lindsley, M.L. Kashon, D.R. Slaughter, D.H. Beezhold, PloS one **8**(2), e57485 (2013)

49. L.C. Marr, J.W. Tang, J. Van Mullekom, S.S. Lakdawala, Journal of The Royal Society Interface **16**(150), 20180298 (2019). https://doi.org/10.1098/rsif.2018.0298. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2018.0298
50. L. Morawska, in *Proceedings of Indoor Air 2005: the 10th International Conference on Indoor Air Quality and Climate* (Springer, 2005), pp. 9–23
51. J. Shaman, V.E. Pitzer, C. Viboud, B.T. Grenfell, M. Lipsitch, PLoS Biology **8**(2), e1000316 (2010)
52. J. Dushoff, J.B. Plotkin, S.A. Levin, D.J.D. Earn, **101**(48), 16915 (2004). https://doi.org/10.1073/pnas.0407293101
53. J. Tamerius, M.I. Nelson, S.Z. Zhou, C. Viboud, M.A. Miller, W.J. Alonso, Environmental health perspectives **119**(4), 439 (2011)
54. D.J. Earn, D. He, M.B. Loeb, K. Fonseca, B.E. Lee, J. Dushoff, Annals of internal medicine **156**(3), 173 (2012)
55. R.A. Neher, R. Dyrdak, V. Druelle, E.B. Hodcroft, J. Albert, Swiss medical weekly **150**(1112) (2020)
56. R. Xu, H. Rahmandad, M. Gupta, C. DiGennaro, N. Ghaffarzadegan, H. Amini, M.S. Jalali, medRxiv (2020). https://doi.org/10.1101/2020.05.05.20092627. URL https://www.medrxiv.org/content/early/2020/05/24/2020.05.05.20092627
57. M. Schell, B.D. Gonzalez, J. Greene, A. Giuliano, Available at SSRN 3579744 (2020). URL https://doi.org/10.2139/ssrn.3579744
58. Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. 2020 Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. Science 369, 315–319.
59. E.S. Rosenberg, J.M. Tesoriero, E.M. Rosenthal, R. Chung, M.A. Barranco, L.M. Styer, M.M. Parker, S.Y.J. Leung, J. Morne, D. Greene, D.R. Holtgrave, D. Hoefer, J. Kumar, T. Udo, B. Hutton, H.A. Zucker, medRxiv (2020). https://doi.org/10.1101/2020.05.25.20113050. URL https://www.medrxiv.org/content/early/2020/05/29/2020.05.25.20113050
60. F.P. Havers, C. Reed, T.W. Lim, J.M. Montgomery, J.D. Klena, A.J. Hall, A.M. Fry, D.L. Cannon, C.F. Chiang, A. Gibbons, I. Krapiunaya, M. Morales-Betoulle, K. Roguski, M. Rasheed, B. Freeman, S. Lester, L. Mills, D.S. Carroll, S.M. Owen, J.A. Johnson, V.A. Semenova, J. Schiffer, N.P. Thornburg, medRxiv (2020). https://doi.org/10.1101/2020.06.25.20140384. URL https://www.medrxiv.org/content/early/2020/06/26/2020.06.25.20140384
61. V. Capasso, G. Serio, Mathematical Biosciences **42**(1–2), 43 (1978)
62. D. Mollison, in *Population Dynamics of Rabies in Wildlife* (1985), pp. 223–234
63. R.J. De Boer, Journal of virology **81**(6), 2838 (2007)
64. A. Farrell, C. Brooke, K. Koelle, R. Ke, BioRxiv p. 547349 (2019)
65. Q.X. Long, X.J. Tang, Q.L. Shi, Q. Li, H.J. Deng, J. Yuan, J.L. Hu, W. Xu, Y. Zhang, F.J. Lv, et al., Nature Medicine pp. 1–5 (2020)
66. J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M.d. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H.C. Lane, New England Journal of Medicine **0**(0), null (2020). URL https://doi.org/10.1056/NEJMoa2007764
67. J. Grein, N. Ohmagari, D. Shin, G. Diaz, E. Asperges, A. Castagna, T. Feldt, G. Green, M.L. Green, F.X. Lescure, E. Nicastri, R. Oda, K. Yo, E. Quiros-Roldan, A. Studemeister, J. Redinski, S. Ahmed, J. Bernett, D. Chelliah, D. Chen, S. Chihara, S.H. Cohen, J. Cunningham, A. D'Arminio Monforte, S. Ismail, H. Kato, G. Lapadula, E. L'Her, T. Maeno, S. Majumder, M. Massari, M. Mora-Rillo, Y. Mutoh, D. Nguyen, E. Verweij, A. Zoufaly, A.O. Osinusi, A. DeZure, Y. Zhao, L. Zhong, A. Chokkalingam, E. Elboudwarej, L. Telep, L. Timbs, I. Henne, S. Sellers, H. Cao, S.K. Tan, L. Winterbourne, P. Desai, R. Mera, A. Gaggar, R.P. Myers, D.M. Brainard, R. Childs, T. Flanigan, New England Journal of Medicine **382**(24), 2327 (2020). URL https://doi.org/10.1056/NEJMoa2007016

68. Y. Wang, D. Zhang, G. Du, R. Du, J. Zhao, Y. Jin, S. Fu, L. Gao, Z. Cheng, Q. Lu, et al., The Lancet (2020)
69. P. Horby, W.S. Lim, J. Emberson, M. Mafham, J. Bell, L. Linsell, N. Staplin, C. Brightling, A. Ustianowski, E. Elmahi, B. Prudon, C. Green, T. Felton, D. Chadwick, K. Rege, C. Fegan, L.C. Chappell, S.N. Faust, T. Jaki, K. Jeffery, A. Montgomery, K. Rowan, E. Juszczak, J.K. Baillie, R. Haynes, M.J. Landray, medRxiv (2020). https://doi.org/10.1101/2020.06.22.20137273. URL https://www.medrxiv.org/content/early/2020/06/22/2020.06.22.20137273

Describing, Modelling and Forecasting the Spatial and Temporal Spread of COVID-19: A Short Review

Julien Arino

1 Introduction

On 21 February 2003, a disease now known as the severe acute respiratory syndrome (SARS) arrived in Hong Kong when a physician from Guandong Province in Mainland China bearing the SARS coronavirus (SARS-CoV), checked in at the Metropole Hotel [54]. The primary case in Hong Kong was by no means the index case: the virus had been circulating in Guandong since at least November 2002 [52]. However, that patient triggered a chain of infections that, together with earlier cases in China, led to 8098 known cases and 774 deaths in 28 countries [42] and was declared a pandemic by the World Health Organisation (WHO).

Similarly, it is not certain at the time of writing that SARS-CoV-2 and its associated disease COVID-19 had its index case in Wuhan, Hubei Province, China. What is certain, on the other hand, is that it is in Wuhan that COVID-19 underwent its first noticeable amplification phase, following which it spread rapidly across the world, to the point that there are now very few top level jurisdictions not having reported COVID-19 cases.

SARS-CoV-2 is the third novel Coronavirus to emerge in the twenty-first century (after SARS and the Middle East respiratory syndrome—MERS [28]) and the second to generate a pandemic (a third pandemic was triggered by the H1N1 influenza outbreak in 2009). COVID-19 is also the most devastating pandemic in over a century in terms of its death toll as well as its economic and societal impact.

Here, I review some aspects of the spatio-temporal spread of COVID-19. Some caveats are in order. Firstly, while spatial epidemiology is not the most popular topic among modellers, it does remain a vast field where a myriad of approaches coexist;

J. Arino (✉)
Department of Mathematics, University of Manitoba, Winnipeg, MB, Canada
e-mail: Julien.Arino@umanitoba.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_2

surveying the work done on the topic would require an entire monograph. While I have tried to be inclusive, it is certain that I have omitted some topics or techniques. I am for instance making the choice to describe mostly mechanistic models of spread, be they deterministic or stochastic, mathematical or computational. Some very good statistical work has been published on the topic of COVID-19, but I focus here to a large extent on models that can explain reality perhaps at the detriment of forecasting power. Secondly, I am aware that many modellers who have worked or are working with public health authorities may not have had the time yet to publish their work. Except for my work, I report here only on papers already published or available on recognised preprint servers. Thirdly, new variants of concern (VoC) were detected while this paper was under review. A short section at the end of this document to describe what little is known in terms of spatio-temporal spread of these variants.

Finally, even though some work makes use of data at a very fine spatial resolution, in keeping with the philosophy of some prior work [13], I focus on models that can be used with publicly available data.

This review is organised as follows. First, in Sect. 2, I provide an overview of the mechanisms that lead to the spatial spread of infections and three of the major types of models that have been used to study it. In Sect. 3, I then describe the spread of COVID-19 from a chronological point of view. Finally, in Sect. 4, I discuss modelling work specific to COVID-19.

2 Spatialised Infections: Mechanisms and Models

Before considering work specific to COVID-19, let me spend some time on the spatialisation of infectious diseases in general. Indeed, while COVID-19 presents specific challenges, it is by no means the first spatial epidemic that humanity is confronted to; for instance, a simple description of the spatio-temporal trajectory of the Plague of Athens can be found in the History of the Peloponnesian War [176], which was written almost 2500 years ago; on a more local scale, spatial epidemiology can be traced to the cholera epidemic of London in 1854 [171]. There is therefore much understanding to be gained about the current crisis by considering what was known prior to its start.

2.1 How Does an Infectious Disease Become Spatial?

Different conceptual models explain the mechanisms that lead to the spatialisation of an infectious disease, leading to potentially different modelling paradigms.

Working at the level of the individual, one can envision spatial spread as the repetition of inter-individual spread events. Individuals are mobile in space and it is their movement while bearing the infectious pathogen that leads to the

disease becoming spatial, when they come into contact with susceptible individuals who are also mobile. This description falls mostly into what have been called Markovian contact processes [146]. When considered at the population level, this leads to models using partial differential equations and is particularly appropriate for describing the spread of a disease where the hosts can move freely, such as epizooties. Such a description can also lead to network or agent-based models.

Work with my collaborators usually instead focuses on locations and adopts a vision of spatial spread articulated in [17]: an infectious disease becomes spatialised by the repetition of processes summarised as *importation*, *amplification*, *exportation* and *transport*. Importation itself is the event when an individual infected with the disease reaches a new location. Importation is *successful* if the imported case leads to at least one local transmission event. This way of thinking about spread is easy to reconcile with data, since locations are jurisdictions in the context of public health. It also matches the *cones of resolution* that some geographers use when they think about the spatial spread of epidemics; see [11] and references therein. See also [110, 111], which consider the roles of the different levels of mobility on the spread of SARS in and to and from Beijing.

2.2 What Are the Main Drivers of Spatial Spread?

Whatever the way one conceptualises the spatialisation process, the main driver of spatial spread is human mobility. Long range fast movements using air travel have considerably changed the way diseases spread and while amplification in a location remains driven by population effects, the initial spread is to a large extent driven by air travel. This was shown for SARS [42], the 2009 H1N1 influenza pandemic [119] and MERS [90], for instance. Long distance high speed train travel has also been associated to spread; see, e.g., [47]. It is interesting to note, though, that despite the highly heterogeneous nature of spatio-temporal spread brought on by modern travel modalities, continental-level effects can still be observed [93].

2.3 How Does One Model a Spatialised Infectious Disease?

As mentioned in the Introduction, I focus here on mechanistic models. There are many ways to model the spatio-temporal spread of infectious diseases. Let me present the main contenders; see an interesting and more complete list in [161] or [147]. I do not detail reaction-diffusion equations, because, to the best of my knowledge, they have seen very little use in modelling the spatio-temporal spread of COVID-19; readers are referred to [158], for instance, for more details on deterministic aspects involving such systems. See [145] for a seminal review of the link between stochastic and deterministic spatial models, as well as interesting overviews in [61].

2.3.1 Agent-Based Models

Agent-based models (ABM) consider populations of autonomous *agents* that interact following some rules [76]. Agents have a set of characteristics that can be modified through their interactions with other agents. Although there are some attempts to mathematise some of the properties of such systems, they remain for the most part computational tools that need to be studied using a large number of simulations. Their strength lies in their realism: an agent can be given realistic behavioural characteristics (schedule, place of residence or of work, etc.). ABM are also easier to implement because they require very little mathematical background. Agent-based models have proved most useful when considering the effect of individual behaviours on the spread of infection in smaller populations. For instance, they have been used to study individual protective behaviour [116], the effect of presenteeism while infected with a disease [126], the risk in small isolated communities [49, 129], the effect of social distancing [169] or the role of avoidance behaviour when vaccines have low effectiveness [177]. Examples of spatialised problems (all about influenza) that were studied using agent-based models include its spread in slums of Delhi [5], the use of a hybrid approach involving networks to describe the social structure and ABM to describe inter-individual spread in Forsyth Country, NC [98], the potential for social structure to generate inequalities in incidence in different areas of a county [127] or the spread within an airport terminal [165]. See also [154, 155], which use a detailed location survey to conduct a simulation of the spread of influenza in Japan. ABM are also useful as a means to model evolution; see, for instance, [96], where an ABM is used to model evolution of virulence at the front line of a spreading epidemic

The area where ABM have proved most informative is when considering spread of infections within areas where movement is constrained and generalised contact is impossible, such as buildings or cruise ships. For instance, when considering nosocomial infections, it is possible to monitor health care personnel movement and use this data to parametrise an ABM of spread within a hospital [109], or to formulate a model of spread between beds within an intensive care unit [109].

However, ABM lose in value when populations become larger, except in rare instances where unexpected *emergent* behaviour occurs. Where the law of large number applies, it is indeed less computationally onerous to use "classic" deterministic or stochastic models. For instance, the model in [18] reproduces almost exactly the behaviour of the agent-based model in [134], but furthermore gives access to explicit expressions of the basic reproduction number $\mathcal{R}_0$ and the final size of the epidemic. See also the comparisons in [7, 66].

Altogether, agent-based models are powerful tools of investigation at the hyper-local scale. Considering agents consisting of groups of individuals instead of individuals also allows to operate at a higher spatial scale, although models then lose some of the interesting properties they have at the finer scale.

2.3.2 Network Models

Network models are very similar to agent-based models, of which they were, essentially, the inspiration. The two types of models are sometimes difficult to distinguish. In epidemiology, work on networks and ABM was popularised in particular by the NIGMS Models of Infectious Disease Agent Study, which led for instance to EpiSimS [56, 77]. In network models, *nodes* are typically simpler than agents in agent-based models; the most straightforward example would be a network consisting of nodes (individuals) that can be in two states: susceptible to the disease or infected and infectious with it. If there is an edge in the network between two nodes, this means the two nodes came into contact; in the case of a network used to model disease spread, this indicates that a contact took place, which could lead to the transmission of the disease. One promising direction of research that has been explored using networks is that of the link between network structure and shape of the epidemic curve; see, e.g., [48, 57]. This is particularly important during the early spread of a disease and has been considered in a variety of contexts using network models.

Because they are simpler than ABM, network models are more amenable to analysis; see, e.g., [34, 38]. Originally, tools used to study the dynamics of network epidemic models originated in statistical mechanics [139, 140]. Because networks allow to incorporate a more realistic description of the contact process while maintaining some level of analytic tractability, comparing their dynamics with that of classical models is useful. In [8], this is done for instance for a two-strains influenza model with vaccination. [156] There has been a move lately towards characterising the dynamics of smaller networks using the properties of individual nodes rather than through distributions of these properties; see, e.g., [32].

While examples of use of network models in mathematical epidemiology abound, their use in situations that are specifically spatial are not as common. Instances include [79], who considered an SEIR model set in a lattice and simulated using a Monte Carlo process, to incorporate both stochasticity and space, [33], who consider the spread of dengue in city blocks or [82], who consider the spread of equine influenza. The latter paper illustrates the strength of the method, in that they have access to extensive data on horse movement between locations and are able to assess the effect of the topology of the networks, both for long range movement and shorter contact patterns with locations, on the spread of the infection.

Networks are also a natural candidate for considering the spread of infections using the air transportation network as a conduit. This was done for instance with SARS [42].

2.3.3 Metapopulation Models

Also known as patch models, metapopulations couple together (typically similar) models, with each model encoding for the dynamics of the disease in a population and coupling representing the movement of individuals between the

populations [16], the average time spent in remote locations [37] or the interactions between populations. Metapopulation models had been used computationally since the early 1970s, for instance to consider the spread of influenza within countries [83, 164]. They have known a resurgence since the beginning of the twenty-first century, because, on the one hand, computing resources made simulating them easier and, on the other hand, papers such as [15, 19] showed that linear algebra techniques could be used to render the study of such systems very similar to that of their constituting systems. See [31] for a list of problems the authors identify as interesting challenges in the field.

Metapopulations are now quite popular and have been used in a variety of settings. A lot of work concerns investigation of properties of spatial models. For instance, with coauthors, I investigated the effect of lowering travel rates between locations [20] and of interconnection between a large urban centre and smaller satellite cities [22]. Other interesting issues studied include the effect of vaccination targeted at high risk areas [29], cooperation between governments on vaccination policy [123]. Geographically targeted vaccination has also been considered at smaller spatial scales; see, e.g., [26, 29, 101, 118, 121, 135]. Other spatial control issues have been considered in [41, 92, 95, 102, 122, 130, 136].

Papers addressing issues that are present also with COVID-19 have considered infection during transport [25], in particular in relation to entry screening [133] as well as exit and entry screening [132]. Exit and entry screening were also considered in [178]. Some work has also considered the effect of media-induced social distancing [87, 173].

Metapopulation models were used to consider specific diseases as well; the spread of SARS [162], age-structured contact patterns during the 2009 H1N1 pandemic [12], chikungunya [50], dengue [141], cholera [73] or malaria [14, 88, 89]; see also [160].

Standard metapopulation models are not well suited to consider the hyperlocal scale, because they assume homogeneity within the constituting units. In [12], an interesting approach is used that allows more heterogeneous contacts within patches. Similarly, in [148], the authors consider behaviour at the hyperlocal scale but still within a metapopulation model. In [7], the behaviour of an ABM is compared with that of a stochastic metapopulation model.

3 Chronology and Characteristics of COVID-19 Spread

To describe the spatio-temporal spread of COVID-19, I use the previously discussed framework of [17]. There is some discrepancy in reporting units, but to some extent, one can think along the lines of the ISO 3166 standard [114]. *Global* spread occurs between ISO 3166-1 codes (countries, dependent territories and special areas of geographical interest). *Local* spread within ISO 3166-1 codes occurs between ISO 3166-2 codes (provinces in Australia and Canada, départements in France, states in Brazil and the USA, etc.). Many countries also report at a finer geographical

scale, which I still call *local*: counties, regional health authorities or cities. Anything below the city level is *hyperlocal*; typically, this corresponds to commuting to work, school or shopping, but can even be mobility within a building. In keeping with my avowed preference for publicly available data, my description involves higher level jurisdictions rather than spread at the hyperlocal scale, which is typically associated to confidential data.

Throughout the description that follows, one should bear in mind that data used to describe the spread is very likely wrong in some instances, or rather, some jurisdictions might be reporting with a delay because of a lower capacity to detect cases. See for instance [100], in which two health security indices, the Global Health Security Index and the Joint External Evaluation, are used to assess the likelihood that countries detected COVID-19 early. (The work also shows that countries with higher values of these indices also saw reduced mortality from the disease to 1 July 2020, although this is likely not true anymore.)

3.1 Chronology and Characteristics of Global Spread

There is evidence that COVID-19 could have started its global spread in December 2019, with reports of a case in France [68] as well as suspicious cases [39] and positive wastewater samples [128] in Italy. However, these retrospective analyses have yet to be confirmed, so at the time of writing, the first ten locations to have confirmed importations are those listed in Table 1. The remainder of January saw cases being confirmed in several other countries. Of note is that China imposed a *cordon sanitaire* in Wuhan on 23 January 2020 and that the first successful importation (in the terminology of [17], i.e., a local transmission event) was reported by Vietnam on 24 January 2020 [182].

Table 1 First ten international locations having reported imported COVID-19. *Date* refers to the date the case was reported. All dates are in 2020. All cases in this table were imported from China, except for Vietnam, which concerned both an imported case and a local contact

Date	Location	Note	Source
13 Jan.	Thailand	Arrived 8 Jan.	[181]
16 Jan.	Japan	Arrived 6 Jan.	[168, 179]
20 Jan.	Republic of Korea	Airport detected on 19 Jan.	[180]
20 Jan.	USA	Arrived Jan. 15	[107]
23 Jan.	Nepal	Arrived 13 Jan.	[35]
23 Jan.	Singapore	Arrived 20 Jan.	[2]
24 Jan.	France	Arrived 22 Jan.	[1]
24 Jan.	Vietnam	Arrived 13 Jan.	[60, 152]
25 Jan.	Australia	Arrived 19 Jan.	[27]
25 Jan.	Malaysia	Arrived 24 Jan.	[74, 151]

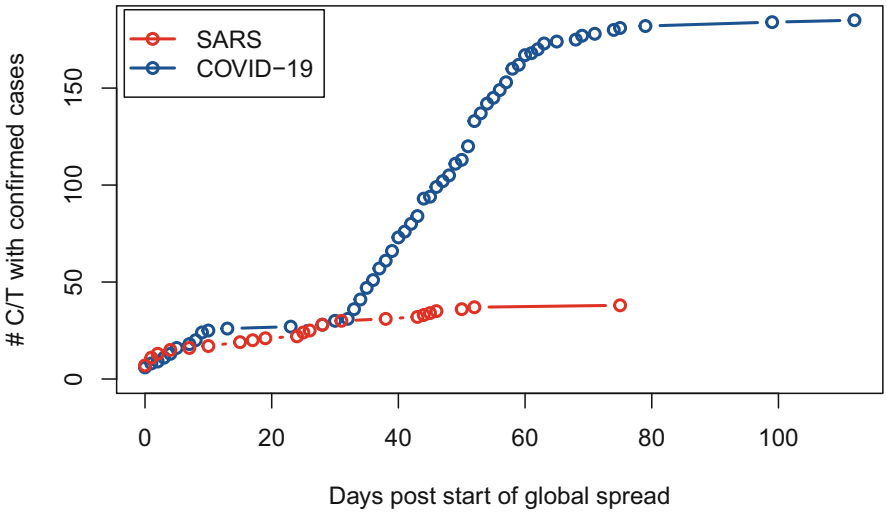


Fig. 1 Number of countries and territories (ISO-3166-1-alpha3 codes) having reported confirmed cases of SARS-CoV (red) and SARS-CoV-2 (blue) as a function of the number of days since importation in HKG (SARS-CoV) and importation in THL (SARS-CoV-2)

Starting in February 2020 and with more and more of the locations having reported importations earlier seeing local transmission chains, global spread accelerated. Figure 1 shows the number of ISO 3166-1-alpha3 codes (top level jurisdictions) reporting their first confirmed case as a function of the time since the first confirmed international exportation event.

While I do not detail them in the modelling section because of my focus there on models able to provide explanations of the phenomena, it is worth noting that interesting time series analyses were performed during the early stages of spread. For instance, [58] used ARIMA analysis of travel data together with disease propagation data to forecast future destinations. The authors find that uncertainty as to the percentage of asymptomatic cases makes previsions complicated; this conclusion is in line with personal work [23]. Likewise, [99, 105] considered spatial autoregressive models. Also, although not global, continental-level spread as documented for Africa in [91] is included here because the focus is not on the transition between the global level to the continental level but on spread within the continent. Finally, [138] use self-organising maps to look for similarities in epidemic curves to identify countries seeing propagation of the same type.

3.2 Attempts to Slow Down the Global Spread

Using the terminology of the conceptual model of spatialisation, when COVID-19 started its international spread, there were very few jurisdictions that were exporters

of COVID-19 and an immense majority of potential importers. Public health authorities in those jurisdictions that did not have cases at that point therefore took measures to try to stop or at least delay importations. To this end, they used three main types of measures: restriction or suspension of travel, entry screening and post-arrival self-isolation measures.

Starting early on and ongoing at the time of writing, various jurisdictions took measures to curtail or even interrupt travel. Passengers themselves also abstained from travelling. The result of this was a precipitous drop in travel volumes. The intensity of this effect can be seen in Fig. 2, which shows the daily number of passengers processed by the United States Transport Security Agency, i.e., the number of individuals undertaking a trip originating in the USA, in 2019 (red) and 2020 (blue). The data shown for 2019 is for a year earlier, but shifted so it corresponds to the same day in the week. At the lowest point, on Thursday 16 April 2020, TSA screened 3.63% of the number of travellers they had screened on Thursday 18 April 2019. The same trend can be observed for instance in tourism, with the United Nations World Tourism Organisation Tourism Dashboard (https://www.unwto.org/international-tourism-and-covid-19) reporting that the number of international tourists arrivals in April and May 2020 was 97% less than the same months in 2019.

In [6], an analysis of the global air transportation network is undertaken using the network distance defined in [44], attempting to tease out the effect of travel interruptions on the spread. While both of these studies provide very interesting insights into the issue, a precise quantification of the effect of such fundamental changes to travel is hard.

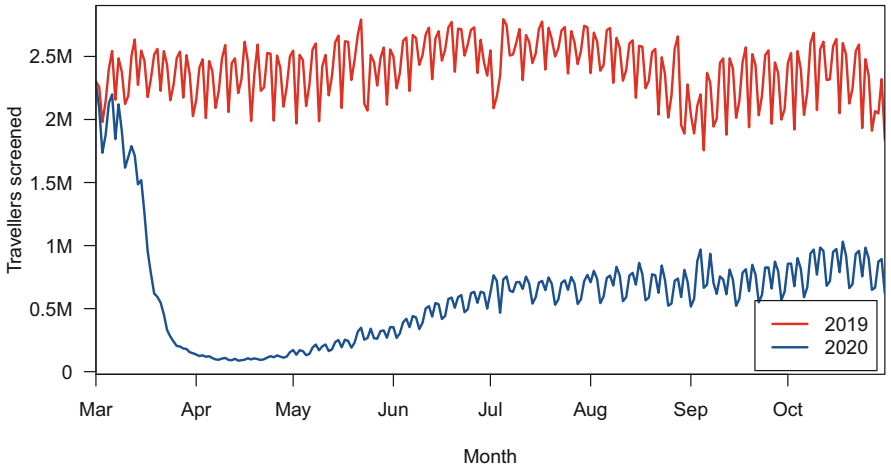


Fig. 2 Daily number of passengers processed by the United States Transport Security Agency (TSA) in 2019 and 2020; data from https://www.tsa.gov/coronavirus/passenger-throughput

Entry screening is typically implemented at ports of entry (ports, airports, border crossings) and seeks to identify individuals who are bearing the disease of concern, in order to isolate them and thereby avoid potential transmission of the disease in the local population. There is some debate about the usefulness of entry screening, especially during the early stages of a global spread event. See [149] for an extensive review. The sensitivity and specificity of the thermal detection equipment used is questionable [40]. In the case of COVID-19, it has also been argued that this low sensitivity would combine with the fact that fever detection would often fail because of the frequency of asymptomatic cases [45]. Entry screening at the beginning of a health crisis also means looking for a needle in a haystack, since the volume of incoming passengers from all locations vastly dominates the volume of passengers coming from the location of interest [120]. Since prevalence is low at the beginning of the event, this further compounds the lack of efficacy and results in poor characteristics for the method [67]. Also, screening protocols themselves vary widely from location to location [86], rendering a general evaluation of the value of a protocol difficult. Despite these reservations, in the case of COVID-19, some of the evidence of early international spread comes through entry screening, so there seems to have been some limited benefit to thermal imaging entry screening.

After the initial few days during which testing was thermal imaging-based, screening switched to using much more reliable PCR tests. This became possible because sequencing of the virus genome was performed remarkably quickly. As a consequence, currently, there are four main attitudes towards screening: no screening at all; "soft" screening, i.e., verbal or written questionnaires; testing on entry; testing prior to entry. Some countries use a combination of approaches, for instance requiring testing only for individuals arriving from regions considered particularly at risk.

A jurisdiction still has one option to combat the risk that successful importations take place: it can recommend or impose that individuals arriving from another jurisdiction spend some time in *quarantine*. Canada, for instance, has insisted on a 2-weeks quarantine period for all incoming travellers since the beginning of the crisis, with exemptions.

3.3 From Global to Local Spread

The first step in switching from a purely global vision of spread to a local one is to consider when COVID-19 could arrive in "one's backyard". In the early stages of the pandemic, before most top-level jurisdictions reported reported human-to-human transmission chains, it was of interest to those jurisdictions having no or few local cases to understand the risks that their connectedness to other jurisdictions carried. Because of evidence gathered during past pandemics and other notable public health events (see Sect. 2.2), this evaluation was mostly carried out by investigating a given jurisdiction's connection to the rest of the world by means of the air transportation network. This was the method used for instance in Mexico [64], India [97] or

Europe [157]. Much practical work on this aspect has come to rely on global airline transportation data such as that provided by the International Air Transport Association (IATA). However, it should be noted that this dataset quickly became unreliable because of the dramatic fall in travel volumes discussed in Sect. 3.2.

3.4 Chronology and Characteristics of Local Spread

Most of the very early work on local spread concerned China, since it was the first country to experience this. The same type of method was used in [78] as is detailed at the start of this section, but with past data on mobility during the Spring Festival of millions of migrant workers residing in Wuhan. The aim was to assess the risk to locations visited by the migrant workers. Since Chinese New Year was on 25 January 2020, while the cordon sanitaire was imposed in Wuhan on 23 January, a lot of individuals did make the trip. This allowed the authors to venture which places were probably under-reporting cases. See also [184], which uses GIS techniques to study the spread within China and the factors contributing to this spread. As do the authors of the previous paper, they find that connection to Wuhan, both in terms of population flow and economically, was the main driver of the initial spread.

Tracing transmission chains originating from importations allowed to better understand the consequences of importations. See, for instance, [36], which breaks down such a transmission chain that started on 27 January 2020 in Bavaria (Germany). In [46], the early spread in Brazil is documented, from importation from Europe (as evidenced by genome typing of the strains) to local spread within states, finally followed by exportation from urban centres. This is confirmed by [84], who consider spread among 604 cities in São Paulo State, Brazil. They show that in the heterogeneous setting they consider, there are two patterns of spread: one spatial, where the disease spreads to the nearest spatial component; the other hierarchical, where within one unit, spread starts with the top level urban centre then makes its way to smaller cities. In [80], the authors use genomic and transportation data to consider the spread within the USA and conclude that quite early on in the spread, importations into uninfected locations in the country were much more likely to originate elsewhere in the country than abroad. Propagation within the USA was also studied by [106], in which the occurrence of space-time clusters is studied. This interestingly shows that as the epidemic took hold, there occurred more and more smaller clusters, confirming in some sense the similar observations in Brazil. Another investigation of continental spread in the USA is carried out in [144] using a multilayer perceptron neural network. The authors use the Moran index computed on the incidence rates and a large number (57) of explanatory variables: socioeconomic, behavioural, environmental, topographic, demographic, age-adjusted mortality rates from several diseases, both infectious and chronic. They find that some of the most important factors predicting COVID-19 incidence rates are the age-adjusted mortality rates of ischemic heart disease, pancreatic cancer and leukemia, median household income and total precipitation.

In [3], the time evolution in several countries and the time evolution within France are investigated using time series methods incorporating spatial components. While mostly methodological, this provides interesting tools to consider the spatio-temporal evolution of the disease across multiple jurisdictions.

Other authors considered mechanisms for slowing down the spatial spread within a country. The authors of [142] advocate for a disconnection between locked-down urban centres and rural areas in India as a means to avoid complete country-wide lockdown. This position is justified; indeed, authors in [63], for instance, found strong correlation between population density and the spread of SARS-CoV-2 in China, so it could be that forbidding movement between locations at high risk (the cities) and those at lower risk is a valid approach. However, to the best of our knowledge, no nation implemented such a system; indeed, during the initial wave, most countries implemented country-level lockdowns that also relied on severely limiting or completely interrupting mobility within their territory. In [70], the authors consider the effect of containment measures on the spread of COVID-19 between provinces in Italy. In [124], the authors used human mobility data in China to consider the spread of COVID-19 within China, in particular in relation to the impact of control measures.

3.5 Chronology and Characteristics of Hyperlocal Spread

Hyperlocal spread was documented early on during the course of the pandemic because of cases that happened onboard cruise ships that were under quarantine. These events, while unfortunate for those involved, have provided a wealth of data. In particular, they were extremely helpful in finding out key epidemiologic parameters such as reproduction number [185], prevalence of asymptomatic infections [75], incidence [153], transmissibility of the disease [143] or case fatality ratio [163]. Because cruise ships have records of who was infected together with the room they were in, it should become possible to build a good understanding of spatial aspects, although to the best of our knowledge, this data has not yet been released.

Many countries faced and are facing outbreaks in long-term care facilities (LTC). There are a variety of reasons for this elevated risk; see, e.g., [172, 175]. This led to tremendous effort to control such outbreaks [175]. Movement within LTC can be documented (and modelled) accurately; see, e.g., [51, 150]. The health of residents is also monitored (usually) well. As a consequence, nosocomial COVID-19 outbreaks also provide valuable data at the hyperlocal level. See documented outbreaks in [113, 125, 137, 166, 167].

Note an interesting "twist" on hyperlocal spread: in [108], the authors conduct a wide-ranging analysis at the hyperlocal scale, in the sense that the consider the movements of individuals at the local scale but over the entire territory of the United States of America. This allows them to consider the effect of spatial heterogeneity of public health orders.

4 COVID-19-Specific Models

I replicate here the hierarchical spatial structure in Sect. 3 rather than the methodological one in Sect. 2.3. Indeed, while most work detailed here falls within one of the three classes of methods in Sect. 2.3, I also report on other methods that gave interesting results.

4.1 Models of Global Spread

To the best of my knowledge, most models for the global spread had in their objectives to study how to slow down the global spread of the infection or considered spread within specific countries or groups of countries; these are discussed in the relevant later sections.

In [21, 24], we set the $SL_1L_2I_1I_2A_1A_2R$ model of [23] in a metapopulation context and focused on the risk of importation in different countries. The model was run daily to provide the Public Health Agency of Canada with an assessment of the most likely countries to import the disease in the coming days. The model includes travel at different levels, which, as pointed out in Sect. 3.3, was a documented feature of spread.

In [170], an SIR-type metapopulation model in the GLEAM framework [30] is used that combines population densities, commute patterns and long-range travel. Used at the early stage of the spread, the authors find that it is likely that the value of the basic reproduction number $\mathcal{R}_0$ and the prevalence are badly estimated in some locations, with estimates in the literature at the time driven by locations with a large population. They conclude that the number of cases was probably underestimated.

4.2 Modelling the Slowing Down of Global Spread

In [55], a metapopulation model for the global spread of COVID-19 is used to consider in particular the role of international travel bans. The authors show that while the cordon sanitaire in Wuhan did little to slow spread within China, its impact internationally was more pronounced. The combined effect of travel restrictions and community effort is also studied, with the interesting finding that travel restrictions alone do not suffice to have an effect on propagation. In [4], a stochastic SEIR metapopulation model is used, together with Official Airlines Guide (OAG) data, to consider the role of travel restrictions taking place after 24 January 2020. The authors found good adequation with the number of imported cases in several countries as of the end of January. They focused in particular on Australia and establish that the travel ban there might have delayed the onset of widespread propagation by 4 weeks.

A stochastic simulation model is used in [69] to consider different scenarios regarding testing (rather than screening) of incoming individuals and the duration of quarantine periods. Similarly, [59] use a stochastic model to quantify the effectiveness of screening and so-called *sensitisation* of travellers, i.e., the provision of health information in an effort to trigger compliance with self-isolation recommendations. In [159], the example of air transportation in Brazil is considered using an SIR-type metapopulation. The speed of spread in relation to network measures such as centrality was explored, with closeness centrality shown to be a good predictor of the vulnerability of a city.

In [17], we considered the risk of disease importation in a location that is seeing little to no local transmission chains. As with most of our work on the subject, we used a modified version of the model in [23]. In this case, we used a stochastic version, which we subjected to stimulations to represent the inflow of infected individuals into a location. The model also allowed us to quantify precisely the effect of quarantine in terms of its effect on the inflow rate.

4.3 Modelling the Transition from Global to Local Spread

In [94], the risk of importation of COVID-19 in African countries was considered using air travel data as well as data from the Monitoring Evaluation Framework (MEF) of the WHO International Health Regulations. The model is quite simple and comprises no dynamic components, meaning that it provides a snapshot evaluation of the risk of importation. As it was formulated at the beginning of the spread event, when most of the exportation was assumed to come from China, it does nonetheless provide meaningful results.

In the already cited [17], we focused on the risk of importation of COVID-19 in locations that are seeing little to no local transmission, thereby considering the interface between the rest of the world and such locations. We showed that the probability of importation was most dependent on the rate at which cases are imported in the locations, but that the outcome of a successful importation was then determined to a large extent by the intensity of public health measures in the locations.

4.4 Models of Local Spread

The location for which data became readily available the soonest was China. Since China is also a very large country, some very interesting work was carried out in the context of spread within that country. The authors of [183] considered the spread of COVID-19 within China using an interesting idea: they estimated the size of the outbreak in Wuhan from known international exportations, then used a metapopulation model with Wuhan as the source of infection to estimate spread within China.

In [131], the role of undocumented infections is investigated in relation with the spread of COVID-19 between 372 Chinese cities using a metapopulation SEIR model incorporating documented and undocumented infections. In [112], spread within Hubei Province and in the rest of China is investigated using statistical tools (the Moran index and a logistic model). Then an ODE SEIR model is used to compute $\mathcal{R}_0$ in the different locations. In [174], an SEIR-type model with additional compartments for diagnosed and confirmed, suspected and infected as well as suspected but uninfected individuals is set in a metapopulation framework with two patches: Hubei Province and the rest of China. The model is used to consider the effect of lifting lockdown measures.

In [10], a simulation platform is used to consider the spread in France. The model operates at the level of subregions (*départements*) and supposes that individuals can be susceptible, asymptomatic, symptomatic, recovered, hospitalised and diseased. An interesting feature of the paper is a comparison between the results of continuous time deterministic and discrete stochastic methods, with the latter showing better adequation with observed data.

The authors of [46] considered spread of SARS-CoV-2 within Brazil. This colossal endeavour considers actual genotyping of the virus and prior to modelling work proper, details importations and the spatio-temporal spread of various genomes of the virus. Spatio-temporal modelling then uses a continuous phylogeographic model. The model is not predictive but sheds light on the spread process: they find that spread was mostly local, i.e., within state borders. Both within-state and between-state spread was also found to have decreased after the implementation of NPI.

In [9], the effect of heterogeneity of policies in the USA is investigated. A model is formulated that is a metapopulation in essence; based on data on people movement to places of gathering such as churches, the model allows the redistribution of individuals between locations following different types of policies. They observe that spatial heterogeneity in measures tends to increase the likelihood of subsequent infection waves. Spatial heterogeneity is also investigated in [65], which uses a metapopulation model to probe the impact of disparity of healthcare capacity in Ohio. In [53], the effect of changing travel rates within and between locations is investigated, with data for Taiwan.

Finally, note that because COVID-19 is spreading globally and that national level jurisdictions (and sometimes even lower level ones) implemented a variety of responses, it is useful to compare the situation in different jurisdictions. Even though this is not spatial modelling *stricto sensu*, such works are worth mentioning here as they provide the underpinning to spatial models. In [104], the authors use an SEIR model to compare transmission patterns in China, South Korea, Italy and Iran. In [103], an age-structured SEIR model is used to compare the dynamics of disease spread in Hubei Province and six European regions. The focus is on the estimation of the case-fatality (CFR), symptomatic case-fatality (sCFR) and infection-fatality (IFR) ratios. The authors find that the latter two indicators are better suited to describe the potential impact of the pandemic and note that they find geographic heterogeneity of the estimated values. This heterogeneity is not only

between Hubei Province and the European locations under consideration, but also between the European locations themselves. With collaborators, we used the model in [23] to provide a daily forecast of spread in several Canadian provinces [21] and found that estimates for some parameters were consistent across provinces while estimates for others varied widely, in particular, the proportion of asymptomatic cases.

4.5 Models for Spread at the Hyperlocal Level

A lot of work during this pandemic has focused implicitly on the hyperlocal level, but recall that here the object is models in which I found an explicit reference to spatial aspects.

In [115], a network model is used to model the spread of SARS-CoV-2 onboard the Diamond Princess cruise ship, with nodes representing individual passengers and crew members. Age-structure was used as well. The model was calibrated to known transmission data and the effect of control measures was then considered. See also the already cited [75, 143, 153, 185] for more modelling work related to spread aboard the Diamond Princess.

The authors of [43] used an SLIAR agent-based model to consider the effect of social distancing, viral shedding and what they call the social distance threshold. They find that the three lead to threshold behaviour ("phase transitions") that have different effects on the course of the epidemic.

In [81], ABM are used to consider in particular the effect of testing policies. Agents are distributed on a map depending on the population density in the areas under consideration. They are also assigned movement patterns that can cover the whole map, a medium range or a small one. Some interesting observations are that when tests have low reliability or that the ability to trace contact is low, a large fraction of the testing capacity remains unused despite an increasing incidence. They also find that mixed testing policies are useful to contain spread.

5 New Variants

SARS-CoV-2 is an RNA virus and as such is subject to high mutation rates leading potentially to variants [72, 117]. Thus the emergence of new variants was expected from the onset of the crisis. At this point, there are several major variants to the original variant that have been detected. This number can be expected to rise: detection of most variants requires genome sequencing, which is performed at different rates in different countries [85], meaning that capacity to detect variants varies greatly globally. Of particular interest at the time of writing is B.1.1.7, which was first detected in the United Kingdom in early December 2020 but is presumed to have been spreading since as early as September 2020. This variant

is particularly concerning as it appears to be more transmissible than the original variant. It seems that this variant should not, however, be detrimental to the ongoing vaccination efforts [62]. Many countries took preventive measures in order to delay the arrival of the variant, essentially forbidding all travel originating from the United Kingdom, but given that circulation probably started several months before these measures, their efficacy is debatable. For instance, [71] estimates that, of 19 countries evaluated, 16 had at least a 50% chance of having already imported the variant by 7 December 2020. The novel variants led some countries to consider exit control measures; some European Union (EU) countries (Belgium and France, for instance) decided late January 2021 or early February 2021 to forbid both entry from and exit to non-EU countries for non-essential travel. In the context of pandemic H1N1 influenza, exit screening was shown to have the potential to be more an efficacious control measure than entry screening [120]. It is therefore interesting to see this type of control finally being applied, although the intent is not the one we were advocating in [120].

Modelling the spatio-temporal spread of these novel variants can be conducted in very much the same manner as was done for the original variant. For instance, metapopulation models for multiple species such as those considered in [19, 20] can be readily adapted to a multiple variant situation. However, it is important to bear in mind that because of the detection issues mentioned earlier, these models are hard to parametrise when considering the initial spread of the variants.

6 Discussion

This is but a brief and very incomplete snapshot of the state of knowledge about the spread of COVID-19 at the time of writing in December 2020, with a few additional details about the new variants added in January 2021. As indicated, it is likely that I omitted a lot of publications on the subject, given the immense amount of literature COVID-19 has generated.

From the perspective of the spatio-temporal spread of the disease, although there is still much to learn, I think we also now have the *luxury of hindsight*: many groups, mine included, have produced a variety of models in the first few months of the crisis, which can and should now be confronted to the reality of the outbreak. Because COVID-19 is so widespread, there is less urgency to consider its spatial spread in the perspective of emergency response and the focus could now evolve, at least in part, to the evaluation of the models we produced. The problem of re-importation of the disease in locations having managed to drive it away remains an important one, so I am not advocating to stop all work regarding spatial spread; I am only pointing out that understanding what worked and what did not during the initial spread would actually help for these subsequent importation events.

Going forward, though, I believe that there is still a lot to be done on one key aspect of spatio-temporal models: most of the work carried out by those of us working in this area has come to rely on one particular dataset, the so-called

IATA air transport data. Figure 2 shows that in the particular instance of COVID-19, the quality of this data leaves a lot to be desired. When the data for 2020 becomes available in 2021, it will be extremely important to scrutinise it in order to understand what type of changes took place. Another important point to ponder will be the use of other data sources to compensate for this loss of relevance of IATA-type data. Cell phone location data is showing promise, but it suffers from several limitations, the most important of which being that it is most useful and detailed at the country level and, more importantly, that it is either proprietary or extremely expensive to acquire.

Acknowledgments I am supported in part by NSERC and by CIHR through the Canadian COVID-19 Mathematical Modelling Task Force. I acknowledge support both financial and logistical from the Public Health Agency of Canada.

References

1. 20 minutes avec AFP. Coronavirus: Trois cas confirmés en France, 15 nouveaux décès en Chine. 20 minutes, January 2020. https://www.20minutes.fr/sante/2703111-20200125-coronavirus-trois-cas-confirmes-france-15-nouveaux-deces-chine.
2. Z. Abdullah and H. Salamat. https://www.channelnewsasia.com/news/singapore/wuhan-virus-pneumonia-singapore-confirms-first-case-12312860. CNA Asia, January 2020. https://www.channelnewsasia.com/news/singapore/wuhan-virus-pneumonia-singapore-confirms-first-case-12312860.
3. P. Abry, N. Pustelnik, S. Roux, et al. Spatial and temporal regularization to estimate COVID-19 reproduction number R(t): Promoting piecewise smoothness via convex optimization. *PloS One*, 15:e0237901, 2020.
4. A. Adekunle, M. Meehan, D. Rojas-Alvarez, et al. Delaying the COVID-19 epidemic in Australia: evaluating the effectiveness of international travel bans. *Australian and New Zealand Journal of Public Health*, July 2020.
5. A. Adiga, S. Chu, S. Eubank, et al. Disparities in spread and control of influenza in slums of Delhi: findings from an agent-based modelling study. *BMJ Open*, 8:e017353, January 2018.
6. A. Adiga, S. Venkatramanan, J. Schlitt, et al. Evaluating the impact of international airline suspensions on the early global spread of COVID-19. *medRxiv*, February 2020.
7. M. Ajelli, B. Gonçalves, D. Balcan, et al. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10:190, June 2010.
8. M. E. Alexander and R. Kobes. Effects of vaccination and population structure on influenza epidemic spread in the presence of two circulating strains. *BMC Public Health*, 11 Suppl 1:S8, February 2011.
9. B. M. Althouse, B. Wallace, B. Case, et al. The unintended consequences of inconsistent pandemic control policies. *medRxiv*, August 2020.
10. P. Amar. Pandæsim: an epidemic spreading stochastic simulator. *Biology*, 9, September 2020.
11. J. J. Angulo, C. K. Takiguti, C. A. Pederneiras, et al. Identification of pattern and process in the spread of a contagious disease. *Social Science & Medicine*, 13D:183–189, November 1979.
12. A. Apolloni, C. Poletto, and V. Colizza. Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC Infectious Diseases*, 13:176, April 2013.
13. J. Arino. Mathematical epidemiology in a data-rich world. *Infectious Disease Modelling*, 5:161–188, 2020.

14. J. Arino, A. Ducrot, and P. Zongo. A metapopulation model for malaria with transmission-blocking partial immunity in hosts. *Journal of Mathematical Biology*, 64(3):423–448, 2012.
15. J. Arino and P. van den Driessche. A multi-city epidemic model. *Mathematical Population Studies*, 10(3):175–193, 2003.
16. J. Arino. Spatio-temporal spread of infectious pathogens of humans. *Infectious Disease Modelling*, 2:218–228, May 2017.
17. J. Arino, N. Bajeux, S. Portet, and J. Watmough. Quarantine and the risk of COVID-19 importation. *Epidemiology and Infection*, 148, E298, 2020. https://doi.org/10.1017/S0950268820002988
18. J. Arino, F. Brauer, P. van den Driessche, et al. Simple models for containment of a pandemic. *Journal of the Royal Society, Interface*, 3:453–457, June 2006.
19. J. Arino, J. R. Davis, D. Hartley, et al. A multi-species epidemic model with spatial dynamics. *Mathematical Medicine and Biology*, 22:129–142, June 2005.
20. J. Arino, R. Jordan, and P. van den Driessche. Quarantine in a multi-species epidemic model with spatial dynamics. *Mathematical Biosciences*, 206:46–60, March 2007.
21. J. Arino, S. Portet, N. Bajeux, and A. Ciupeanu. Investigation of global and local COVID-19 importation risks. Technical report, Public Health Agency of Canada, 2020.
22. J. Arino and S. Portet. Epidemiological implications of mobility between a large urban centre and smaller satellite cities. *Journal of Mathematical Biology*, 71(5):1243–1265, 2015.
23. J. Arino and S. Portet. A simple model for COVID-19. *Infectious Disease Modelling*, 5:309–315, 2020.
24. J. Arino, S. Portet, and E. Rees. A method for assessing the most likely next destinations of COVID-19. In preparation.
25. J. Arino, C. Sun, and W. Yang. Revisiting a two-patch SIS model with infection during transport. *Mathematical Medicine and Biology*, 33(1):29–55, 2016.
26. E. Asano, L. J. Gross, S. Lenhart, and L. A. Real. Optimal control of vaccine distribution in a rabies metapopulation model. *Mathematical Biosciences and Engineering*, 5:219–238, April 2008.
27. Australian Government Department of Health. First confirmed case of novel coronavirus in Australia. Media release, January 2020. https://www.health.gov.au/ministers/the-hon-greg-hunt-mp/media/first-confirmed-case-of-novel-coronavirus-in-australia.
28. E. I. Azhar, S. Lanini, G. Ippolito, and A. Zumla. The Middle East Respiratory Syndrome coronavirus – a continuing risk to global health security. *Advances in Experimental Medicine and Biology*, 972:49–60, 2017.
29. A. S. Azman and J. Lessler. Reactive vaccination in the presence of disease hotspots. *Proceedings. Biological Sciences*, 282:20141341, January 2015.
30. D. Balcan, B. Gonçalves, H. Hu, et al. Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *Journal of Computational Science*, 1(3):132–145, 2010.
31. F. Ball, T. Britton, T. House, et al. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, March 2015.
32. F. Ball and T. House. Heterogeneous network epidemics: real-time growth, variance and extinction of infection. *Journal of Mathematical Biology*, 75:577–619, September 2017.
33. D. H. Barmak, C. O. Dorso, M. Otero, and H. G. Solari. Dengue epidemics and human mobility. *Physical Review E*, 84:011901, July 2011.
34. R. C. Barnard, L. Berthouze, P. L. Simon, and I. Z. Kiss. Epidemic threshold in pairwise models for clustered networks: closures and fast correlations. *Journal of Mathematical Biology*, 79:823–860, August 2019.
35. A. Bastola, R. Sah, A. J. Rodriguez-Morales, et al. The first 2019 novel coronavirus case in Nepal. *The Lancet Infectious diseases*, 20:279–280, March 2020.
36. M. M. Böhmer, U. Buchholz, V. M. Corman, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases*, 20:920–928, August 2020.

37. D. Bichara and A. Iggidr. Multi-patch and multi-group epidemic models: a new framework. *Journal of Mathematical Biology*, 77:107–134, July 2018.
38. N. Bifolchi, R. Deardon, and Z. Feng. Spatial approximations of network-based individual level infectious disease models. *Spatial and Spatio-Temporal Epidemiology*, 6:59–70, September 2013.
39. L. I. Birtolo, V. Maestrini, P. Severino, et al. Coronavirus disease 2019 in Rome: was it circulating before December? *Journal of Cardiovascular Medicine*, 21:835–836, October 2020.
40. D. Bitar, A. Goubar, and J. C. Desenclos. International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers. *Euro Surveillance*, 14(2):1–5, February 2009.
41. W. Bock and Y. Jayathunga. Optimal control of a multi-patch Dengue model under the influence of Wolbachia bacterium. *Mathematical Biosciences*, 315:108219, September 2019.
42. J. T. Bowen and C. Laroe. Airline networks and the international diffusion of severe acute respiratory syndrome (SARS). *The Geographical Journal*, 172:130–144, June 2006.
43. B. Braun, B. Taraktas, B. Beckage, and J. Molofsky. Simulating phase transitions and control measures for network epidemics caused by infections with presymptomatic, asymptomatic, and symptomatic stages. *PLoS One*, 15:e0238412, 2020.
44. D. Brockmann and D. Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342:1337–1342, December 2013.
45. G. M. Bwire and L. S. Paulo. Coronavirus disease-2019: is fever an adequate screening for the returning travelers? *Tropical Medicine and Health*, 48:14, 2020.
46. D. S. Candido, I. M. Claro, J. G. de Jesus, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*, 369:1255–1260, September 2020.
47. Z. Cao, D. Zeng, X. Zheng, et al. Spatio-temporal evolution of Beijing 2003 SARS epidemic. *Science China. Earth Sciences*, 53:1017–1028, 2010.
48. N. B. Carnegie. Effects of contact network structure on epidemic transmission trees: implications for data required to estimate network structure. *Statistics in Medicine*, 37:236–248, January 2018.
49. C. Carpenter and L. Sattenspiel. The design and use of an agent-based model to simulate the 1918 influenza epidemic at Norway House, Manitoba. *American Journal of Human Biology*, 21:290–300, 2009.
50. S. Chadsuthi, B. M. Althouse, S. Iamsirithaworn, et al. Travel distance and human movement predict paths of emergence and spatial spread of chikungunya in Thailand. *Epidemiology & Infection*, 146:1654–1662, October 2018.
51. D. Champredon, M. Najafi, M. Laskowski, et al. Individual movements and contact patterns in a Canadian long-term care facility. *AIMS Public Health*, 5:111–121, 2018.
52. M. Chan-Yeung and R.-H. Xu. SARS: epidemiology. *Respirology*, 8 Suppl:S9–14, November 2003.
53. M. C. Chang, R. Kahn, Y. A. Li, et al. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. *BMC Public Health*, 21:226, 2021. https://doi.org/10.1186/s12889-021-10260-7
54. J. D. Cherry. The chronology of the 2002-2003 SARS mini pandemic. *Paediatric Respiratory Reviews*, 5:262–269, December 2004.
55. M. Chinazzi, J. T. Davis, M. Ajelli, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368:395–400, April 2020.
56. G. Chowell, J. M. Hyman, S. Eubank, and C. Castillo-Chavez. Scaling laws for the movement of people between locations in a large city. *Physical Review E*, 68:066102, December 2003.
57. G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud. Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18:66–97, September 2016.
58. P. Christidis and A. Christodoulou. The predictive capacity of air travel patterns during the global spread of the COVID-19 pandemic: risk, uncertainty and randomness. *International Journal of Environmental Research and Public Health*, 17, May 2020.

59. S. Clifford, C. A. B. Pearson, P. Klepac, et al. Effectiveness of interventions targeting air travellers for delaying local outbreaks of SARS-CoV-2. *Journal of Travel Medicine*, 27, August 2020.
60. J. Coleman. Vietnam reports first coronavirus cases. The Hill, January 2020.
61. V. Colizza, M. Barthélemy, A. Barrat, and A. Vespignani. Epidemic modeling in complex realities. *Comptes rendus Biologies*, 330:364–374, April 2007.
62. P. Conti, A. Caraffa, C. E. Gallenga, et al. The British variant of the new coronavirus-19 (Sars-Cov-2) should not create a vaccine problem. *Journal of Biological Regulators and Homeostatic Agents*, 35, December 2020.
63. S. Copiello and C. Grillenzoni. The spread of 2019-nCoV in China was primarily driven by population density. Comment on "Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China" by Zhu et al. *The Science of the Total Environment*, 744:141028, November 2020.
64. G. Cruz-Pacheco, J. F. Bustamante-Castañeda, J. G. Caputo, et al. Dispersion of a new coronavirus SARS-CoV-2 by airlines in 2020: temporal estimates of the outbreak in Mexico. *Revista de Investigacion Clinica*, 72:138–143, 2020.
65. D. F. Cuadros, Y. Xiao, Z. Mukandavire, et al. Spatiotemporal transmission dynamics of the covid-19 pandemic and its impact on critical healthcare capacity. *Health & Place*, 64:102404, July 2020.
66. O. O. Dalgiç, O. Y. Özaltin, W. A. Ciccotelli, and F. S. Erenay. Deriving effective vaccine allocation strategies for pandemic influenza: Comparison of an agent-based simulation and a compartmental model. *PLoS One*, 12:e0172261, 2017.
67. M. Dell'Omodarme and M. C. Prati. The probability of failing in detecting an infectious disease at entry points into a country. *Statistics in Medicine*, 24:2669–2679, September 2005.
68. A. Deslandes, V. Berti, Y. Tandjaoui-Lambotte, et al. SARS-CoV-2 was already spreading in France in late December 2019. *International Journal of Antimicrobial Agents*, 55:106006, June 2020.
69. B. L. Dickens, J. R. Koo, J. T. Lim, H. Sun, H. E. Clapham, A. Wilder-Smith, and A. R. Cook. Strategies at points of entry to reduce importation risk of COVID-19 cases and reopen travel. *Journal of Travel Medicine*, 27(8):taaa141, December 2020. https://doi.org/10.1093/jtm/taaa141
70. M. M. Dickson, G. Espa, D. Giuliani, et al. Assessing the effect of containment measures on the spatio-temporal dynamic of COVID-19 in Italy. *Nonlinear Dynamics*, pages 1–14, August 2020.
71. Z. Du, L. Wang, B. Yang, S. T. Ali, T. K. Tsang, S. Shan, P. Wu, E. Lau, B. J. Cowling, and L. A. Meyers. Risk for international importations of variant SARS-CoV-2 originating in the United Kingdom. *Emerging Infectious Diseases*, 27(5):1527–1529, 2021. https://doi.org/10.3201/eid2705.210050
72. E. A. Duarte, I. S. Novella, S. C. Weaver, et al. RNA virus quasispecies: significance for viral disease and epidemiology. *Infectious Agents and Disease*, 3:201–214, August 1994.
73. M. C. Eisenberg, Z. Shuai, J. H. Tien, and P. van den Driessche. A cholera model in a patchy environment with water and human movement. *Mathematical Biosciences*, 246:105–112, November 2013.
74. A. Elengoe. COVID-19 outbreak in Malaysia. *Osong Public Health and Research Perspectives*, 11:93–100, June 2020.
75. J. C. Emery, T. W. Russell, Y. Liu, et al. The contribution of asymptomatic SARS-CoV-2 infections to transmission on the Diamond Princess cruise ship. *eLife*, 9, August 2020.
76. J. M. Epstein. *Generative Social Science: Studies in Agent-Based Computational Modeling*, volume 13. Princeton University Press, 2006.
77. S. Eubank. Network based models of infectious disease spread. *Japanese Journal of Infectious Diseases*, 58:S9–13, December 2005.
78. C. Fan, L. Liu, W. Guo, et al. Prediction of epidemic spread of the 2019 novel coronavirus driven by Spring Festival transportation in China: a population-based study. *International Journal of Environmental Research and Public Health*, 17, March 2020.

79. H. Fang, J. Chen, and J. Hu. Modelling the SARS epidemic by a lattice-based Monte-Carlo simulation. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2005:7470–7473, 2005.
80. J. R. Fauver, M. E. Petrone, E. B. Hodcroft, et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell*, 181:990–996.e5, May 2020.
81. V. G. Fiore, N. DeFelice, B. S. Glicksberg, O. Perl, A. Shuster, K. Kulkarni, M. O'Brien, M. A. Pisauro, D. Chung, and X. Gu. Containment of COVID-19: Simulating the impact of different policies and testing capacities for contact tracing, testing, and isolation. *PLoS One*, 16(3):e0247614, 2021. https://doi.org/10.1371/journal.pone.0247614
82. S. M. Firestone, M. P. Ward, R. M. Christley, and N. K. Dhand. The importance of location in contact networks: Describing early epidemic spread using spatial social network analysis. *Preventive Veterinary Medicine*, 102:185–195, December 2011.
83. A. Flahault, S. Letrait, P. Blin, et al. Modelling the 1985 influenza epidemic in France. *Statistics in Medicine*, 7:1147–1155, November 1988.
84. C. M. C. B. Fortaleza, R. B. Guimarães, G. B. de Almeida, et al. Taking the inner route: spatial and demographic factors affecting vulnerability to COVID-19 among 604 cities from inner São Paulo State, Brazil. *Epidemiology & Infection*, 148:e118, June 2020.
85. Y. Furuse. Genomic sequencing effort for SARS-CoV-2 by country during the pandemic. *International Journal of Infectious Diseases*, 103:305–307, 2021.
86. W. Gaber, U. Goetsch, R. Diel, et al. Screening for infectious diseases at international airports: the Frankfurt model. *Aviation, Space, and Environmental Medicine*, 80:595–600, July 2009.
87. D. Gao and S. Ruan. An SIS patch model with variable transmission coefficients. *Mathematical Biosciences*, 232:110–115, August 2011.
88. D. Gao and S. Ruan. A multi-patch malaria model with logistic growth populations. *SIAM Journal on Applied Mathematics*, 72:819–841, January 2012.
89. D. Gao, P. van den Driessche, and C. Cosner. Habitat fragmentation promotes malaria persistence. *Journal of Mathematical Biology*, 79:2255–2280, December 2019.
90. L. M. Gardner, A. A. Chughtai, and C. R. MacIntyre. Risk of global spread of Middle East respiratory syndrome coronavirus (MERS-CoV) via the air transport network. *Journal of Travel Medicine*, 23, June 2016.
91. E. Gayawan, O. O. Awe, B. M. Oseni, et al. The spatio-temporal epidemic dynamics of COVID-19 outbreak in Africa. *Epidemiology & Infection*, 148:e212, September 2020.
92. K. Gaythorpe and B. Adams. Disease and disaster: Optimal deployment of epidemic control facilities in a spatially heterogeneous population with changing behaviour. *Journal of Theoretical Biology*, 397:169–178, May 2016.
93. J. L. Geoghegan, A. F. Saavedra, S. Duchêne, et al. Continental synchronicity of human influenza virus epidemics despite climatic variation. *PLoS Pathogens*, 14:e1006780, January 2018.
94. M. Gilbert, G. Pullano, F. Pinotti, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet*, 395:871–877, March 2020.
95. K. Glass and B. Barnes. Eliminating infectious diseases of livestock: a metapopulation model of infection control. *Theoretical Population Biology*, 85:63–72, May 2013.
96. Q. Griette, G. Raoul, and S. Gandon. Virulence evolution at the front line of spreading epidemics. *Evolution*, 69:2810–2819, November 2015.
97. S. S. Gunthe and S. S. Patra. Impact of international travel dynamics on domestic spread of 2019-nCoV in India: origin-based risk assessment in importation of infected travelers. *Globalization and Health*, 16:45, May 2020.
98. D. Guo, K. C. Li, T. R. Peters, et al. Multi-scale modeling for the transmission of influenza and the evaluation of interventions toward it. *Scientific Reports*, 5:8980, March 2015.
99. C. M. Hafner. The spread of the Covid-19 pandemic in time and space. *International Journal of Environmental Research and Public Health*, 17, May 2020.

100. N. Haider, A. Yavlinsky, Y.-M. Chang, et al. The Global Health Security Index and Joint External Evaluation score for health preparedness are not correlated with countries' COVID-19 detection response time and mortality outcome. *Epidemiology & Infection*, 148:e210, September 2020.
101. I. M. Hall, J. R. Egan, I. Barrass, et al. Comparison of smallpox outbreak control strategies using a spatial metapopulation model. *Epidemiology & Infection*, 135:1133–1144, October 2007.
102. P. Harvim, H. Zhang, P. Georgescu, and L. Zhang. Transmission dynamics and control mechanisms of vector-borne diseases with active and passive movements between urban and satellite cities. *Bulletin of Mathematical Biology*, 81:4518–4563, November 2019.
103. A. Hauser, M. J. Counotte, C. C. Margossian, et al. Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe. *PLoS Medicine*, 17:e1003189, July 2020.
104. J. He, G. Chen, Y. Jiang, et al. Comparative infection modeling and control of COVID-19 transmission patterns in China, South Korea, Italy and Iran. *The Science of the Total Environment*, 747:141447, August 2020.
105. A. Hernandez-Matamoros, H. Fujita, T. Hayashi, and H. Perez-Meana. Forecasting of COVID-19 per regions using ARIMA models and polynomial functions. *Applied Soft Computing*, 96:106610, November 2020.
106. A. Hohl, E. M. Delmelle, M. R. Desjardins, and Y. Lan. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-Temporal Epidemiology*, 34:100354, August 2020.
107. M. L. Holshue, C. DeBolt, S. Lindquist, et al. First case of 2019 novel coronavirus in the United States. *The New England Journal of Medicine*, 382:929–936, March 2020.
108. D. Holtz, M. Zhao, S. G. Benzell, et al. Interdependence and the cost of uncoordinated responses to covid-19. *Proceedings of the National Academy of Sciences of the United States of America*, 117:19837–19843, August 2020.
109. T. Hornbeck, D. Naylor, A. M. Segre, et al. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *The Journal of Infectious Diseases*, 206:1549–1557, November 2012.
110. B. Hu, J. Gong, J. Sun, and J. Zhou. Exploring the epidemic transmission network of SARS in-out flow in mainland China. *Chinese Science Bulletin*, 58:1818–1831, 2013.
111. B. Hu, J. Gong, J. Zhou, et al. Spatial-temporal characteristics of epidemic spread in-out flow-using SARS epidemic in Beijing as a case study. *Science China. Earth Sciences*, 56:1380–1397, 2013.
112. R. Huang, M. Liu, and Y. Ding. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *Journal of Infection in Developing Countries*, 14:246–253, March 2020.
113. O. Iritani, T. Okuno, D. Hama, et al. Clusters of COVID-19 in long-term care hospitals and facilities in Japan from 16 January to 9 May 2020. *Geriatrics & Gerontology International*, 20:715–719, July 2020.
114. ISO. ISO 3166 code lists, 2013.
115. S. M. Jenness, K. S. Willebrand, A. A. Malik, B. A. Lopman, and S. B. Omer. Dynamic network strategies for SARS-CoV-2 control on a cruise ship. *medRxiv* [Preprint], October 2020. http://dx.doi.org/10.1101/2020.08.26.20182766
116. E. Karimi, K. Schmitt, and A. Akgunduz. Effect of individual protective behaviors on influenza transmission: an agent-based model. *Health Care Management Science*, 18:318–333, September 2015.
117. T. F. Kautz and N. L. Forrester. RNA virus fidelity mutants: a useful tool for evolutionary biology or a complex challenge? *Viruses*, 10, November 2018.

118. M. R. Kelly, J. H. Tien, M. C. Eisenberg, and S. Lenhart. The impact of spatial arrangements on epidemic disease dynamics and intervention strategies. *Journal of Biological Dynamics*, 10:222–249, 2016.
119. K. Khan, J. Arino, W. Hu, et al. Spread of a novel influenza A (H1N1) virus via global airline transportation. *The New England Journal of Medicine*, 361:212–214, July 2009.
120. K. Khan, R. Eckhardt, J. S. Brownstein, et al. Entry and exit screening of airline travellers during the A(H1N1) 2009 pandemic: a retrospective evaluation. *Bulletin of the World Health Organization*, 91:368–376, May 2013.
121. A. Khatua, T. K. Kar, S. K. Nandi, et al. Impact of human mobility on the transmission dynamics of infectious diseases. *Energy, Ecology & Environment*, pages 1–18, May 2020.
122. J. E. Kim, H. Lee, C. H. Lee, and S. Lee. Assessment of optimal strategies in a two-patch dengue transmission model with seasonality. *PloS One*, 12:e0173673, 2017.
123. P. Klepac, I. Megiddo, B. T. Grenfell, and R. Laxminarayan. Self-enforcing regional vaccination agreements. *Journal of the Royal Society, Interface*, 13:20150907, January 2016.
124. M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 368:493–497, May 2020.
125. C. Kuhn and A. Rose. Tall Pines Healthcare COVID-19 outbreak experience in rural Waldo County, Maine, April 2020. *The Journal of Ambulatory Care Management*, 43:294–300, 2020.
126. S. Kumar, J. J. Grefenstette, D. Galloway, et al. Policies to reduce influenza in the workplace: impact assessments using an agent-based model. *American Journal of Public Health*, 103:1406–1411, August 2013.
127. S. Kumar, K. Piper, D. D. Galloway, et al. Is population structure sufficient to generate area-level inequalities in influenza rates? An examination using agent-based models. *BMC Public Health*, 15:947, September 2015.
128. G. La Rosa, P. Mancini, G. Bonanno Ferraro, et al. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. *The Science of the Total Environment*, 750:141711, August 2020.
129. M. Laskowski, V. R. Duvvuri, D. L. Buckeridge, et al. Influenza H3N2 variant viruses with pandemic potential: preventing catastrophe in remote and isolated Canadian communities. *Preventive Medicine*, 57:910–913, December 2013.
130. S. Lee and C. Castillo-Chavez. The role of residence times in two-patch dengue transmission dynamics and optimal strategies. *Journal of Theoretical Biology*, 374:152–164, June 2015.
131. R. Li, S. Pei, B. Chen, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368:489–493, May 2020.
132. X. Liu, X. Chen, and Y. Takeuchi. Dynamics of an SIQS epidemic model with transport-related infection and exit-entry screenings. *Journal of Theoretical Biology*, 285:25–35, September 2011.
133. X. Liu and Y. Takeuchi. Spread of disease with transport-related infection and entry screening. *Journal of Theoretical Biology*, 242:517–528, September 2006.
134. I. M. Longini, A. Nizam, S. Xu, et al. Containing pandemic influenza at the source. *Science*, 309:1083–1087, August 2005.
135. L. Matrajt, M. E. Halloran, and I. M. Longini. Optimal vaccine allocation for the early mitigation of pandemic influenza. *PLoS Computational Biology*, 9:e1002964, 2013.
136. L. Matthews, D. T. Haydon, D. J. Shaw, et al. Neighbourhood control policies and the spread of infectious diseases. *Proceedings. Biological Sciences*, 270:1659–1666, August 2003.
137. T. M. McMichael, S. Clark, S. Pogosjans, et al. COVID-19 in a Long-Term Care Facility - King County, Washington, February 27-March 9, 2020. *Morbidity and Mortality Weekly Report*, 69:339–342, March 2020.
138. P. Melin, J. C. Monica, D. Sanchez, and O. Castillo. Analysis of spatial spread relationships of coronavirus (COVID-19) pandemic in the world using self organizing maps. *Chaos, Solitons, and Fractals*, 138:109917, September 2020.
139. L. A. Meyers, M. E. J. Newman, and B. Pourbohloul. Predicting epidemics on directed contact networks. *Journal of Theoretical Biology*, 240:400–418, June 2006.

140. L. A. Meyers, B. Pourbohloul, M. E. J. Newman, et al. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*, 232:71–81, January 2005.
141. A. Mishra, B. Ambrosio, S. Gakkhar, and M. A. Aziz-Alaoui. A network model for control of dengue epidemic using sterile insect technique. *Mathematical Biosciences and Engineering*, 15:441–460, April 2018.
142. S. Mishra, A. Mohapatra, R. Kumar, et al. Restricting rural-urban connect to combat infectious disease epidemic as India fights COVID-19. *Journal of Family Medicine and Primary Care*, 9:1792–1794, April 2020.
143. K. Mizumoto and G. Chowell. Transmission potential of the novel coronavirus (COVID-19) onboard the Diamond Princess Cruises ship. *Infectious Disease Modelling*, 5:264–270, 2020.
144. A. Mollalo, K. M. Rivera, and B. Vahedi. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *International Journal of Environmental Research and Public Health*, 17, June 2020.
145. D. Mollison. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3):283–313, 1977.
146. D. Mollison. Markovian contact processes. *Advances in Applied Probability*, 10(1):85–108, 1978.
147. D. Mollison and K. Kuulasmaa. Spatial epidemic models: theory and simulations. *Population Dynamics of Rabies in Wildlife*, 8:291–309, 1985.
148. R. Moss, E. Naghizade, M. Tomko, and N. Geard. What can urban mobility data reveal about the spatial distribution of infection in a single city? *BMC Public Health*, 19:656, May 2019.
149. V. A. Mouchtouri, E. P. Christoforidou, M. An der Heiden, et al. exit and entry screening practices for infectious diseases among travelers at points of entry: looking for evidence on public health impact. *International Journal of Environmental Research and Public Health*, 16, November 2019.
150. M. Najafi, M. Laskowski, P. T. de Boer, et al. The effect of individual movements and interventions on the spread of Influenza in long-term care facilities. *Medical Decision Making*, 37:871–881, November 2017.
151. New Straits Times. 3 coronavirus cases confirmed in Johor Baru. New Straits Times, January 2020. https://www.nst.com.my/news/nation/2020/01/559563/breaking-3-coronavirus-cases-confirmed-johor-baru.
152. T. H. D. Nguyen and D. C. Vu. Summary of the COVID-19 outbreak in Vietnam - Lessons and suggestions. *Travel Medicine and Infectious Disease*, 37:101651, 2020.
153. H. Nishiura. Backcalculating the incidence of infection with COVID-19 on the Diamond Princess. *Journal of Clinical Medicine*, 9, February 2020.
154. Y. Ohkusa and T. Sugawara. Application of an individual-based model with real data for transportation mode and location to pandemic influenza. *Journal of Infection and Chemotherapy*, 13:380–389, December 2007.
155. Y. Ohkusa and T. Sugawara. Simulation model of pandemic influenza in the whole of Japan. *Japanese Journal of Infectious Diseases*, 62:98–106, March 2009.
156. B. Pourbohloul, A. Ahued, B. Davoudi, et al. Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in North America. *Influenza and Other Respiratory Viruses*, 3:215–222, September 2009.
157. G. Pullano, F. Pinotti, E. Valdano, et al. Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. *Euro Surveillance*, 25, January 2020.
158. L. Rass and J. Radcliffe. *Spatial Deterministic Epidemics*, volume 102 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2003.
159. S. P. Ribeiro, A. Castro E Silva, W. Dáttilo, et al. Severe airport sanitarian control could slow down the spreading of covid-19 pandemics in brazil. *PeerJ*, 8:e9446, 2020.
160. S. Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316:1298–1301, June 2007.
161. D. A. Robertson. Spatial transmission models: A taxonomy and framework. *Risk Analysis*, 39:225–243, January 2019.

162. S. Ruan, W. Wang, and S. A. Levin. The effect of global travel on the spread of SARS. *Mathematical Biosciences and Engineering*, 3:205–218, January 2006.
163. T. W. Russell, J. Hellewell, C. I. Jarvis, 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 Surveillance*, 25, March 2020.
164. L. A. Rvachev and I. M. Longini Jr. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75(1):3–22, 1985.
165. Q. Shao and M. Jia. Influences on influenza transmission within terminal based on hierarchical structure of personal contact network. *BMC Public Health*, 15:257, March 2015.
166. S. M. Shi, I. Bakaev, H. Chen, et al. Risk factors, presentation, and course of Coronavirus Disease 2019 in a large, academic long-term care facility. *Journal of the American Medical Directors Association*, 21:1378–1383.e1, October 2020.
167. C. D. Shrader, S. Assadzandi, C. S. Pilkerton, and A. M. Ashcraft. Responding to a COVID-19 outbreak at a long-term care facility. *Journal of Applied Gerontology*, page 733464820959163, September 2020.
168. W. Sim. Japan confirms first case of infection from Wuhan coronavirus; Vietnam quarantines two tourists. The Straits Times, January 2020.
169. M. Singh, P. Sarkhel, G. J. Kang, et al. Impact of demographic disparities in social distancing and vaccination on influenza epidemics in urban and rural regions of the United States. *BMC Infectious Diseases*, 19:221, March 2019.
170. M. Siwiak, P. Szczesny, and M. Siwiak. From the index case to global spread: the global mobility based modelling of the COVID-19 pandemic implies higher infection rate and lower detection ratio than current estimates. *PeerJ*, 8:e9548, 2020.
171. J. Snow. *On the Mode of Communication of Cholera*. John Churchill, second edition, 1855.
172. N. M. Stall, A. Jones, K. A. Brown, et al. For-profit long-term care homes and the risk of COVID-19 outbreaks and resident deaths. *Canadian Medical Association Journal*, 192:E946–E955, August 2020.
173. C. Sun, W. Yang, J. Arino, and K. Khan. Effect of media-induced social distancing on disease transmission in a two patch setting. *Mathematical Biosciences*, 230:87–95, April 2011.
174. X. Sun, Y. Xiao, and X. Ji. When to lift the lockdown in Hubei province during COVID-19 epidemic? an insight from a patch model and multiple source data. *Journal of Theoretical Biology*, 507:110469, August 2020.
175. C. T. Telford, U. Onwubiko, D. P. Holland, et al. Preventing COVID-19 outbreaks in long-term care facilities through preemptive testing of residents and staff members - Fulton County, Georgia, March-May 2020. *Morbidity and Mortality Weekly Report*, 69:1296–1299, September 2020.
176. Thucydides. *The History of the Peloponnesian War*. Circa 430 BCE. http://classics.mit.edu/Thucydides/pelopwar.html.
177. T. N. Vilches, M. Jaberi-Douraki, and S. M. Moghadas. Risk of influenza infection with low vaccine effectiveness: the role of avoidance behaviour. *Epidemiology & Infection*, 147:e75, January 2019.
178. X. Wang, S. Liu, L. Wang, and W. Zhang. An epidemic patchy model with entry-exit screening. *Bulletin of Mathematical Biology*, 77:1237–1255, July 2015.
179. World Health Organization. Novel Coronavirus - Japan (ex-China). Online, January 2020.
180. World Health Organization. Novel Coronavirus - Republic of Korea (ex-China). Online, January 2020.
181. World Health Organization. Novel Coronavirus - Thailand (ex-China). Online, January 2020.
182. World Health Organization. Novel Coronavirus (2019-nCoV) – Situation report 4. Online, January 2020. 24 January 2020.
183. J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395:689–697, February 2020.

184. Z. Xie, Y. Qin, Y. Li, et al. Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. *The Science of the Total Environment*, 744:140929, November 2020.
185. S. Zhang, M. Diao, W. Yu, et al. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *International Journal of Infectious Diseases*, 93:201–204, April 2020.

A Logistic Growth Model with Logistically Varying Carrying Capacity for Covid-19 Deaths Using Data from Ontario, Canada

Georges Bucyibaruta, C. B. Dean, and E. M. Renouf

1 Introduction

Covid-19 infections have increased rapidly worldwide for the last several months. By mid-June, 2020, worldwide confirmed cases of Covid-19 were setting records for daily growth. This growth has continued into late July, with newly reported infections surpassing one quarter of a million daily, driven largely by new infections in the United States, Brazil, India, and South Africa. By July 30, 2020, worldwide recorded infections have surpassed 17 million. Canada announced its first case of Covid-19 in Toronto, Ontario on January 25, 2020 in an international traveller from Wuhan. By July 2020, Canada surpassed 115,000 cases of recorded infections, and almost nine thousand deaths. Within Canada, most cases have occurred within the provinces of Ontario and Quebec. Both provinces adopted more stringent public health measures to reduce infections after major issues arose within long term care facilities. Within Ontario, the number of cases grew rapidly by the end of March, at more than 400 daily new cases, reaching a peak of over 600 daily new cases in the third week of April. By the third week of May, Ontario saw a decline in daily deaths. The Ontario provincial database available through the provincial data portal records onset dates as early as January 01, 2020 in both travellers and non-travellers. While no death dates are recorded in this publicly available database, the earliest deaths recorded show 17 cases with onset dates on or before March 10, 2020 over a geographically dispersed set of health units (Windsor, Chatham, London, Niagara, Waterloo, Peel, Haliburton, Simcoe, Durham, Toronto, York and more).

While Ontario may be past the first peak of the epidemic, the lifting of public health restrictions and social distancing measures in a series of three planned stages

G. Bucyibaruta · C. B. Dean (✉) · E. M. Renouf
University of Waterloo, Waterloo, ON, Canada
e-mail: gbucyiba@uwaterloo.ca; vpri@uwaterloo.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_3

may result in a rise in cases and hospitalizations due to Covid-19. Approximately 2 months after the initial lockdown in March 2020, the province of Ontario unveiled its three stage re-opening plan. Stage 1 of re-opening began on May 19, 2020, with a limited set of businesses allowed to re-open in accordance with strict public health guidelines. Two weeks after the first stage of re-opening, Ontario cases remained stubbornly flat with new cases persisting between 300 and 450 daily. Approximately two-thirds of the health regions in Ontario were allowed to progress to Stage 2 re-opening on June 12, 2020, where businesses such as hair and nail salons, restaurants with outdoor patios, campgrounds and community pools were allowed to re-open. At that time, new recorded cases had fallen to just below 200 per day in Ontario. All remaining health regions progressed to Stage 2 by July 7. On July 17, 2020, 24 of Ontario's 34 health regions, who were among the first health units to progress to Stage 2, were allowed to progress to Stage 3 of re-opening, which allows most businesses and workplaces to re-open but with limits on capacity and with measures in place such as wearing of masks indoors. By the final week of July, reported new infections hovered between approximately 100 and 170 per day, with some of the higher numbers driven by large outbreaks in agricultural farm workers. Meanwhile, health officials across the province remain alert to warning signs of resurgence.

Our approach here uses a logistic growth model for the cumulative number of deaths from Covid-19. The logistic growth model includes a carrying capacity parameter that is meant to reflect an upper limit in the number of deaths. Here we consider, conceptually, that deaths would be limited by the number of (true) cases. Since the number of cases is changing over time, so should the carrying capacity parameter for death in the logistic model. The logistic growth model we employ therefore allows the carrying capacity parameter to change over time by incorporating a logistic growth function for the carrying capacity parameter.

Throughout the course of the Covid-19 pandemic, Ontario and Canadian data have been modelled in a series of important papers using a broad set of epidemiological methods [4, 13, 14]. One of the first papers using Canadian data fit an exponential curve to the number of daily cases and estimated growth rates by fitting a linear regression model to the logarithm of the data [11]. Growth rates were estimated for two separate windows of time, to allow for changes in trend as a result of a significant public health intervention.

In terms of analysis of Ontario data, simple exponential or logistic growth models have been considered in a report published by a resource management group in collaboration with researchers at academic institutions across Canada. Growth rate curves of time series data of Covid-19 deaths for Canada are modeled to incorporate changes in rates under different public health interventions [5]. From the time series of daily deaths, the authors back-calculate the cumulative number of infections. Using a time series growth model the number of future infections are predicted. The model utilizes either a simple exponential growth model or a logistic growth model. In a report published May 17, 2020, the authors demonstrate how to use their modeling framework and provide software to generate forecasts of cases and deaths for Ontario and Canada under the assumption of continued health policy measures as they currently stand.

There have been other studies of Ontario data beyond growth curve modeling. For example, in a recently published article, Wu and colleagues [15] use a generalized Susceptible Exposed Infected Recovered (SEIR) model that allows incorporation of asymptomatic infectious, quarantined susceptible and isolated exposed model compartments. They evaluated trends in transmission and the effect of social distancing measures based on data up to March 29, 2020, showing an increasing effectiveness of public health interventions in lowering the reproductive number of Covid-19.

Analyses utilizing logistic growth curve models to forecast epidemic growth of infections have been published using data outside of Canada as well. These models use the standard form of the logistic growth curve to model infections. To model early growth of the epidemic in Hubei, China, [10] applied the generalized logistic growth model to produce short-term forecasts of cases using data up to February 29, 2020. [16] also use a generalized logistic growth model to forecast cases in mainland China excluding Hubei province. They compare the classical logistic growth model, a generalized model, and a generalized Richards model using data up to March 10, 2020. [3] uses a logistic growth model as well as an SEIR model to estimate final epidemic size worldwide.

A recent paper using publicly available data from Nigeria applied a logistic growth curve as one element of a larger model forecasting epidemic growth [1]. Using the daily number of new cases of Covid-19 in Nigeria, the authors implemented an ensemble of forecast models. One of these models included as a component a logistic growth model with time-varying carrying capacity. This implementation allows the carrying capacity to vary as a function of time, rather than a logistically varying carrying capacity as in our model.

The logistic growth model has been used extensively, either in traditional or extended form, to model new infections for Covid-19. A model accommodating a varying carrying capacity parameter as considered in this paper has not been used in logistic growth models for Covid-19. In our application which follows, we demonstrate the utility of the logistic growth model to model deaths, with the modification that the parameter identified with carrying capacity varies over time. We will also provide confidence bands for potential indicators through Monte Carlo methods.

2 Data Description

The aggregate data used here were obtained from the daily epidemiological summaries released by Public Health Ontario on their website. In a table under the heading ‘Severity’, these daily reports show the reported cumulative deaths and the daily change from the previous report, with the caveats that only deaths for lab-confirmed cases are included here, and also that there is a reporting delay for deaths. Figure 1 shows deaths in Ontario as reported by Public Health Ontario.

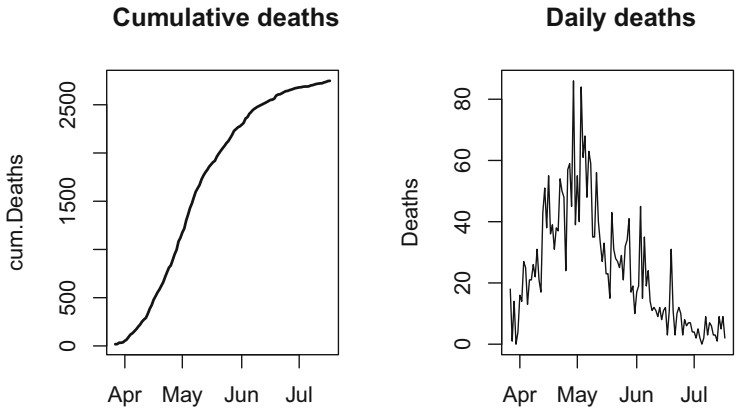


Fig. 1 Cumulative and daily number of deaths in Ontario from 27/03/2020 until 17/07/2020 (data from Public Health Ontario daily epidemiological summaries)

3 Methods

Our focus is on modeling deaths in the province of Ontario. This approach uses a growth model for the cumulative number of deaths from Covid-19, and our growth model accommodates, conceptually, that deaths have a carrying capacity that would be limited by, for example, the number of cases or hospitalizations which change over time. This is not a process model, but an empirical model. A series of publications, [2, 8, 12], portray an array of logistically developing and diffusing social mechanisms. They compare technological innovations as a social epidemic by arguing that the former do not usually distribute themselves evenly through time. They consider a model where the carrying capacity of the system increases dynamically, but in a distinct pulse. Conceptually, we adapt this approach and we allow the carrying capacity of our logistic growth model to vary as a logistic growth curve.

3.1 Logistic Growth Model for the Mean

As the carrying capacity is meant to conceptually reflect the number of cases or hospitalizations, and since these values are changing over time, so should the carrying capacity for death in the logistic model. Let $N(t)$ be the cumulative number of deaths at time t, where $t = 0$ is the recorded date of the first death in Ontario. The logistic growth curve model can be represented in the following way:

$$\frac{dN(t)}{dt} = rN(t)\left[1 - \frac{N(t)}{K(t)}\right] \tag{1}$$

where

r is the growth rate and
$K(t)$ is the carrying capacity parameter for modeling $N(t)$ at time t,
$N(t=0) = N_0$ is the cumulative number of deaths at the initial time.

Time here is recorded as rounded to days, so N_0 is the number of deaths on the first day of recorded deaths. The general solution of (1) is

$$N(t) = \frac{N_0 \exp(rt)}{1 + rN_0 \int_0^t \frac{\exp(rx)}{K(x)} dx} \tag{2}$$

with $K(t)$ also modeled as a logistic growth curve:

$$\frac{dK(t)}{dt} = \alpha K(t)\left[1 - \frac{K(t)}{G}\right] \tag{3}$$

with α being the growth rate for $K(t)$, and G being the carrying capacity parameter for modeling $K(t)$. The analytical solution of $K(t)$ follows as

$$K(t) = \frac{G}{1 + (\frac{G}{G_0} - 1)\exp(-\alpha t)} \tag{4}$$

where G_0 is the initial value of the carrying capacity for modeling $K(t)$. Substituting (4) in (2) yields the solution:

$$N(t) = \frac{G}{1 + (A_1 \exp(-\alpha t)) + (A_2 \exp(-rt))} \tag{5}$$

where

$$A_1 = \left(\frac{G}{G_0} - 1\right)\left(\frac{r}{r - \alpha}\right) \tag{6}$$

and

$$A_2 = \left(\frac{G}{N_0} - 1\right) - \left(\frac{G}{G_0} - 1\right)\left(\frac{r}{r - \alpha}\right). \tag{7}$$

3.2 Non-linear Least Squares Estimation

We employ non-linear least squares estimation for the parameters. The function $N(t)$ is known up to a set of $p = 4$ unknown parameters $\theta = (\theta_1, ..., \theta_p) = (G, G_0, r, \alpha)$ which also must be estimated. Under the assumption that both the

predictor and the response are observed without error, the relationship in (5) will hold to define cumulative counts of deaths over time if the model is correct.

In reality, measurement errors will arise. Non-linear least squares estimation proceeds by finding $\hat{\theta}$ that minimizes

$$RSS(\theta) = \sum_{i=1}^{n} (y_i - N(t_i))^2 \tag{8}$$

where $y_i = N(t_i) + \epsilon_i$, and $\epsilon_i \sim N(0, \sigma^2)$.
An estimate of the measurement error is obtained as

$$\hat{\sigma^2} = \frac{RSS(\hat{\theta})}{n}$$

where $RSS(\hat{\theta})$ is the residual sum of squares.

4 Introducing Stochasticity into the Daily Counts

The logistic growth model provides an estimate of the future mean cumulative deaths. We also consider stochasticity in the daily counts, required for short-term analyses of the behaviour of the disease progression and assume that the daily number of deaths follow a negative binomial distribution with mean derived from the fitted values of the logistic growth model. We estimate the dispersion parameter $\hat{\kappa}$ using maximum likelihood estimation and this allows us to incorporate stochasticity in the daily counts. We then utilize a Monte Carlo approach for obtaining future daily predictions by generating future daily data using a negative binomial distribution with mean derived from the fitted logistic model for cumulative deaths and dispersion $\hat{\kappa}$. Using $B = 1000$ simulations, under the negative binomial distribution, we predict future cumulative deaths. In a single peak epidemic wave, an indicator of lack of control could be based on the cumulative number of deaths $N(t)$, and the rate of change of deaths $\frac{dN(t)}{dt}$, as described in the next section.

5 Short-Term Predictions and Beyond

We discuss here potential tools that could be utilized as short-term predictors, as applicable more broadly for pandemic monitoring in various settings. For example, we may calculate the probability that the total number of deaths observed in the next l days, after a reference point t_0, indicating current time, will exceed that observed in the past l days, where l monitors short-term activity, for example, $l = 3$ or $l = 5$. We also examine the probability that the current growth rate of deaths exceeds that

seen during the beginning of the first phase of re-opening when the pandemic was seen to be sufficiently under control.

Looking beyond the short-term predictors, a few indicators of a second wave, or resurgence, have been proposed however most are ad hoc apart from the risk ratio recently proposed by Noorbhai [9], who offers a model based on the ratio of total recoveries to cases. This is problematic for Ontario where recoveries are tracked with different methodologies within each of the 34 health units. Freitag et al. [6] has proposed a spatio-temporal model of mobility levels, taking into account population density, as an indicator of resurgence. An immediate issue for Ontario, however, is that indicators for resurgence cannot be based upon growth models, such as the one used here, which are meant for modeling an epidemic with a single peak.

6 Results

The logistic growth model with a logistically varying carrying capacity parameter was fitted to cumulative deaths. Figure 2 shows the results of the analysis with both fitted and observed values displayed. The time-varying carrying capacity curve is shown in blue. The plot illustrates how the logistic curve approaches its upper limit over time. As the logistic growth model reaches the asymptote, the difference between the carrying capacity and the logistic growth curve diminishes. Figure 3 displays the confidence bands associated with the fit are also provided. The model

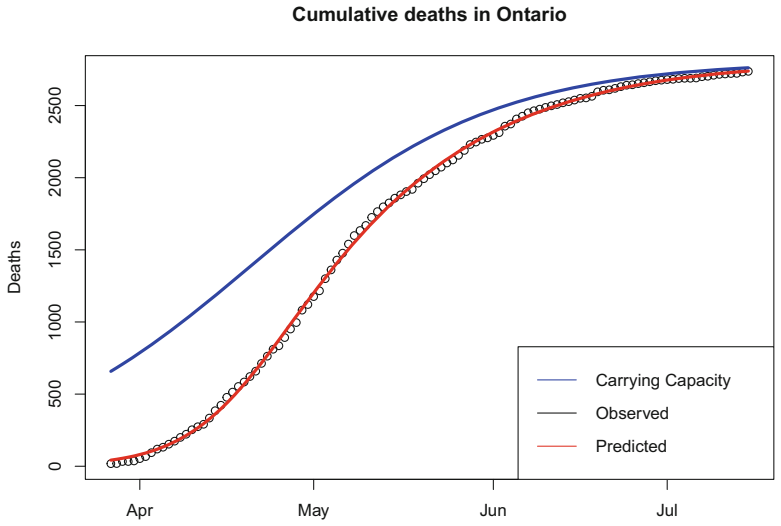


Fig. 2 Fitted logistic growth model with logistically varying carrying capacity from 27/03/2020 to 17/07/2020

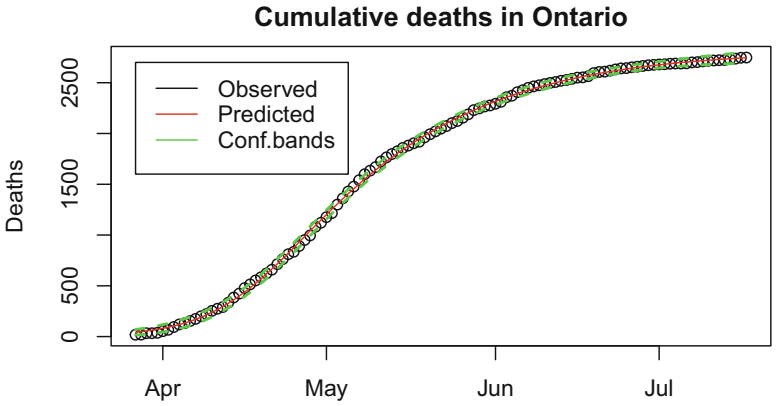


Fig. 3 Fitted Logistic growth model with logistically varying carrying capacity from 27/03/2020 to 17/07/2020, showing confidence bands in green

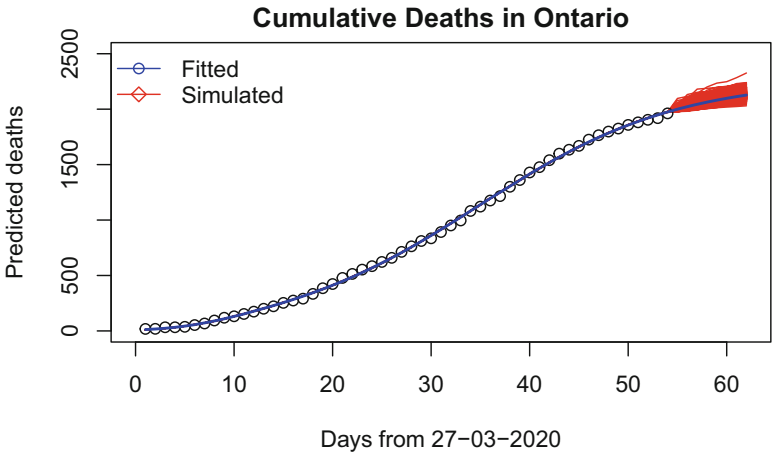


Fig. 4 Total number of Deaths predicted under logistic growth curve (blue) with 95% confidence bands(green) and simulated curves under negative binomial distribution (red)

show no obvious lack of fit. Figure 4 displays the logistic growth curve with 95% confidence bands and simulated curves under negative binomial distribution.

In order to assess our model we present some comparisons between the give day ahead forecasted number of deaths for various dates versus what was actually observed, in Table 1.

To examine the future trajectory of predicted deaths, we consider the probability that the total number of deaths observed in the next l days after day $t_0 = 54$ in our dataset (19 May, 2020), exceeds that observed in the last l days, where $l = 3$

Table 1 Observed cumulative deaths at time t_0 (second column), predicted cumulative deaths after 5 days (3rd column) and observed cumulative deaths after 5 days (last column)

Date (t_0)	$N(t_0)$	Predicted$_{N(t_0+5)}$	Observed$_{N(t_0+5)}$
2020-05-10	1669	1952	1858
2020-05-19	1962	2081	2012
2020-06-04	2372	2447	2475
2020-06-12	2507	2599	2553
2020-07-04	2689	2731	2710
2020-07-17	2748	2758	2755

or 5. Using the Monte Carlo approach described in Sect. 4, these probabilities are estimated as,

$$\hat{P}(N(t_0+3) > N(t_0) + \{N(t_0) - N(t_0-3)\}) = .29$$

$$\hat{P}(N(t_0+5) > N(t_0) + \{N(t_0) - N(t_0-5)\}) = .14$$

It is useful to note that the probability declines as l increases. In reality, starting at the reference time, the number of deaths in the next l days slightly exceeded the number of deaths observed during the previous l days, for both $l = 3$ and 5, by 5 deaths and 3 deaths respectively. As another example, when $t_0 = 84$, corresponding to 18 June, 2020, we have,

$$\hat{P}(N(t_0+5) > N(t_0) + \{N(t_0) - N(t_0-5)\}) = 0.72$$

Starting at the reference time, the number of deaths in the subsequent 5 days exceeded the number of deaths observed during the previous 5 days by 22 deaths. The probability values themselves give some indication of the strength of evidence concerning the prediction, yet a threshold is required to form an alarm system. This could be developed through a receiver operating characteristic curve analysis.

Figure 5 shows the receiver operating characteristic curve for the model. This curve was developed through a Monte Carlo simulation, similar to the method described in Sect. 4, where we calculated the probability of exceeding the l-day-ahead target for deaths, the target being the total number of deaths in the previous l days, for different t_0's. Then we assessed whether the observed number of deaths actually exceeded the target: for each t_0 an *outcome* of 1 was assigned if the number of deaths in the next l days exceeded the number of deaths observed during the previous l days, otherwise an outcome of 0 was assigned. By comparing the estimated probabilities of exceeding the l-day-ahead target to a series of thresholds between 0 and 1, we obtain a *prediction* of whether the number of deaths l days ahead will exceed the target. For each of these probability thresholds, we are then able to compare *outcomes* with *predictions* to calculate the true positive rate (TPR) and false positive rate (FPR) associated with each of these thresholds.

On the ROC curve, the point at the top left corner of the curve is identified as providing the level of best performance. This occurs at a TPR of 0.9 and FPR of 0.3, as shown in Fig. 5. This, in turn, corresponds to a threshold probability of 0.6. At

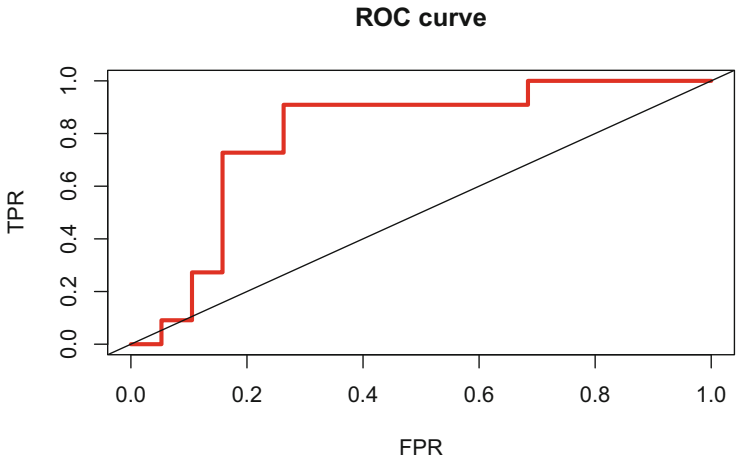


Fig. 5 Optimal threshold for predicting an increase in deaths over the next l days, compared to the previous l days, where $l = 5$

this point, we therefore expect 90% of true positives will be well classified, and 30% of false positives will be misclassified. Therefore, when considering predictions of deaths 5 days ahead, we would sound an alarm indicating an expected increase in deaths when the probability described above is greater than the threshold of 0.6. A probability of 0.72, as obtained in our most recent example, would be considered high enough to sound an alarm, according to the threshold determined by means of our ROC curve analysis.

We can also develop an assessment of risk level given by $dN(t)$, the rate of change. Since $dN(t)$ is the derivative of N(t), these indicators provide the same probability measure, however $dN(t)$ gives us the ability to compare to earlier phases of the pandemic, with a straightforward visual representation. Figure 6 gives a visual representation of the values of $dN(t)$ over time since late March, 2020.

We present in Table 2 a scale of low, medium and high risk values for the predicted rate of change as measured at time t with the intention being that rates of change should be decreasing over time, and where they are not, could support concerns about resurgence.

If we use the date of June 18, 2020 (day 84 in our dataset) given in our previous example, we calculate the predicted average growth rate over the next 5 days to be 11.4. We can compare this value to the start of Phase 2 on June 12, shown in Fig. 6, where the instantaneous growth rate was 13, and we see therefore, that growth rate is declining.

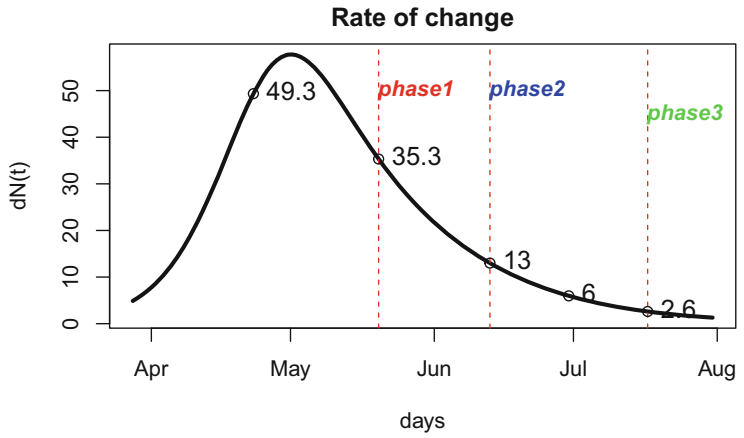


Fig. 6 An indicator based on $dN(t)$, the rate of change of deaths

Table 2 Risk levels based on the rate of change, $dN(t)$

$dN(t)$	Risk level
>35	Very high
13–34	High
7–12	Medium
2–6	Low
<2	Very low

7 Discussion

In this paper we have used a simple model to predict future deaths in the short term. Our model fits reasonably well, with narrow confidence bands. We consider an extension of the classic logistic growth model which allows the carrying capacity for deaths to change logistically over time. This model does not take into account the mechanisms of transmission of Covid-19, such as health interventions and human behavior. As the situation evolves, anomalous values or rapidly changing trends could upend any prediction efforts. Worse still, sudden shocks that permanently affect a time series could also render all past data as irrelevant. We also note that phenomenological growth models such as the logistic growth curve model presented here are meant for predicting growth trajectories during a single peak epidemic. However, multiple peak epidemic trajectories caused by factors such as increasing contacts and releasing of public health interventions are much more challenging to model.

Our study has some important limitations to acknowledge with respect to the data used in our model. This pandemic in particular has highlighted challenges in data collection and management in Ontario. The deaths, as reported, do have a lag from their actual death date as can be gleaned from graphs provided in the Daily Epidemiological Summary published by Public Health Ontario. While deaths are

a late indicator, they are still a valuable marker to track as part of a composite surveillance plan. In addition to monitoring the growth rate of deaths as described in this paper, policy makers would consider a broad context, examining other metrics such as hospitalization numbers, test percent positivity, current reproduction number, and number of new cases per 100,000 population. These indicators for the future number of deaths which we present here are meant as an additional layer of insight to combine with other important metrics.

In future work we will explore modeling the carrying capacity parameter as a function of hospitalizations. Hospitalizations may indeed provide conceptually a better upper limit for modeling deaths as many cases may be mild whereas hospitalizations would be strongly linked to the more severe cases that progress to death. In future work we hope to address the reporting lag for deaths and to incorporate a method to adjust for these lags in our predictions based on hospitalization rates.

Provincial analyses could be supplemented by regional analyses in order to detect regional trends. As noted in [7], using aggregate level data as an indicator for all of Ontario can obscure what might be happening at the local level. The indicators in this paper could easily be extended for use at the regional level and we intend to model this in future work.

At the time of completing this study, Ontario has allowed all health regions to progress to Stage 3 of re-opening, where bars and restaurants with indoor seating have re-opened, as well as gyms, personal care services such as hair salons, and also places of worship. Most remaining workplaces and businesses are allowed to re-open with some precautions in place. On September 8, 2020, schools in the province will re-open with some additional public health measures such as mandatory masks for grade 4 and up, but for the most part will proceed as usual except in 24 out of 76 school boards, where secondary school class sizes will be limited to 15 students who will attend on a rotating schedule. Physical distancing requirements are set at only 1 m by the Ministry of Education, even in classes where masks are not mandatory.

This return to school will result in increased contact for the population of Ontario, and in conjunction with the upcoming influenza season, the fall season may be a period of increased risk for Covid-19. Formal indicators of public health interventions may be useful for managing risk, and we intend to investigate extensions of this type of model that will allow modeling of future epidemic peaks in Ontario. In this paper we presented a model to predict future deaths in the short-term with an appropriate measure of uncertainty.

8 Declaration of Competing Interest

We have no conflict of interest.

Acknowledgments The authors are supported in part by a CIHR grant for the Canadian 2019 Novel Coronavirus (COVID-19) Rapid Research Funding Opportunity.

References

1. Abdulmajeed K, Adeleke M, Popoola L (2020) Online forecasting of Covid-19 cases in Nigeria using limited data. Data in Brief p 105683
2. Banks R (2013) Growth and diffusion phenomena: Mathematical frameworks and applications. Springer Science and Business Media 14
3. Batista M (2020) Estimation of the final size of the Covid-19 epidemic. MedRxiv https://doi.org/10.1101/2020.02.16.20023606
4. Bignami S, Ghio D, Van Assche A (2020) Estimates of Covid-19 case-fatality risk from individual-level data. medRxiv
5. Daniel C, Frid L, Rayfield B, Embrey A (2020) A scenario-based forecasting framework for Covid-19: Case study application for Canada https://apexrms.com/wp-content/uploads/download3iles/covid-19-canada-framework-whitepaper.pdf
6. Freitag MO, Schmude J, Siebenschuh C, Stolovitzky G, Hamann H, Lu S (2020) Critical mobility, a practical criterion and early indicator for regional Covid-19 resurgence. medRxiv
7. Harris JE (2020) Reopening under Covid-19: What to watch for. Tech. rep., National Bureau of Economic Research
8. Meyer PS, Ausubel JH (1999) Technological forecasting and social change. Technological forecasting and social change 61(3)
9. Noorbhai H (2020) A mathematical model to guide the re-opening of economies during the Covid-19 pandemic. Annals of Medicine and Surgery 57:5
10. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, Yan P, Chowell G (2020) Real-time forecasts of the Covid-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling 5:256–263
11. Scarabel F, Pellis L, Bragazzi NL, Wu J (2020) Canada needs to rapidly escalate public health interventions for its Covid-19 mitigation strategies. Infectious Disease Modelling 5:316–322
12. Tsoularis A, Wallace J (2002) Analysis of logistic growth models. Mathematical Biosciences 179(1):21–55
13. Tuite AR, Fisman DN, Greer AL (2020) Mathematical modelling of Covid-19 transmission and mitigation strategies in the population of Ontario, Canada. CMAJ 192(19):E497–E505
14. Tuite AR, Greer AL, De Keninck S, Fisman DN (2020) Risk for Covid-19 resurgence related to duration and effectiveness of physical distancing in Ontario, Canada. Annals of Internal Medicine 173(8):675–678
15. Wu J, Tang B, Bragazzi NL, Nah K, McCarthy Z (2020) Quantifying the role of social distancing, personal protection and case detection in mitigating Covid-19 outbreak in Ontario, Canada. Journal of mathematics in industry, 10, 1–12.
16. Wu K, Darcet D, Wang Q, Sornette D (2020) Generalized logistic growth modeling of the Covid-19 outbreak in 29 provinces in China and in the rest of the world. arXiv preprint arXiv:200305681

COVID-19 in Ontario: Modelling the Pandemic by Age Groups Incorporating Preventative Rapid-Testing

Roie Fields, Lia Humphrey, Edward W. Thommes, and Monica G. Cojocaru

1 Introduction

Since its first reported outbreak in Wuhan, China [1], the novel betacoronavirus SARS-COV-2 and its associated disease COVID-19 has spread to virtually all corners of the globe. In absence of vaccination and effective antiviral drugs, over 47 million people worldwide have been confirmed as having contracted COVID-19, resulting in over 1.2 million deaths as of November 2, 2020 [2].

Some early models for the COVID-19 outbreak were purely curve-fitting [3]. Much of the epidemiological modelling is now performed with variants of the classic SIR model [4–7] which fits the number of individuals in a population who are susceptible to, infected with, or have recovered from a communicable disease. The model can be extended to incorporate different phases of a disease outbreak, [8] or population subtypes [9, 10]. Studies have stratified Ontario populations by categorising gender [11] or profession (such as healthcare workers) [12] separately. Little work has been done to differentiate age groups in COVID-19 models, which we show is an important factor to consider (see also [13]).

Although contact rates between individuals have previously been studied [14, 15], due to imposed social distancing and sheltering measures, absolute contact rates during the pandemic have changed drastically from their projected baseline as was demonstrated by a paper that was published shortly after our manuscript was submitted [16]. Presently, governments have imposed measures in an attempt to reduce the spread of COVID-19. These include sheltering in place and isolation,

R. Fields (✉) · L. Humphrey · M. G. Cojocaru
University of Guelph, Guelph, ON, Canada
e-mail: rfields@uoguelph.ca; humphrel@uoguelph.ca; mcojocar@uoguelph.ca

E. W. Thommes
Sanofi Pasteur Global, Toronto, ON, Canada
e-mail: Edward.W.Thommes@sanofi.com

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_4

travel restrictions, social distancing protocols, and the closure of public and commercial spaces. Specifically in Ontario, the government has mandated a slow reopening of the economy [17], and provided guidance on protective measures such as hand-washing and the use of face coverings. In fact, in the face of the second wave, most recent government measures have reversed to stronger social distancing ones in provincial areas that are more affected.

It is known that the contact rates between individuals of different age groups will vary significantly [18], so it is important to take this into account. An Ontario paper uses an SEIR-type model with compartments by age, but only modifies the severity of infection per age group [9]. A Brazilian study uses data from nine distinct age groups, but uses identical contact rates within and between groups [19]. The work in [20] uses an SEIR model with age stratification to investigate the effect of various management strategies. Unfortunately, the study suffers from a lack of empirical data instead arriving at exposure rates by "making educated estimates of the effects of hygiene restrictions and specific social interactions in each place". Research is now suggesting that children only play a minor role in transmitting COVID-19 [21]. The spread of the disease is greatest between adults in the same age group, and less frequently between parents and children. These results will have an impact on the effective contact rates between individuals [22].

In this chapter we use a SEIR-type compartmental model and provincial case data from Ontario, Canada, to build upon our previous work [13] to depict COVID-19 case progression from February 14 to October 5, 2020. We fit our differential equations model describing COVID-19 transmission to Ontario Public Health onset case data. We then estimate the effects of preventative measures (social distancing, mask-wearing, etc.) on age-stratified contact rates. Furthermore, to address the implementation of population-wide diagnostic testing by the provincial government on May 24, 2020, we retroactively investigate the effects of different levels of testing for a given test success rate and infer potential cases prevented. Finally, we estimate levels of large-scale rapid testing that, in combination with other measures, maintain an effective reproduction number (R_{eff}) below 1.

The presentation below includes a section on building our model, as well as a detailed discussion of our data sources. We then continue in Sect. 3 with our estimates on fitting the pandemic contact rates to onset cases, and then we show how differing levels of large-scale testing may help in controlling the pandemic.

2 Methods and Materials

2.1 Adapted SEIR Model

We begin by presenting an adapted SEIR model that modifies variants of the SEILR model presented in [23, 24] and [13]). In Sects. 2 and 3, we summarize model findings from [13] to lay the groundwork for a revised SEILR(pas)

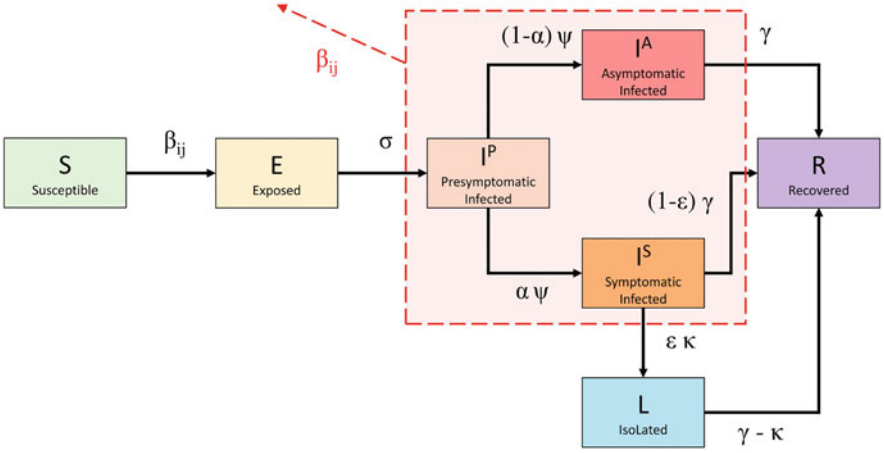


Fig. 1 Schematic of SEILR(pas) model for COVID-19 progression and control measures for a Group i, $i \in \{1, 2, 3\}$ as in [13]. The outward arrow labelled β_{ij} implies that the encircled compartments (Presymptomatic, Asymptomatic, and Symptomatic infected) in each age group have an impact on new transmissions to the other age groups $\{1, \ldots, 3\}$

system in Sect. 4, incorporating diagnostic COVID-19 tests at various levels of uptake, to explore sensitivity effects on the total number of cases. We differentiate the **extended SEILR(pas) model** from other compartmental models through the inclusion of additional compartments to discern various stages of symptom progression. Seen in Fig. 1, these are Susceptible (S), Exposed (E)—not yet contagious, Infected (p)resymptomatic (I^P)—contagious, infected (a)symptomatic (I^A), Infected (s)ymptomatic (I^S), Recovered (R), and iso(L)ated—symptomatic cases isolated to prevent spread.

In order to account for differences in disease susceptibility and infection outcomes, we divided the population of Ontario in three age-stratified subgroups: Group 1, denoted by N_1: 0–19 years old, Group 2, denoted by N_2: 20–59 years old, and finally Group 3, denoted by N_3: (≥ 60) years old.

Contact rates (c_{ij}) within and between the age groups were inferred from Canadian contact data estimated by Prem et al. [15] and we continue to use those here. Since contact rates are a factor in determining the size of the effective contact rates, typically denoted by β, in the context of age-stratified transmission, the effective rates will be denoted by β_{ij}. Before we outline the equations governing our model, we include first the flow chart between the above compartments, together with classic notations of the flow rates:

Considering the i-group model for $i, j \in \{1, \ldots, K\}$, the differential equations governing each of the groups are given by:

$$\begin{cases} \dfrac{dS_i}{dt} = -\displaystyle\sum_{j=1}^{K} \beta_{ij} S_i(t) \frac{I_j^P(t) + I_j^A(t) + I_j^S(t)}{N_j(t)}, \\ \dfrac{dE_i}{dt} = \displaystyle\sum_{j=1}^{K} \beta_{ij} S_i(t) \frac{I_j^P(t) + I_j^A(t) + I_j^S(t)}{N_j(t)} - \sigma E_i(t), \\ \dfrac{dI_i^P}{dt} = \sigma E_i(t) - \psi I_i^P(t), \\ \dfrac{dI_i^A}{dt} = (1-\alpha) \cdot \psi \cdot I_i^P(t) - \gamma I_i^A(t), \\ \dfrac{dI_i^S}{dt} = \alpha \cdot \psi \cdot I_i^P(t) - (1-\varepsilon)\gamma I_i^S(t) - \varepsilon \cdot \kappa \cdot I_i^S(t), \\ \dfrac{dL_i}{dt} = \varepsilon \cdot \kappa \cdot I_i^S(t) - (\gamma - \kappa) L_i(t), \\ \dfrac{dR_i}{dt} = \gamma(I_i^A(t) + (1-\varepsilon) I_i^S(t)) + (\gamma - \kappa) L_i(t). \end{cases}$$

Individuals are all initialized in the S compartment, with the exception of a single infected individual from the largest group (2) seeded in the I^P compartment. Susceptible individuals exposed to the virus enter the E category for an average of $1/\sigma$ days before they become contagious, at which point they move into the I^P category. After an average of $1/\psi$ days, α proportion of individuals will develop symptoms and move into the I^S compartment, with the remaining $1-\alpha$ moving into the I^A compartment. Individuals in I^P, I^A and I^S compartments are all contagious. ε proportion of individuals in I^S choose to self-isolate to prevent further disease transmission, and as a result move into compartment L (isolated) and do not interact with others. They do so with a delay of $1/\kappa$, accounting for test result waiting time and individuals who may disregard minor symptoms initially. In $1/\gamma$ days after entering the I^A and I^S compartment, individuals all recover from (or succumb to) their infection and move into the R compartment, where they remain permanently. The model was implemented and numerically solved with *R* (version 4.0.2) using the packages *ggplot2* and *deSolve*.

2.2 Ontario Specific Parameter Values and Data Sources

We assume that the outbreak of COVID-19 in Ontario begins on February 14th, and that the implementation of intervention and control strategies begins on March 15th.

Control responses in Ontario were officially enacted on Monday, March 17th [25], but commensurate with earlier institutional responses, we adjust for the fact that the public began to alter behaviour before this date. For further discussion, see [13]. We assume individuals who recover from infection remain in the R compartment, with immunity from COVID-19 for the duration of the simulation. Given the short time frame for simulation, our model does not include births or natural deaths, and any COVID-19 deaths are captured in R. Furthermore, we assume an isolation delay of 1 day on average, due to individuals not always taking initial symptoms seriously enough to isolate. Additionally, we assume an isolation compliance rate of 95%.

An early study of the virus in quarantined cruise ship passengers finds that 17.9% of infected individuals were asymptomatic, and suggests the true proportion could be up to 39.9% depending on the latency period of the virus [26]. A meta-analysis of six studies estimates the asymptomatic infection rate to be anywhere in the range of 18.4% to 78.3%, and cites 46% as the most likely value [27]. Public Health Ontario states estimates ranging from 8.2% to 87.8%, depending on a variety of factors [28]. The true proportion of asymptomatic individuals may be difficult to determine accurately, because those without symptoms are unlikely to be tested. As such, we approximate a value of 0.5 as the proportion of symptomatic cases.

Next, assume that the only cases reported by Ontario's public health units (PHUs) are symptomatic cases, and that every symptomatic case will be identified and reported. In addition, we assume that asymptomatic and presymptomatic cases are not tested and therefore do not self-isolate.

The model presented above is now tailored to Ontario using the following parameter values, as well as Ontario data sources for the pandemic evolution to date (Table 1).

3 Fitting the Age Stratified Model to iPHIS Data

In this section we present a brief recap of our fitting results for the model (2.1) to the existing iPHIS data. We have presented these in detail in [13] (where we include discussions about the effects of mobility on the effective contact rates β_{ij}). Here we focus the presentation on the needed details for ease of reading.

In [15] (a study of country-specific contact rates), a 16×16 contact matrix provides contact rates between Canadians aged 0–80 in 5-year intervals with the each row representing the age interval of an individual, and the age interval of the group being contacted listed on the columns. The highly stratified contact rates in this 16×16 matrix were then combined in [13] to create a condensed 3×3 contact matrix in accordance to our selected age groups in the following manner:

Table 1 Parameters and initial values of the SEILR(pas) system (2.1)

Symbol	Definition	Initial value	Reference
K	Number of age groups	3	
N_i	Subgroup population size		
	N_{total}	14,566,547	[29]
	N_1: (0–19)	3,141,693	[29]
	N_2: (20–59)	7,977,131	[29]
	N_3: (60+)	3,447,723	[29]
c_{ij}	Number of daily contacts with Group j per member of Group i	Table 2	
p	Transmission probability per contagious contact	Section 3.1	
β_{ij}	Effective contact-rate, $i, j \in \{1, \ldots, K\}$	$c_{ij} * p$	
σ	Time from exposure until contagious (days)	1/2.5	[9]
ψ	Time from contagious until symptomatic (days)	1/3.5[a, b]	[9, 30]
α	Proportion of permanently asymptomatic cases	0.5	[26–28]
ε	Proportion of compliance with isolation	0.95	
γ	Recovery/removal rate	1/7	[8]
z	Sensitivity of COVID-19 tests	0.7	[31]
δ	Daily population testing rate	0	
		0.0025	
		0.005	
		0.01	

[a] The parameter ψ is estimated from the relationship $1/\sigma + 1/\psi$ = incubation period, where $1/\sigma = 2.5$ days and incubation period = 6 days

[b] This value pre-dates October 1st, 2020

Table 2 Contact rates calculated for the three Ontario population subgroups. Contact rates are taken from Canadian estimates by Prem et al. [15] and weighted according to census data from Statistics Canada [29]

	Group 1	Group 2	Group 3	Total contacts
Group 1	$c_{11} = 8.565645854$	$c_{12} = 4.661272358$	$c_{13} = 0.304014985$	13.5309332
Group 2	$c_{21} = 2.987996842$	$c_{22} = 11.49063721$	$c_{23} = 0.522285468$	15.00091952
Group 3	$c_{31} = 1.248771202$	$c_{32} = 3.686037351$	$c_{33} = 1.982952539$	6.917761092

1. Columns are partitioned into the three age groups with which individuals come into contact: Group 1 (0–19), Group 2 (20–59), and Group 3 (60–80).[1]
2. Contacts are summed in each row in accordance with the new column partitions, yielding a 16×3 matrix, representing each of the 16 age groups' combined daily contacts with each of the 3 new age groups created

[1] Since [15] estimates contact rates only up to 80 years of age, we assume that Ontarians aged 80+ have identical contact rates to those aged 60–80 years.

3. A population-weighted average of each respective cumulative contact-rate for all members of each group was taken to generate our final 3×3 matrix in accordance with our new age groups, outlined below. Each interval's population was taken from [29].

We preserve our contact structure by adopting Canadian relative contact rates from [15] and assume they apply to Ontario for our time period with uniform contact scaling across all interactions. We assume our age-stratified population subgroups are homogeneously mixed, with heterogeneous intergroup mixing.

3.1 Dampening Contact Rates to Replicate Behavioural Changes and Preventative Measures

The effective contact rate for a member of Group i with members of Group j is given by

$$\beta_{ij} = c_{ij} \cdot p, \tag{1}$$

As p for COVID-19 is not well established, we must solve for it implicitly by fitting our model to the onset data presented by iPHIS for our pre-intervention period (Feb 14th–Mar 15th). We do so using a derivative-free optimization method (the Golden Section Search method) to minimize the SSE between our model output and iPHIS data (see [32]). We conclude that pre-lockdown, the transmission rate is $p = 0.045085$. We note that this transmission success rate is much lower than the estimate of 0.145 given by Wu et al. [33], yet higher than the estimates of 0.041, 0.0365, 0.018 and 0.0045 for presymptomatic, severe symptomatic, mild symptomatic and asymptomatic cases respectively given by Abdollahi et al. [34].

The following depicts our simulated data from Feb 14th–March 15th (Fig. 2).

After March 15, 2020, both the social behaviour and the mobility of the population have changed, as the province adopted various preventative measures to help curb the spread of COVID-19. (For more detail on date selection, see [13].) We therefore need our model to reflect:

- changes in the frequency of contacts due to social distancing measures intra- and inter-groups
- changes in the transmission probability due to mask wearing and other hygiene practices.

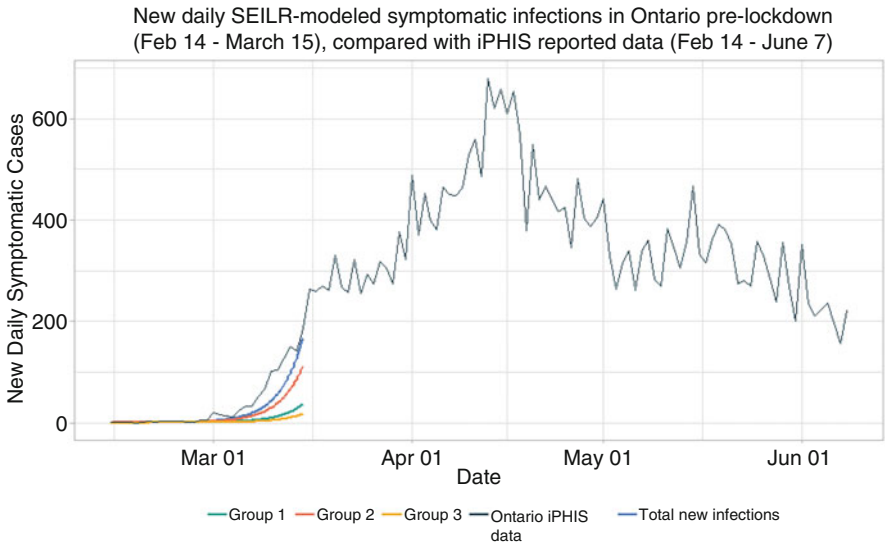


Fig. 2 Pre-lockdown iPHIS data (black) and simulated symptomatic infections in each of the 3 groups up to March 15. The blue curve represents symptomatic infected in the overall Ontario population

To do so, we introduce a dampening variable, denoted by $q_{int} \in [0, 1]$ during a given time interval and generically denoted by int, such that

$$\beta_{ij}(int) = c_{ij} \cdot p \cdot q_{int}, \quad \forall\, int \in interval \tag{2}$$

where q_{int} is constant during time interval int, $interval$ is the number of time intervals used in the model, and int is defined as the ceiling of the number of weeks since preventative measures are introduced. Since no preventative measures were present pre-lockdown (Feb 14th–March 15th), we will define this time period as $int\ 0 :=$ [Feb 14, March 15] and $q_0 := 1$.

We use again the Golden Section Search to fit our model to iPHIS data, but this time we fix $p = 0.045085$. We solve for q_{int} and obtain the following results (Fig. 3).

Our model yields 56,865 total cases from February 14th to October 5th, 2020, 3.2% fewer than the 58,716 cases reported by iPHIS during the same time period. Of the 58,716 cases reported by iPHIS: 9.01% were among Group 1, 64.1% were among Group 2, and 26.88% were among Group 3. Of the 56,865 cases modeled: 22.78% were among Group 1, 68.34% Group 2 and 10.59% Group 3. While our

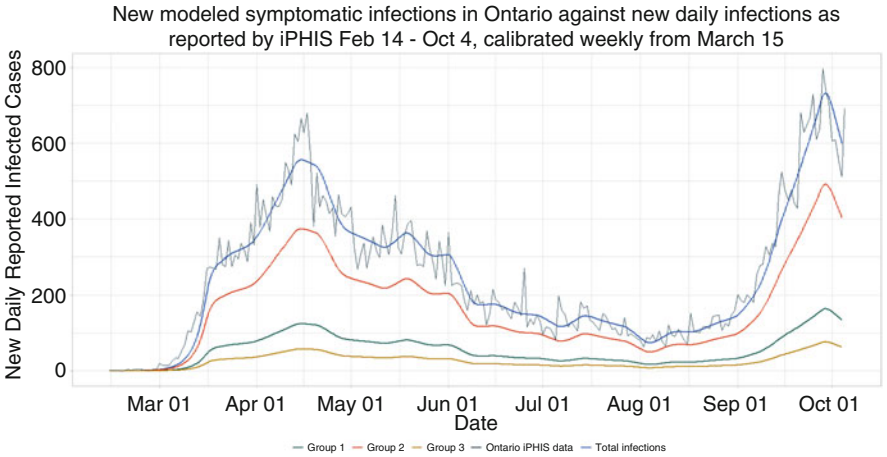


Fig. 3 Simulated new daily symptomatic infections in each of the 3 age groups according to the SEILR(pas) model

modeled Group 2 cases are very similar to the iPHIS-reported Group 2 cases, the model predicts too many Group 1 cases, and too few Group 3 cases. We can explain the excess of Group 1 cases over Group 3 since both have similar population sizes, by noticing that Group 1 has almost double the total daily contacts.

To better understand the behaviour of the epidemic curve, we estimate R_{eff}, the effective reproductive number of the virus. To estimate R_{eff} near the disease-free equilibrium pre-lockdown, we compute the spectral radius of our model's next generation matrix (which is a 12×12 matrix) while taking $q = 1$. Doing[2] so yields an $R_{eff} = 4.78$. Consequently, we find that with $p = 0.045085$ and $q = 0.20918$, $R_{eff} = 1$. This implies that 0.20918 is our threshold value for q, which we denote from here onward by $\bar{q}$: this value prevents total new cases from increasing. At q values above this threshold we expect the disease to proliferate, and at q values below this threshold we expect case numbers to decrease over time. For this reason, it is interesting to investigate the behaviour of q over time (Fig. 4). When comparing the placement of q relative to $\bar{q}$ to the behaviour of disease spread over the same time intervals, it can be observed that when $q < \bar{q}$, case numbers decrease, when $q > \bar{q}$, case numbers rise, and when $q \approx \bar{q}$, case numbers remain relatively consistent.

[2] We used MAPLE 19 to derive a closed form solution of the next generation matrix of our model.

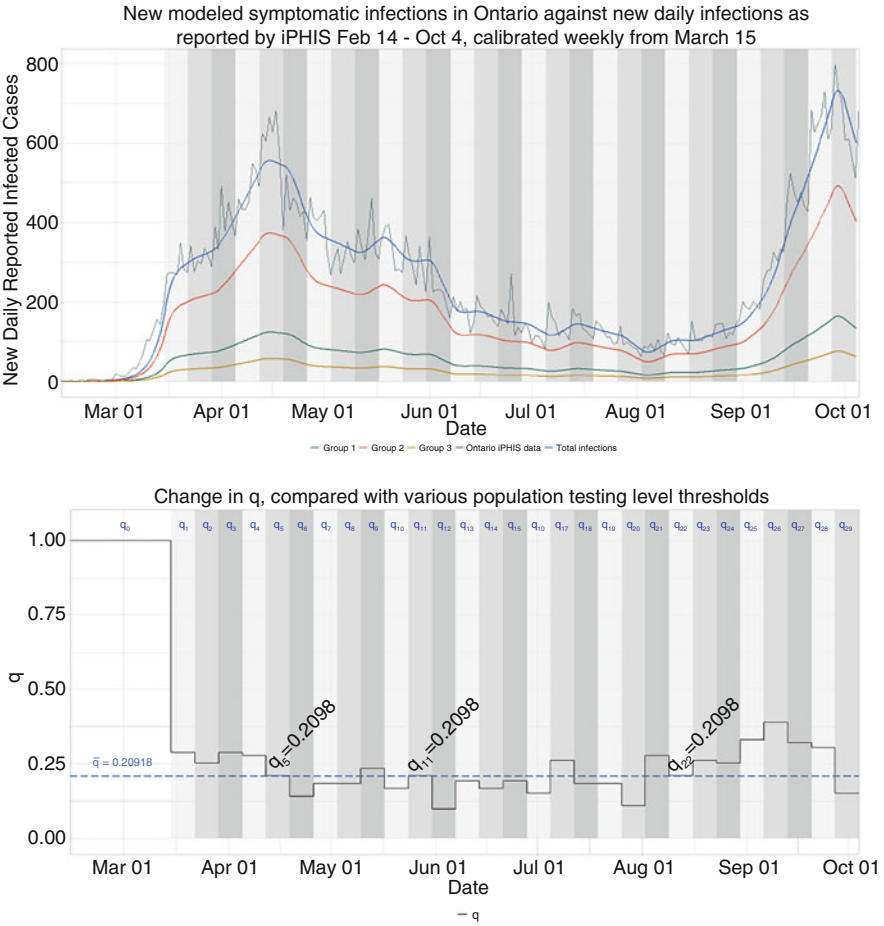


Fig. 4 Change in q over time as compared to threshold, $\bar{q} = 0.20918$

4 Preventative Testing and Its Effects on the Evolution of the Pandemic in ON

Our model considers the combination of physical distancing, mask-wearing, handwashing, and other contact-limiting measures as the standard for preventing the spread of COVID-19, as was common in the early stages of the pandemic without wide availability of diagnostic tests. By May 24, Ontario had opened municipal testing centers with sufficient resources to extend testing not only to essential workers and those with symptoms, but to suspected asymptomatic cases as well. To date,

however, voluntary testing levels still fall 36% short of provincial testing capacity (up to 50,000 tests daily), and reportedly fewer people are getting tested than before [35]. Supposing the testing uptake had been higher (e.g. due to employers incentives, more rigorous contact tracing, etc.) we investigate the potential effect of widespread testing and isolation on infection rates in Ontario.

We introduce a variation of our model in which we test δ of the population of Ontario every day (with replacement). We assume a uniform probability of being tested among all age groups. Thus, on any given day, each individual within the population has probability δ of being tested. If an individual tests positive they are removed from the population and asked to isolate (see new compartment L^T) for 13 days, the total number of days for the course of the virus in our model. We assume a uniform test accuracy of $z = 0.7$. [31]. Thus $z\delta$ of all infected individuals (those in compartments E, I^P, I^A, and I^S) will be removed from the population daily and placed in the L^T compartment where they isolate for 13 days and are thus unable to transmit the disease further. Figure 5 describes the new flow between compartments as introduced here.

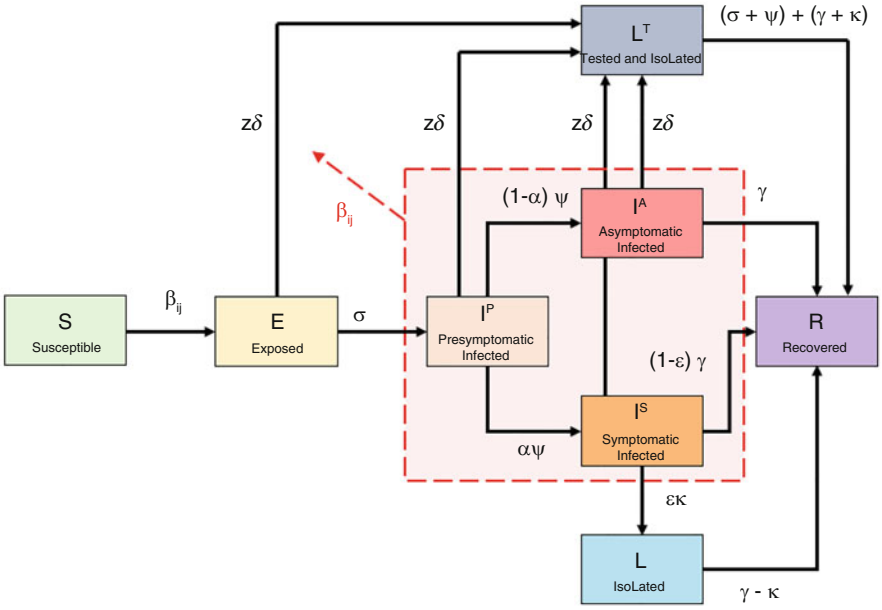


Fig. 5 Schematic of the SEILR(pas) model, now including uniform random population testing and isolation of infected cases via a new compartment L^T, for Group $i \in 1, 2, 3$

Similarly, the associated system of differential equations governing this modified model is given by:

$$
\begin{cases}
\dfrac{dS_i}{dt} = -\displaystyle\sum_{j=1}^{K} \beta_{ij} S_i(t) \frac{I_j^P(t) + I_j^A(t) + I_j^S(t)}{N_j(t)}, \\
\dfrac{dE_i}{dt} = \displaystyle\sum_{j=1}^{K} \beta_{ij} S_i(t) \frac{I_j^P(t) + I_j^A(t) + I_j^S(t)}{N_j(t)} - \sigma E_i(t) - z\delta E_i(t), \\
\dfrac{dI_i^P}{dt} = \sigma E_i(t) - \psi I_i^P(t) - z\delta I_i^P(t), \\
\dfrac{dI_i^A}{dt} = (1-\alpha)\cdot\psi\cdot I_i^P(t) - \gamma I_i^A(t) - z\delta I_i^A(t), \\
\dfrac{dI_i^S}{dt} = \alpha\cdot\psi\cdot I_i^P(t) - (1-\varepsilon)\gamma I_i^S(t) - \varepsilon\cdot\kappa\cdot I_i^S(t) - z\delta I_i^S(t), \\
\dfrac{dL_i}{dt} = \varepsilon\cdot\kappa\cdot I_i^S(t) - (\gamma-\kappa)L_i(t), \\
\dfrac{dL_i^T}{dt} = z\delta E_i(t) + z\delta I_i^P(t) + z\delta I_i^A(t) + z\delta I_i^S(t) - (\sigma+\omega-\gamma-\kappa)L_i^T, \\
\dfrac{dR_i}{dt} = \gamma(I_i^A(t) + (1-\varepsilon)I_i^S(t)) + (\gamma-\kappa)L_i(t) + (\sigma+\omega-\gamma-\kappa)L_i^T.
\end{cases}
$$

On May 24th, Ontario announced open testing to all Ontario residents, regardless of symptoms. We assume that by this date, Ontario had the ability to conduct as many tests as desired. We would like to investigate the scenario where the Government of Ontario actively conducted random tests instead of waiting for people to volunteer. As such, we introduce our new model on May 25th.

We run trials with 3 daily testing levels and compare them to the test-free model we simulated previously. These test rates are 0.25%, 0.5%, and 1% of the total Ontario population tested per day, beginning on May 25, with no daily random testing before this date. Over this 128 day time period, Ontario conducted a total of 4,415,630 tests, an average of 34,497 tests per day. While this does amount to a daily testing rate of 0.2368% of the population, this refers to voluntary and targeted testing, biased towards individuals who are more likely to have been exposed. The effects of this biased testing are encapsulated in the q-values obtained in the initial fitting of our model to Ontario data. The testing we are modeling here is uniform random testing that would serve supplementary to the testing that currently exists. In other words, we assume that the testing that currently occurs at a rate of 34,497 per day still exists, and the δ proportion of the population modeled here is an *additional* number of daily tests, taken with a uniform random distribution of individuals (Fig. 6).

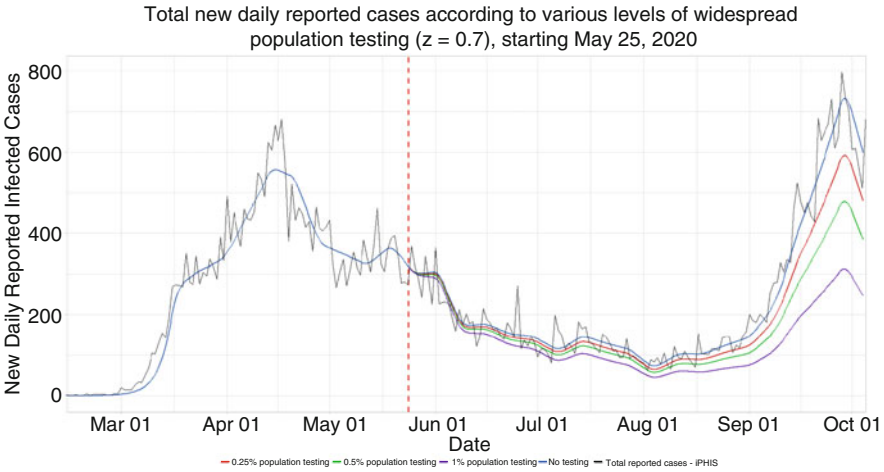


Fig. 6 Simulated Ontario case number trajectories at 0%, 0.25%, 0.5%, and 1% daily testing of the Ontario population starting from May 25. New daily reported infected cases decrease with increased population testing

Table 3 Summary of results for various simulated levels of testing performed on the Ontario population from May 25 to October 5

$\delta \times 10^2$ (daily testing rate) (% of pop.)	0	0.25	0.5	1
Daily tests	0	36,416	72,832	145,665
Total tests	0	4,843,328	9,686,656	19,373,445
Case total	29,834	26,246	23,230	18,541
Case difference from 0%	N/A	3588	6604	11,293
% decrease	N/A	0.12027	0.22136	0.37853
Case decrease/test	N/A	0.000740813	0.000681763	0.000582911

As expected, when increasing daily testing levels and isolating infected individuals, we see a clear decrease in new daily infected cases reported when increasing testing rates, as shown in Table 3:

As shown in Table 3, Ontario could have decreased total cases since May 25th by almost 40% by conducting an additional 145,665 tests per day (1% of population), roughly 5 times as many as were conducted daily over this time period. This level of testing results in a 5.82911×10^{-4} decrease in cases per test conducted, or one fewer case for every $1,715.5$ tests. Even by only conducting an additional 26,246 daily tests (0.25% of population), slightly more than double the current rate, we estimate Ontario could have avoided 3,588 cases, or 12.027% of total cases since May 25-th. At this testing rate, each test removes an average of 7.40813×10^{-4} cases, or one case for every $1,349.9$ tests. It is important to note there are diminishing returns on cases removed per test as testing rate increases, as the fraction of total cases removed each day is taken from a lower total number (Fig. 7).

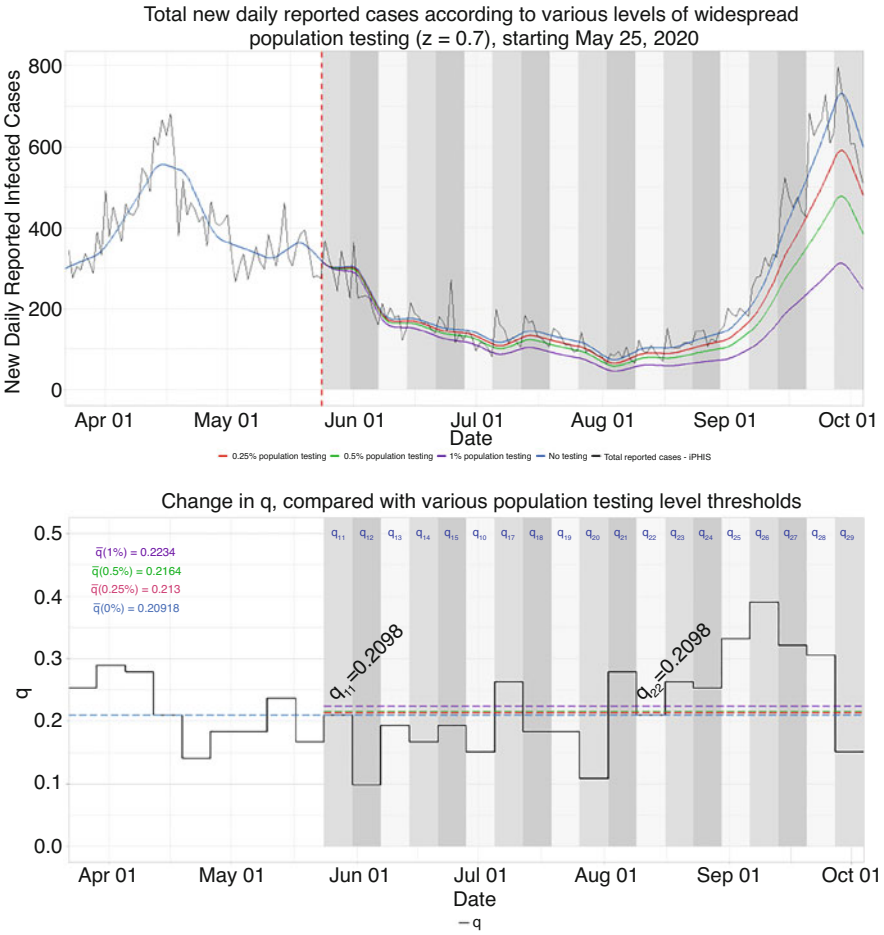


Fig. 7 The relationship between simulated Ontario case number trajectories at 0%, 0.25%, 0.5%, and 1% daily testing of the Ontario population starting from May 25 (upper panel), and the behaviour constant q alongside the respective $\bar{q}$ thresholds for each level of testing (0.20918, 0.213, 0.2164, 0.2234), representing the minimum reduction in contacts required to maintain $R_{eff} \leq 1$ (lower panel)

5 Discussion and Conclusions

Despite drastically lowering total cases, introducing random uniform testing at rates up to 1% of total population did not have a major impact on $\bar{q}$, the behaviour threshold for $R_{eff} \leq 1$. In order to lower $\bar{q}$ by a significant number, a large proportion of cases would have to be prevented altogether. For instance, at a 1% daily testing rate and a 13 day disease cycle, there is a $(1 - (0.7 \cdot 0.99)^{13}) = 8.727\%$ chance of each infected individual being removed from the population.

The theoretically most effective timepoint to minimize new cases via isolation of infected individuals is immediately upon contraction.

Thus, if all individuals removed due to random uniform testing (L^T) were instead removed from the system on day 1 of infection, it would have the same effect as scaling q down by a factor of 0.9127, allowing for a q value 9.56% higher for the same R_{eff}. Since under our model individuals are removed not only on day 1, but throughout the 13 day infected period, the effect of uniform random testing at 1% daily rate will be lesser than a hypothetical 9.56% threshold increase. Indeed, $\bar{q}(0\%) = 0.9363 \cdot \bar{q}(1\%)$, or a threshold increase of about 6.803%, a lesser effect than the ideal scenario in which $\bar{q}(0\%) = 0.9127 \cdot \bar{q}(1\%)$.

Unfortunately, testing is clearly bounded in its utility. In order to increase our threshold q such that $\bar{q} = 1$, we would need to implement a testing rate of approximately 38% of the remaining susceptible total population per day, a logistically daunting task. While testing can clearly help control outbreaks and reduce spread of COVID-19, it is currently limited to an aid in the control of COVID, and cannot by itself allow us to return to pre-COVID-19 behaviour.

The approval and advent of rapid-testing can dramatically improve the testing rate needed above. The 15 min response tests (by Abbott) or the ones developed in British Columbia and at the University of Guelph, Ontario, can be scaled up and used in formalized settings, such as: employers being able to test employees, schools, retails and restaurants could screen students and patrons, etc. This would be consistent with our previous work [23] where we assess that testing everyone every 4–6 days would suffice to control the pandemic (i.e., keep $R_{eff} \approx 1$) without the need of social distancing measures of various intensities. In our opinion, the cost of implementing rapid testing on a larger scale, in conjunction with PCR tests and public education on the value of preventative testing would compare favourably to the current economic losses incurred by playing "whack-a-mole" with regional pandemic outbreaks and public fatigue.

References

1. L. Chen, W. Liu, Q. Zhang, K. Xu, G. Ye, W. Wu, Z. Sun, F. Liu, K. Wu, B. Zhong, Y. Mei, W. Zhang, Y. Chen, Y. Li, M. Shi, K. Lan, and Y. Liu, "RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak," *Emerging Microbes & Infections*, vol. 9, no. 1, pp. 313–319, 2020. PMID: 32020836.
2. Johns Hopkins Coronavirus Resource Center, "COVID-19 Map."
3. W. Zhang, W. G. W. Zhao, D. Wu, and Y. Yang, "Predicting COVID-19 trends in Canada: A tale of four models," *Cognitive Computation and Systems*, May 2020.
4. L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of covid-19 in china by dynamical modeling," *arXiv preprint arXiv:2002.06563*, 2020.
5. B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao, and J. Wu, "Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions," *Journal of clinical medicine*, vol. 9, no. 2, p. 462, 2020.

6. C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, *et al.*, "The effectiveness of quarantine of Wuhan city against the corona virus disease 2019 (covid-19): A well-mixed SEIR model analysis," *Journal of medical virology*, 2020.
7. L. López and X. Rodo, "A modified SEIR model to predict the covid-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics," *Available at SSRN 3576802*, 2020.
8. J. Wu, B. Tang, N. L. Bragazzi, K. Nah, and Z. McCarthy, "Quantifying the role of social distancing, personal protection and case detection in mitigating COVID-19 outbreak in Ontario, Canada," *Journal of Mathematics in Industry*, vol. 10, May 2020.
9. A. R. Tuite, D. N. Fisman, and A. L. Greer, "Mathematical modelling of covid-19 transmission and mitigation strategies in the population of Ontario, Canada," *CMAJ*, vol. 192, no. 19, pp. E497–E505, 2020.
10. A. Asgary, M. G. Cojocaru, M. M. Najafabadi, and J. Wu, "Simulating preventative testing of sars-cov-2 in schools: Policy implications," *submitted to: BMC Public Health*, 2020.
11. N. M. Stall, W. Wu, L. Lapointe-Shaw, D. Fisman, M. Hillmer, and P. A. Rochon, "Sex-specific differences in COVID-19 testing, cases and outcomes: a population-wide study in Ontario, Canada," May 2020.
12. K. L. Schwartz, C. Achonu, S. A. Buchan, K. A. Brown, B. Lee, M. Whelan, J. H. Wu, and G. Garber, "COVID-19 infections among healthcare workers and transmission within households," June 2020.
13. R. Fields, L. Humphrey, M. Nahirniak, D. Flynn-Primrose, E. W. Thommes, and M. G. Cojocaru, "Age-stratified transmission model of covid-19 in ontario with human mobility," *submitted to: Heliyon Mathematics*, 2020.
14. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, *et al.*, "Social contacts and mixing patterns relevant to the spread of infectious diseases," *PLoS Med*, vol. 5, no. 3, p. e74, 2008.
15. K. Prem, A. R. Cook, and M. Jit, "Projecting social contact matrices in 152 countries using contact surveys and demographic data," *PLoS computational biology*, vol. 13, no. 9, p. e1005697, 2017.
16. Z. McCarthy, Y. Xiao, F. Scarabel, B. Tang, N. L. Bragazzi, K. Nah, J. M. Heffernan, A. Asgary, V. K. Murty, N. H. Ogden, *et al.*, "Quantifying the shift in social contact patterns in response to non-pharmaceutical interventions," *Journal of Mathematics in Industry*, vol. 10, no. 1, pp. 1–25, 2020.
17. Government of Ontario, "Reopening ontario," 2020.
18. J. Mossong, M. Jit, N. HENS, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, G. Scalia-Tomba, J. Wallinga, M. Sadkowska-Todys, *et al.*, "Social contact and mixing patterns relevant to the spread of infectious diseases: a multi-country population-based survey," Oxford Univ Press, 2007.
19. W. Lyra, J. D. do Nascimento, J. Belkhiria, L. de Almeida, P. P. Chrispim, and I. de Andrade, "COVID-19 pandemics modeling with SEIR (+ CAQH), social distancing, and age stratification. The effect of vertical confinement and release in Brazil," *medRxiv*, 2020.
20. A. Radulescu and K. Cavanagh, "Management strategies in a SEIR model of covid 19 community spread," *arXiv preprint arXiv:2003.11150*, 2020.
21. W. van der Hoek, J. A. Backer, R. Bodewes, I. Friesema, A. Meijer, R. Pijnacker, D. F. Reukers, C. Reusken, I. Roof, N. Rots, M. J. te Wierik, A. R. van Gageldonk-Lafeber, C. T. Waegemaekers, and S. van den Hof, "De rol van kinderen in de transmissie van SARS-CoV-2," *Ned Magazine Medical*, vol. 164, no. D5140, 2020.
22. Rijksinstituut voor Volksgezondheid en Milieu (National Institute for Public Health and the Environment), "Children and covid-19," 2020.
23. L. Humphrey, E. W. Thommes, R. Fields, N. Hakim, A. Chit, and M. G. Cojocaru, "A path out of covid-19 quarantine: an analysis of policy scenarios," *medRxiv*, 2020.
24. F. Brauer, "The Kermack–McKendrick epidemic model revisited," *Mathematical biosciences*, vol. 198, no. 2, pp. 119–131, 2005.

25. Government of Ontario, "News release: Ontario enacts declaration of emergency to protect the public."
26. K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, Yokohama, Japan, 2020," *Eurosurveillance*, vol. 25, Mar. 2020.
27. W. He, G. Y. Yi, and Y. Zhu, "Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: Meta-analysis and sensitivity analysis," *Journal of Medical Virology*, June 2020.
28. Public Health Ontario, "Covid-19 – what we know so far about. . . asymptomatic infection and asymptomatic transmission," 2020.
29. Statistics Canada, "Table 17-10-0005-01. population estimates on July 1st, by age and sex," 2017.
30. J. A. Backer, D. Klinkenberg, and J. Wallinga, "Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from Wuhan, China, 20–28 January 2020," *Eurosurveillance*, vol. 25, no. 5, p. 2000062, 2020.
31. L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler, "Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based sars-cov-2 tests by time since exposure," *Annals of Internal Medicine*, 2020.
32. D. Luenberger, *Linear and nonlinear programming*. New York: Springer, 2008.
33. J. Wu, B. Tang, N. L. Bragazzi, K. Nah, and Z. McCarthy, "Quantifying the role of social distancing, personal protection and case detection in mitigating covid-19 outbreak in Ontario, Canada," *Journal of Mathematics in Industry*, vol. 10, no. 1, pp. 1–12, 2020.
34. E. Abdollahi, M. Haworth-Brockman, Y. Keynan, J. M. Langley, and S. M. Moghadas, "Simulating the effect of school closure during covid-19 outbreaks in Ontario, Canada," *BMC medicine*, vol. 18, no. 1, pp. 1–8, 2020.
35. L. Powers, "Ontario reports 790 more covid-19 cases, testing still below target," *CBC News*, Oct 2020.

Sub-epidemic Model Forecasts During the First Wave of the COVID-19 Pandemic in the USA and European Hotspots

Gerardo Chowell, Richard Rothenberg, Kimberlyn Roosa, Amna Tariq, James M. Hyman, and Ruiyan Luo

1 Introduction

Throughout history, many of the large-scale infectious disease outbreaks, such as pandemic influenza (1918; 2009–2010), measles, and Ebola (2014–2016), have resulted in a single peak followed by a steady decline; therefore, many of the mathematical models commonly used to model outbreaks tend to follow this pattern. Other outbreaks, like HIV/AIDS and Ebola (2018–2020), however, result in more complicated trajectories that require models that can extend beyond the standard single-peak trends. The ongoing COVID-19 pandemic is accompanied by complicating factors in many countries, like delayed action for social distancing measures and premature relaxation of these measures. The sub-epidemic model presented here is flexible to these complex patterns, while remaining relatively simple. The model results and forecasts provide information for data scientists and policy makers to inform future decisions regarding relaxation of intervention measures or tighter controls for social distancing. Further, the model can infer the start of another outbreak wave from the data, which allows hospitals and health care settings to prepare for another potential increase of cases.

The asynchronicity of the infection patterns of the current coronavirus disease 2019 (COVID-19) pandemic illustrates the need for models that can capture complex dynamics beyond a single-peak trajectory to forecast the worldwide spread. This is also true for the spread within nations and within other sub-regions at various

G. Chowell (✉) · R. Rothenberg · K. Roosa · A. Tariq · R. Luo
Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA
e-mail: gchowell@gsu.edu

J. M. Hyman
Department of Mathematics, Center for Computational Science, Tulane University, New Orleans, LA, USA

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_5

geographic scales. The infections in these asynchronous transmission networks underlie the reported infection data and need to be accounted for in forecasting models.

We analyze the COVID-19 pandemic assuming that the total number of new infections is the sum of all the infections created in multiple asynchronous outbreaks at differing spatial scales. We assume there are weak ties across sub-populations, and we represent the overall epidemic as an aggregation of *sub-epidemics*, rather than a single, universally connected outbreak. The sub-epidemics can start at different time points and affect different segments of the population in different geographic areas. Thus, we model sub-epidemics associated with transmission chains that are asynchronously triggered and that progress somewhat independently from the other sub-epidemics.

Jewell et al. [1] review the difficulties associated with long-term forecasting of the ongoing COVID-19 pandemic using statistical models that are not based on transmission dynamics. They also describe the limitations of models that use established mortality curves to calculate the pace of growth, the most likely inflection point, and subsequent diminution of the epidemic. The review analyzes the need for broad uncertainty bands, particularly for sub-national estimates. It also addresses the unavoidable volatility of both reporting and estimates based on reports. The analysis, delivered in the spirit of caution rather than remonstration, implies the need for other approaches that depend on overall transmission dynamics or large-scale agent-based simulations. Our sub-epidemic approach addresses this need in both the emerging and endemic stages of an epidemic.

This approach is analogous to the model used by Blower et al. [2] to demonstrate how the rise and endemic leveling of tuberculosis outbreaks could be explained by dynamical changes in the transmission parameters. A related multi-stage approach was used by Garnett [3] to explain the pattern of spread for sexually transmitted diseases and changes in the reproductive number during the course of an epidemic. Rothenberg et al. [4] demonstrated that the national curve of Penicillinase-Producing Neisseria gonorrhoeae occurrence resulted from multiple asynchronous outbreaks.

As with HIV/AIDS, which has now entered a phase of intractable endemic transmission in some areas [5], COVID-19 is likely to become endemic. New vaccines and pharmacotherapy might mitigate the transmission, but the disease will not be eradicated in the foreseeable future. Some earlier predictions based on mathematical models predicted that COVID-19 would soon disappear or approach a very low-level endemic equilibrium determined by herd immunity. To avoid unrealistic medium-range projections, some investigators artificially truncate the model projections before the model reaches these unrealistic forecasts.

Here, we demonstrate a five-parameter sub-epidemic wave modeling framework that provides a simple characterization of unfolding trajectories of COVID-19 epidemics that are progressing across the world at different spatial scales [6]. We systematically assess calibration and forecasting performance for the ongoing COVID-19 pandemic in hotspots located in the USA and Europe using the sub-epidemic wave model, and we compare results with those obtained using the

Richards model, a well-known three-parameter single-peak growth model [7]. The sub-epidemic approach captures the rise to an initial peak followed by a wide range of post-peak behavior, ranging from a typical decline to a steady incidence level to repeated small waves for sub-epidemic outbreaks. This framework yields excellent short- and intermediate-term forecasts that are not attainable with other single-peak transmission models of similar complexity, whether mechanistic or phenomenological. We illustrate how this view of the epidemic could help data scientists and policymakers better understand and predict the underlying transmission dynamics of COVID-19.

2 Methods

2.1 Country-Level Data

We retrieved daily reported cumulative case data of the COVID-19 pandemic for France, the United Kingdom (UK), and the United States of America (USA) from the World Health Organization (WHO) website [8] and for Spain and Italy from the corresponding governmental websites [9, 10] from early February to May 24, 2020. We calculated the daily incidence from the cumulative trajectory and analyzed the incidence trajectory for the 5 countries.

2.2 State-Level US Data

We also retrieved daily cumulative case count data from The COVID Tracking Project [11] from February 27, 2020 to May 24, 2020 for five representative COVID-19 hotspot states in the USA, namely New York, Louisiana, Georgia, Arizona and Washington.

3 Methodology Overview: Parameter Estimation and Short-term Forecasts with Quantified Uncertainty

3.1 Parameter Estimation

Given a model, parameter estimation is the process of finding the parameter values and their uncertainty that best explain empirical data. In this section we briefly describe the parameter quantification method described in refs. [12]. First, we define the general form of a dynamic model composed by a system of h ordinary differential equations as follows:

$$\begin{aligned}
\dot{x}_1(t) &= g_1\left(x_1, x_2, \ldots, x_h, \Theta\right) \\
\dot{x}_2(t) &= g_2\left(x_1, x_2, \ldots, x_h, \Theta\right) \\
&\vdots \\
\dot{x}_h(t) &= g_h\left(x_1, x_2, \ldots, x_h, \Theta\right)
\end{aligned}$$

where $\dot{x_i}$ denotes the rate of change of the system state x_i where $i = 1, 2, \ldots, h$ and $\Theta = (\theta_1, \theta_2, \ldots, \theta_m)$ is the set of model parameters.

To calibrate dynamic models describing the trajectory of epidemics, researchers require temporal data for one or more states of the system (e.g., daily number of new outpatients, inpatients and deaths). In this paper, we consider the case with only one state of the system

$$\dot{x} = g\left(x, \Theta\right)$$

The temporal resolution of the data typically varies according to the time scale at which relevant processes operate (e.g., daily, weekly, yearly) and the frequency at which the state of the system is measured. We denote the time series of n longitudinal observations of the single state by

$$y_{t_i} = y_{t_1}, y_{t_2}, \ldots, y_{t_n}$$

where $i = 1, 2, \ldots, n$where t_i are the time points of the time series data and n is the number of observation time points. Let $f(t, \Theta)$ denote the mean of observed incidence series y_t over time, which corresponds to $\dot{x}(t)$if $x(t)$denotes the cumulative count at time t. Usually the incidence series $y_{t_i}$is assumed to have a Poisson distribution with mean $\dot{x}(t)$or a negative binomial distribution if over-dispersion is present. Modeling an error structure using a negative binomial distribution would require the estimation of the over-dispersion coefficient. In the same way we estimate other model parameters, one would need to estimate the extra parameter from data.

Model parameters are estimated by fitting the model solution to the observed data via nonlinear least squares [13]. This is achieved by searching for the set of parameters $\hat{\Theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m\right)$that minimizes the sum of squared differences between the observed data $y_{t_i} = y_{t_1}, y_{t_2} \ldots . y_{t_n}$and the model mean which corresponds to $f(t_i, \Theta)$. That is, $\Theta = (\theta_1, \theta_2, \ldots, \theta_m)$ is estimated by $\hat{\Theta} = \arg\min \sum_{i=1}^{n} \left(f\left(t_i, \Theta\right) - y_{t_i}\right)^2$.

Then, $\hat{\Theta}$ is the parameter set that yields the smallest differences between the data and model. This parameter estimation method gives the same weight to all of the data points. This method does not require a specific distributional assumption for y_t, except for the first moment $E[y_t] = f(t_i; \Theta)$; meaning, the mean at time t is

equivalent to the count (e.g., number of cases) at time t [14]. Moreover, this method yields asymptotically unbiased point estimates regardless of any misspecification of the variance-covariance error structure. Hence, the model mean $f\left(t_i, \hat{\Theta}\right)$ yields the best fit to observed data y_{t_i} in terms of squared L2 norm. More generally, this objective function can be extended to simultaneously fit more than one state variable to their corresponding observed time series.

If we assume a Poisson error structure in the data, the parameters can be estimated via maximum likelihood estimation (MLE). Consider the probability density function (PDF) that specifies the probability of observing data y_t given the parameter set Θ, or $f(y_t|\,\Theta)$; given a set of parameter values, the PDF can show which data are more probable, or more likely [14]. MLE aims to determine the values of the parameter set that maximizes the likelihood function, where the likelihood function is defined as $L(\Theta|\,y_t) = f(y_t|\,\Theta)$ [14, 15]. The resulting parameter set is called the MLE estimate, the most likely to have generated the observed data. Specifically, the MLE estimate is obtained by maximizing the corresponding log-likelihood function. For count data with variability characterized by the Poisson distribution, we utilize Poisson-MLE [16, 17], where the log-likelihood function is given by:

$$L\left(\Theta|y_{t_i}\right) = \sum_{i=1}^{n}\left[y_{t_i}\,\log\left(f\left(t_i;\Theta\right)\right) - f\left(t_i;\Theta\right)\right]$$

and the Poisson-MLE estimate is expressed as

$$\hat{\Theta}{=}argmax\sum_{i=1}^{n}\left[y_{t_i}\,\log\left(f\left(t_i;\Theta\right)\right) - f\left(t_i;\Theta\right)\right]$$

.

In Matlab, we can use the *fmincon* function to set the optimization problem.

To quantify parameter uncertainty, we follow a parametric bootstrapping approach which is particularly powerful, as it allows the computation of standard errors and related statistics in the absence of closed-form formulas [18, 19]. As previously described in ref. [12], we generate S replicates from the best-fit model $f\left(t_i, \hat{\Theta}\right)$ by assuming an error structure in the data (e.g., Poisson) in order to quantify the uncertainty of the parameter estimates and construct confidence intervals. Specifically, using the best-fit model $f\left(t_i, \hat{\Theta}\right)$, we generate S-times replicated simulated datasets, where the observation at time $t_i$is sampled from the Poisson distribution with mean $f\left(t_i, \hat{\Theta}\right)$. Next, we refit the model to each of the S simulated datasets to re-estimate parameters for each of the S-simulated realizations. The new parameter estimates for each realization are denoted by $\hat{\Theta}_i$where $i = 1, 2, \ldots, S$. Using the sets of re-estimated parameters $\left(\hat{\Theta}_i\right)$,it is possible to characterize the empirical distribution of each estimate and construct confidence intervals for each parameter. Moreover, the resulting uncertainty around

the model fit is given by $f\left(t,\hat{\Theta}_1\right)$, $f\left(t,\hat{\Theta}_2\right),\ldots,f\left(t,\hat{\Theta}_S\right)$. It is worth noting that a Poisson error structure is the most common for modeling count data where the mean of the distribution equals the variance. In situations where the time series data show over-dispersion, a negative binomial distribution can be employed instead [12].

3.2 *Model-Based Forecasts with Quantified Uncertainty*

Forecasting from a given model $f\left(t,\hat{\Theta}\right)$, h units of time ahead is straight forward: $f\left(t+h,\hat{\Theta}\right)$. We can use the RMSE to quantify the forecasting performance of the models. The uncertainty of the forecasted value can be obtained using the previously described parametric bootstrap method. Let

$$f\left(t+h,\hat{\Theta}_1\right),f\left(t+h,\hat{\Theta}_2\right),\ldots,f\left(t+h,\hat{\Theta}_S\right)$$

denote the forecasted value of the current state of the system propagated by a horizon of h time units, where $\hat{\Theta}_i$ denotes the estimation of parameter set Θ from the ith bootstrap sample. We can calculate the bootstrap variance of the estimates to measure the uncertainty of the forecasts, and the 2.5% and 97.5% percentiles to construct 95% confidence intervals.

4 Generalized Growth Model (GGM)

The generalized growth model (GGM) is a simple model that characterizes the early ascending phase of the epidemic. Previous studies have highlighted the occurrence of early sub-exponential growth patterns in various infectious disease outbreaks. This model allows for the relaxation of exponential growth by modulating a "scaling of growth parameter", p, which allows the model to capture a range of epidemic growth profiles [20]. The GGM is given by the following differential equation:

$$\frac{dC(t)}{dt}=C'(t)=rC(t)^p$$

In this equation $C^{'}(t)$ describes the incidence curve over time t, $C(t)$ describes the cumulative number of cases at time t, $p\in[0,1]$ is a "deceleration or scaling of growth" parameter and r is the growth rate. This model represents constant incidence over time if $p=0$ and exponential growth for cumulative cases if $p=1$. If p is in the range $0<p<1$, then the model indicates sub-exponential or polynomial growth dynamics [20–25].

5 Generalized Logistic Growth Model

The generalized logistic growth model (GLM) is an extension of the simple logistic growth model that allows for capturing a range of epidemic growth profiles, including sub-exponential and exponential growth dynamics. The GLM characterizes epidemic growth through the intrinsic growth rate r, a dimensionless "deceleration of growth" parameter p, and the final epidemic size, k_0. The deceleration parameter modulates the epidemic growth patterns including sub-exponential growth ($0 < p < 1$), constant incidence ($p = 0$) and exponential growth dynamics ($p = 1$). The GLM is given by the following differential equation:

$$\frac{dC(t)}{dt} = rC^p(t)\left(1 - \frac{C(t)}{k_0}\right)$$

where $\frac{dC(t)}{dt}$ describes the incidence over time t, and the cumulative number of cases at time t is given by (t) [25–28].

6 Richards Growth Model

The Richards growth model is also an extension of the simple logistic growth model and relies on three parameters. It extends the simple logistic growth model by incorporating a scaling parameter, a, that measures the deviation from the symmetric simple logistic growth curve [7, 21, 29]. The Richards model is given by the differential equation:

$$\frac{dC(t)}{dt} = rC(t)\left[1 - \left(\frac{C(t)}{k_0}\right)^a\right]$$

where $c(t)$ represents the cumulative case count at time t, r is the growth rate, a is a scaling parameter and k_0 is the final epidemic size.

6.1 Sub-epidemic Wave Modeling Motivation

The concept of weak ties was originally proposed by Granovetter in 1973 [30] to form a connection between microevents and macro events. We use this idea to link the person-to-person viral transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to the trajectory of the COVID-19 epidemic. The transient connection between two people with different personal networks that results in the transference of the virus between the networks can be viewed as a weak tie. This event can cause asynchronous epidemic curves within the overall

network. The events can spread the infection between sub-populations defined by neighborhoods, zip codes, counties, states, or countries. The resulting epidemic curve can be modeled as the sum of asynchronous sub-epidemics that reflect the movement of the virus into new populations.

In the absence of native immunity, specific viricidal treatment, or a working vaccine, our non-pharmacological preventive tools—testing, contact tracing, social separation, isolation, lockdown—are the key influences on sub-epidemic spread [31–34]. The continued importation of new cases will result in low-level endemic transmission. A model based on sub-epidemic events can forecast the level of endemic spread at a steady state. This can then be used to guide intervention efforts accounting for the continued seeding of new infections.

6.2 Sub-epidemic Modeling Approach

We use a five-parameter epidemic wave model that aggregates linked overlapping sub-epidemics [6]. The strength (e.g., weak vs. strong) of the overlap determines when the next sub-epidemic is triggered and is controlled by the onset threshold parameter, C_{thrs}. The incidence defines a generalized-logistic growth model (GLM) differential equation for the cumulative number of cases, C_t, at time t:

$$\frac{dC(t)}{dt} = rC^p(t)\left(1 - \frac{C(t)}{K_0}\right).$$

Here, r is the fixed growth rate, and p *is* the scaling of growth parameter, and K_0 is the final size of the initial sub-epidemic. The growth rate depends on the parameter . If $p = 0$, then the early incidence is constant over time, while if $p = 1$ then the early incidence grows exponentially. Intermediate values of $(0 < p < 1)$ describe early sub-exponential (e.g. polynomial) growth patterns.

The sub-epidemics are modeled by a system of coupled differential equations:

$$\frac{dC_i(t)}{dt} = rA_{i-1}(t)C_i(t)^p\left(1 - \frac{C_i(t)}{K_i}\right),$$

Here $C_i(t)$ is the cumulative number of infections for sub-epidemic i, and K_i is the size of the i_{th} sub-epidemic where $i = 1, \ldots, n$. Starting from an initial sub-epidemic size K_0, the size of consecutive sub-epidemics K_i decline at the rate q following an exponential or power-law function.

The onset timing of the $(i + 1)_{th}$ sub-epidemic is determined by the indicator variable $A_i(t)$. This results in a coupled system of sub-epidemics where the $(i + 1)_{th}$ sub-epidemic is triggered when the cumulative number of cases for the i_{th} sub-epidemic exceeds a total of C_{thr} cases. The sub-epidemics are *overlapping* because the C_{thr} sub-epidemic takes off before the i_{th} sub-epidemic completes its course. That is,

$$A_i(t) = \begin{cases} 1 & C_i(t) > C_{thr} \\ 0 & Otherwise \end{cases}, \quad i = 1, 2, \ldots, n-1$$

The threshold parameters are defined so $1 \leq C_{th} < K_0$ and $A_0(t) = 1$ for the first sub-epdemic.

This framework allows the size of the i_{th} sub-epidemic (K_i) to remain steady or decline based on the factors underlying the transmission dynamics. These factors could include a gradually increasing effect of public health interventions or population behavior changes that mitigate transmission. We consider both exponential and inverse decline functions to model the size of consecutive sub-epidemics.

6.3 Exponential Decline of Sub-epidemic Sizes

If consecutive sub-epidemics decline exponentially, then K_i is given by:

$$K_i = K_0 e^{-q(i-1)}$$

where K_0 is the size of the initial sub-epidemic ($K_1 = K_0$). If $q = 0$, then the model predicts an epidemic wave comprising sub-epidemics of the same size. When $q > 0$, then the total number of sub-epidemics n_{tot} is finite and depends on C_{thr}, q, and K_0 The sub-epidemic is only triggered if $C_{thr} \leq K_i$, resulting in a finite number of sub-epidemics,

$$n_{tot} = \left\lfloor -\frac{1}{q} \ln\left(\frac{C_{thr}}{K_0}\right) + 1 \right\rfloor$$

The brackets $\lfloor * \rfloor$ denote the largest integer that is smaller than or equal to *. The total size of the epidemic wave composed of n_{tot} overlapping sub-epidemics has a closed-form solution:

$$K_{tot} = \sum_{i=1}^{n_{tot}} K_0 e^{-q(i-1)} = \frac{K_0\left(1 - e^{-qn_{tot}}\right)}{1 - e^{-q}}$$

6.4 Inverse Decline of Sub-epidemic Sizes

The consecutive sub-epidemics decline according to the inverse function given by:

$$K_i = K_0 \left(\frac{1}{i}\right)^q.$$

When $q > 0$, then the total number of sub-epidemics n_{tot} is finite and is given by:

$$n_{tot} = \left[\left(\frac{C_{thr}}{K_0} \right)^{-\frac{1}{q}} \right].$$

The total size of an epidemic wave is the sum of n overlapping sub-epidemics,

$$K_{tot} = \sum_{i=1}^{n_{tot}} K_0 \left(\frac{1}{i} \right)^q.$$

In the absence of control interventions or behavior change ($q = 0$), the total epidemic size depends on a given number n of sub-epidemics,

$$K_{tot} = nK_0.$$

The initial number of cases is given by $C_1(0) = I_0$ where I_0 is the initial number of cases in the observed case data. The cumulative cases, $C(t)$, is the sum of all cumulative infections over the n overlapping sub-epidemics waves:

$$C_{tot}(t) = \sum_{i=1}^{n} C_i(t)$$

6.5 Parameter Estimation

Fitting the model to the time series of case incidence requires estimating up to five model parameters $\Theta = (C_{thr}, q, r, p, K)$. If a single sub-epidemic is sufficient to fit the data, then the model is simplified to the three-parameter generalized-logistic growth model. The model parameters were estimated by a nonlinear least square fit of the model solution to the observed incidence data [13]. This is achieved by searching for the set of parameter $\hat{\Theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m \right)$ that minimizes the sum of squared differences between the observed incidence data $y_{t_i} = y_{t_1}, y_{t_2}, \ldots, y_{t_N}$ and the corresponding mean incidence curve denoted by $f(t_i, \Theta)$. That is, the parameters are estimated by

$$\hat{\Theta} = \text{argmin} \sum_{i=1}^{N} \left(f\left(t_i, \Theta\right) - y_{t_i} \right)^2$$

where t_i are the time points at which the time-series data are observed, and N is the number of data points available for inference. Hence, the model solution $f\left(t_i, \hat{\Theta}\right)$ yields the best fit to the time series data y_{t_i} where $\hat{\Theta}$ is the vector of parameter estimates.

We solve the nonlinear least squares problem using the trust-region reflective algorithm. We used parametric bootstrap, assuming an error structure described in the next section, to quantify the uncertainty in the parameters obtained by a nonlinear least squares fit of the data, as described in refs. [12, 35]. Our best-fit model solution is given by $f\left(t, \hat{\Theta}\right)$ where $\hat{\Theta}$ is the vector of parameter estimates. Our MATLAB (The MathWorks, Inc) code for model fitting along with outbreak datasets is publicly available [36].

The confidence interval for each estimated parameter and 95% prediction intervals of the model fits were obtained using parametric bootstrap [12]. Let S denote the number of bootstrap realizations and $\hat{\Theta}_i$ denote the re-estimation of parameter set Θ from the i^{th} bootstrap sample. The variance and confidence interval for $\hat{\Theta}$ are estimated from $\hat{\Theta}_1, \ldots, \hat{\Theta}_S$. Similarly, the uncertainty of the model forecasts, $f\left(t, \hat{\Theta}\right)$, is estimated using the variance of the parametric bootstrap samples

$$f\left(t, \hat{\Theta}_1\right), f\left(t, \hat{\Theta}_2\right), \ldots, f\left(t, \hat{\Theta}_s\right).$$

where $\hat{\Theta}_i$ denotes the estimation of parameter set Θ from the i_{th} bootstrap sample. The 95% prediction intervals of the forecasts in the examples are calculated from the 2.5% and 97.5% percentiles of the bootstrap forecasts.

6.6 Error Structure

We model a negative binomial distribution for the error structure and assume a constant variance/mean ratio over time (i.e., the overdispersion parameter). To estimate this constant ratio, we group every four daily observations into a bin across time, calculate the mean and variance for each bin, and then estimate a constant variance/mean ratio by calculating the average of the variance/mean ratios over these bins. Exploratory analyses indicate that this ratio is frequently stable across bins, except for 1–2 extremely large values, which could result from a sudden increase or decrease in the number of reported cases. These sudden changes could result from changes in case definition or a weekend effect whereby the number of reported cases decreases systematically during weekends. Hence, these extreme large values of variance/mean ratio are excluded when estimating the constant variance/mean ratio.

6.7 Model Calibration and Forecasting Approach

For each of the ten regions, we analyzed six weekly sequential forecasts, conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020, and assessed the calibration and forecasting performances at increasing time horizons of 2, 4, 6, ..., and 20 days ahead. The models were sequentially re-calibrated each week using the most up-to-date daily curve of COVID-19 reported cases. That is, each sequential forecast included one additional week of data than the previous forecast. For comparison, we also generated forecasts using the Richards model, a well-known single-peak growth model with three parameters [7, 37].

6.8 Model Performance

To assess both the quality of the model fit and the short-term forecasts, we used four performance metrics: the mean absolute error (MAE), the mean squared error (MSE), the coverage of the 95% prediction intervals, and the mean interval score (MIS) [38]. The *mean absolute error* (MAE) is given by:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{N}\left| f\left(t_i, \hat{\Theta}\right) - y_{t_i}\right|$$

Here y_{t_i} is the time series of incident cases describing the epidemic wave where t_i are the time points of the time series data [39]. Similarly, the *mean squared error* (MSE) is given by:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{N}\left(f\left(t_i, \hat{\Theta}\right) - y_{t_i}\right)^2$$

In addition, we assessed the *coverage of the 95% prediction interval*, e.g., the proportion of the observations that fell within the 95% prediction interval as well as a metric that addresses the width of the 95% prediction interval as well as coverage via the *mean interval score* (MIS) [38, 40] which is given by:

$$\text{MIS} = \frac{1}{h}\sum_{i=1}^{h}\left(U_{t_i} - L_{t_i}\right) + \frac{2}{0.05}\left(L_{t_i} - y_{t_i}\right) I\left\{y_{t_i} < L_{t_i}\right\} + \frac{2}{0.05}\left(y_{t_i} - U_{t_i}\right) I\left\{y_{t_i} > U_{t_i}\right\}$$

where L_t and U_t are the lower and upper bounds of the 95% prediction interval and I{} is an indicator function. Thus, this metric rewards for narrow 95% prediction intervals and penalizes at the points where the observations are outside the bounds specified by the 95% prediction interval where the width of the prediction interval adds up to the penalty (if any) [38].

The mean interval score (MIS) and the coverage of the 95% prediction intervals take into account the uncertainty of the predictions whereas the mean absolute error (MAE) and mean squared error (MSE) only assess the closeness of the mean trajectory of the epidemic to the observations [41]. These performance metrics have also been adopted in international forecasting competitions [40].

For comparison purposes, we compare the performance of the sub-epidemic wave model with that obtained from the 3-parameter Richards model [7], a well-known single-peak growth model given by:

$$\dot{C}(t) = rC(t)\left(1 - \left(\frac{C(t)}{K}\right)^{\theta}\right),$$

where θ determines the deviation from symmetry, and again r is the growth rate, and K is the final epidemic size.

7 Results

7.1 Model Parameters and Calibration Performance

A five-parameter dynamic model, postulating sub-exponential growth in linked sub-epidemics, captures the aggregated growth curve in diverse settings (Figs. 1, 2, and 3 and Figs. S3, S4, S5, S6, S7, S8, and S9). Using national-level data from five countries, we estimate the initial sub-exponential growth parameter (p) with a mean ranging from 0.7 to 0.9. Our analysis of five representative hotspot states in the USA indicates that early growth was sub-exponential in New York, Arizona, Georgia, and Washington (mean p ~ 0.5–0.9) and exponential in Louisiana (Table 1). Moreover, the rate of sub-epidemic decline that captures the effects of interventions and population behavior changes is shown in Fig. S10. The decay rate was fastest for Italy, followed by France, with the lowest decline rate in the USA (Table 1). Within the USA, the decline rate was the fastest for New York and Louisiana and more gradual for Georgia and Washington (Fig. S10).

The calibration performance across all regions presented in Figs. S1 and S2 is substantially better for the overlapping sub-epidemic model compared to the Richards model based on each of the performance metrics (for MAE, MSE, and MIS, smaller is better; for 95% PI coverage, larger is better). An informative example of the model fit to the trajectory of the COVID-19 epidemic in Spain (Fig. 1) shows the early growth of the epidemic in a single large sub-epidemic followed by a smaller sub-epidemic (blue in row 2, column 2 of Fig. 1), which is then followed by a much smaller sub-epidemic (green). In row 1 (Fig. 1), the parameter distributions demonstrate relatively small confidence intervals. Thus, the model captures a common phenomenon in epidemic situations: an initial steep rise, followed by a leveling or decline, then a second rise, and a subsequent repeat of

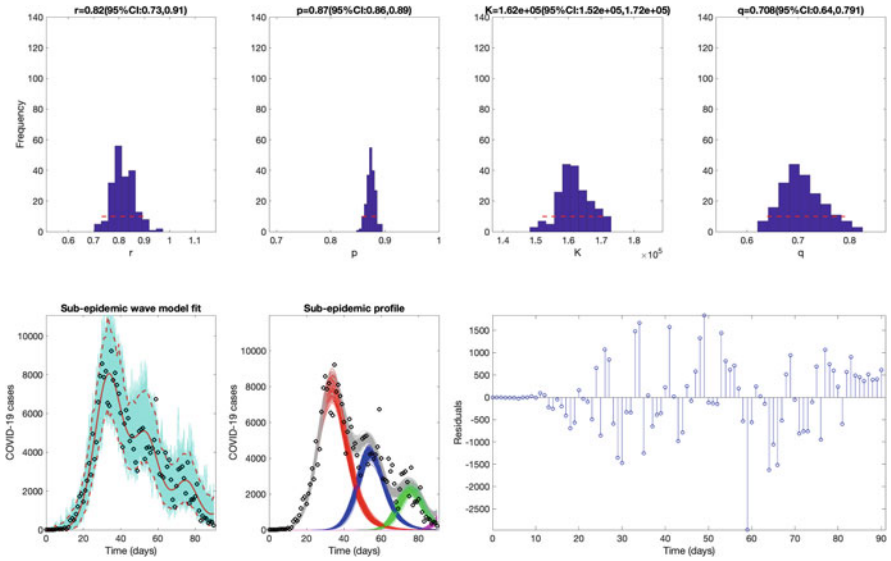


Fig. 1 The best fit of the sub-epidemic model to the COVID-19 epidemic in Spain. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of estimated parameters. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

the same pattern. A somewhat different pattern is observed in the USA, which experienced sustained transmission with high mortality for a long period (Fig. 2). A single epidemic wave failed to capture the early growth phase and the later leveling off; whereas, the aggregation of multiple sub-epidemics produces a better fit to the observed dynamics. In comparison, New York, the early epicenter of the pandemic in the USA, displays a similar sub-epidemic profile, while the sub-epidemic sizes decline at a much faster rate (Fig. 3).

Similar composite figures for the remaining regions (Figs. S3, S4, S5, S6, S7, S8, and S9) demonstrate diverse patterns of underlying sub-epidemic waves. For example, Italy experienced a single peak, largely the result of an initial sub-epidemic (in red), that was quickly followed by several rapidly declining sub-epidemics that slowed the downward progression (Fig. S3). The UK's sub-epidemic profile resembles that of the USA, but the sub-epidemics decline at a faster rate (Fig. S5; Table 1).

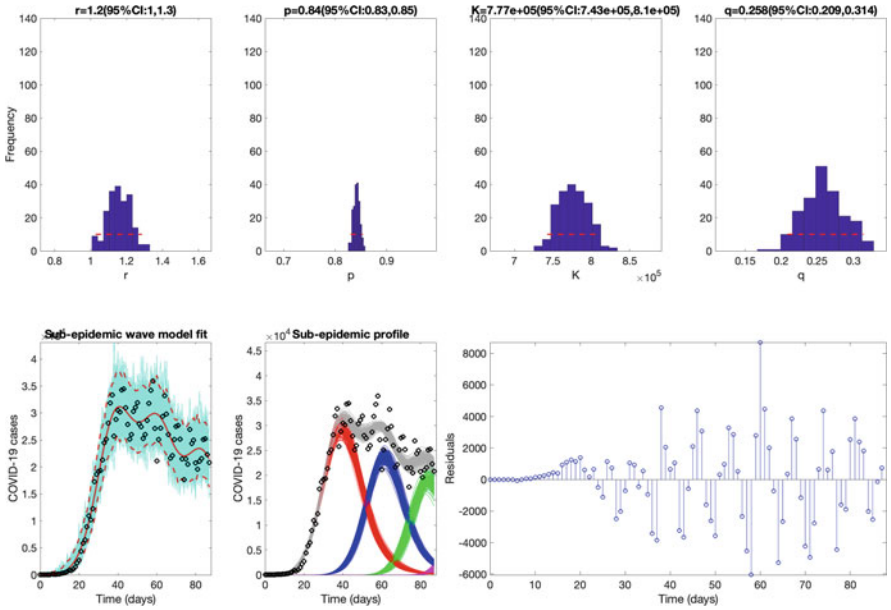


Fig. 2 The best fit of the sub-epidemic model to the COVID-19 epidemic in the USA. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of estimated parameters. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

7.2 Forecasting Performance

The sub-epidemic wave model outperformed the simpler Richards model in most of the 2-20 day ahead forecasts (see Fig. 4 and Figs. S11, S12, S13, S14, S15, S16, S17, S18, and S19). We observe that the sub-epidemic model forecasting accuracy increases as evidence for the second sub-epidemic appears in the data. For instance, the initial forecasts for the USA using the sub-epidemic model (Figs. 5 and S20) underestimate reported incidence for the 20 days after April 7th, which is likely attributable to the unexpected leveling off of the epidemic wave. However, this model provided more accurate forecasts in subsequent 20-day forecasts.

Similarly, sub-national models of the USA state trajectories confirm the general findings of fit and 20-day forecasting (see supplementary materials). Among the most striking of these is the sub-epidemic structure modeled for New York state (Fig. S25). When the sub-epidemic model is calibrated by April 7, 2020, a single sub-epidemic is observed; however, subsequent weeks of data helped infer an under-

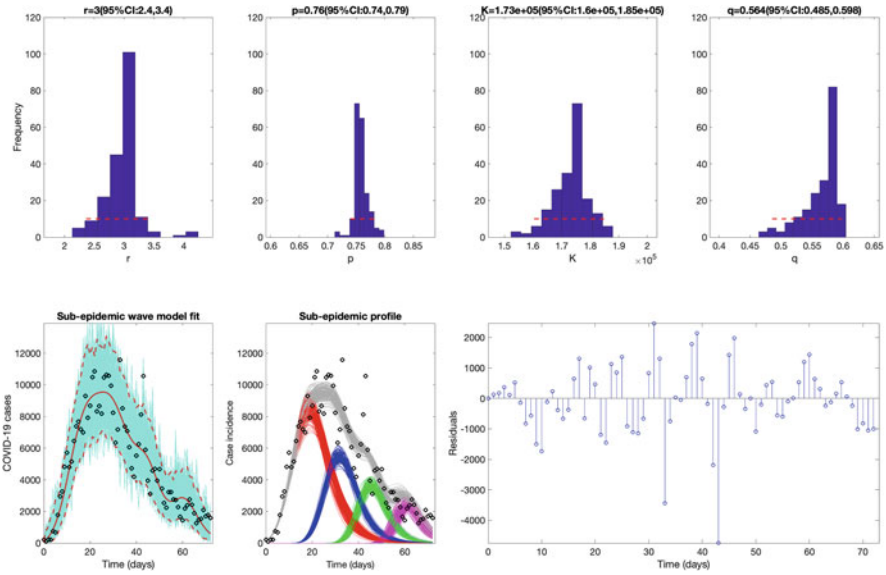


Fig. 3 The best fit of the sub-epidemic model to the COVID-19 epidemic in New York State. The sub-epidemic wave model successfully captures the overlapping sub-epidemic growth pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of estimated parameters. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

lying overlapping sub-epidemic structure and correctly forecasted the subsequent downward trend. With variation, other states shown in the supplementary materials provided similar confirmation of the method.

8 Discussion

Our sub-epidemic modeling framework is based on the premise that the aggregation of regular sub-epidemic dynamics can determine the shape of the trajectory of epidemic waves observed at larger spatial scales. This framework has been particularly suitable for forecasting the spatial wave dynamics of the COVID-19 pandemic, where the trajectory of the epidemic at different spatial scales does not display a single peak followed by a "burnout" period, but instead follows more complex transmission patterns including leveling off, plateaus, and long-tail decline periods. The model overwhelmingly outperformed a standard growth model that only allows for single-peak transmission dynamics. Model parameters also inform

Table 1 Mean and corresponding 95% confidence intervals of estimated parameters for each COVID-19 hotspot using data up to May 24, 2020

Region	Growth rate (r)	Scaling of growth (p)	Initial sub-epidemic size (K0)	Decline rate (q)	Decline function
Italy	2.5 (2.2, 2.3)	0.69 (0.66, 0.71)	162000 (134000, 182000)	2.1 (1.61, 2.51)	Power law
Spain	0.82 (0.73, 0.91)	0.87 (0.86, 0.89)	162000 (152000, 172000)	0.71 (0.64, 0.79)	Exponential
France	0.72 (0.43, 1.4)	0.87 (0.79, 0.92)	101000 (83000, 120000)	1.91 (1.42, 2.47)	Power law
UK	0.74 (0.64, 0.88)	0.84 (0.82, 0.86)	166000 (152000, 179000)	0.53 (0.38, 0.68)	Exponential
USA	1.2 (1.0, 1.3)	0.84 (0.83, 0.85)	777000 (743000, 810000)	0.26 (0.21, 0.31)	Exponential
New York	3 (2.4, 3.4)	0.76 (0.74, 0.79)	173000 (160000, 185000)	0.56 (0.49, 0.60)	Exponential
Louisiana	0.37 (0.35, 0.40)	0.99 (0.97, 1.0)	18400 (15900, 21300)	1.20 (0.96, 1.47)	Power law
Georgia	0.56 (0.43, 0.96)	0.88 (0.77, 0.93)	13900 (11100, 17300)	0.20 (0.00003, 0.60)	Power law
Arizona	3.3 (1.9, 5.1)	0.54 (0.47, 0.63)	16900 (13300, 25000)	0.26 (0.0004, 1.2)	Exponential
Washington	0.93 (0.76, 1.2)	0.79 (0.74, 0.85)	4330 (3360, 5390)	0.46 (0.21, 0.64)	Power law

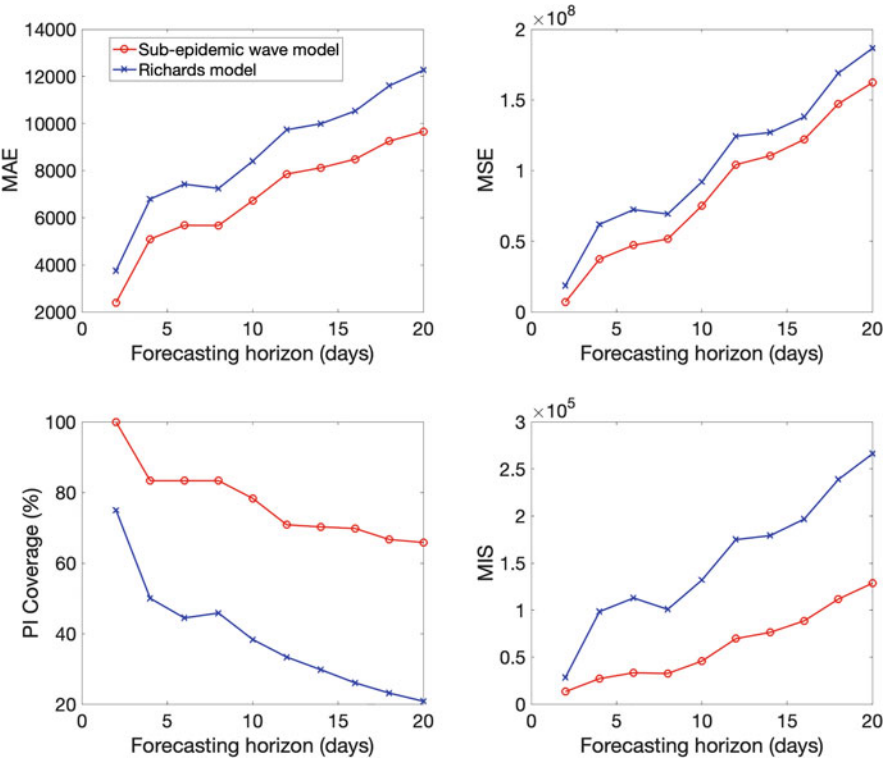


Fig. 4 Mean performance of the sub-epidemic wave (red) and the Richards (blue) models in 2-20 day ahead forecasts conducted during the epidemic in the USA. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons.

the effect of interventions and population behavior changes in terms of the sub-epidemic decay rate.

Overall, this approach predicts that a relaxation of the tools currently at our disposal—primarily aimed at preventing person-to-person and person-to-surface contact—would result in continuing sub-epidemics and ongoing endemic transmission. If we add widespread availability of testing, contact tracing, and cluster investigation (e.g. nursing homes, meatpacking plants, and other sites of unavoidable congregation), early suppression of sub-epidemics may be possible. The United States leads in the total number of tests performed, but it is currently ranked 25th among all nations in testing per capita [42]. The sub-epidemic description of COVID-19 transmission provides a rationale for substantial increases in testing.

Parsimony in model construction is not an absolute requirement, but it has several advantages. With fewer parameters to estimate, the joint simulations are more efficient and more understandable. Degenerate results are more easily avoided, and, when properly constructed, confidence intervals for the key parameters are more constrained. In our projections, we fit five parameters to the data:

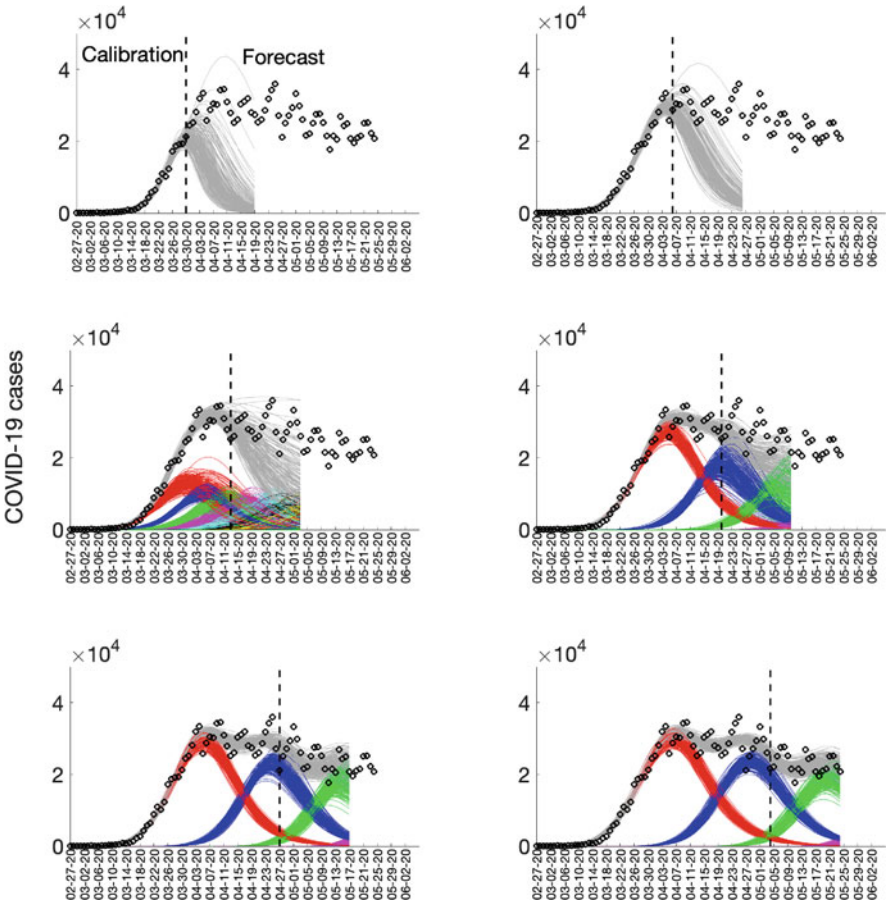


Fig. 5 Sub-epidemic profiles of the sequential 20-day ahead forecasts for the COVID-19 epidemic in the USA. Different colors represent different sub-epidemics of the epidemic wave profile. The aggregated trajectories are shown in gray, and black circles correspond to the data points. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

1. The onset threshold parameter, C_{thr}, that triggers the onset of a new sub-epidemic and determines if the overlap is weak or strong,
2. The new epidemic starting size, K_0,
3. The size of consecutive sub-epidemics decline rates q,
4. The positive parameter r denoting the growth rate of a sub-epidemic, and
5. The "scaling of growth" parameter $p \in [0,1]$(exponential or sub-exponential).

As shown in Fig. 1, the confidence limits for these parameters are narrow, and the scaling of growth parameter is constantly in the 0.8–0.9 range (Table 1).

Short-term forecasting is an important attribute of the model. Though long-term forecasts are of value, their dependability varies inversely with the time horizon. The 20-day forecasts are most valuable for the monitoring, management, and relaxation of the social distancing requirements. The early detection of potential sub-epidemics can signal the need for strict distancing controls, and the reports of cases can identify the geographic location of incubating sub-epidemics. No single model or method can provide an unerring approach to epidemic control. The multiplicity of models now available can be viewed as a source of confusion, but it is better thought of as a strength that provides multiple perspectives [43–45]. The sub-epidemic approach adds to the current armamentarium for guiding us through the COVID-19 pandemic.

Acknowledgments

Funding: GC was supported by grants NSF 1414374 as part of the joint NSF-National Institutes of Health NIH-United States Department of Agriculture USDA Ecology and Evolution of Infectious Diseases program; UK Biotechnology and Biological Sciences Research Council [grant BB/M008894/1] and RAPID NSF 2026797.

Author contributions: GC conceived the study. KR and AT contributed to data analysis. All authors contributed to the interpretation of the results. GC and RR wrote the first draft of the manuscript. All authors contributed to writing subsequent drafts of the manuscript. All authors read and approved the final manuscript.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data are publicly available.

Appendix

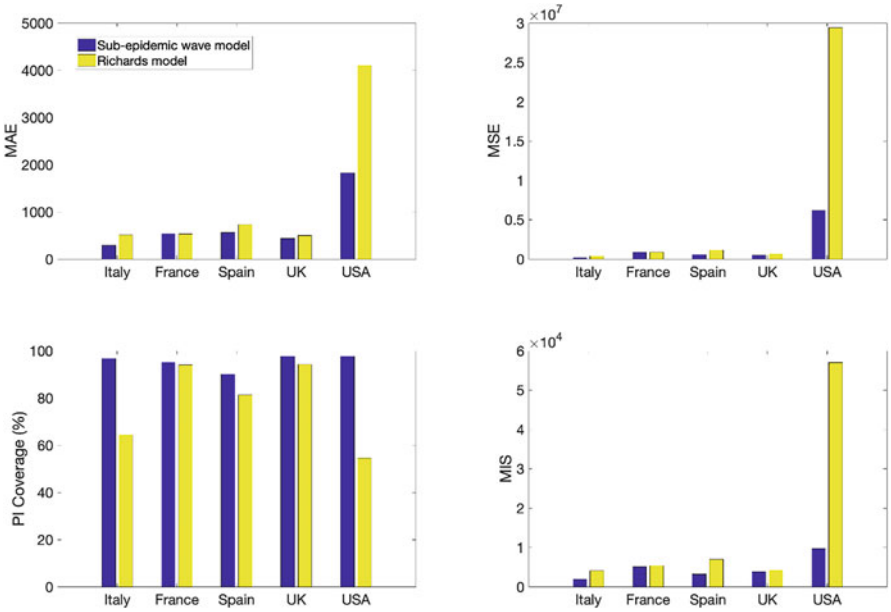


Fig. S1 The calibration performance metrics across five countries are uniformly better for the overlapping sub-epidemic models (for MAE, MSE, and MIS, smaller is better; for % covered, larger is better)

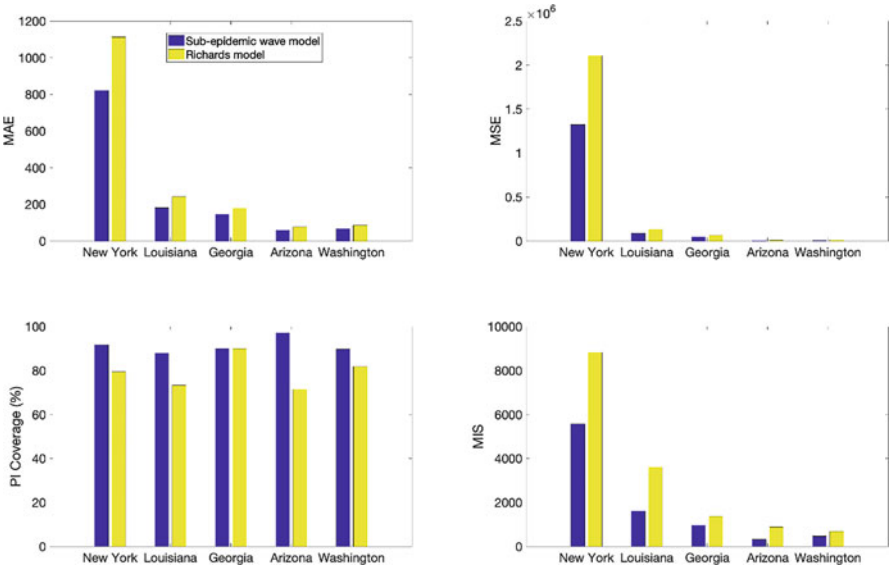


Fig. S2 The calibration performance metrics across five hotspots in the USA are uniformly better for the overlapping sub-epidemic models (for MAE, MSE, and MIS, smaller is better; for % covered, larger is better)

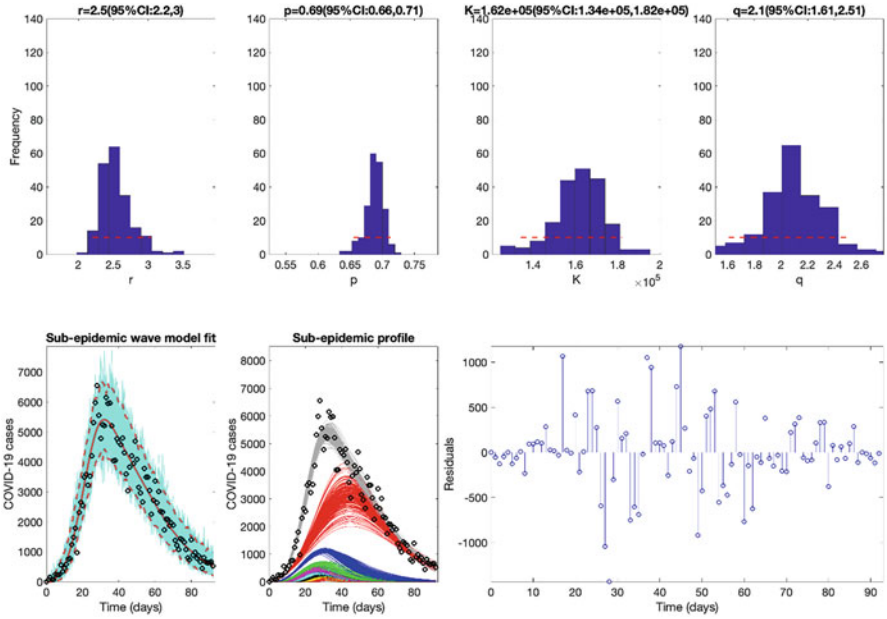


Fig. S3 The best fit of the sub-epidemic model to the COVID-19 epidemic in Italy. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

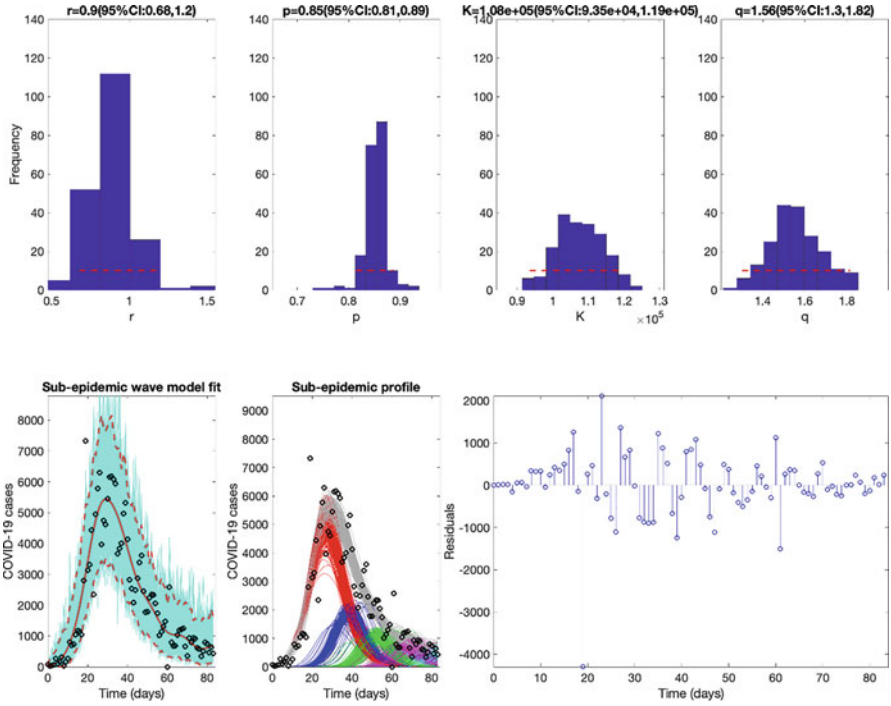


Fig. S4 The best fit of the sub-epidemic model to the COVID-19 epidemic in France. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

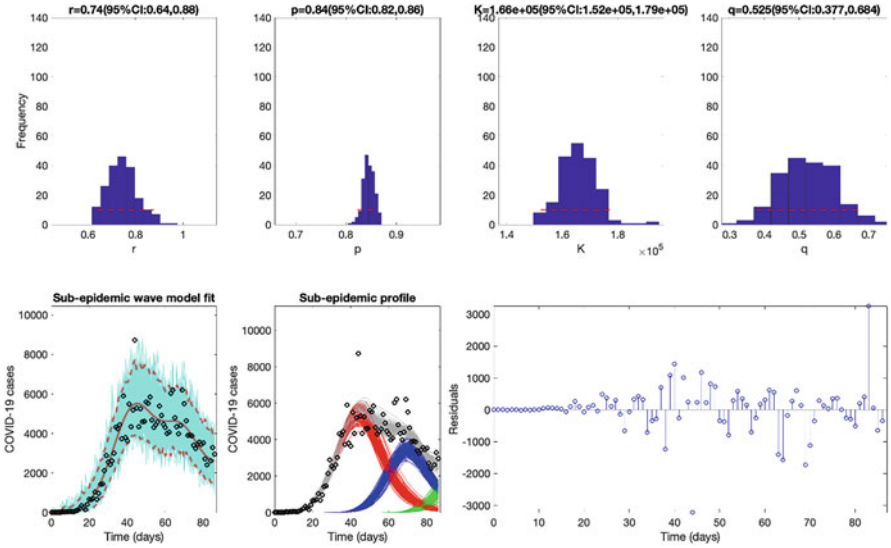


Fig. S5 The best fit of the sub-epidemic model to the COVID-19 epidemic in the United Kingdom. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

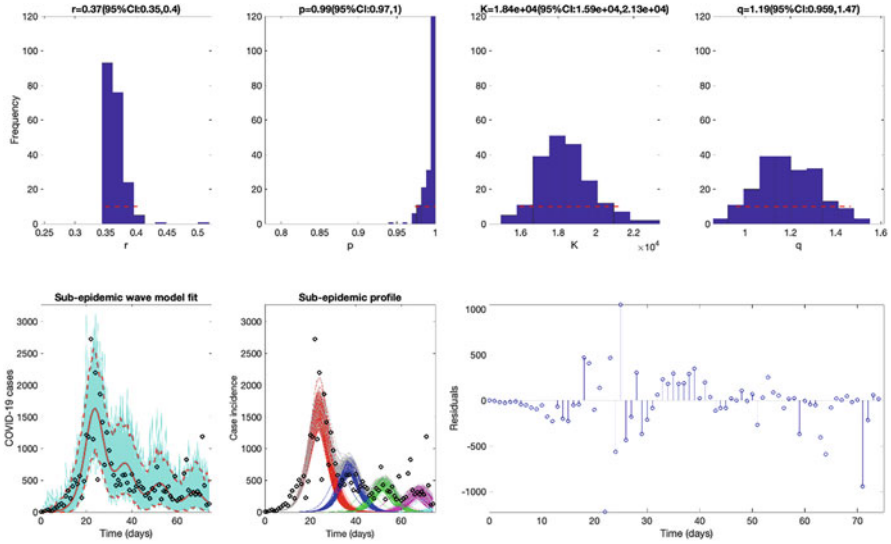


Fig. S6 The best fit of the sub-epidemic model to the COVID-19 epidemic in Louisiana, USA. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

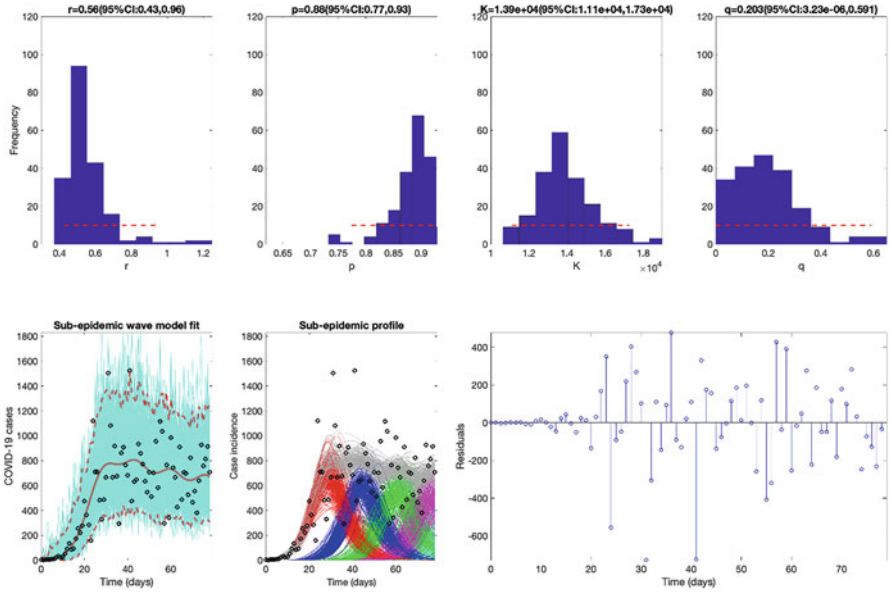


Fig. S7 The best fit of the sub-epidemic model to the COVID-19 epidemic in Georgia, USA. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

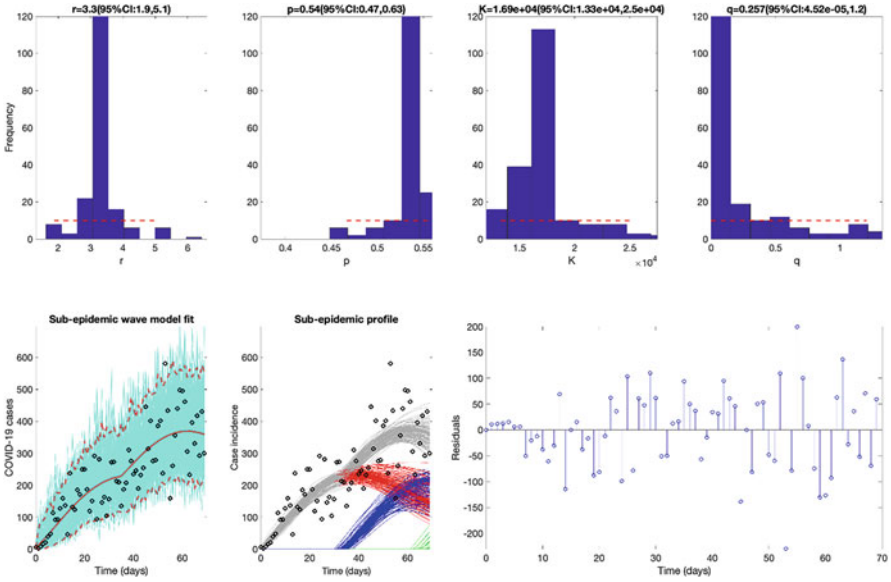


Fig. S8 The best fit of the sub-epidemic model to the COVID-19 epidemic in Arizona, USA. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different colors

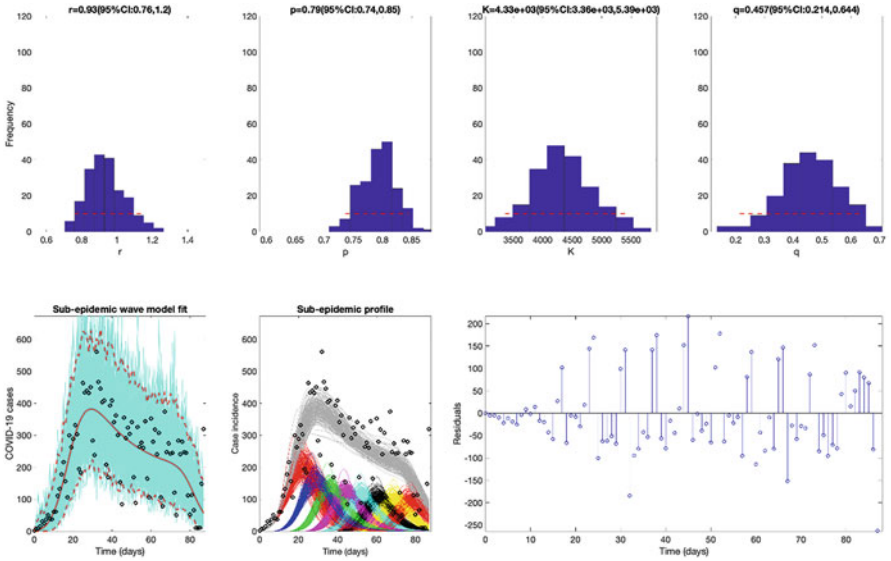


Fig. S9 The best fit of the sub-epidemic model to the COVID-19 epidemic in Washington. The sub-epidemic wave model successfully captures the multimodal pattern of the COVID-19 epidemic. Further, parameter estimates are well identified, as indicated by their relatively narrow confidence intervals. The top panels display the empirical distribution of r, p, K, and q. Bottom panels show the model fit (left), the sub-epidemic profile (center), and the residuals (right). Black circles correspond to the data points. The best model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. Cyan curves are the associated uncertainty from individual bootstrapped curves. Three hundred realizations of the sub-epidemic waves are plotted using different color

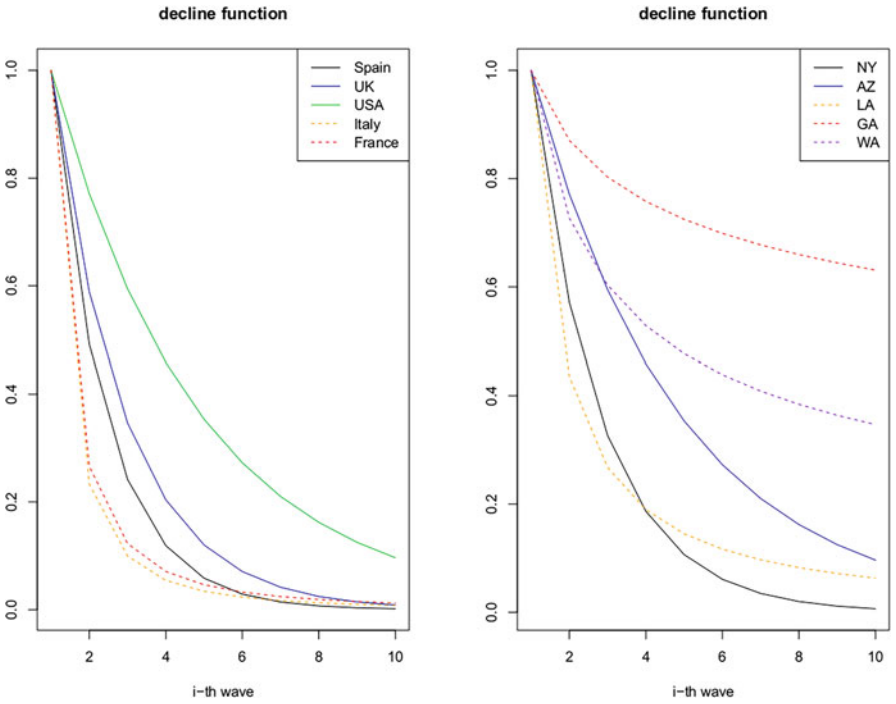


Fig. S10 The sub-epidemic decline function across countries and USA states based on results presented in Table 1

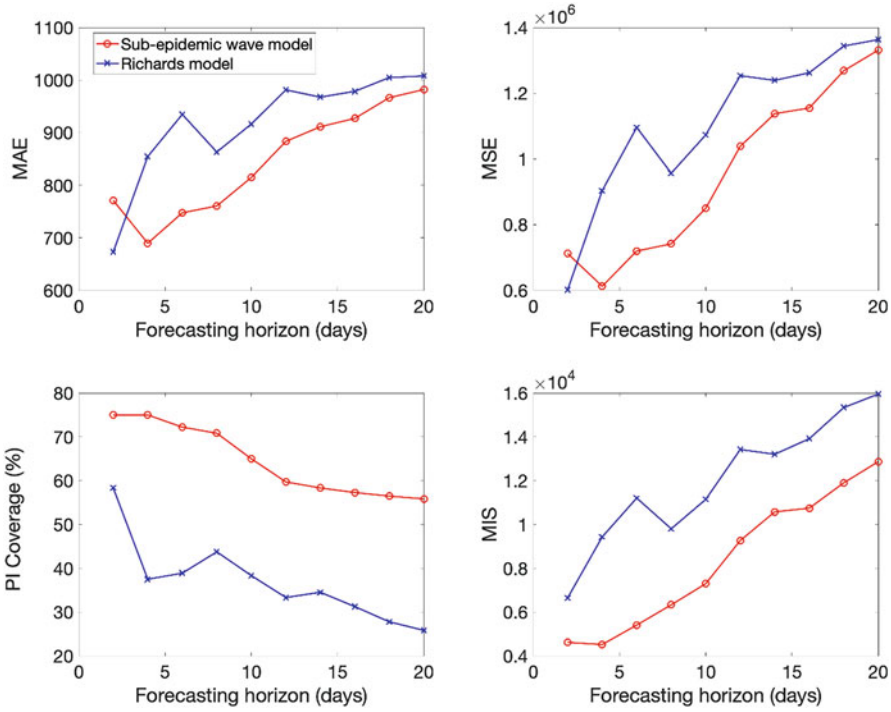


Fig. S11 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in Italy. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons except for 2-day ahead forecasts based on the MAE and the MSE

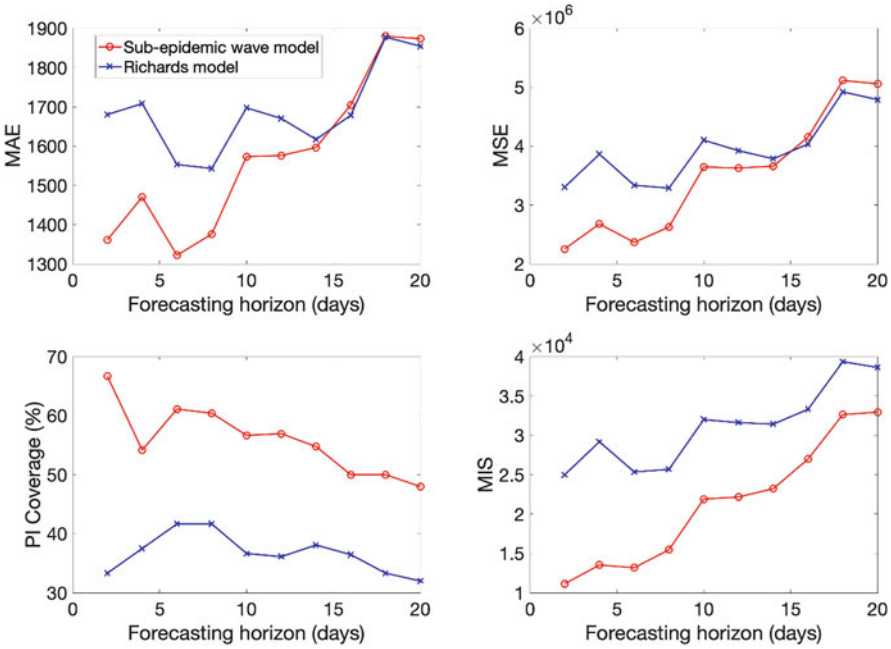


Fig. S12 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in Spain. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons, but the MSE and MAE reached similar values at longer forecasting horizons

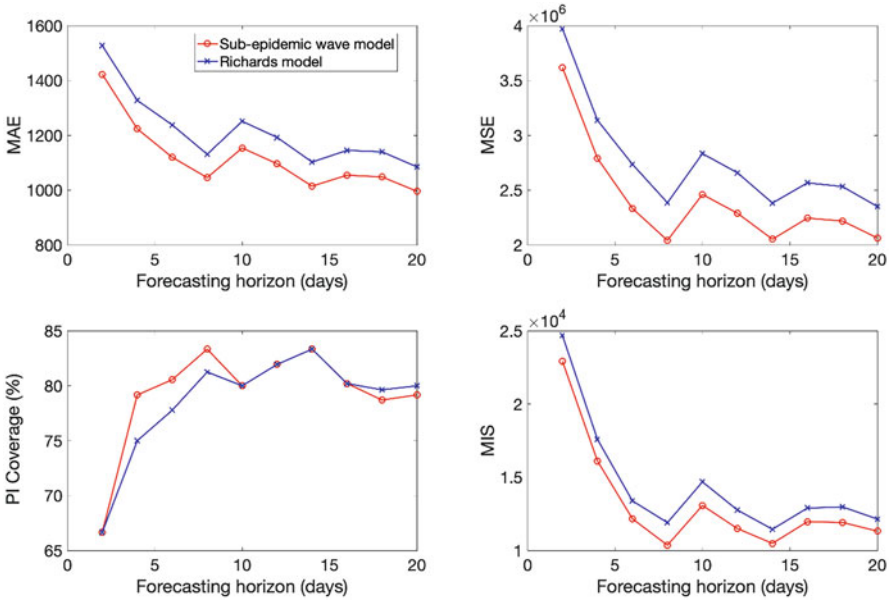


Fig. S13 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in France. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons

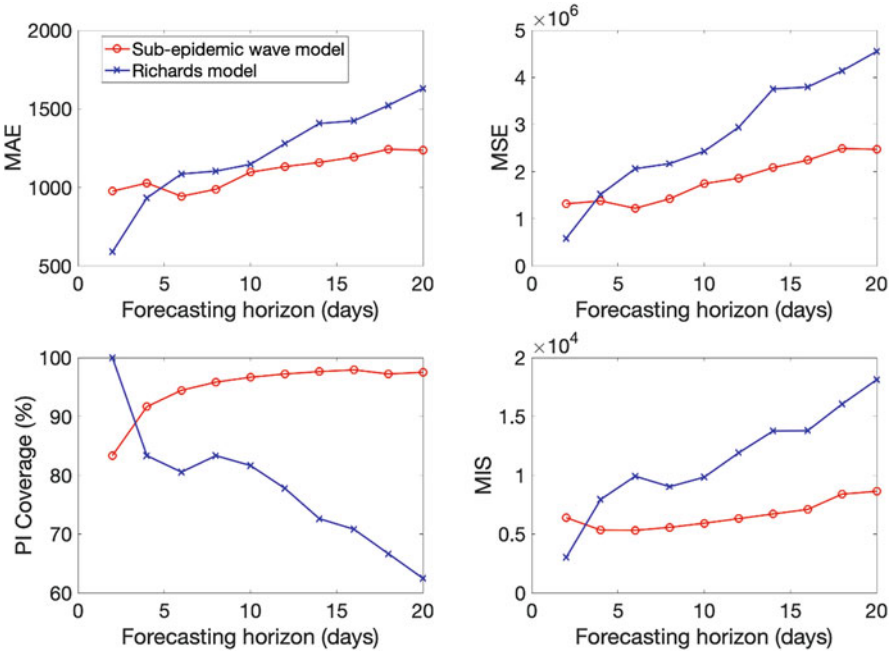


Fig. S14 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in the UK. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons except for 2-day ahead forecasts for which the Richards model reached somewhat better performance

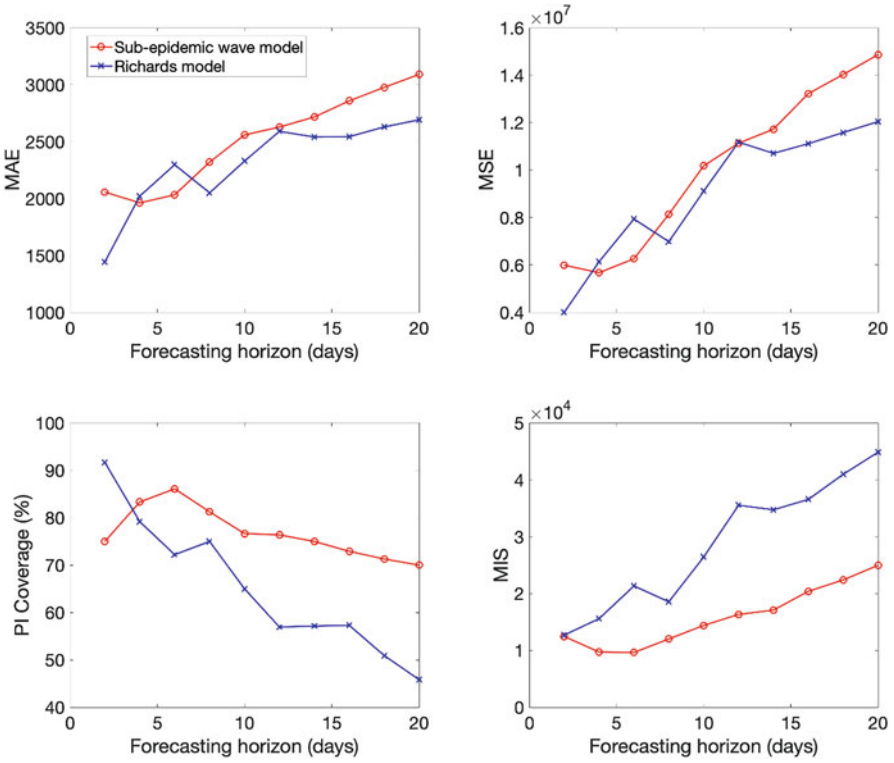


Fig. S15 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in New York. The sub-epidemic model outperformed the Richards model across all forecasting horizons based on the PI Coverage and the MIS except for 2-day ahead forecasts. However, the Richards model more frequently outperformed the sub-epidemic wave model based on the MAE and MSE

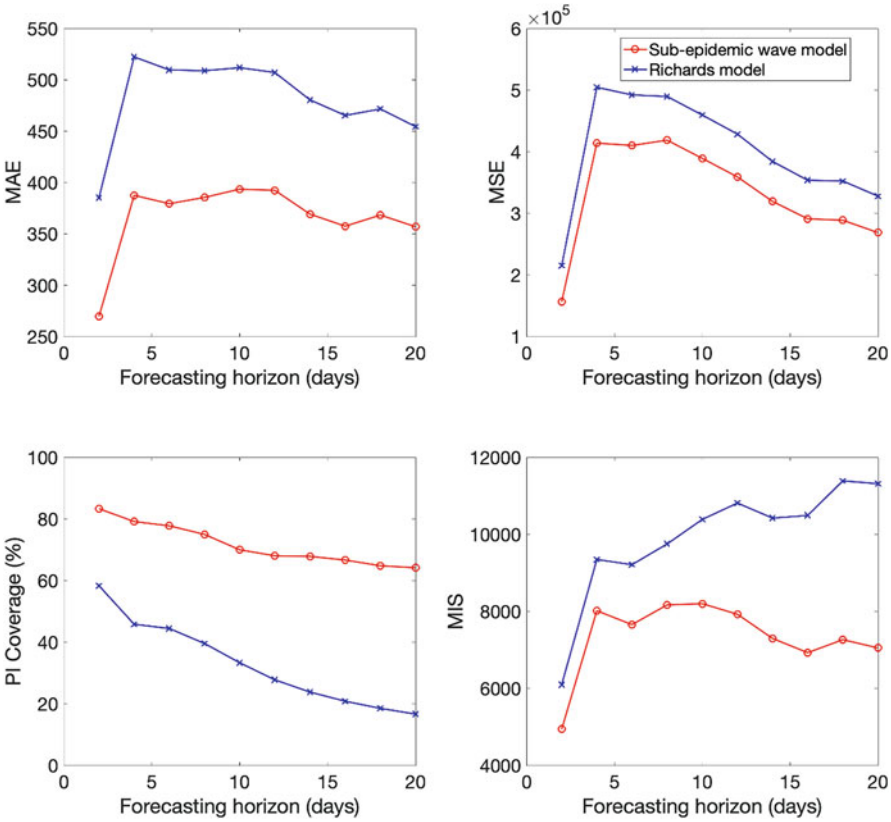


Fig. S16 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in Louisiana. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons

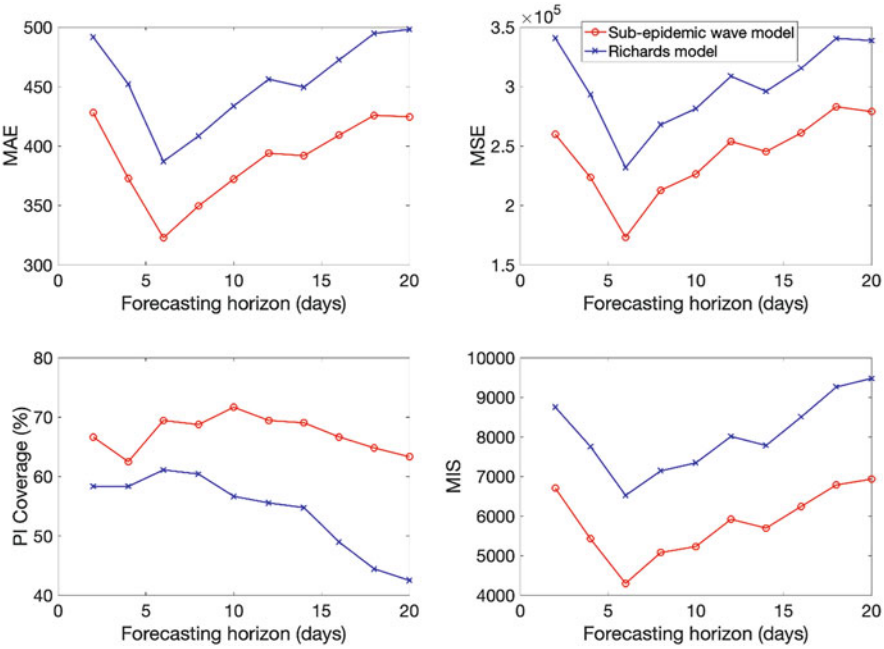


Fig. S17 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in Georgia. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons

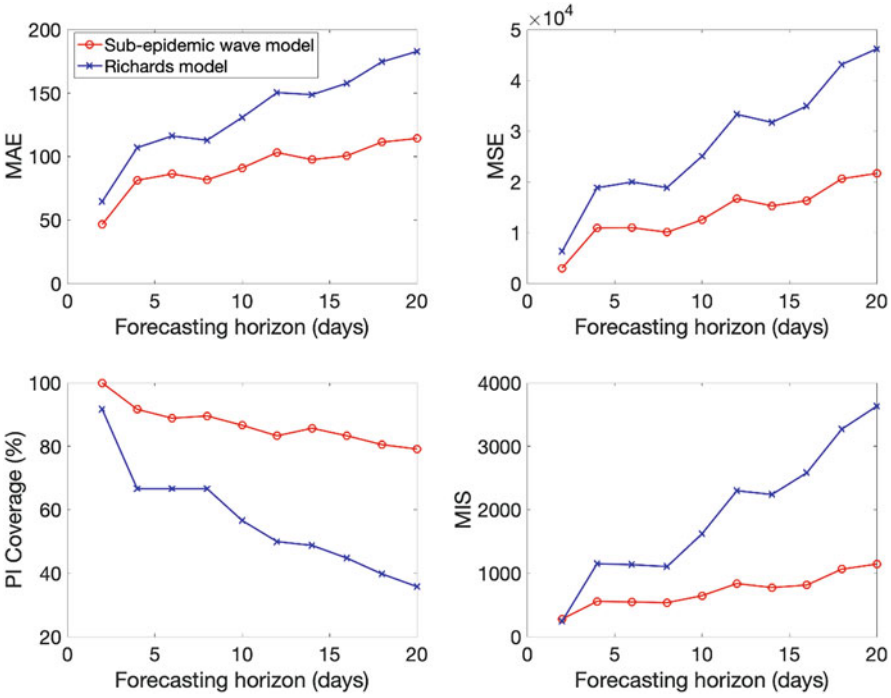


Fig. S18 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in Arizona. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons

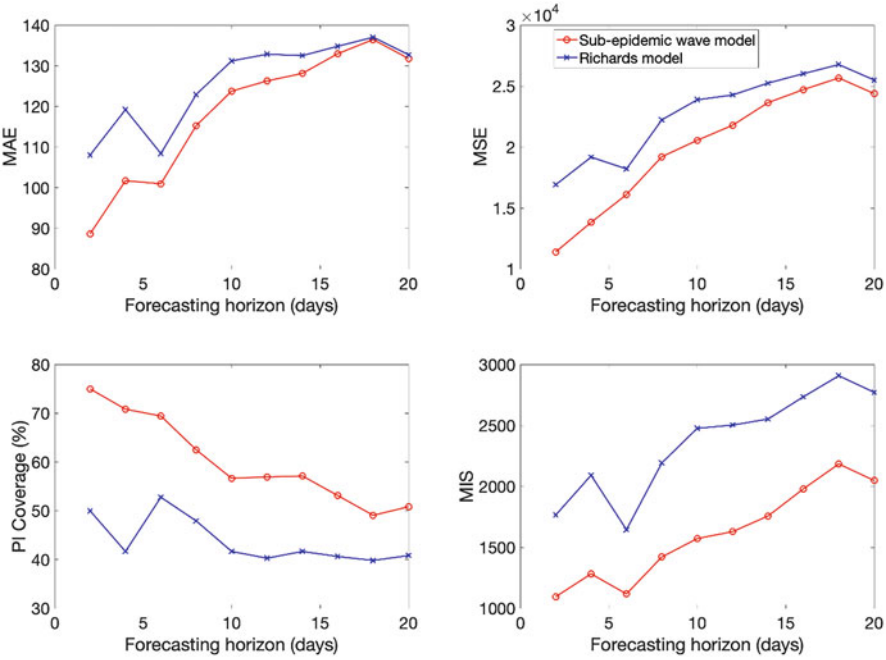


Fig. S19 Mean performance of the sub-epidemic wave and the Richards models in 2–20 day ahead forecasts conducted during the epidemic in Washington. The sub-epidemic model outperformed the Richards model across all metrics and forecasting horizons

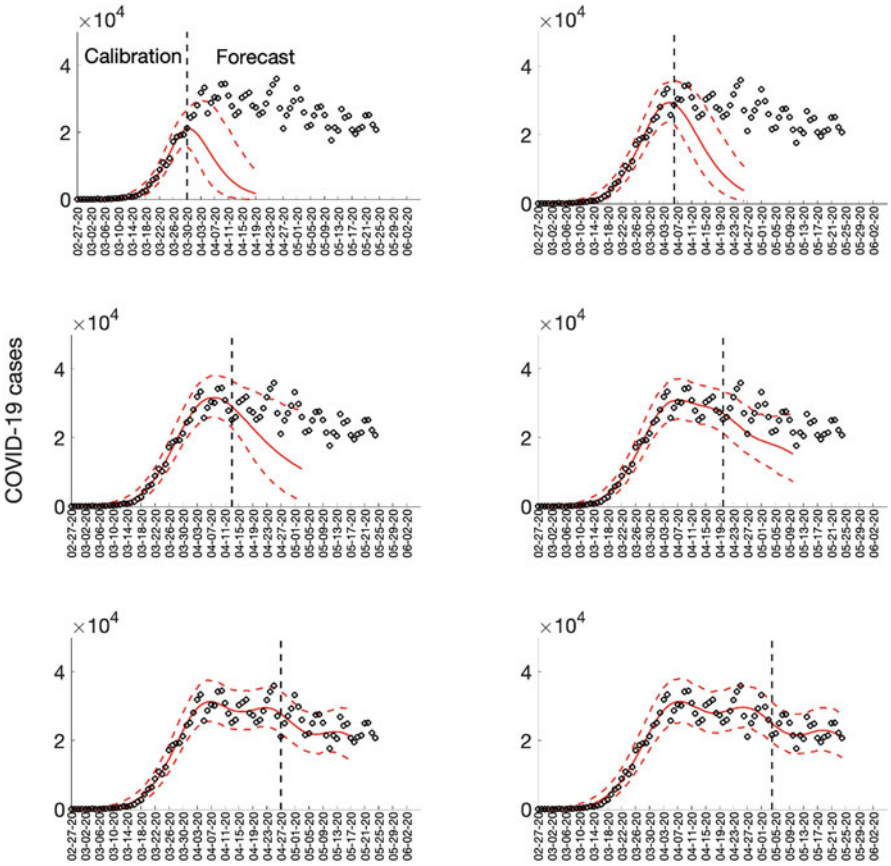


Fig. S20 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in the USA. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

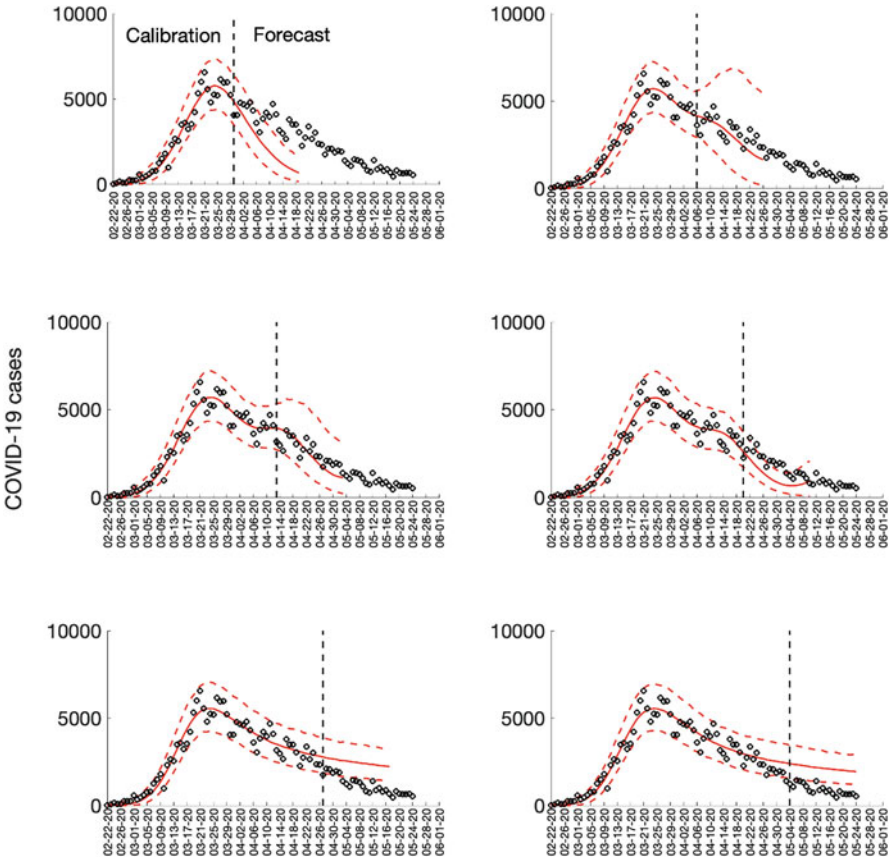


Fig. S21 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in Italy. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

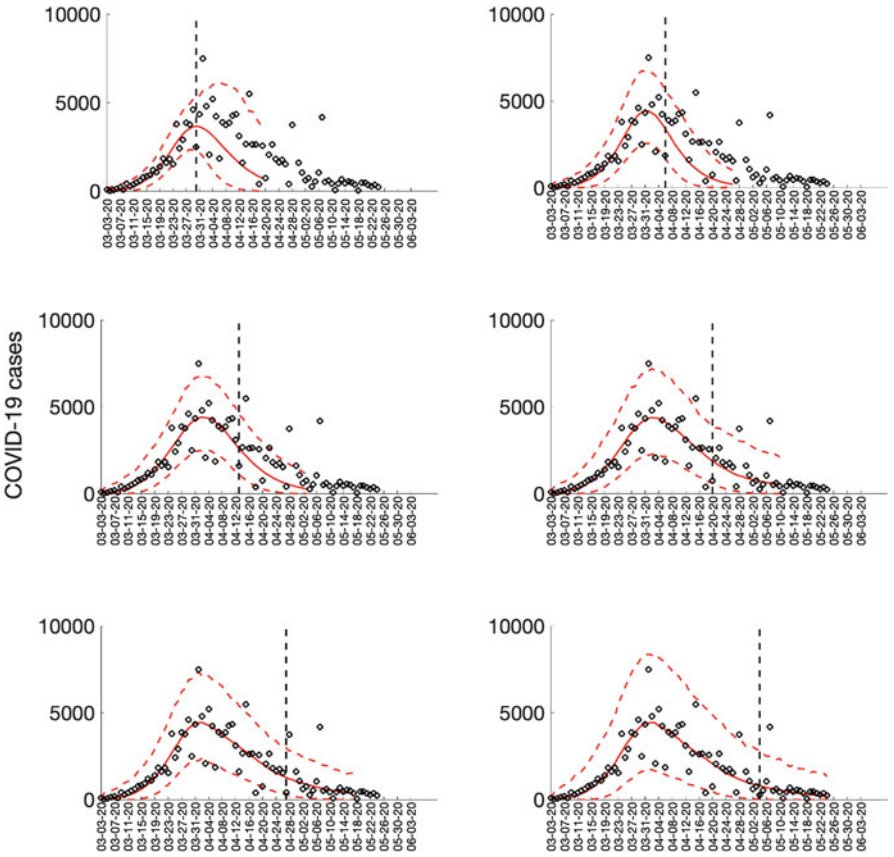


Fig. S22 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in France. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

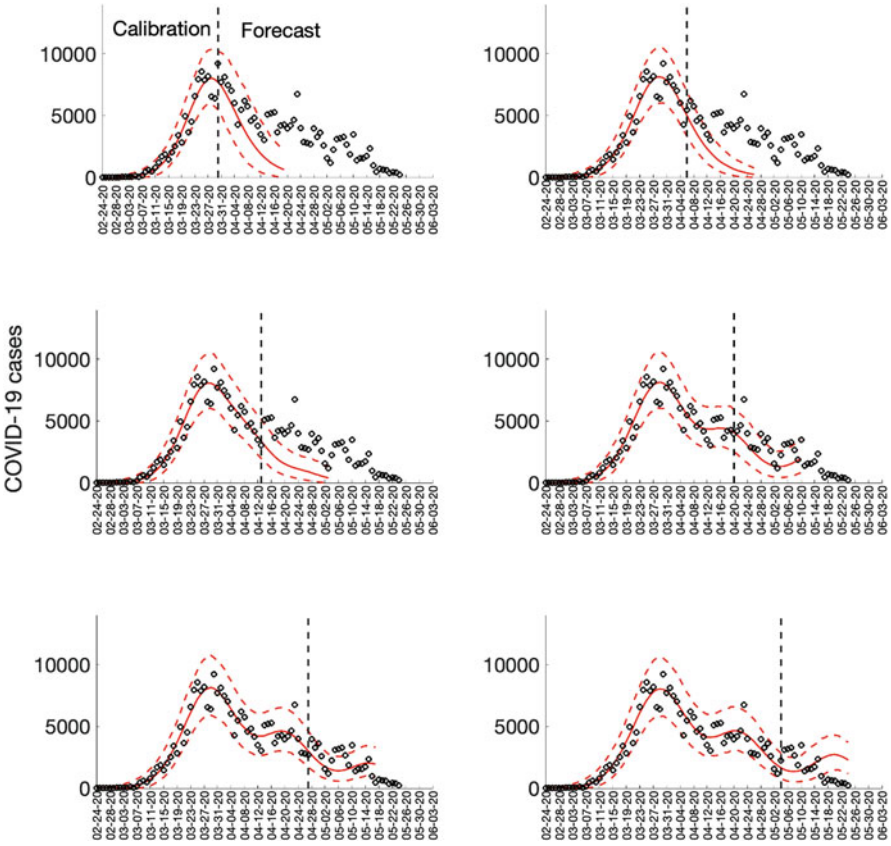


Fig. S23 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in Spain. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

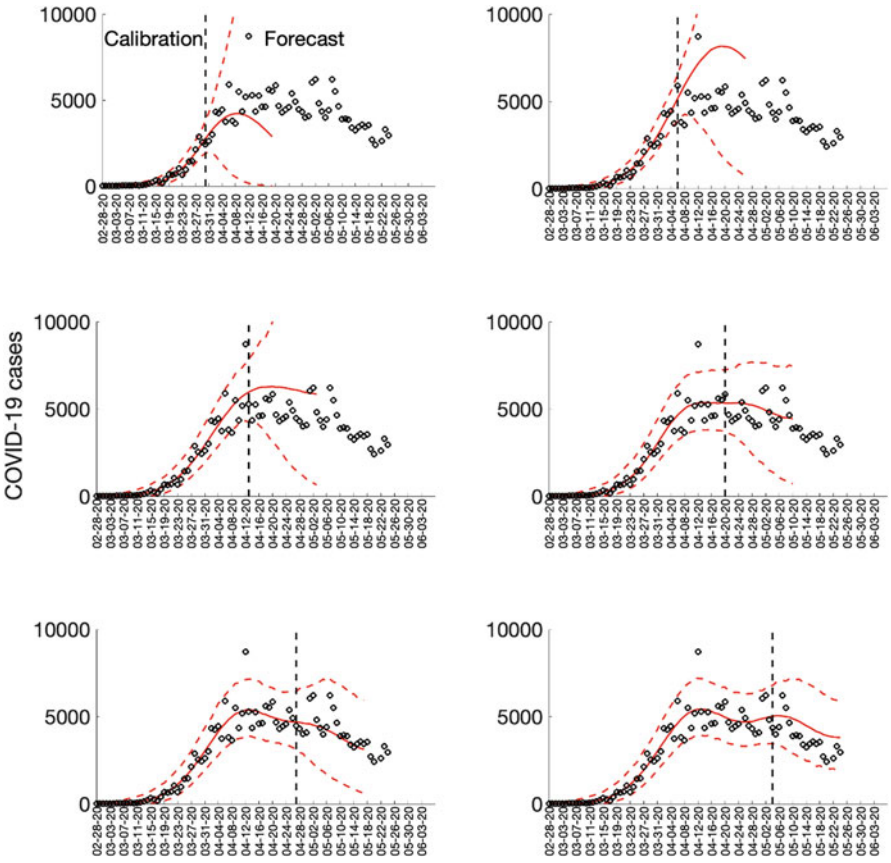


Fig. S24 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in the UK. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

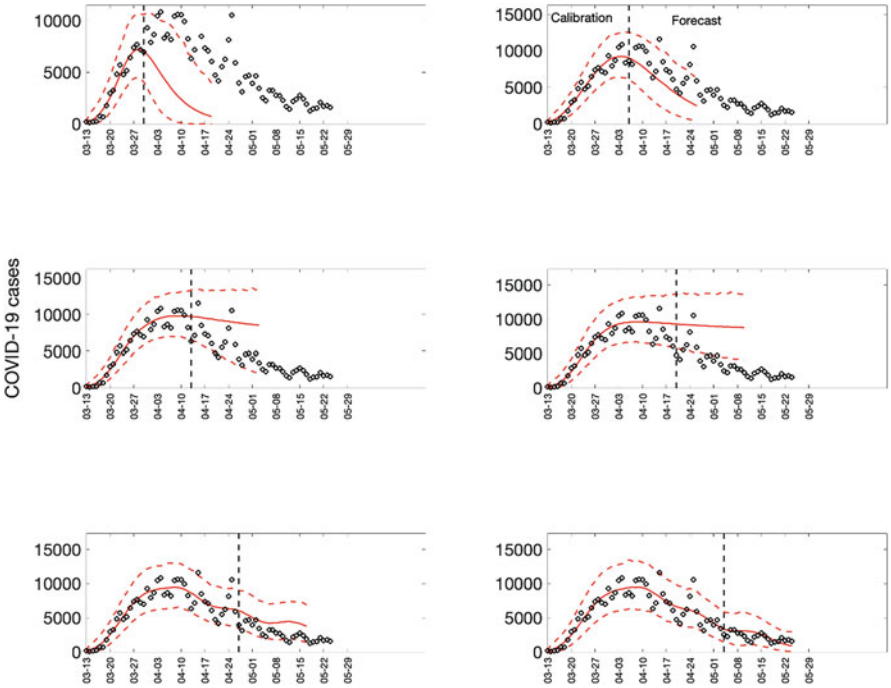


Fig. S25 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in New York State. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

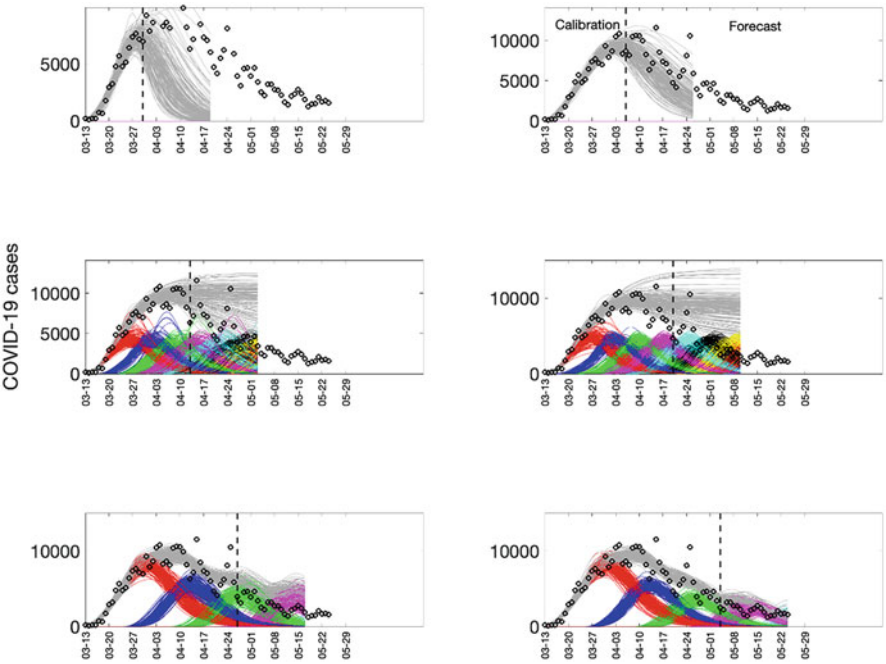


Fig. S26 Sub-epidemic profiles of the sequential 20-day ahead forecasts for the COVID-19 epidemic in New York. Different colors represent different sub-epidemics of the epidemic wave profile. The aggregated trajectories are shown in gray and black circles correspond to the data points. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

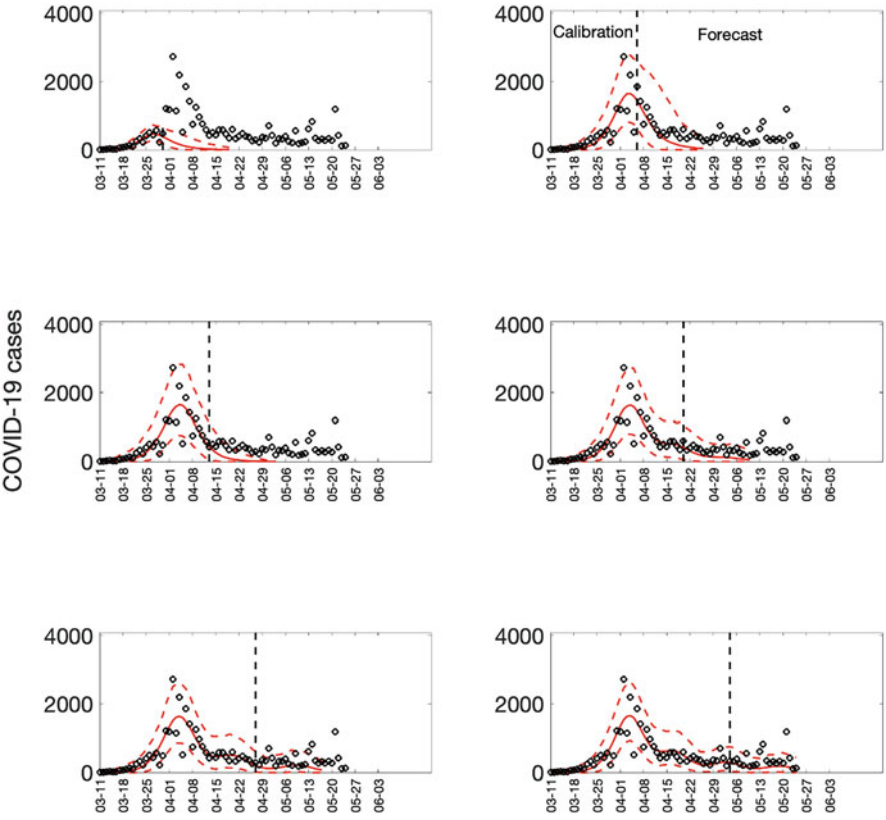


Fig. S27 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in Louisiana. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

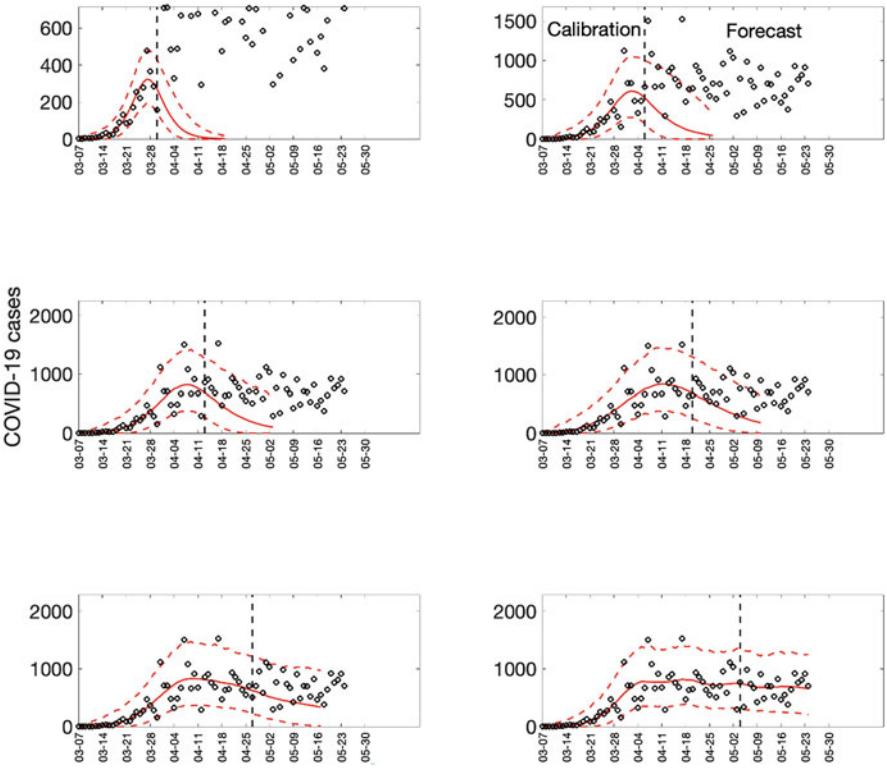


Fig. S28 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in Georgia. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

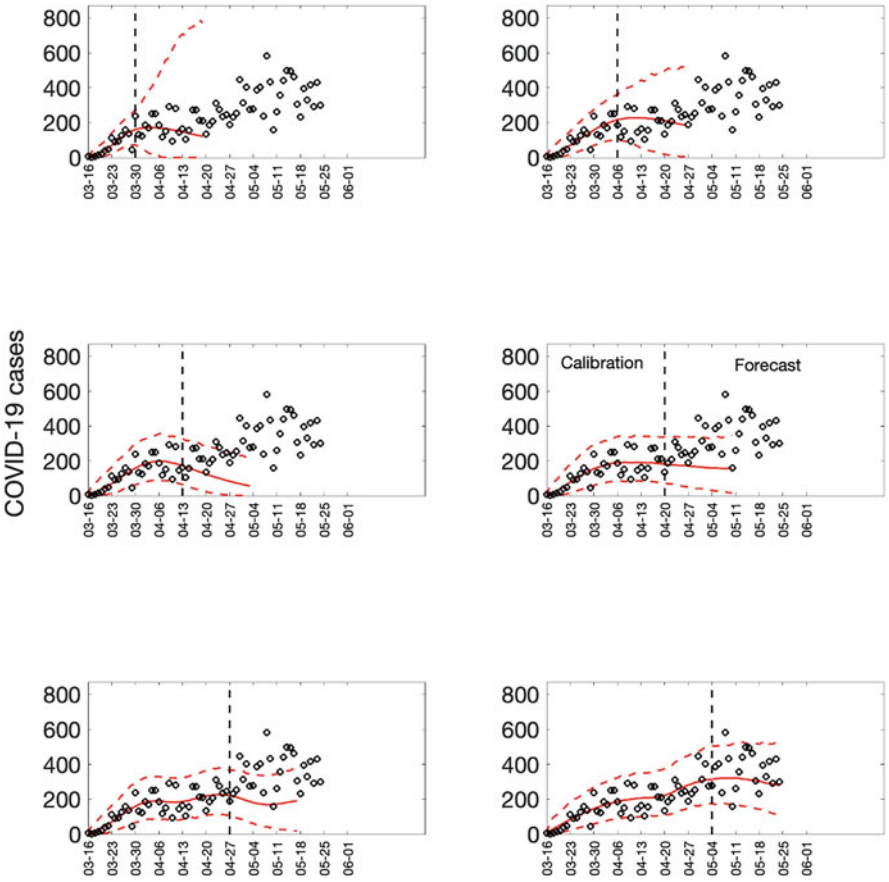


Fig. S29 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in Arizona. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

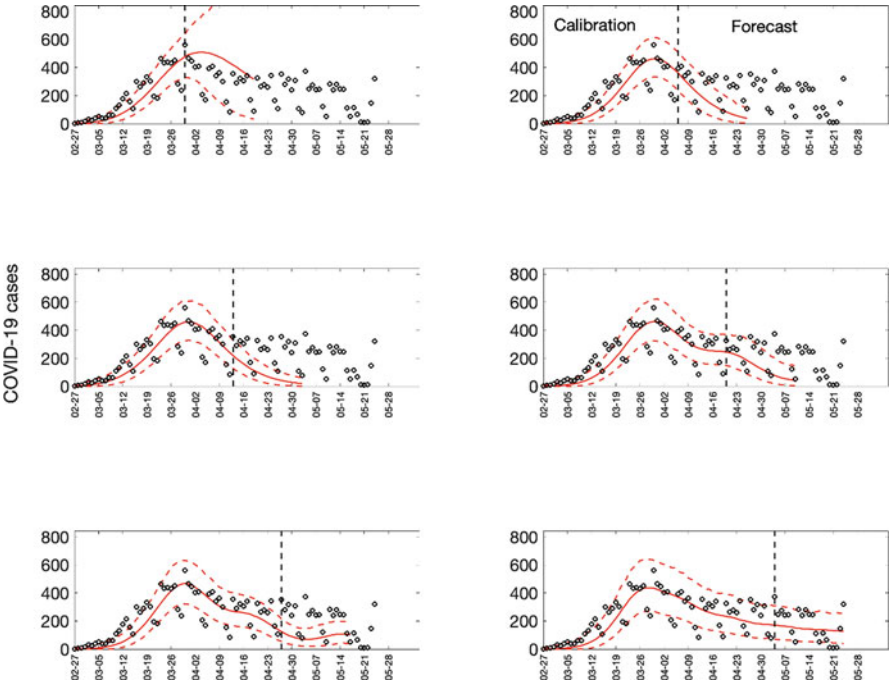


Fig. S30 Sequential 20-day ahead forecasts of the sub-epidemic wave model for the COVID-19 epidemic in Washington. Black circles correspond to the data points. The model fit (solid red line) and 95% prediction interval (dashed red lines) are also shown. The vertical line separates the calibration period (left) from the forecasting period (right). The sequential forecasts were conducted on March 30, April 6, April 13, April 20, April 27, and May 4, 2020

References

1. Jewell NP, Lewnard JA, Jewell BL. Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic. Annals of Internal Medicine 2020;173:xxx-xxx. https://doi.org/107326/M20-1565.
2. Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, et al. The intrinsic transmission dynamics of tuberculosis epidemics [see comments]. Nature Medicine. 1995;95(8):815-21.
3. Garnett GP. The geographical and temporal evolution of sexually transmitted disease epidemics. Sexually Transmitted Infections. 2002;78(Suppl 1):14-9.
4. Rothenberg R, Voigt R. Epidemiologic Aspects of Control of Penicillinase-Producing Neisseria gonorrhoeae. Sexually Transmitted Diseases. 1988;15(4):211-6.
5. Rothenberg R, Dai D, Adams MA, Heath JW. The HIV endemic: maintaining disease transmission in at-risk urban areas. Sexually Transmitted Diseases. 2017;44(2):71-8.
6. Chowell G, Tariq A, Hyman JM. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. BMC Medicine. 2019;17(1):164.
7. Wang XS, Wu J, Yang Y. Richards model revisited: validation by and application to infection dynamics. Journal of theoretical biology. 2012;313:12-9.
8. Nossiter A. Male reports first death from Ebola. New York Times [2014 Oct 24]. Available from: http://www.nytimes.com/2014/10/25/world/africa/mali-reports-first-death-from-ebola.html (accessed on 2015 Jan 13). 2014.
9. Onishi N, Santora M. Ebola patient in Dallas lied on screening form, Liberian airport official says. New York Times [2014 Oct 2]. Available from: http://www.nytimes.com/2014/10/03/world/africa/dallas-ebola-patient-thomas-duncan-airport-screening.html (accessed on 2015 Feb 28). 2014.
10. Onishi N. Last known Ebola patient in Liberia is discharged. New York Times [2015 Mar 5]. Available from: http://www.nytimes.com/2015/03/06/world/africa/last-ebola-patient-in-liberia-beatrice-yardolo-discharged-from-treatment.html?ref=topics&_r=0 (accessed on 2015 Mar 6). 2015.
11. The COVID Tracking Project [Available from: https://covidtracking.com/data.
12. Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A Primer for parameter uncertainty, identifiability, and forecasts. Infect Dis Model. 2017;2(3):379-98.
13. Banks HT, Hu S, Thompson WC. Modeling and inverse problems in the presence of uncertainty: CRC Press; 2014.
14. Myung IJ. Tutorial on maximum likelihood estimation. Journal of Mathematical Pyschology; 2003. p. 90-100.
15. Kashin K. Statistical Inference: Maximum Likelihood Estimation. 2014.
16. Roosa K, Luo R, Chowell G. Comparative assessment of parameter estimation methods in the presence of overdispersion: a simulation study. Mathematical biosciences and engineering : MBE. 2019;16(5):4299-313.
17. Yan P, Chowell G. Quantitative methods for investigating infectious disease outbreaks. Switzerland: Springer Nature; 2019.
18. Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY.: Springer-Verlag New York; 2009.
19. 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(2):268-75.
20. 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.
21. Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infectious Disease Modelling. 2017;2(3):379-98.

22. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review. Physics of Life Reviews. 2016;18:66-97.
23. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model. 2020;5:256-63.
24. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13-23, 2020. J Clin Med. 2020;9(2).
25. Pell B, Kuang Y, Viboud C, Chowell G. Using phenomenological models for forecasting the 2015 Ebola challenge. Epidemics. 2018;22:62-70.
26. Shanafelt DW, Jones G, Lima M, Perrings C, Chowell G. Forecasting the 2001 Foot-and-Mouth Disease Epidemic in the UK. Ecohealth. 2018;15(2):338-47.
27. Chowell G, Hincapie-Palacio D, Ospina J, Pell B, Tariq A, Dahal S, et al. Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. PLoS currents. 2016;8:ecurrents.outbreaks.f14b2217c902f453d9320a43a35b583.
28. Shanafelt DW, Jones G, Lima M, Perrings C, Chowell G. Forecasting the 2001 Foot-and-Mouth Disease Epidemic in the UK. EcoHealth. 2017.
29. Richards FJ. A Flexible Growth Function for Empirical Use. Journal of Experimental Botany. 1959;10(2):290-301.
30. Granovetter MS. The strength of weak ties. American Journal of Sociology. 1973;78(6):1360-80.
31. Cheng VCC, Wong S-C, To KKW, Ho PL, Yuen K-Y. Preparedness and proactive infection control measures against the emerging Wuhan coronavirus pneumonia in China. Journal of Hospital Infection. 2020.
32. Pan J, Yao Y, Liu Z, Li M, Wang Y, Dong W, et al. Effectiveness of control strategies for Coronavirus Disease 2019: a SEIR dynamic modeling study. https://doi.org/10.1101/2020.02.19.200253872020.
33. Prem K, Liu Y, Russell T, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies that reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China. 2020. https://doi.org/10.1101/2020.03.09.20033050.
34. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak: an observational and modelling study. 2020. https://doi.org/10.1101/2020.03.03.20029843.
35. Chowell G, Ammon CE, Hengartner NW, Hyman JM. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. Journal of theoretical biology. 2006;241(2):193-204.
36. Chowell G, Tariq A, Hyman JM. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves: Datasets and fitting code. figshare. Available from: https://doi.org/10.6084/m9.figshare.8867882. 2019.
37. Hsieh YH, Cheng YS. Real-time forecast of multiphase outbreak. Emerging infectious diseases. 2006;12(1):122-7.
38. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. 2007;102(477):359-78.
39. Kuhn M, Johnson K. Applied predictive modeling: New York: Springer; 2013.
40. M4Competition. Competitor's Guide: Prizes and Rules. Available from: https://www.m4.unic.ac.cy/wp-content/uploads/2018/03/M4-Competitors-Guide.pdf (accessed 04/01/2019) [
41. Funk S, Camacho A, Kucharski AJ, Lowe R, Eggo RM, Edmunds WJ. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. PLoS computational biology. 2019;15(2):e1006785.
42. COVID-19 coronavirus / cases [Internet]. 2020. Available from: https://www.worldometers.info/coronavirus/coronavirus-cases/.

43. Roosa K, Tariq A, Yan P, Hyman JM, Chowell G. Multi-model forecasts of the ongoing Ebola epidemic in the Democratic Republic of Congo, March-October 2019. Journal of the Royal Society, Interface/the Royal Society. 2020;17(169):20200447.
44. Chowell G, Viboud C, Simonsen L, Merler S, Vespignani A. Perspectives on model forecasts of the 2014-2015 Ebola epidemic in West Africa: lessons and the way forward. BMC medicine. 2017;15(1):42.
45. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, et al. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics. 2018;22:13-21. Worldometer. (Accessed May 11, 2020, at https://www.worldometers.info/coronavirus/.)

A Model on the Large Scale Use of Convalescent Plasma to Treat Patients with Severe Symptoms

Xi Huo

1 Background

Convalescent plasma (sera) refers to the processed blood product, which contains neutralized antibodies against a specific pathogen. Convalescent plasma therapy (CP therapy), also known as passive antibody therapy, refers to the treatment that transfuses convalescent plasma donated by recent survivors of infectious disease into the ills [1]. This therapy can be dated back to the 1890s and was once the only possible treatment for several infectious diseases before the antibiotic era. In the past two decades, convalescent plasma therapy has been used to treat patients with severe illness and life-threatening conditions during several emerging infectious disease outbreaks, such as severe acute respiratory syndrome (SARS-CoV-1) in 2002-2003, H1N1 in 2009, Middle East respiratory syndrome (MERS) in 2012, Ebola in 2014-2016, and the ongoing COVID-19 outbreak [2].

On March 24th 2020, the US Food and Drug Administration (FDA) allowed healthcare providers to administer CP as an investigational treatment for patients with severe COVID-19 symptoms. From April 3 to June 2, 2020, the FDA Expanded Access Program for COVID-19 convalescent plasma transfused a convenience sample of 20,000 hospitalized patients with COVID-19 convalescent plasma, and demonstrated the safty of this therapy [3]. On August 23rd, 2020, a new guidance was issued to provide updated recommendations to health care providers on the use of COVID-19 convalescent plasma during public health emergency. Indeed, the efficacy of convalescent plasma therapy would be distinct from pathogen to pathogen and still needs to be determined by clinical studies [4]. However, this therapy might be the only readily available treatment strategy for emerging

X. Huo (✉)
Department of Mathematics, University of Miami, Coral Gables, FL, USA
e-mail: x.huo@math.miami.edu

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_6

infectious diseases, and the large-scale use of it will help to buy time for the development and testing of potential treatments and vaccines.

The population-wide use of convalescent plasma therapy involves several logistics managements: convalescent patients should be gathered and screened to select potential donors, donated blood products should be stored and stockpiled appropriately, the blood bank should distribute convalescent plasma based on both treatment requests and deployment strategy. Thus the more donations are made, the higher stockpile is achieved, and the more lives could be saved. The logistic coordination of each step should be quantified to estimate the feasibility, and the effectiveness in mortality reduction should be projected to adjust our expectations. This chapter will discuss the use of an ordinary differential equation model framework that reflects the treatment-donation-stockpile dynamics arising from the population-wide implementation of convalescent plasma therapy during an infectious disease outbreak. We also refer to the limited number of previous works on modeling the use of convalescent plasma in treating patients under critical conditions of H1N1 in Hong Kong [5], Ebola in West Africa [6], and COVID-19 in Italy [7].

2 Model

To get a complete estimation of the resources needed to conduct convalescent plasma therapy, one will need to consider both the transmission dynamics and the treatment-donation-stockpile dynamics. Simulations of the transmission dynamics will provide the essential inflow to the treatment-donation-stockpile model: projections on the number of patients under severe conditions, recovered individuals, and potential donors are all driven factors of the treatment implementation. On the other side, the treatment dynamics might not influence the spread of the virus—outbreaks can only be contained via a series of non-pharmaceutical interventions (such as isolation, quarantine, testing, contact tracing, personal protective equipment, etc.) with the absence of vaccines. Therefore, the treatment-donation-stockpile dynamics can be decoupled from the transmission dynamics, and the compartmental diagram is illustrated in Fig. 1.

Treatment Dynamics The infected population can be stratified into three classes—patients with mild symptoms treated under palliative care (P_m), patients with severe symptoms treated under palliative care (P_s), and patients treated by convalescent plasma therapy (T). Patients in each compartment are assumed to recover ($\gamma_p^m, \gamma_p^s, \gamma_T$) or die ($\mu_p^m, \mu_p^s, \mu_T$) at constant rates. All recovered individuals are stratified into compartment R to select for qualified donors.

The inflow of new patients can be simulated by the transmission dynamics model, and we denote $A(t)$ as the rate of newly identified patients at time t, and assume a proportion p of them are under severe conditions. Patients identified initially with mild symptoms could gradually develop severe symptoms, and we assume this

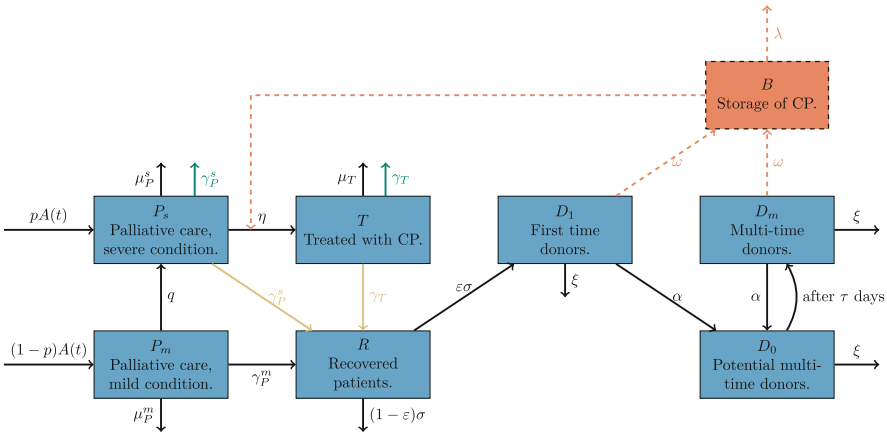


Fig. 1 Compartmental diagram of the treatment-donation-stockpile dynamics. Blue compartments and solid arrows refer to the population dynamics of patients and donors, red compartment and dashed arrows refer to the convalescent plasma dynamics. Depending on the specific disease, donors would either consist of all recovered patients (illustrated in green arrows) or consist of patients with mild symptoms only (represented in yellow arrows)

happens at a linear rate of q. The fraction of severe cases upon identification p and the symptom deterioration rate q would vary significantly from disease to disease. In the Ebola outbreak, most of the cases are under severe conditions, and then one can assume $p = 1$, $q = 0$, and omit the compartment P_m. While in the case of COVID-19, symptom deterioration is uncommon, and one may have to estimate p and q based on a combination of case report data, hospitalization data, and intensive care unit data.

In most cases, only those patients with severe symptoms will be considered for convalescent therapy, where the rate of receiving such therapy would highly depend on the stockpile of plasma at the time. Thus we model the rate for severe patients to receive convalescent therapy as a function $\eta(P_s, B)$ with its value determined by both the demand and supply. There is more than one way to formulate the need-supply function. When the stockpile is abundant all the time, one can consider the saturation function

$$\eta(P_s, B) = \frac{\eta P_s B}{B + K}, \tag{1}$$

with η as the treatment arrangement and approval rate for blood transfusion therapy, and K as the threshold of stockpile level where only half of the demands could be met. When the stockpile is scarce, one can also consider using the step-wise function

$$\eta(P_s, B) = \eta \min\{P_s, B\}, \tag{2}$$

with η^{-1} as the mean turnaround time for treatment arrangements and approvals. The model equations for the treatment-related compartments then follow below:

$$\begin{aligned} P_s'(t) &= pA(t) - (\mu_P^c + \gamma_P^c)P_s(t) - \eta(P_s(t), B(t)), \\ P_m'(t) &= (1-p)A(t) - (\mu_P^m + \gamma_P^m)P_m(t), \\ T'(t) &= \eta(P_s(t), B(t)) - (\mu_T + \gamma_T)T(t). \end{aligned} \tag{M.1}$$

Donation Dynamics In order to model the donation process, all recovered patients are stratified into classes for convalescent patients who are considered as potential donors (R), donors ready for their first-time donations (D_1), donors preparing for their next-time donations (D_0), and donors ready to make their next-time donations (D_m). The criteria for potential donors might defer from disease to disease: for example, the majority of COVID-19 and H1N1 patients would experience mild symptoms and would provide sufficient donations to save the critical patients, whereas for fatal illness as Ebola all survivors must be considered for donations.

Further, among all potential donors, only a fraction ε of them would be qualified, and we assume the rate of the screening process as σ. Qualified donors who are ready for donations would donate their convalescent plasma at a rate of α, and donors at all stages could quit at a constant rate of ξ. The dynamics between available donors and those at rest can be modeled using a delay term to reflect the mandatory resting period of τ between two consecutive donations.

Depending on the actual screening and plasmapheresis resources, the donor screening rate σ and donation rate α can be formulated in many different ways. Under abundant resources, one can model σ and α as constant linear rates. While under limited resources, for example, the number of donor applications and donations far exceeds the number of screening and plasmapheresis throughput U_s and U_p, one can formulate the two rates via the step-wise functions

$$\sigma(R) = \sigma \min\{R, U_s\}, \tag{3}$$

$$\alpha(D_1) = \alpha \min\{D_1 + D_m, U_p\} \cdot \frac{D_1}{D_1 + D_m}, \quad \alpha(D_m) = \alpha \min\{D_1 + D_m, U_p\} \cdot \frac{D_m}{D_1 + D_m}, \tag{4}$$

with σ, α represent the mean turnaround time for screening and plasmapheresis. The equations relating to the donation population dynamics are as following:

$$\begin{aligned} R'(t) &= \gamma_P^m P_m(t) + \gamma_P^s P_s(t) + \gamma_T T(t) - \sigma(R(t)), \\ D_1'(t) &= \varepsilon\sigma(R(t)) - \alpha(D_1(t)) - \xi D_1(t), \end{aligned}$$

$$D_0'(t) = \alpha(D_1(t)) + \alpha(D_m(t)) - \xi D_0(t) - e^{-\xi\tau}[\alpha(D_1(t-\tau)) + \alpha(D_m(t-\tau))], \tag{M.2}$$

$$D_m'(t) = e^{-\xi\tau}[\alpha(D_1(t-\tau)) + \alpha(D_m(t-\tau))] - \alpha(D_m(t)) - \xi D_m(t),$$

where the terms $\gamma_P^s P_s(t) + \gamma_T T(t)$ are only needed when modeling diseases with high mortality rates (such as Ebola).

Blood Bank The number of convalescent blood products in the unit of treatment is denoted as B, that is, one unit of blood product from the bank can be used to treat one patient. The inflow rate to compartment B is proportionally related to the donation rate $\alpha(D_1) + \alpha(D_m)$. The proportionality constant, ω, could vary accordingly to the actual donation-treatment protocols, for example, the Expanded Access to Convalescent Plasma for the Treatment of Patients with COVID-19 program led by Mayo Clinic utilized $1 \sim 2$ units of convalescent plasma to treat each patient, thus, in this case, ω is regarded to vary between 0.5 and 1. Further, one can assume a linear expiration rate of λ for the unused donations, where this value can be estimated differently based on the types of blood donations in storage. In the 2014–2016 Ebola outbreak, WHO suggested that the donations should either be stored in the format of whole blood or plasma, where the average expiration period is 35 days for whole blood and 40 days for plasma. Based on the above discussions, the equation for blood bank storage is shown below

$$B'(t) = \omega[\alpha(D_1) + \alpha(D_m)] - \eta(P_s(t), B(t)) - \lambda B(t). \tag{M.3}$$

3 Parameterization

Death and recovery rates for each compartment should be parameterized together so as to reflect the correct mortality rate for patients under different situations. Let $P(t)$ denote the number of patients at time t and let μ and γ respectively represent the death and recovery rates. Then the instantaneous outflow rates of patients due to death and recovery are respectively modeled as $\mu P(t)$ and $\gamma P(t)$. And the case fatality ratio (CFR) can be calculated as

$$\text{CFR} = \frac{\int_0^{+\infty} \mu P(t)dt}{\int_0^{+\infty} \mu P(t)dt + \int_0^{+\infty} \gamma P(t)dt} = \frac{\mu}{\mu + \gamma}.$$

Therefore, one only needs to estimate two quantities out of the above three. In the treatment-donation dynamics, it is more vital to estimate the instantaneous number of recovered individuals. Also, the average recovery period and the mortality rate

are relatively easier to determine in practice. Thus one might fix γ as the reciprocal of mean recovery time and simply set $\mu = \frac{\text{CFR}}{1-\text{CFR}}\gamma$.

It is worth mentioning that for the problem modeled, the method discussed above is more suitable for the compartments into which newly identified patients are stratified—such as P_s and P_m in our model. While patients in the blood transfusion compartment T have already spent an average time η^{-1} in the compartment P_s, one has to pay extra attention when estimating the recovery rate γ_T. The best-case scenario is knowing the average time for patients to recover after blood transfusion, but this information could be hard to perceive given insufficient knowledge about CP therapy for emerging diseases. One way to proceed is to assume that it takes the same amount of time for patients under CP therapy to recover compared to patients with severe symptoms. Then we have $\eta^{-1} + \gamma_T^{-1} = {\gamma_P^s}^{-1}$ for a possible estimation of γ_T and therefore μ_T.

Rates of making and quitting donations can be estimated similarly. Let $D(t)$ be the compartment of donors and assume that the donors would leave the compartment by either making a donation or quitting at linear rates $\alpha D(t)$ and $\xi D(t)$. Assume p_d as the probability for a donor to donate during the first day of qualification, and we have $p_d = 1 - e^{-\alpha}$. Assume p_c as the probability for a person to quit in the first day of qualification, and $p_c = 1 - e^{-\xi}$. The ranges of the probability parameters p_d and p_c are relatively easier to perceive and can be therefore used to estimate α and ξ. Notably, from the definitions of p_d and p_c, it is not necessary to suppose $p_d + p_c = 1$.

4 Outcome Measurements

Given projections on the time-dependent case data, one can measure the following essential aspects of CP therapy (here we use a time unit of days).

Time-dependent outcomes on the treatment-donation-stockpile dynamics include: $(O1)$ patients under each therapy (palliative care or CP therapy) can be measured by evaluating $P_s(t) + P_m(t)$ and $T(t)$; $(O2)$ cumulative deaths can be evaluated by $\int_0^t \mu_P^m P_m(x) + \mu_P^s P_s(x) + \mu_T T(x)dx$; $(O3)$ daily screening burden for potential donors as $\int_{t-1}^t \varepsilon R(x)dx$; $(O4)$ daily number of donations (hence the daily demands for plasmapheresis) as $\int_{t-1}^t \alpha(D_1(x) + D_m(x))dx$; $(O5)$ blood bank storage as $B(t)$; and $(O6)$ cumulative expired CP donations $\int_0^t \lambda B(x)dx$.

Overall outcomes that evaluate the feasibility and effectiveness of the CP therapy implementation can be reported in four major aspects. The population-level effectiveness of therapy can be assessed by computing the overall CFR after the implementation, which is obtained by $(O2)$ and the total case count projected by the transmission model. The feasibility of the population-wide use of CP therapy can be evaluated by the maximal demands of plasmapheresis and that of potential donor screenings, which can be obtained by evaluating the maximal values of $(O3)$ and $(O4)$. Further, based on the choice of function $\eta(P_s, B)$, the distribution strategy of

CP stockpiles should be evaluated not only by CFR but also by the total number of unused donations, which is estimated from $(O6)$.

5 Simulations

A simple transmission model is used in this section to generate the outbreak data with the following equations:

$$\begin{aligned} S'(t) &= -\beta(t)S(t)I(t), \\ L'(t) &= \beta(t)S(t)I(t) - \rho L(t), \\ I'(t) &= \rho I(t) - \kappa I(t), \end{aligned} \tag{M.0}$$

where S refers to the susceptible population, L refers to the latent population, and I refers to the infectious population that are not yet hospitalized or isolated. The parameters and initial conditions of the transmission model are listed in Table 1, and were used to simulate the Ebola outbreak in Guinea, which emerged in December 2013 [9]. In the early stage of this outbreak, the case count increased exponentially, and the basic reproductive number was estimated between 1.50 and 1.52. With the local leaders' help and carefully designed policy implementations, the outbreak was eventually contained and resulted in 3814 total cases and 2544 deaths. The nonpharmaceutical interventions' impacts were reflected by considering a time-dependent transmission rate estimated to decay exponentially over time. Thus, it was formulated as $\beta(t) = 0.3e^{-0.0023t}$.

By the time of the outbreak, no licensed medicine or vaccine exists for Ebola. In September 2014, WHO issued interim guidance on the large-scale use of convalescent blood products collected from Ebola survivors for blood transfusion therapy[8]. The parameters used in these simulations are directly adopted from the WHO guidance and are listed in Table 1. As discussed before, most Ebola patients would experience life-threatening symptoms. Thus one should assume $p = 1, q = 0$ and $P_m(t) = 0$ for all $t \geq 0$, and assume all recovered patients would be considered for potential donors. In the simulations presented, the donor screening and plasma donation dynamics are assumed to be linear processes, where the maximal demands of these two processes are estimated to inform the capacity requirements. For all simulations presented here, the collection and enforcement of blood transfusion therapy are initiated on the 60th day from the beginning of the outbreak. Numerical simulations are conducted by the software Python version 3.7.3 and utilized the ODE and DDE solvers *odeint* and *ddeint*.

Choice of the CP distribution function $\eta(P_s, B)$ as a saturation function in (1) or a step-wise function in (2) represents different deployment philosophies of CP stockpiles. When the CP storage meets the treatment demand ($B > P_s$), the choice of saturation function is equivalent to a conservative use of blood donations, resulting in fewer patients treated. On the other way around, when the treatment

Table 1 Parameter summary and values used in simulations

Parameter	Definition	Values
Transmission parameters		
N	Total population	12×10^6
β	Transmission rate	without control: $0.3/N$
		with control: $0.3e^{-rt}/N$
ρ^{-1}	Latent period	11.4 days
κ	Hospitalization rate	$1/5$ day^{-1}
p	Progression rate from mild to critical	NA
γ_P^c	Recover rate of critical case	0.05 day^{-1}
γ_P^m	Recover rate of mild case	NA
μ_P^c	Death rate of critical case	$7\gamma_P^c/3$[a]
μ_P^m	Death rate of mild case	NA
Treatment parameters		
γ_T	Recover rate with CP	$(1/\gamma_P^c - 1)^{-1}$
μ_T	Death rate with CP	$\frac{f}{1-f}\gamma_T$
η	Arrangement rate of CP therapy	1 day^{-1}
K	Threshold of blood stock	5000
f	Fatality ratio with CP	0.1
Donation parameters		
ε	Probability of becoming a donor	0.5
σ	Transition rate from discharged patients to potential donors	$1/28$ day^{-1}
α	Rate of donation	$-\ln(1 - p_d)$
ω	Inflow rate of donation to blood bank	Same as α
p_d	Probability of making donation	0.9
ξ	Loss rate of donors	$-\ln(1 - p_c)$
p_c	Probability of quitting donation	0.1
τ	Donor recovery period between consecutive donations	14 days
λ	CP expiration rate	$1/40$ day^{-1}
Initial values for transmission dynamics		
$S(0)$	Susceptible population	12×10^6
$L(0)$	Latent population	0
$I(0)$	Infectious population	21
$P_c(0)$	Critically ill cases	0
$P_m(0)$	Mild cases	NA
$R(0)$	Recovered population	0
Initial values for donation dynamics		
$T(0)$	Patients under CP treatment	0
$D_1(0)$	Potential first-time donors	0
$D_0(0)$	Donors under recovery	0
$D_m(0)$	Potential multi-time donors	0
$B(0)$	Blood bank stockpile	0

[a] The estimation of the death rate for EVD is based on a 70% mortality rate

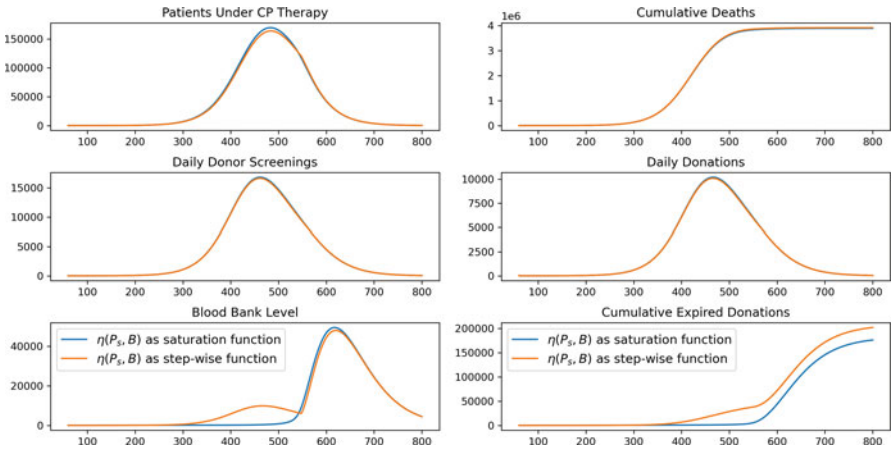


Fig. 2 Treatment-donation-stockpile dynamics under no intervention. All parameters are fixed according to Table 1 except for $\beta = 0.3/N$

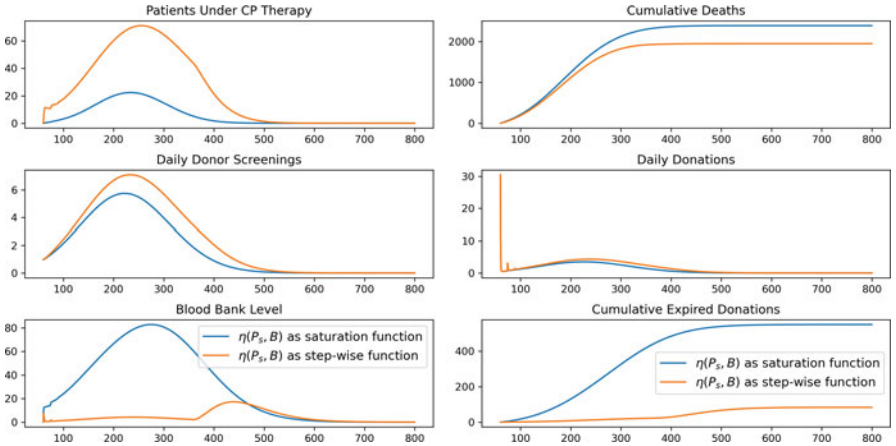


Fig. 3 Treatment-donation-stockpile dynamics under nonpharmaceutical interventions. All parameters are fixed according to Table 1 except for $\beta(t) = 0.3e^{-0.0023t}/N$

demand exceeds the available storage ($P_s < B$), the model with the saturation function would experience a faster CP consumption rate comparing to that with the step-wise function ($\eta \frac{P_s}{B+K} B > \eta B$), resulting in a faster depletion of blood donations and hence the model will project more treatments.

Such phenomenon is illustrated respectively in Figs. 2 and 3. In Fig. 2, the transmission rate is fixed as a constant by assuming no transmission intervention strategies were enforced, and the CP demand far exceeds the stockpiles throughout the outbreak—hence more treatments are projected by the model with saturation

distribution function. In Fig. 3, the transmission rate is formulated according to the results from the data fitting analyses literature, which represents an implementation of strong nonpharmaceutical interventions. The available stockpiles are sufficient throughout the outbreak—hence more patients are projected to be saved by the model with step-wise function.

However, in reality, neither the CP therapy deployment nor the treatment request delivery can be processed simultaneously or in every second. Depending on the health system turnaround frequency, it might be more realistic to reduce the model into a discrete-time system and refresh the system status in a unit of days in the simulations, and we refer to the work by Wu et al. on H1N1 for more information [5].

Feasibility and possible benefits of the large-scale use of CP therapy depend on a few aspects. The simulations in Figs. 4 and 5 discuss the impacts of the efficacy of CP therapy and the effectiveness of nonpharmaceutical interventions. All simulations are performed by utilizing the step-wise CP distribution function $\eta(P_s, B)$. Based on both simulations, CP therapy's potential benefits in reducing the overall CFR are mostly impacted by the therapy's efficacy. While the feasibility of the population-wide implementation of the donor screening and blood donation collection depends significantly on the containment of the outbreak.

Figure 4 estimated the peak-time plasmapheresis demand as 10,000 when no infection control strategies were enforced. Such a demand is not manageable in almost any part of the world: the estimated number of plasmapheresis machines that would become available for such treatment was 4.9 per 10^6 population in Houston, Texas [5]. Thus for a local healthcare system as advanced as that in Houston, there are less than 60 machines available for conducting blood donations for the simulated 12-million-population community, which means each machine has to process 166 donations every day during the peak time. On the other hand, with a reasonable level of infection control, which is not necessarily as efficient as that in Guinea's Ebola outbreak, the demands for donor screening and plasma collection appear to be much more realistic.

6 Discussion

In the event of a disease outbreak caused by novel pathogens, no drugs or other potential therapy might be immediately available to prevent or treat the new disease. Convalescent plasma therapy is one of the few rapidly available treatments since recovered individuals could be tracked and identified as potential donors even in the early stage of the outbreak. Implementing the population-wide use of this therapy could buy time for the development of vaccines and other promising treatment options, whereas the logistic operation is complex and involves the coordination of many parts of the public health and medical system. In the past few years, the use of convalescent plasma has been suggested and investigated by medical authorities during almost every large-scale infectious disease outbreak. It is still highly likely

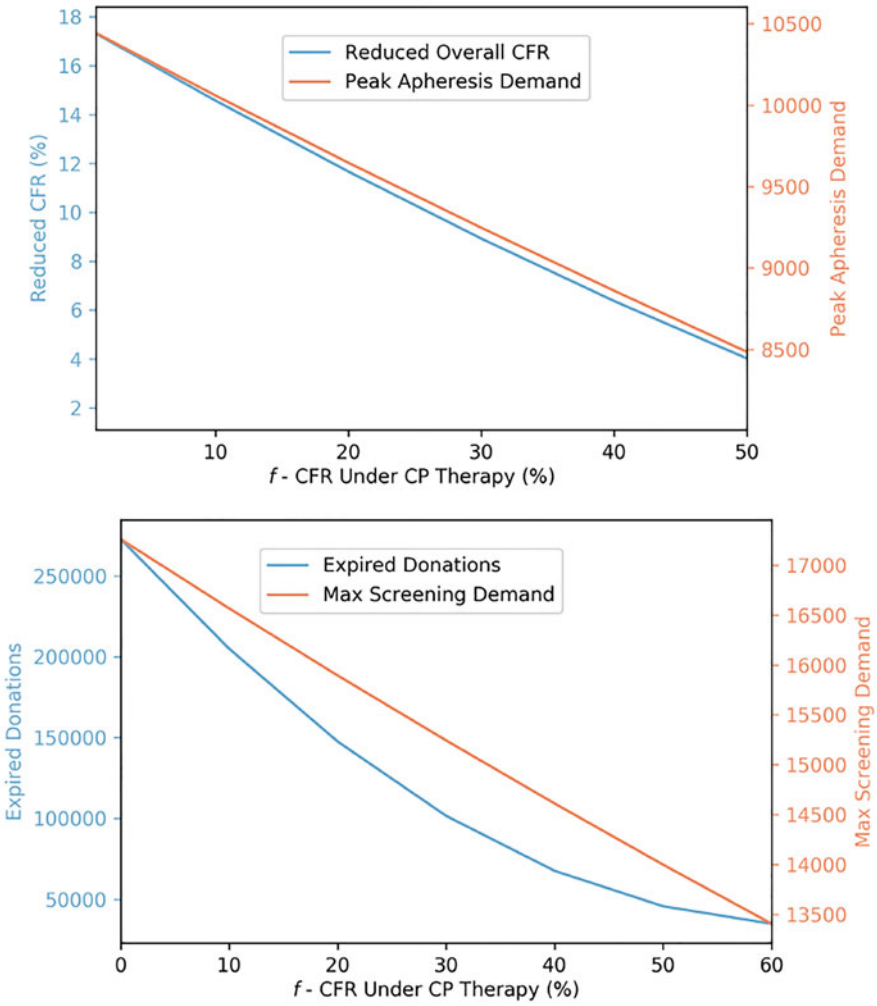


Fig. 4 Effectiveness and feasibility with respect to different clinical efficacies of CP therapy without nonpharmaceutical interventions. All parameters are fixed according to Table 1 except for the CFR of CP therapy (hence μ_T) being varied at different levels among 0∼60%

for this therapy to be proposed for many more times in future disease outbreaks. Therefore, integrating strategic plans that balance the needs of treatment delivery, donor screening and documentation, and plasma bank coordination is essential to save lives. The model presented in this chapter can be applied to quantify and validate such integrative plans for regions with detailed information on their apheresis and serology testing facilities. Further, the modelling framework can be easily extended to take into consideration many other important aspects about plasma therapy. For example, the timing of infusing a severe patient with convalescent

Fig. 5 Effectiveness and feasibility with respect to different transmission control efforts. All parameters are fixed according to Table 1 except for $\beta(t) = 0.3e^{-rt}/N$ with r varying from 0.001 to 0.0025

plasma is considered to be important and could affect the pharmacokinetics of the neutralizing antibodies [10]. Thus an extended model structured with the age-since-infection of the severely ill patients would help address this issue.

Acknowledgments This work was partially supported by the National Science Foundation (grant number: DMS-1853622) and Direct and Indirect Effects of COVID-19 Program from College of Arts and Sciences, University of Miami.

References

1. Montelongo-Jauregui, D., Vila, T., Sultan, A.S., Jabra-Rizk, M.A.: Convalescent serum therapy for COVID-19: A 19th century remedy for a 21st century disease. PLoS Pathogens. **16**(8), p.e1008735 (2020)
2. Casadevall, A., Pirofski, L.A.: The convalescent sera option for containing COVID-19. J. Clin. Investig. **130**(4), 1545–1548 (2020)
3. Joyner, M.J., Bruno, K.A., Klassen, S.A., Kunze, K.L., Johnson, P.W., Lesser, E.R., Wiggins, C.C., Senefeld, J.W., Klompas, A.M., Hodge, D.O., Shepherd, J.R.: Safety update: COVID-19 convalescent plasma in 20,000 hospitalized patients. Mayo Clin. Proc. **95**(9), 1888–1897 (2020)
4. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines: Convalescent Plasma. National Institutes of Health. https://www.covid19treatmentguidelines.nih.gov/. Accessed Jan. 18, 2021.
5. Wu, J.T., Lee, C.K., Cowling, B.J., Yuen, K.Y.: Logistical feasibility and potential benefits of a population-wide passive-immunotherapy program during an influenza pandemic. Proc. Natl. Acad. Sci. U.S.A. **107**(7), 3269–3274 (2010)
6. Huo, X., Sun, X., Lan, K., Wu, J.: Treatment–donation-stockpile dynamics in Ebola convalescent blood transfusion therapy. J. Theor. Biol. **392**, 53–61 (2016)
7. Huo, X., Sun, X., Bragazzi, N.L., Wu, J.: Effectiveness and feasibility of convalescent blood transfusion to reduce COVID-19 fatality ratio. R. Soc. Open Sci. **8**(2), 202248 (2021)
8. World Health Organization. Use of convalescent whole blood or plasma collected from patients recovered from Ebola virus disease for transfusion, as an empirical treatment during outbreaks: interim guidance for national health authorities and blood transfusion services (No. WHO/HIS/SDS/2014.8). World Health Organization (2014)
9. Althaus, C.L.: Estimating the reproduction number of Zaire ebolavirus (ebov) during the 2014 outbreak in West Africa. PLOS Curr. Outbreaks. Sep2, Edition 1 (2014)
10. Focosi, D., Farrugia, A.: The art of the possible in approaching efficacy trials for COVID-19 convalescent plasma. Int. J. Infect. Dis. **102**, 244–246 (2021)

Generalized Additive Models to Capture the Death Rates in Canada COVID-19

Farzali Izadi

2000 Mathematics Subject Classification Primary 68Txx; Secondary 65Y10, 68W20

1 Introduction

In late December 2019, some severe pneumonia cases of unknown cause were reported in Wuhan, Hubei province, China. These cases epidemiologically related to a seafood wholesale market in Wuhan, although many of the first 41 cases were later reported to have no known exposure to the market. In early January 2020, the World Health Organisation (WHO) named this novel coronavirus as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and the associated disease as COVID-19 [14]. Following this report, there has been a rapid increase in the number of cases as on 24th June 2020, there were over 9.2 million confirmed cases and almost 475,000 deaths worldwide, while the confirmed cases in Canada was 103,000, with 8500 deaths mostly distributed in three provinces of Quebec, Ontario, and Alberta. The death rates for these cases which are available from https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html
are shown in Fig. 1.

From the box plot one can see that there are a few outliers in the Alberta data only, see Fig. 2. Also the trend, seasonal and random components of each case can be obtained by decomposition function of time series where in Fig. 3 we plotted only the Canada deaths. To see the normality of the data, we can plot the histogram as well as the density of the deaths, Fig. 4. The plots clearly show that there are substantial overdispersion relative to the Poisson for all cases except possibly for the Alberta data.

F. Izadi (✉)
Mathematics Department, Urmia University, Urmia, Iran
e-mail: f.izadi@utoronto.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_7

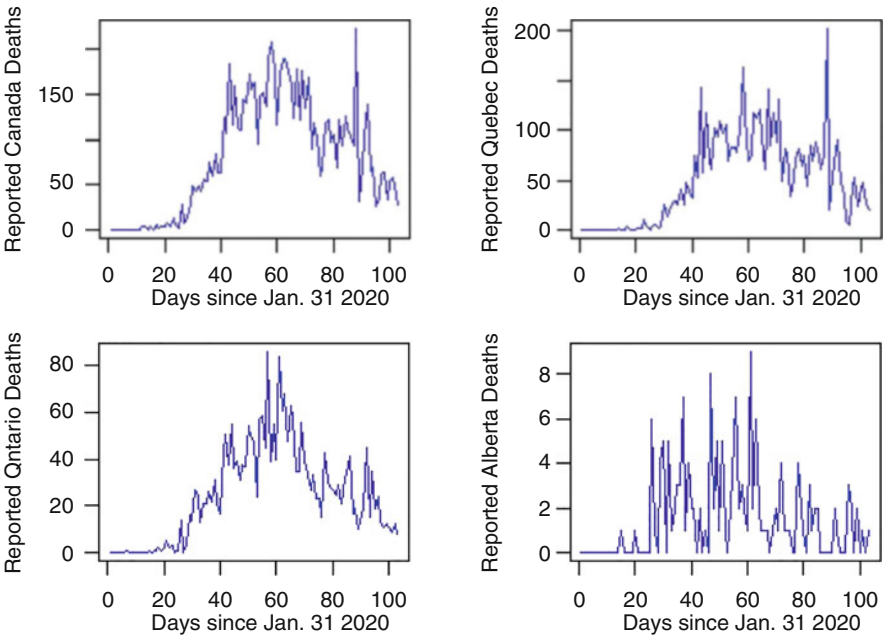


Fig. 1 Reported deaths with COVID-19 in the Canada, Quebec, Ontario and Alberta since January 31, 2020 The Canada lock down started on March 17, 2020

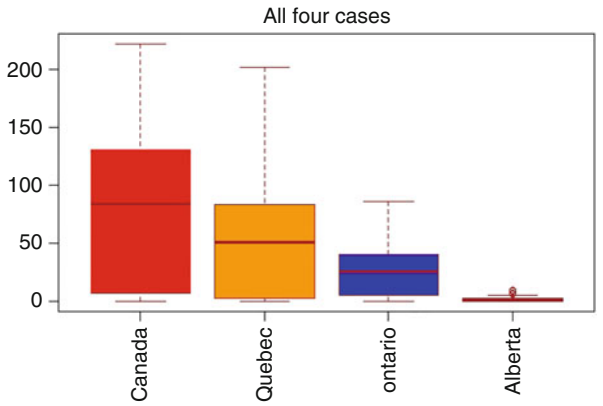


Fig. 2 Note that there are some outliers in Alberta deaths

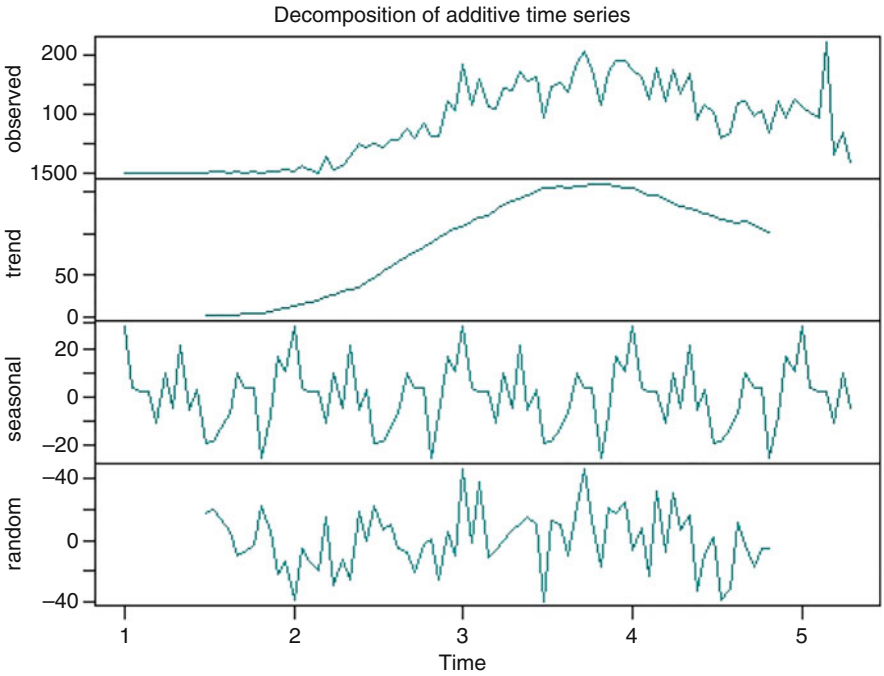


Fig. 3 Decomposition of Canada death series into trend, seasonal and random components

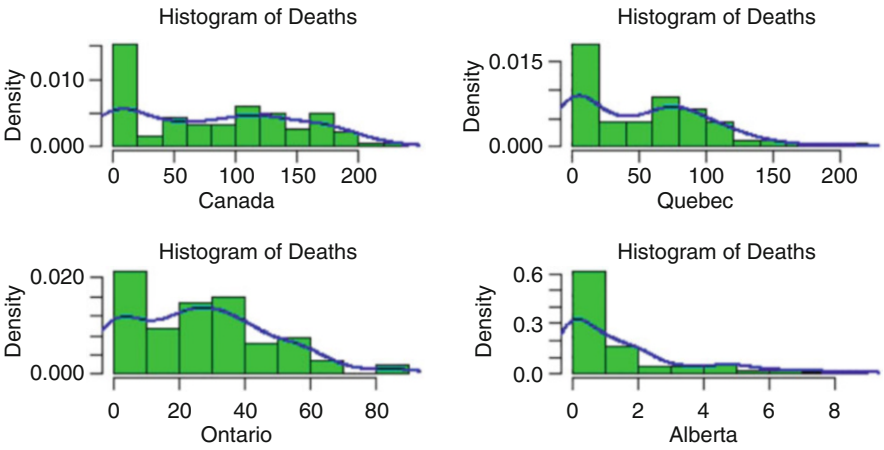


Fig. 4 Histogram and the density of the deaths series. Note that the substantial overdispersion in all cases

2 Generalized Additive Models

In general, generalized additive models (GAMs), see e.g. S. Wood [10] are a semi-parametric extension of the generalized linear models (GLMs), used often for the case when there is no a priori reason for choosing a particular response function (such as linear, quadratic, etc.). GAMs first introduced by Hastie and Tibshirani [7] and Hastie and Tibshirani [8] and are widely used in practice [1–3, 6]. On the other hand, GLMs themselves are extension of the linear models (LMs). To understand the thing better, let us start with LMs. For a univariate response variable of multiple predictors one may simply write

$$y = \alpha X + \varepsilon = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m + \varepsilon, \quad \varepsilon \sim N(\mu, \sigma^2). \tag{1}$$

It is clear that the response variable y is normally distributed with mean μ, and variance σ^2 and the linearity of the model is apparent from the equation. One of the issues with this model is that, the assumptions about the data generating process are limited. One remedy for this is to consider other types of distribution, and include a link function $g(.)$ relating the mean μ, i.e.,

$$E(y) = \mu, \quad g(\mu) = \alpha X. \tag{2}$$

In fact the typical linear regression model is a generalized linear model with a Gaussian distribution and identity link function. To further illustrate the generalization, we may consider a distribution other than the Gaussian for example Poisson or a negative binomial distribution for a count data where the link function is natural log function.

$$g(y) = \alpha X + \varepsilon, \quad \varepsilon \sim D(\mu, \theta), \tag{3}$$

where D is any exponential family distribution. We can still generalize more to add nonlinear terms in the above equation namely

$$g(\mu_i) = X_i \alpha + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \cdots,$$

where $\mu_i = \mathbb{E}(y_i)$ and $y_i \sim$ *an exponential or non-exponential family distribution* and $f_j s$ are any univariate or multivariate functions of independent variables called smooth and nonparametric part of the model, which mean that the shape of predictor functions are fully characterized by the data as opposed to parametric terms that are defined by a set of parameters like the parameter vector α in the linear part in which both to be estimated.

Having said that, the ordinary least square problem generalizes into

$$\mathcal{L}(\alpha) = (\|y - \mathbf{X}\alpha\|^2 + \sum_{j=1}^{n} \lambda_j \int_0^1 [f_j^{(2)}]^2(x)) \tag{4}$$

where λ_j, controlling the extent of penalization and establishes a trade-off between the goodness of fit and the smoothness of the model.

On the other hand, these functions are represented using appropriate intermediate rank spline basis expansions of modest rank k_j, as

$$f_j(x) = \sum_{j=1}^{k_j} \beta_{ji} b_{ji}(x). \tag{5}$$

Substituting the function under the integral sign with the above equation we get

$$\int_0^1 [f_j^{(2)}]^2(x) = \beta^T S_j \beta \tag{6}$$

where the right hand side is a quadratic form with respect to the known matrix S_j.

Collecting both the parametric and non-parametric coefficients into a double parameter (α, β), we obtain

$$\mathcal{L}(\alpha, \beta) = (\|y - \mathbf{X}\alpha\|^2 + \frac{1}{2} \sum_{j=1}^{n} \lambda_j \beta^T S_j \beta). \tag{7}$$

Writing $S_\lambda = \sum_{j=1}^{n} \lambda_j S_j$ we get the more compact form

$$\mathcal{L}(\alpha, \beta) = (l(\alpha) + \frac{1}{2} \lambda_j \beta^T S_\lambda \beta). \tag{8}$$

Let $\hat{\beta}$ be the maximizer of $\mathcal{L}$ and $\mathcal{H}$ the negative Hessian of $\mathcal{L}$ at $\hat{\beta}$. From a Bayesian viewpoint $\hat{\beta}$ is a posterior mode for β. The Bayesian approach views the smooth functions as intrinsic Gaussian random fields with prior given by $N(0, S_\lambda^-)$, where S_λ^- is a Moore–Penrose or pseudoinverse of S_λ. Furthermore in the large sample limit

$$\beta | y \sim N(\hat{\beta}, (\mathcal{H} + S_\lambda)^{-1}). \tag{9}$$

Writing the density in (2.8) as $\mathcal{D}_g$, and the joint density of y and β as $\mathcal{D}(y, \beta)$, the Laplace approximation to the marginal likelihood for the smoothing parameters λ and α is $\mathcal{D}(\lambda, \alpha) = \mathcal{D}(y, \beta)/\mathcal{D}_g(\beta, y)$. Finally, nested Newton iterations are used to find the values of $\log \lambda$ and α; maximizing $\mathcal{D}(\lambda, \alpha)$ and the corresponding $\hat{\beta}$ [11].

3 Application of GAMs for Fatal Cases of Canada

Let y_k denotes the deaths reported on day k. We will study the Canada deaths in general and the provinces of Quebec, Ontario, and Alberta in particular cases. Due to overdispersion seen in the deaths, we assume that y_k follows a negative binomial distribution with mean μ and variance
$\mu + \mu^2/\theta$. Note that $NB(\theta, p)$, where $p = \mu/(\mu + \theta)$.
Then let

$$g(\mu) = f(t_k) + f_w(d_k) + f_b(d_k) + f_m(d_k), \tag{10}$$

where g is a link function log in our discussion, f is a smooth function of time, t_k, measured in days, f_w is a zero mean cyclic smooth function of day of the week $d_k \in \{1, 2, \cdots, 7\}$, set up so that $f_w^{(n)}(0) = f_w^{(n)}(7)$, f_b is a zero mean cyclic smooth function of day of the biweek $d_k \in \{1, 2, \cdots, 14\}$, set up so that $f_b^{(n)}(0) = f_b^{(n)}(14)$, f_m is a zero mean cyclic smooth function of day of the month $d_k \in \{1, 2, \cdots, 30\}$, set up so that $f_m^{(n)}(0) = f_m^{(n)}(30)$, and $n \in \{0, 1, 2\}$ denotes the order of the derivative in f_w, f_b and f_m, see S. Wood 2020 [13]. Based on the discussion and notations in Sect. 1, f, f_w, f_b and f_m are basis functions representing the underlying death rate and the strong weekly, biweekly and monthly cycles seen in the data respectively.

From (2.8) we can easily compute the confidence intervals for each f and make inferences about when the peak in f occurs. This can be done by executing gam function of mgcv library in R code. The fitted models to the reported deaths in Canada, Quebec, Ontario, and Alberta are shown in Figs. 5, 6, 7, and 8 respectively.

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.47710    0.06161   56.43   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
         edf Ref.df  Chi.sq p-value
s(day) 7.675  9.478 727.830  <2e-16 ***
s(dow) 1.671  5.000   5.504  0.0242 *
s(dom) 0.187 28.000   0.203  0.0606 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.856   Deviance explained = 92.8%
-REML =  422.9  Scale est. = 1         n = 103
```

Fig. 5 Results of the fitted model to the Canada deaths. Note that all variable are statistically significant

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.6953     0.1238   21.78   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
         edf Ref.df  Chi.sq p-value
s(day) 6.987  8.704 369.422  <2e-16 ***
s(dow) 1.676  5.000   5.217  0.0301 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.73   Deviance explained =   89%
-REML = 390.08  Scale est. = 1         n = 103
```

Fig. 6 Results of the fitted model to the Quebec deaths. Note that all variable are statistically significant

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.31512    0.09301   24.89   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
         edf Ref.df  Chi.sq  p-value
s(day) 6.315  7.625 371.971  < 2e-16 ***
s(dow) 1.497  5.000   4.156 0.046717 *
s(dom) 1.960 28.000   4.828 0.000793 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.82   Deviance explained = 89.3%
-REML =  324.9  Scale est. = 1         n = 103
```

Fig. 7 Results of the fitted model to the Ontario deaths. Note that all variable are highly significant

In the Canada case, we model the deaths with respect to day, day of week and day of month. As the results of Fig. 5 shows, all the variables are statistically significant with 0.856 as R-squared.

For the deaths of Quebec, we see that the day, day of week is an appropriate choice for the predicted variables. As the results of Fig. 6 shows, all the variables are statistically significant with 0.75 as R-squared.

In the case of the Ontario we see that also the day, day of week and day of month is an appropriate choice for the predicted variables. As the results of Fig. 7 shows, all the variables are highly significant with 0.82 as R-squared.

```
Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.7061     0.1216   22.25   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
          edf Ref.df Chi.sq p-value
s(day)  6.7943  8.438 368.07  <2e-16 ***
s(biw)  0.7526 12.000   2.12  0.0726 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.722   Deviance explained = 88.5%
-REML = 390.72  Scale est. = 1         n = 103
```

Fig. 8 Results of the fitted model to the Alberta deaths. Note that all variable are statistically significant

We figured out that the day, day of biweek as predictors for the Alberta case is an appropriate choice for the model. As the results of Fig. 8 shows, all the variables are statistically significant with 0.722 as R-squared.

The next statistics is the results of the gam.check consists of Q-Q plot where for a good fit the data should lie on the red line. The second is the histogram of the residuals. In this plot, the histogram must be symmetric with respect to the line $x = 0$. Third plot is the Residuals vs. Linear predictors. This plot must be symmetric with respect to the line $y = 0$. Finally, the last plot is the Response vs. Fitted values. The more closer the data to the line $y = x$, the better the fitted model. These plots for the corresponding data of Canada, Quebec, Ontario and Alberta are given in Figs. 9, 10, 11, and 12 respectively.

The other statistics of the model fits to the reported deaths in all cases are shown in Fig. 13, 14, 15, and 16 respectively. The posterior modes (solid) and 95% confidence intervals for the model functions as well as auto-correlation functions and the deviance residuals against day for the Quebec and Alberta.

4 Inference About the Peak of the Day of Each Four Cases

With gam model it is also straightforward to make inferences about when the peak in f occurs. To this end, it is enough to use the model matrix by removing cyclic part, the coefficients and variance-covariance matrix of the model to estimate the model functions and 95% confidence intervals. To find the day of occurrence of the

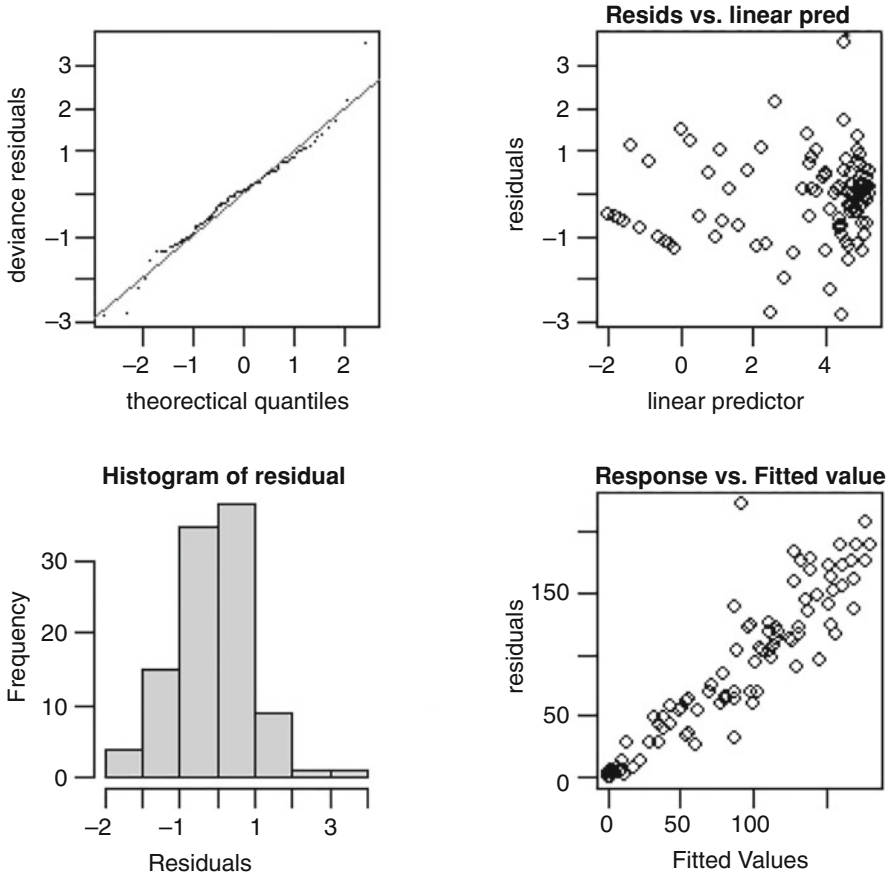


Fig. 9 Canada: Q-Q plot, Histogram of residual, Residuals vs. Linear predictors and Response vs. Fitted values

peak for each corresponding underlying death rate function, f, simply we generates multivariate normal random deviates using 'rmvn' function from 'mgcv' package in which it takes 3 arguments as number of simulations, the coefficients and variance-covariance matrix of the model. The results for all 4 cases are shown on the Figs. 17 and 18 respectively.

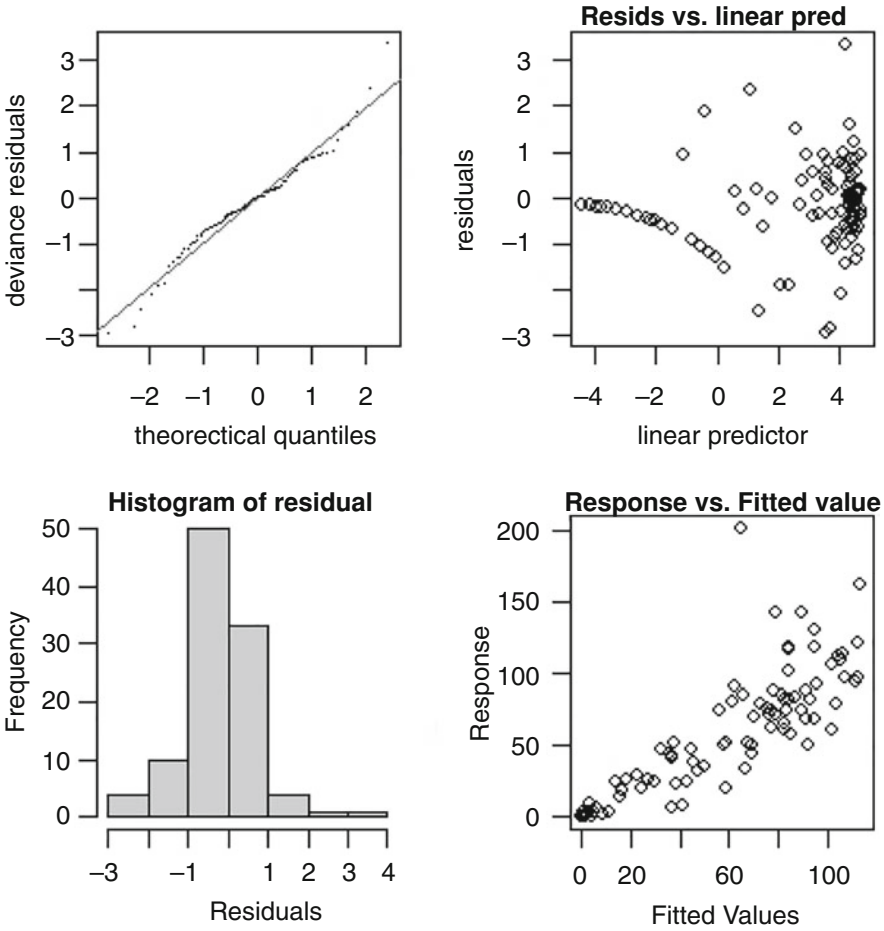


Fig. 10 Quebec: Q-Q plot, Histogram of residual, Residuals vs. Linear predictors and Response vs. Fitted values

5 Inference About the Past Fatal Infection Cases

To obtain the model priors with auto-generated code and associated data is to simulate jagam from rjags library in GAMs. We also load glm to improve samplers for GLMs. This is useful for inference about models with complex random effects structure. The new mgcv function jagam is designed to be called in the same way that the modeling function gam would be called. That is, a model formula and family

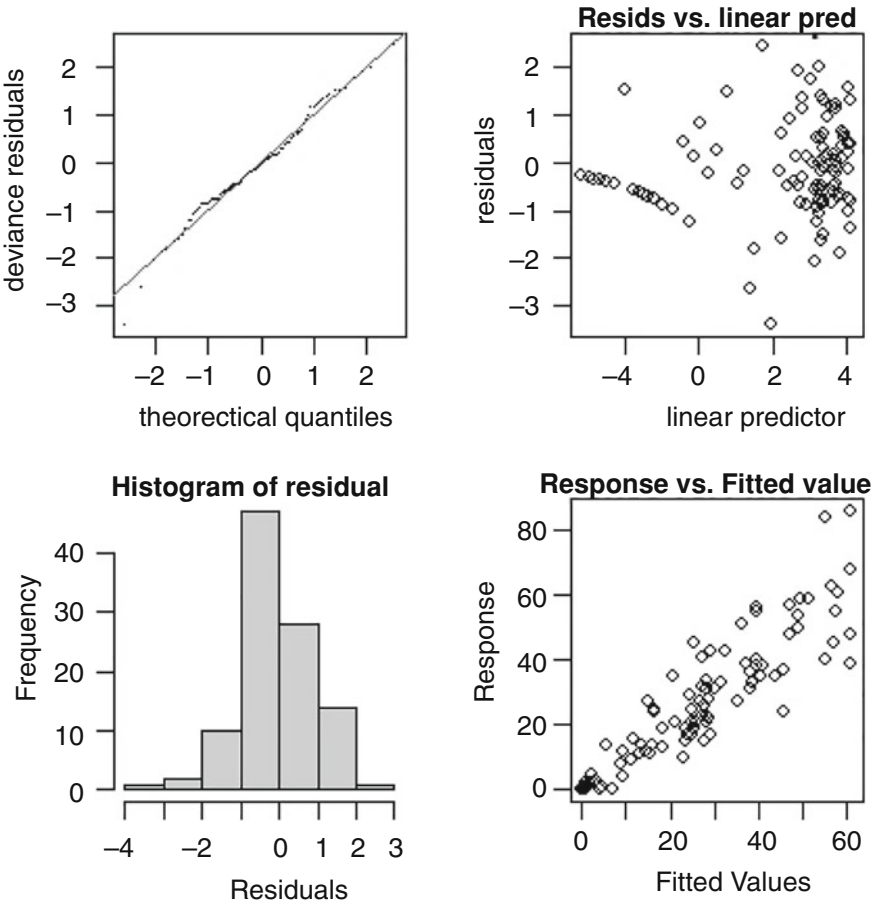


Fig. 11 Ontario: Q-Q plot, Histogram of residual, Residuals vs. Linear predictors and Response vs. Fitted values

object specify the required model structure, while the required data are supplied in a data frame or list or on the search path. However, unlike gam, jagam does no model fitting. Rather it writes JAGS code to specify the model as a Bayesian graphical model for simulation with JAGS, and produces a list containing the data objects referred to in the JAGS code, suitable for passing to JAGS via the rjags function jags.model [5]. To infer the sequence of past fatal infections one needs to produce the observed sequence of deaths. Verity et al. [9] show that the distribution of time from onset of symptoms to death for fatal cases can be modelled by a gamma density

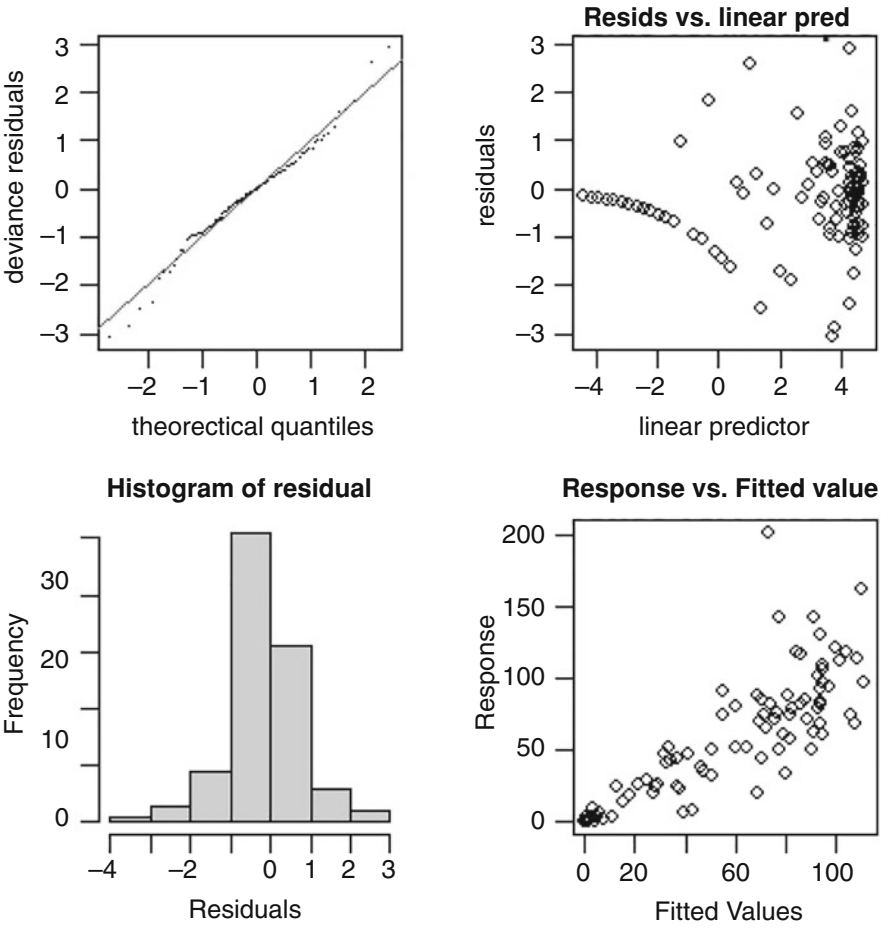


Fig. 12 Alberta: Q-Q plot, Histogram of residual, Residuals vs. Linear predictors and Response vs. Fitted values

with mean 17.8 and variance 71.2 (s.d. 8.44). Let $f_v(t)$ be the function describing the variation in the number of eventually fatal cases over time. Let $\mathbf{B}$ be the lower triangular square matrix of order n, describing the onset-to-death gamma density, given above where n is the number of day of pandemic under consideration. Then we have $\mathbf{B}_{ij} = \gamma(i - j + 1)$ if $i \geq j$ and 0 otherwise. Now $\mathbf{B}f_v = h$, where $h(k)$ is the expected number of deaths on day k. Then $\log(f_v(k))$ can be represented using an intermediate rank spline, again with a cubic spline penalty. We can then

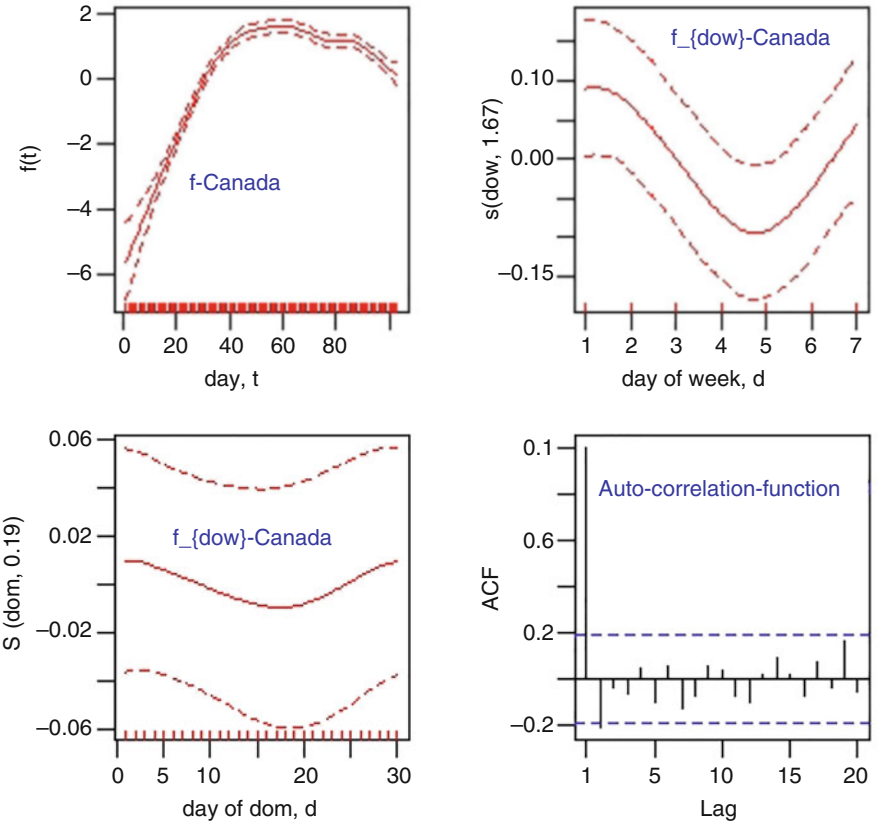


Fig. 13 Posterior modes (solid) and 95% confidence intervals for the model functions for the Canada as well as auto-correlation functions of the deviance residuals. Lag 1 is just significantly negative

employ exactly the model of the previous section but setting $f(k) = \log(h(k))$. The only difference is that we need to infer f_v over a considerable period before the first death occurs where 15 days is clearly sufficient given the form of deaths data. After executing the jagam code for the data frame of deaths, day, day of week, day of biweek and day of month variables—note that we have used different combination of the above variables to see which one is appropriate of Canada deaths in general and Quebec, Ontario and Alberta in particular cases. Jagam code with Poisson family distribution produces a model containing all the priors. Then by adding the matrix **B** and a bit of extra regularization to the output we pass

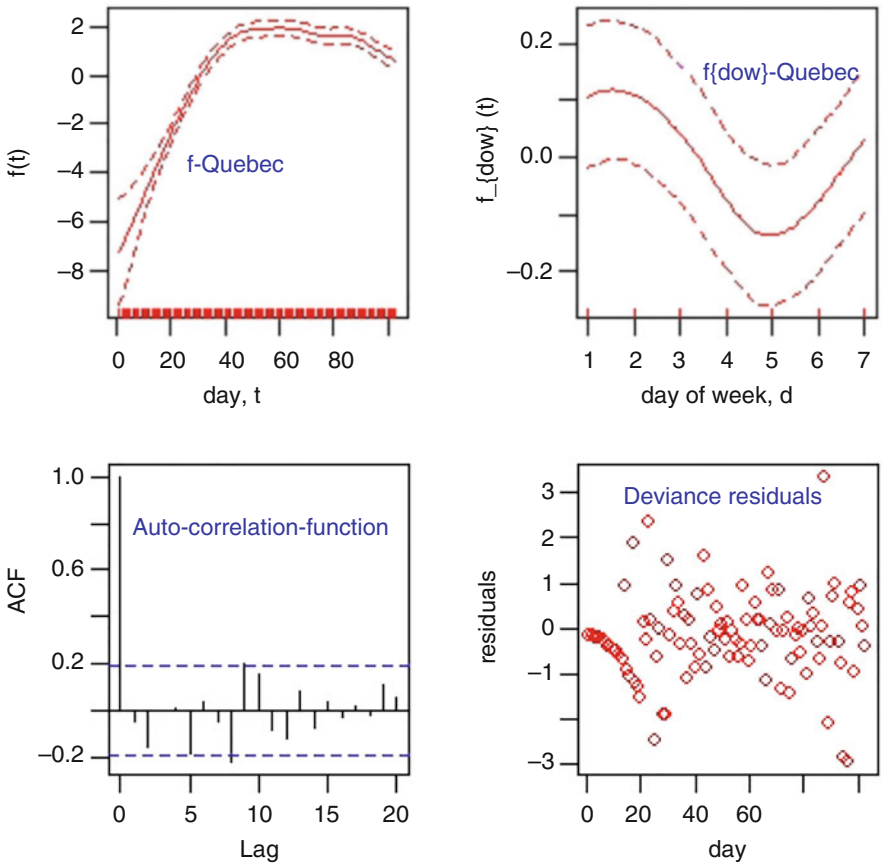


Fig. 14 Posterior modes (solid) and 95% confidence intervals for the model functions for the Quebec as well as auto-correlation functions of the deviance residuals and deviance residuals against day. Lag 7 is just significantly negative

the model to jags.model function JAGS is Just Another Gibbs Sampler. It is a program for the analysis of Bayesian models using Markov Chain Monte Carlo (MCMC) simulation. Then by passing the result to the jags.samples function with the parameters of thin = 300 and iteration = 1,000,000 we get the values of θ, ρ, and the Monte Carlo array b of three dimension with the first dimension equals to the number of parameters in the gam model. As in the gam model, we take the data

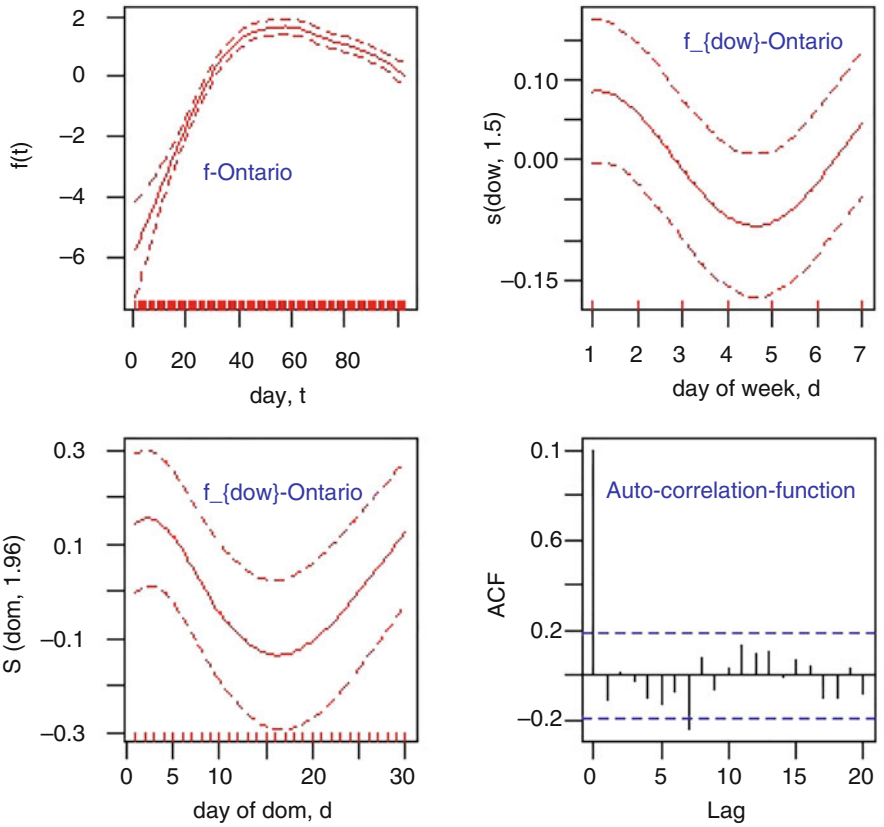


Fig. 15 Posterior modes (solid) and 95% confidence intervals for the model functions for the Ontario as well as auto-correlation functions of the deviance residuals. Lag 7 is just significantly negative

in the first component of the jags model as the model.matrix **X** and the data in first two dimension of the b to simulate the fatal infection profiles $f = \exp(Xb)$ and get all the necessary statistics such as median with confidence intervals, the peak points of the median profile, the squared second difference of the median infection profile on the log scale which is proportional to the smoothing penalty and the absolute gradient of the infection profile. See Wood [12] These are depicted in Fig. 19.

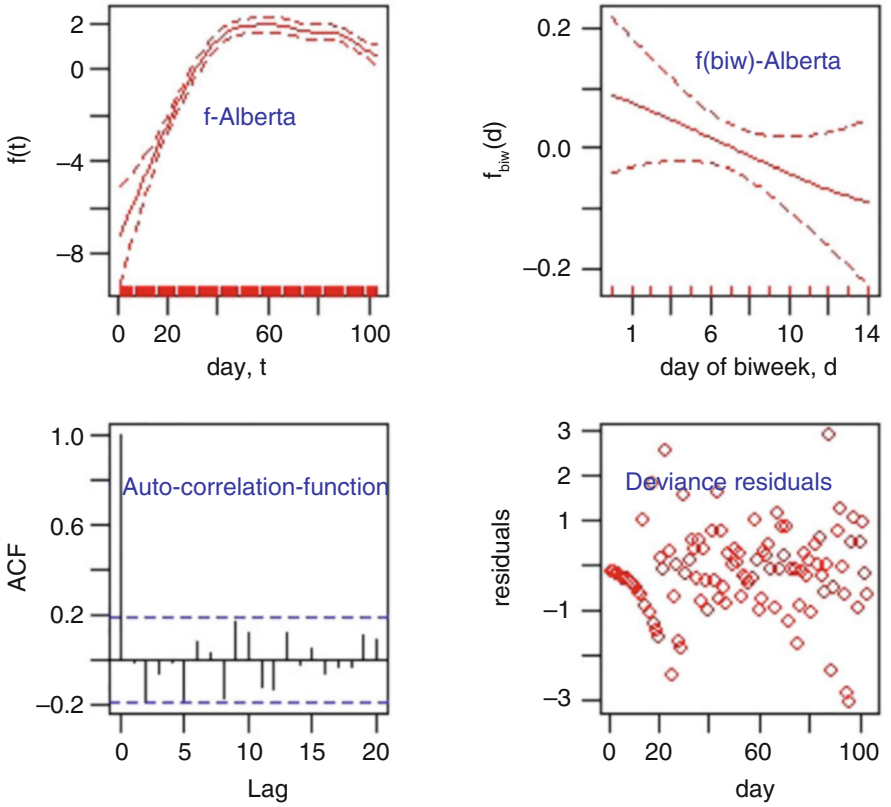


Fig. 16 Posterior modes (solid) and 95% confidence intervals for the model functions for the Alberta as well as auto-correlation functions of the deviance residuals against day. No significant lag

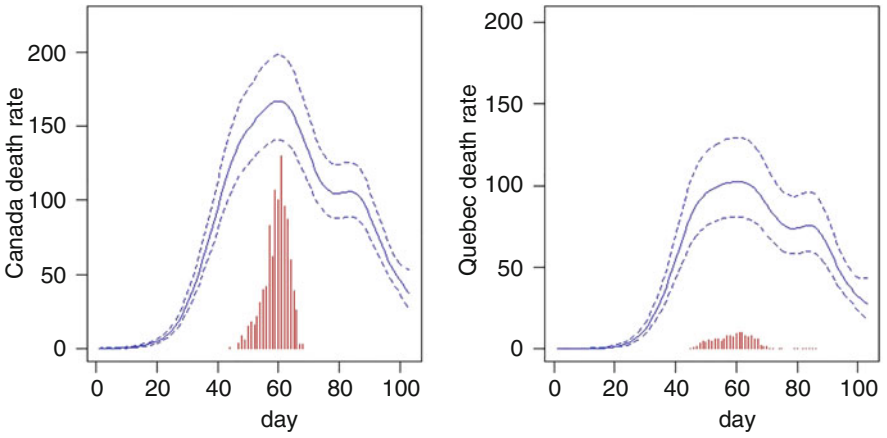


Fig. 17 Underlying death rates for the Canada (left) and Quebec (right) with 95% confidence intervals. Day 0 is March 17th 2020 the Canada lock down. The scaled red bar chart illustrates the distribution of the location of the peak of the death rate curve, obtained by simulation from the approximate posterior for each model

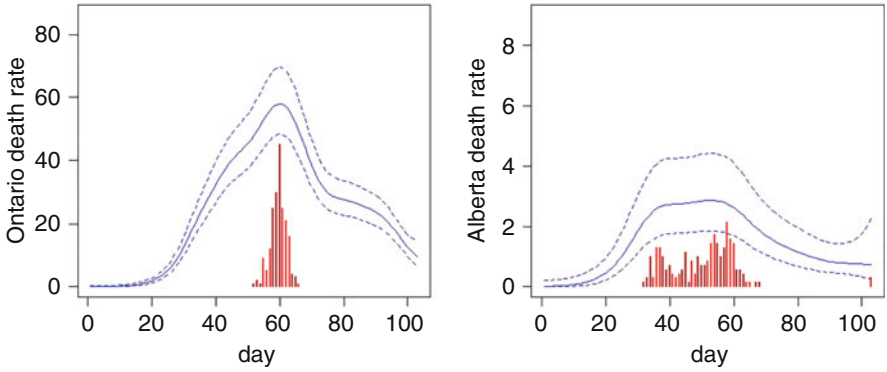


Fig. 18 Underlying death rates for the Ontario (left) and Alberta (right) with 95% confidence intervals. Day 0 is March 17th 2020 the Canada lock down. The scaled red bar chart illustrates the distribution of the location of the peak of the death rate curve, obtained by simulation from the approximate posterior for each model

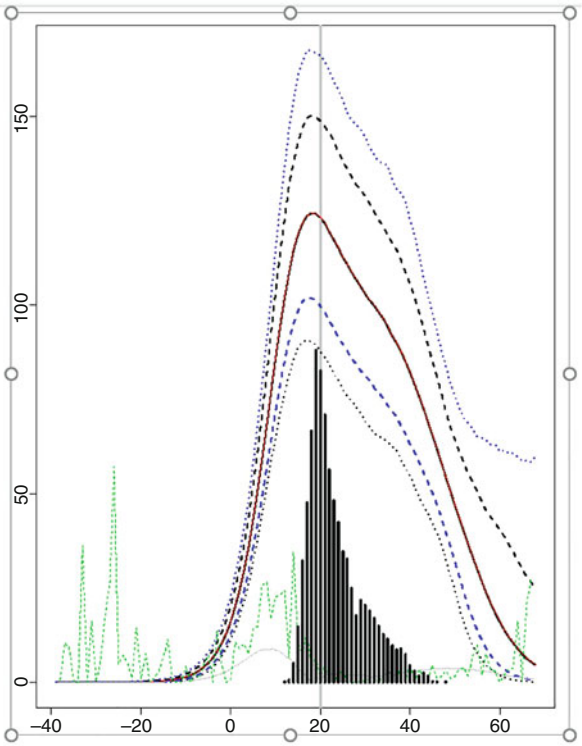


Fig. 19 The inferred Canada fatal infection profile over time (day 0 is March 17th 2020). The red curve is the posterior median, the dashed curves delimit an 80% confidence interval and the dotted curves a 95% confidence interval. The scaled black bar chart shows the posterior distribution of the day of peak fatal infection. The dashed grey curve is proportional to the squared second difference of the median infection profile on the log scale which is proportional to the smoothing penalty. The green curve is proportional to the absolute gradient of the infection profile

Acknowledgments I am very grateful to Simon S. Wood as his paper [13] was very inspirational to my work.

References

1. Fahrmeir, L., and Lang, S. 2001, Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors, Applied Statistics, 50, 201–220.
2. Fahrmeir, L., Kneib, T., and Lang, S. 2004, Penalized Structured Additive Regression for Space-TimeData: ABayesian Perspective, Statistica Sinica, 14, 731–761.
3. Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. 2013, Regression Models, New York: Springer.
4. Green, P. J. and B. W. Silverman 1994. Nonparametric Regression and Generalized Linear Models. Chapman & Hall.

5. Plummer M 2016. rjags: Bayesian Graphical Models Using MCMC. R package version 4-6, URL https://CRAN.R-project.org/package=rjags.
6. Ruppert, D.,Wand, M. P., and Carroll, R. J. 2003, Semiparametric Regression, New York: Cambridge University Press.
7. Hastie, T., and Tibshirani, R. 1986, Generalized Additive Models (with discussion), Statistical Science, 1, 297–318.
8. Hastie, T., and Tibshirani, R. 1990, Generalized AdditiveModels, London: Chapman & Hall.
9. Verity, R., L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, et al. 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases.
10. Wood, S. N. 2017. Generalized Additive Models: An Introduction with R (2 ed.). Boca Raton, FL: CRC press.
11. Wood, S. N., N. Pya, and B. *Säfken* 2016. Smoothing parameter and model selection for general smooth models (with discussion). Journal of the American Statistical Association 111, 1548–1575.
12. Wood, S. N. 2016. Just another Gibbs additive modeller: Interfacing JAGS and mgcv. Journal of Statistical Software 75(7).
13. Wood, S. N. Simple models for COVID-19 death and fatal infection profiles arXiv:2005.02090v1 [stat.AP] 5 May 2020.
14. World Health Organization. www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed 15 Mar. 2020.

Real-Time Prediction of the End of an Epidemic Wave: COVID-19 in China as a Case-Study

Quentin Griette, Zhihua Liu, Pierre Magal, and Robin N. Thompson

1 Introduction

The COVID-19 pandemic has now spread worldwide, causing over one million deaths and 40 million reported cases so far (as of 25 October, 2020 [37]). SARS-CoV-2, the virus that causes COVID-19, emerged in China at the end of 2019. In early 2020, the Chinese government imposed strong public health measures, including enhanced epidemiological surveys and surveillance, travel restrictions, quarantine, contact tracing and isolation [27]. These intense interventions were sufficient to bring the epidemic wave under control, and since mid-March case numbers have remained low.

A key challenge in infectious disease epidemiology is forecasting the progression of an epidemic. Significant attention has been directed towards developing methods for estimating future numbers of cases and deaths, as well as forecasting the timing of the epidemic peak [2–5, 16–20, 28, 32]. Predicting the ends of epidemic waves, on the other hand, has received considerably less attention [23], despite the fact that the end of an epidemic wave signals an opportunity to relax costly public health measures. Some previous studies have estimated the probability that an epidemic is over as a function of the time since the last observed case using renewal equation

Q. Griette · P. Magal (✉)
University of Bordeaux, IMB, Talence, France

CNRS, IMB, Talence, France
e-mail: pierre.magal@u-bordeaux.fr

Z. Liu
School of Mathematical Sciences, Beijing Normal University, Beijing, People's Republic of China

R. N. Thompson
Mathematics Institute, University of Warwick, Coventry, United Kingdom

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_8

models [14, 22] or stochastic compartmental models [31]. However, predicting the end of the first COVID-19 epidemic wave in China was particularly challenging for two key reasons. First, evidence emerged early in the COVID-19 pandemic that infected individuals could transmit the virus prior to displaying symptoms ("presymptomatic infection"). Second, some infected individuals never display symptoms or display only mild symptoms, and therefore do not report disease ("asymptomatic infection"). It is now widely accepted that these presymptomatic and asymptomatic hosts play a significant role in SARS-CoV-2 transmission [6, 12, 29, 30, 33].

Early evidence for asymptomatic transmission included a study by Nishiura et al. [24], which reported early in the pandemic that 13 evacuees on charter flights from Wuhan (China) were infected and four of these individuals never developed symptoms. Chowell et al. [21] estimated the proportion of asymptomatic infections to be 17.9%. Research by Li et al. [15] generated an estimate that 86% of all infections were undocumented (95% CI: [82%–90%]) prior to the introduction of travel restrictions in China on January 23, 2020, and a team in China [35] suggested that there were 37,400 cases in Wuhan that authorities were unaware of by February 18, 2020. More recently, Ferretti et al. [6] split the reproduction number into components corresponding to transmission from symptomatic, presymptomatic and asymptomatic infectious individuals, as well as environmental transmission. Unreported cases, largely due to presymptomatic and asymptomatic infections, were a key driver of the rapid geographic spread of SARS-CoV-2 and explain why early containment of the virus was impossible (compared to, e.g. SARS [7]). In [4], we consider the symptomatic reported and unreported patients and we prove that it is hopeless to estimate the fraction of reported (or unreported) patients by using SI models. In other words, several values of the fraction of reported symptomatic patients give the exact same fit to the data. Finally, a study based on several cohorts of patients was conducted in Oran et al. [26].

Here, we consider a compartmental model characterising SARS-CoV-2 transmission, and parameterise it using data from the first (yet unique) epidemic wave in China. Our model incorporates key features of this epidemic wave, including explicit inclusion of public health measures designed to mitigate the severity of the epidemic, as well as presymptomatic and asymptomatic infections. When we conducted our analysis in real-time, the proportion of infected individuals that were symptomatic and reported disease was unknown (and, in fact, the precise value remains uncertain even now), so we consider a range of values of that parameter (f). We derive an analytic expression for predicting when an epidemic wave is likely to end, under the assumption that public health measures that are in place remain fixed until the epidemic wave is over. We use this expression to show how the predicted end of epidemic wave date changed as the epidemic wave continued, and compare these results to equivalent results obtained using model simulations. Not only do we provide a framework for predicting the ends of epidemic waves, but we also show that the times at which epidemic waves end depend on the proportion of detected cases. This emphasises the importance of intense surveillance to find infectious cases, including those who do not display clear symptoms.

2 Methods

2.1 Data

We use cumulative data describing daily numbers of cases in mainland China from January 20, 2020 to March 18, 2020, obtained from the National Health Commission of the People's Republic of China and Chinese Center for Disease Control and Prevention [38, 39]. Up until February 10 2020, cases in the dataset were only those that were confirmed by laboratory testing. From February 11 to February 15, data were available not only for cases confirmed by laboratory testing, but also for cases that were clinically diagnosed based on medical imaging. From February 16 onwards, these two data types were combined in the dataset, so that it was impossible to distinguish between laboratory confirmed and clinically diagnosed cases. Changing case definitions in response to changes in case numbers is necessary and commonplace [34], however such changes make inferring epidemic trends based on case numbers challenging. To account for this and remove the substantial jump in cases on February 16 due to changes in testing practices, we calculated the cumulative number of clinically diagnosed cases between February 11 and February 15, and subtracted this from the cumulative numbers of cases from February 16 onwards. We therefore obtained approximate numbers of confirmed cases throughout the period from January 20 to March 18, 2020. The dataset, accounting for this adjustment, is shown in the Supplementary Information (Table 4).

We note that, on January 23, mainland China began implementing lockdowns, beginning with a lockdown in the city of Wuhan.

2.2 Mathematical Model

To characterise changes in observed case numbers from January 20 to March 18 in mainland China, we considered a compartmental model in which we track the number of individuals that are either susceptible to the virus ($S(t)$), in early infection and infectious ($I(t)$) and in later infection and reporting disease ($R(t)$) or in later infection and not reporting disease ($U(t)$) [11, 17]. Individuals that are in later infection and not reporting disease include those that are asymptomatic and those who develop only mild symptoms and so do not adhere to interventions targeting symptomatic individuals. The model is therefore given by:

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

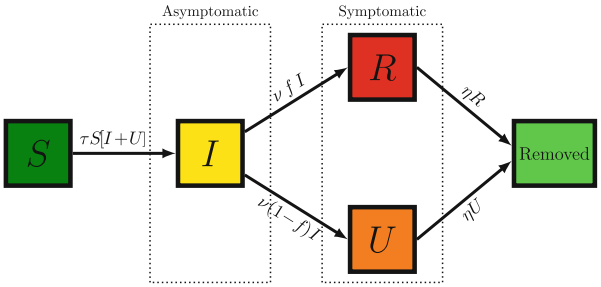


Fig. 1 Schematic showing the different compartments and transition rates in the model given by system of Eqs. (1)

Table 1 Parameters and initial conditions of the model

Symbol	Interpretation	Method
t_0	Epidemic start time	Fitted
S_0	Number susceptible at time t_0	Fixed
I_0	Number in early infection and infectious at time t_0	Fitted
U_0	Number in later infection and not reporting disease at time t_0	Fitted
$\tau(t)$	Transmission rate at time t, accounting for public health measures	Fitted
$1/\nu$	Average duration of early infection	Fixed
f	Fraction of infected individuals that go on to report disease	Fixed
$1/\eta$	Average duration of later infection	Fixed

with initial data

$$S(t_0) = S_0 > 0,\ I(t_0) = I_0 > 0,\ R(t_0) = R_0 \geq 0 \text{ and } U(t_0) = U_0 \geq 0. \tag{2}$$

In this model, $t \geq t_0$ is time in days and t_0 is the start date of the epidemic wave. A schematic illustrating the different model compartments is shown in Fig. 1 and the model parameters—including whether the parameter values were assumed or obtained via model fitting—are listed in Table 1. It has previously been demonstrated that the latent period for COVID-19 is short [18], and COVID-19 patients have been found to have high viral loads early in infection [13, 36], so we do not include individuals who are presymptomatic and not yet infectious in the model. However, explicit inclusion of individuals who are infected but not yet infectious would be a straightforward extension of our model [19].

Early infection (which corresponds to the incubation period, for individuals who develop clear symptoms) is assumed to last for an average period of $1/\nu$ days. The infectious period is assumed to be $1/\nu + 1/\eta$ days, although we assume that individuals that report disease do not transmit the virus during their symptomatic infectious period (i.e. they adhere to public health measures that are effective at reducing transmission). A fraction, f, of infected hosts report disease, whereas a fraction $1 - f$ do not report disease at any stage of their infection.

In the model, the transmission rate at time t, accounting for public health measures in place at that time, is denoted by $\tau(t)$. During the exponential growth phase, we assume that $\tau(t) \equiv \tau_0$ is constant. We then use a time-dependent decreasing transmission rate $\tau(t)$ to incorporate the effects of the strong measures taken by Chinese authorities to control the epidemic wave (see Introduction for a description of the different measures that were introduced):

$$\begin{cases} \tau(t) = \tau_0, \ 0 \leq t \leq N, \\ \tau(t) = \tau_0 \exp\left(-\mu\,(t-N)\right), \ t > N. \end{cases} \tag{3}$$

The date N and the value of μ are chosen so that daily numbers of cumulative reported cases in the numerical simulation of the epidemic align with the analogous values in the dataset.

The cumulative number of reported cases at time t is given by

$$CR(t) = \nu f \int_{t_0}^{t} I(\sigma) d\sigma, \ \text{for } t \geq t_0, \tag{4}$$

and the cumulative number of unreported cases at time t is given by

$$CU(t) = \nu(1-f) \int_{t_0}^{t} I(\sigma) d\sigma, \ \text{for } t \geq t_0. \tag{5}$$

The daily number of reported cases can be obtained by computing the solution of the following equation:

$$DR'(t) = \nu\, f\, I(t) - DR(t), \ \text{for } t \geq t_0 \text{ and } DR(t_0) = 0. \tag{6}$$

2.3 Parameter Values

Since there is substantial uncertainty surrounding the proportion of cases that are symptomatic and report disease for COVID-19, the value of f is unknown. Since intense interventions were introduced in China during the first epidemic wave, and the full extent of asymptomatic transmission was unknown, we assume in the baseline version of our analysis that $f = 0.8$. However, we checked the robustness of our results to this assumption by also considering different values ($f = 0.2, 0.4$ and 0.6).

We assume that the durations of early and late infection are $\nu = 1/7$ days and $\eta = 1/7$ days, respectively. By assuming that the mean duration of early infection (i.e. duration of infection prior to symptoms, for individuals that go on to develop symptoms) is 7 days, the expected generation time for individuals that develop symptoms might be expected to be around 3.5 days. This lies within the range

of estimated generation times for COVID-19 (see e.g. [8]). COVID-19 patients have been found to shed virus up to around 1 week after hospitalisation, thereby motivating our assumed value of η [36].

To determine the initial conditions (Eqs. (2)), we assumed that in the initial exponential growth phase of the epidemic wave (the earliest stages of the epidemic, which is assumed to be between January 19 and January 26, 2020), $CR(t)$ took the form:

$$CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3, \ t \geq t_0. \tag{7}$$

Following [16], expressions for I_0, U_0, R_0 can be obtained:

$$I_0 = \frac{\chi_2}{f(\nu f + \nu_2)}, \ U_0 = \left(\frac{(1-f)(\nu f + \nu_2)}{\eta + \chi_2}\right) I_0, \ R_0 = 0. \tag{8}$$

Furthermore, the transmission rate during this exponential growth phase of the epidemic wave is given by the constant value

$$\tau(t) = \tau_0 = \left(\frac{\chi_2 + \nu f + \nu_2}{S_0}\right)\left(\frac{\eta + \chi_2}{\nu(1-f) + \eta + \chi_2}\right), \tag{9}$$

the epidemic start time is

$$t_0 = \frac{1}{\chi_2}\Big(\log(\chi_3) - \log(\chi_1)\Big), \tag{10}$$

and the value of the basic reproductive number is

$$\mathcal{R}_0 = \left(\frac{\tau_0 S_0}{\nu f + \nu_2}\right)\left(1 + \frac{\nu_2}{\eta}\right). \tag{11}$$

In the above, the value of $\chi_3 = 30$ is assumed and the values of χ_1 and χ_2 are obtained by fitting Eq. (7) to data on the cumulative numbers of cases per day using least squares estimation. Specifically, we use the "polyfit" Matlab function to estimate χ_1 and χ_2. The population size is assumed to be large, so that the initial number of susceptible individuals, S_0, corresponds to the total population size.

3 Results

3.1 Fitting the Model to Data

We first estimated the values of χ_1 and χ_2 using data on the cumulative number of confirmed cases in the earliest stages of the epidemic wave (January 19 to January

26, 2020). The values of τ_0 and the initial conditions (I_0, U_0 and t_0) are then obtained using formulae (8)–(10). The fitted parameter values are shown in Table 2. Analogous results for different values of the reporting fraction, f, are also shown.

We then used the mathematical model (1) with these parameter values and initial conditions to project the cumulative number of reported cases forwards (black line in Fig. 2 (left)), choosing μ so that $CR(t)$ matched the observed data (red dots in Fig. 2 (left)). The inferred cumulative numbers of unreported cases are also shown in Fig. 2 (left), assuming that $f = 0.8$. Daily numbers of reported cases corresponding to this forward projection are shown in Fig. 2 (right).

3.2 *Predicting the End of the Epidemic Wave*

To predict the end of the epidemic wave, we are particularly interested in the time period in which cases are fading out and very few new infections are occurring. We consider a scenario in which the current time is day t_1, and we are attempting to predict when the epidemic will end. As long as t_1 is sufficiently long after the peak of the epidemic wave that the quantity $\tau(t)S(t) \leq \tau(t)S_0$ is small, the approximation

$$I'(t) \simeq -\nu I(t),$$

can be used instead of the second equation in system (1) when $t > t_1$. For the parameter values used in our model, temporal changes in $S_0\tau(t)$ are shown in the Supplementary Information (Fig. 6), highlighting that $\tau(t)S(t)$ is small from the second half of March, 2020, onwards).

Hence, to obtain an analytic expression describing the predicted end of the epidemic wave, we considered the following approximate system of equations whenever $t \geq t_1$:

$$\begin{cases} I'(t) = -\nu I(t), \\ R'(t) = \nu f\, I(t) - \eta R(t), \\ U'(t) = \nu(1-f)\, I(t) - \eta U(t). \end{cases} \tag{12}$$

This system is supplemented by the initial data

$$I(t_1) = I_1, U(t_1) = U_1 \text{ and } R(t_1) = R_1. \tag{13}$$

where I_1, U_1 and R_1 are the values of the solutions of the original system (1)–(2) on day t_1. A schematic for the approximate model (12) is shown in the Supplementary Information (Fig. 7).

The error between the original model and the approximate model is shown in the Supplementary Information (Fig. 8), where the error is given by

Table 2 Values of parameters obtained by fitting to cumulative data from the initial exponential phase of the mainland China epidemic wave. The values of I_0 U_0, τ_0, and t_0 are obtained using formulae (8)–(10). Here we take $\chi_3 = 30$ in order to obtain non-zero integer values of I_0 and U_0

χ_1	χ_2	χ_3	t_0	f	μ	N	I_0	U_0	S_0	τ_0
0.2601	0.3553	30	13.3617	0.8	0.1480	Jan. 26	93.2785	5.3494	1.40005×10^9	3.3655×10^{-10}
0.2601	0.3553	30	13.3617	0.6	0.1531	Jan. 26	124.3550	14.2646	1.40005×10^9	3.1920×10^{-10}
0.2601	0.3553	30	13.3617	0.4	0.1574	Jan. 26	186.5325	32.0953	1.40005×10^9	3.0358×10^{-10}
0.2601	0.3553	30	13.3617	0.2	0.1612	Jan. 26	373.0650	85.5875	1.40005×10^9	2.8942×10^{-10}

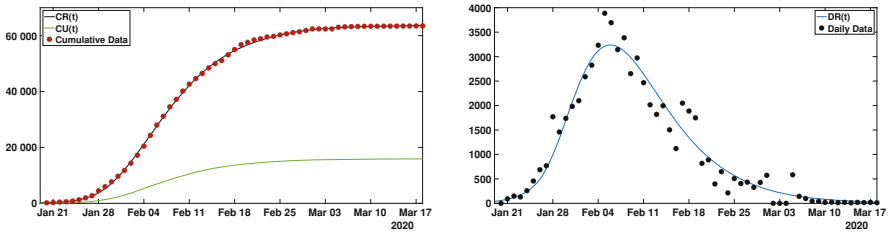


Fig. 2 Comparison of the model output with the data for mainland China. The parameter values and initial conditions are listed in Table 2, and $f = 0.8$. On the left hand side we plot the cumulative data (red dots), the simulated cumulative reported cases $CR(t)$ (black line) and unreported cases $CU(t)$ (green line). On the right hand side, we plot data on the daily numbers of cases (black dots) and the inferred daily number of cases using the model, $DR(t)$ (blue line)

$$\text{err}(t_1) = \sup_{t \geq t_1} \max\left(|I(t) - I_1(t)|, |U(t) - U_1(t)|\right). \tag{14}$$

In this expression, $I(t)$ and $U(t)$ are the solutions of the original system (1), and $I_1(t)$ and $U_1(t)$ are solutions of the approximate model. In both cases, the models are fitted to observed data on cumulative numbers of reported cases (hence, this error formula does not involve $R(t)$ which is very similar for the two models). When applied in the later stages of the epidemic wave, the approximate model is more accurate than earlier in the epidemic wave.

By considering the analogous continuous-time Markov chain to the approximate model (12), the probability that the epidemic is over on different future dates can be estimated analytically (see Supplementary Information Sect. 5 for additional details). Specifically, the probability that no individuals remain in the I or U compartments can be calculated at different times in future:

$$\mathbb{P}(I(s) + U(s) = 0 \text{ for } s \geq t \mid I(t_1) = I_1, U(t_1) = U_1) = \left(1 - e^{-\eta(t-t_1)}\right)^{U_1} \times \left(1 - e^{-\nu(t-t_1)} - (1-f)\nu(t-t_1)e^{-\eta(t-t_1)}\right)^{I_1}. \tag{15}$$

The predictions generated by Eq. (15) for different values of t_1 are shown in Fig. 3. We note that, as t_1 increases, the probability distribution of the date of extinction converges to a limit profile.

Furthermore, we also computed the earliest dates that corresponded to at least 90%, 95% and 99% probabilities that the epidemic was over, for different values of t_1, using Eq. (15). The results of this analysis are shown in Fig. 4.

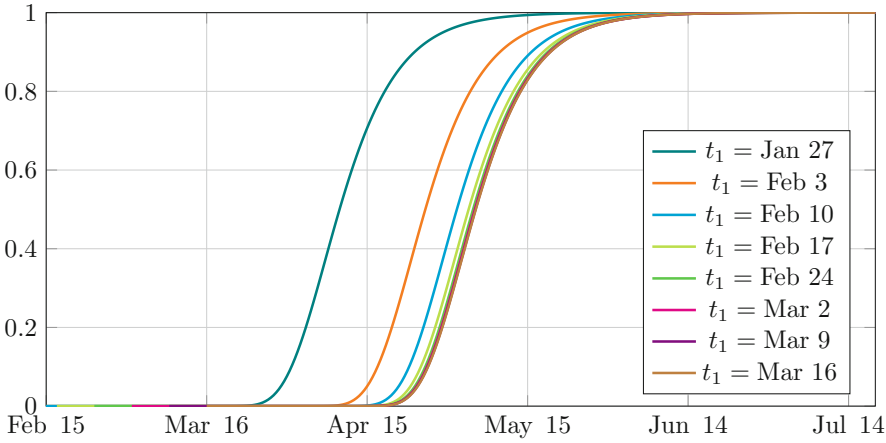


Fig. 3 Estimated extinction probabilities (using Eq. (15)). The numerical values for I_1 and U_1 were computed from the ODE model, considering t_1 values at 7 day intervals. In this figure, we assume that $f = 0.8$ (other parameter values are listed in Table 2)

3.3 Comparing the Analytic Predictions with Stochastic Simulations

To investigate the accuracy of our analytic predictions, we also estimated the end of epidemic time using simulations of the analogous stochastic model to the system of Eqs. (1). Specifically, as before, the deterministic model was fitted to the data on cumulative numbers of confirmed cases and used up until time t_1. Then from time t_1 onwards, stochastic simulations were run using the direct method version of the Gillespie stochastic simulation algorithm [9].

In Fig. 5, we plot the cumulative distribution for the epidemic wave extinction probability obtained using the stochastic simulations. As can be seen in that figure, since the stochastic simulations involve using the exact model (Eqs. (1)) rather than the approximate model, the predicted end dates of the epidemic wave are independent of t_1. The graph in Fig. 5 corresponds to the limit profile discussed at the end of the previous section (i.e. the analytic prediction when t_1 is sufficiently late in the epidemic that the analytic prediction is accurate). From Fig. 3, it can be seen that that this approximation is accurate when t_1 is February 17, 2020, or later.

We also computed the error between the analytic end of epidemic time prediction and the analogous quantity using the stochastic simulations. More precisely, we computed the quantity (Tables 3, 4, and 5)

$$\mathrm{diff}(t_1) = \sup_{t \geq t_1} |f_{IBM}(t) - f_{\mathrm{analytic}}(t)| \tag{16}$$

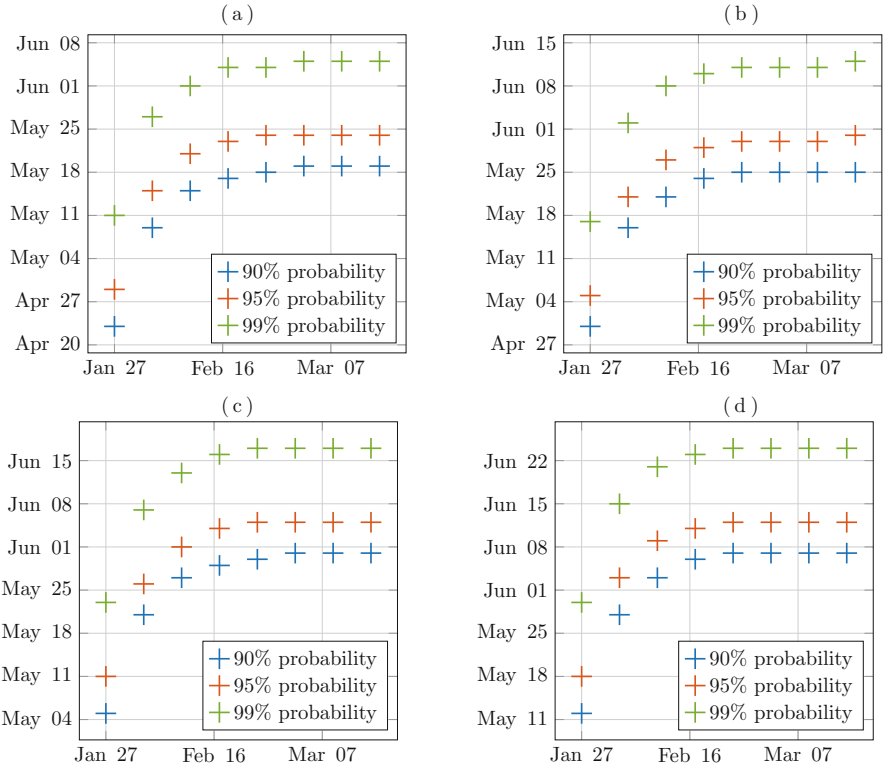


Fig. 4 For each panel, the x-axis corresponds to the day t_1 and the y-axis corresponds to the dates of epidemic wave extinction at different probability levels (90%, 95% and 99%) computed by using (15). Different panels correspond to different values of the parameter f ((**a**) $f = 0.8$; (**b**) $f = 0.6$; (**c**) $f = 0.4$; (**d**) $f = 0.2$). The values of I_1 and U_1 are computed by solving system of Eqs. (1) numerically up to the time $t = t_1$. Parameter values are listed in Table 2

for each value of t_1 presented in Figs. 3 and 5, where f_{IBM} is the cumulative distribution computed by stochastic simulations (Fig. 5) and f_{analytic} is the cumulative distribution given by Eq. (15) (Fig. 3). The results are shown in the Supplementary Information (Table 6, Figs. 6, 7, and 8).

Finally, we compared the mean outputs from the stochastic simulations to the numerical solutions of the original model (system of Eqs. (1)). Unsurprisingly, these quantities match closely (Fig. 9). In Fig. 10, we show the variability between different stochastic simulations obtained when the stochastic simulations are run throughout the epidemic (i.e. starting on day t_0). This high variability observed between different simulations is largely due to the small number of individuals infected initially; when instead stochastic simulations were run from day t_1 onwards, the variability between different stochastic simulations reduced (see Supplementary Information, Table 7).

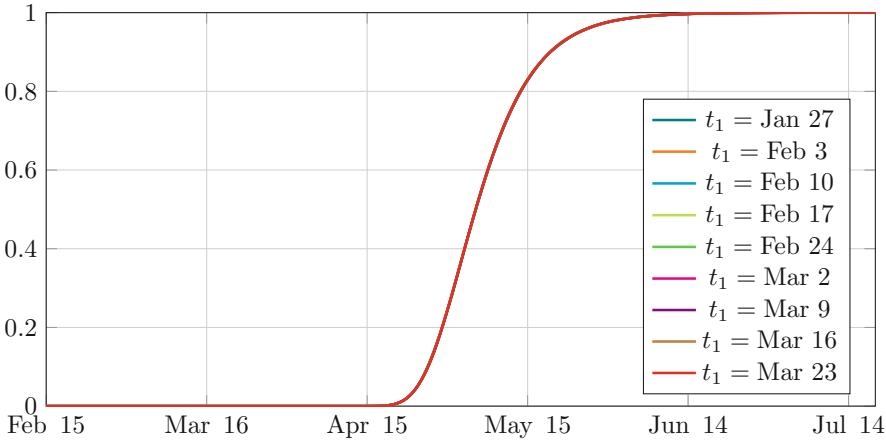


Fig. 5 Estimated cumulative probability distribution for the end of epidemic wave date obtained using stochastic simulations. Results are shown for different values of t_1, although as expected the different lines in this graph lie on top of each other. Initial conditions for the stochastic simulations were computed by rounding the solutions of Eqs. (1) at $t = t_1$ to the nearest integers. 150,000 simulations were run for each value of t_1. In this figure, $f = 0.8$. Other parameter values are shown in Table 2

4 Discussion

Despite receiving surprisingly little attention from epidemiological modellers, predicting the ends of epidemic waves is important for estimating how long intense interventions are likely to be required [14, 22, 23, 31]. In this study, we developed a framework for predicting the ends of epidemic waves using compartmental epidemiological models. This involving fitting a compartmental model to case notification data and using an analytic expression to estimate when the epidemic wave is likely to end. We also compared our analytic prediction to analogous results obtained via model simulations, thereby demonstrating that our results are accurate whenever the underlying epidemiological model provides a realistic reflection of pathogen transmission.

In Table 3, we show the results that we obtained using this framework in real-time to predict the end of the first COVID-19 epidemic wave in China. Specifically, the results in this table correspond to those shown in Fig. 3, after the end of epidemic wave probability converged to the limit profile (i.e. using values of t_1 from approximately mid-February onwards). Importantly, the predicted epidemic wave end date depended on the assumed proportion of infectious cases that report disease (f). Since this quantity was unknown, and remains uncertain even now, we conclude that accurate estimation of the reporting fraction is essential to forecast the ends of epidemic waves accurately.

Our intention here was to develop a basic modelling approach for predicting when an epidemic wave is likely to end. To improve the accuracy of predictions,

Table 3 The predicted end of epidemic wave date inferred when t_1 was March 16, 2020, for different levels of risk aversion. For example, assuming $f = 0.8$, our model predicted a 10% chance that the epidemic wave would persist beyond May 19, 2020

Level of risk	10%	5%	1%
Extinction date ($f = 0.8$)	May 19	May 24	June 5
Extinction date ($f = 0.6$)	May 25	May 31	June 12
Extinction date ($f = 0.4$)	May 31	June 5	June 17
Extinction date ($f = 0.2$)	June 7	June 12	June 24

Table 4 Cumulative data describing confirmed cases in mainland China from January 20, 2020 to March 18, 2020

January						
19	20	21	22	23	24	25
198	291	440	571	830	1287	1975
26	27	28	29	30	31	
2744	4515	5974	7711	9692	11,791	
February						
1	2	3	4	5	6	7
14,380	17,205	20,438	24,324	28,018	31,161	34,546
8	9	10	11	12	13	14
37,198	40,171	42,638	44,653	46,472	48.467	49,970
15	16	17	18	19	20	21
51,091	70,548 − 17,409	72,436 − 17,409	74,185 − 17,409	75,002 − 17,409	75,891 − 17,409	76,288 − 17,409
22	23	24	25	26	27	28
76,936 − 17,409	77,150 − 17,409	77,658 − 17,409	78,064 − 17,409	78,497 − 17,409	78,824 − 17,409	79,251 − 17,409
29						
79,824 − 17,409						
March						
1	2	3	4	5	6	7
79,824 − 17,409	79,824 − 17,409	79,824 − 17,409	80,409 − 17,409	80,552 − 17,409	80,651 − 17,409	80,695 − 17,409
8	9	10	11	12	13	14
80,735 − 17,409	80,754 − 17,409	80,778 − 17,409	80,793 − 17,409	80,813 − 17,409	80,824 − 17,409	80,844 − 17,409
15	16	17	18			
80,860 − 17,409	80,881 − 17,409	80,894 − 17,409	80,928 − 17,409			

this approach would require adjustments to account for important features of real-world epidemic waves. As well as uncertainty in the reporting fraction, another key assumption was that public health measures remained in place until the end of the epidemic wave. Of course, if measures such as isolation of infectious cases

Table 5 Absolute difference between the cumulative distribution given by the stochastic simulations and the reference simulation $t_1 = 82$. For each t_1 we computed the error as $\text{diff}(t_1) = \sup_{t \geq t_1} |f_{t_1}(t) - f_{81}(t)|$, where f_{t_1} is the estimated distribution computed simulations, for which the initial condition correspond to the components of (1) at $t = t_1$ rounded to the closest integer

t_1	26	33	40	47	54	61
Date	Jan. 27	Feb. 3	Feb. 10	Feb. 17	Feb. 24	Mar. 2
$\text{diff}(t_1)$	2.9×10^{-3}	2.1×10^{-3}	2.9×10^{-3}	1.8×10^{-3}	2.5×10^{-3}	1.4×10^{-3}
t_1	68	75	82			
Date	Mar. 9	Mar. 16	Mar. 23			
$\text{diff}(t_1)$	1.6×10^{-3}	1.2×10^{-3}	0.00			

Table 6 Absolute difference between the cumulative distribution given by the stochastic simulations and the analytic approximation using the approximate model (12), computed using Eq. (16)

t_1	26	33	40	47	54	61
Date	Jan. 27	Feb. 3	Feb. 10	Feb. 17	Feb. 24	Mar. 2
$\text{diff}(t_1)$	8.6×10^{-1}	4.4×10^{-1}	1.7×10^{-1}	6.4×10^{-2}	2.5×10^{-2}	8.1×10^{-3}
t_1	68	75	82			
Date	Mar. 9	Mar. 16	Mar. 23			
$\text{diff}(t_1)$	3.5×10^{-3}	8.5×10^{-4}	5.7×10^{-4}			

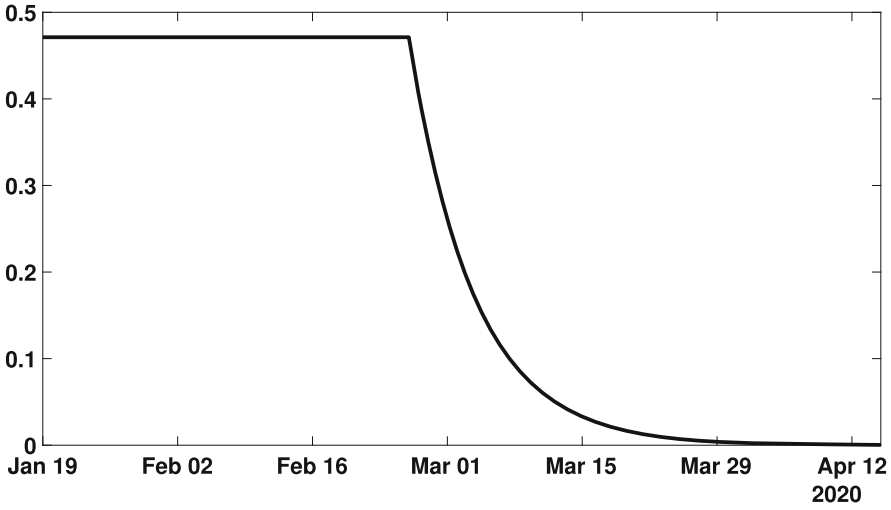


Fig. 6 Graph of $\tau(t)S_0 = \tau_0 S_0 \exp(-\mu \max(t - N, 0))$ with $S_0 = 1.40005 \times 10^9$, $\tau_0 = 3.3655 \times 10^{-10}$, $N =$ Jan. 26, and $\mu = 0.148$. The transmission rate is very small in the second half of March onwards. The parameter values correspond to the baseline case that we considered ($f = 0.8$) see Table 2

are relaxed before an epidemic wave has ended, then the epidemic end date is likely to be different to the one predicted using our modelling framework. In that scenario, relaxation of interventions could in theory be integrated explicitly into the

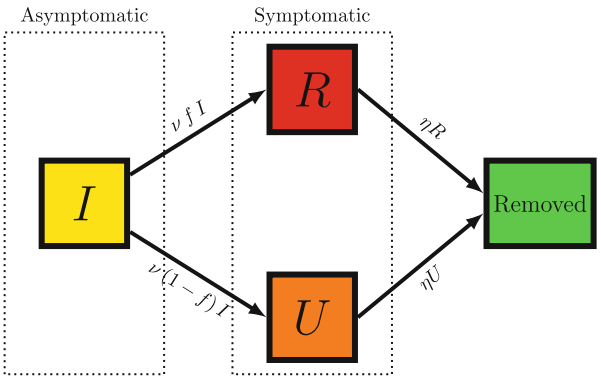


Fig. 7 Schematic of the model (12)

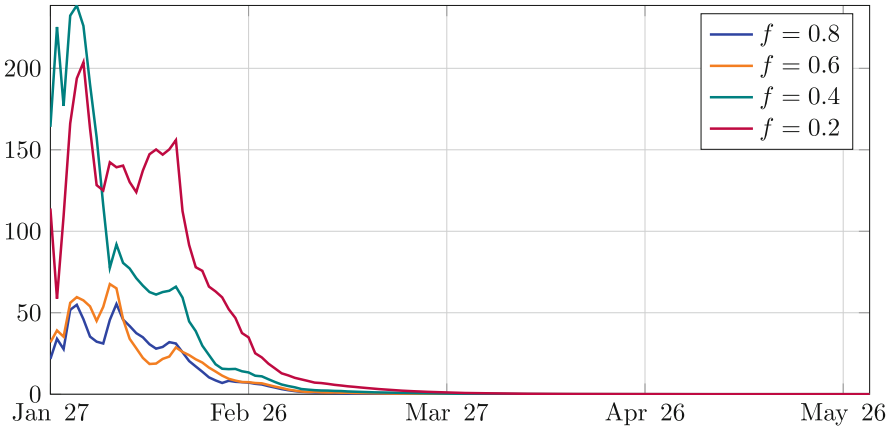


Fig. 8 In this figure the x-axis corresponds to t_1 and the y-axis corresponds to the error $\text{err}(t_1)$ defined in (14). We observe that the smaller f, the larger the error. Parameter values are listed in Table 2

underlying model, and model simulations used to predict the end of epidemic waves. Since interventions are often included in compartmental models [3, 5, 28, 32], this is a straightforward extension of the research presented here. We also note that, if interventions are relaxed following the end of an epidemic wave, then additional cases could begin a second wave—a phenomenon that is now arguably being observed in a range of countries worldwide for COVID-19.

We note that there were very few cases in mainland China after mid-March, 2020. As a result, our modelling framework tended to estimate later end of epidemic wave dates than turned out to be the case. The most likely explanation for this is that, by characterising the impacts of control interventions using Eq. (3), public health measures did not have a sufficiently strong effect in the model. Testing the effects

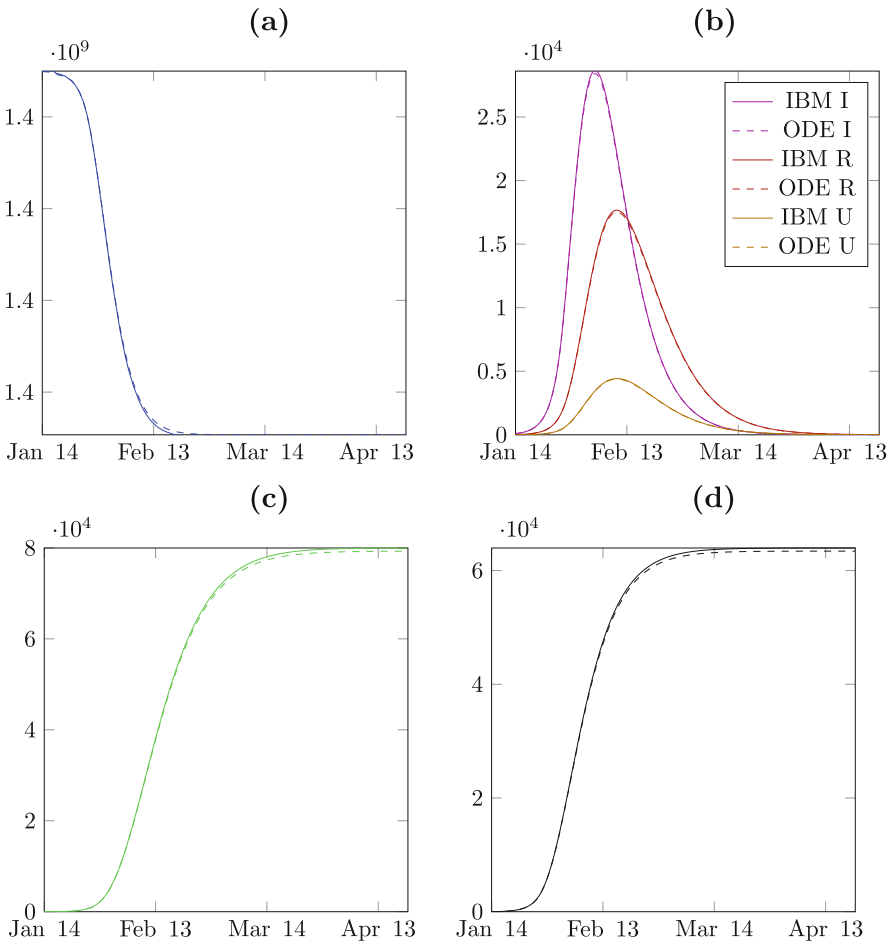


Fig. 9 In figure (**a**) we plot a comparison between the average S (susceptible) computed from the IBM and the S component of the solution of (1). In figure (**b**) we plot a comparison between the average I (asymptomatic), R (reported) and U (unreported) computed from the IBM and the components I, R and U of the solution of (1). In figure (**c**) we plot a comparison between the average RR (removed) computed from the IBM and the components RR of the solution of (1). In figure (**d**) we plot a comparison between the average CR (cumulative reported cases) computed from the IBM and the curve CR computed by (1)–(4). In this figure 500 independent runs of the IBM simulations are used and the corresponding components of the ODE model start from the same initial condition (at $t = t_0$). The parameters we used for both computations are the following: $I_0 = 93$, $U_0 = 5$, $S_0 = 1.40005 \times 10^9 - (I_0 + U_0)$, $R_0 = RR_0 = CR_0 = 0$ and $f = 0.8$, $\tau_0 = 3.3655 \times 10^{-10}$, $N = 26$, $\mu = 0.148$, $\nu = \frac{1}{7}$, $\eta = \frac{1}{7}$, $t_0 = 13.3617$

of different possible characterisations of the effects of public health measures is left as future work.

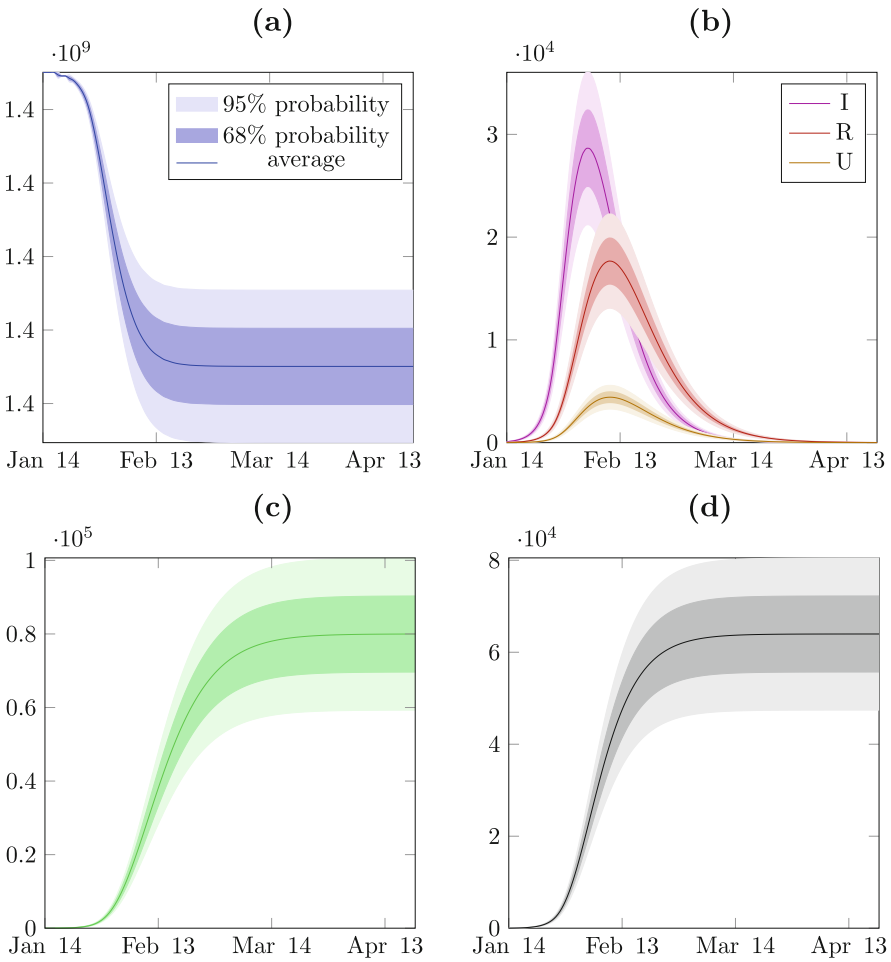


Fig. 10 In figure (**a**) we plot the mean value and variance of S (susceptible) computed from the IBM. The dark blue area contains 68% of the trajectories, and the light blue area 95%. In figure (**b**) we plot the mean value and variance of I (infected), R (reported) and U (unreported) computed from the IBM. The dark areas contains 68% of the trajectories, and the light areas 95%. In figure (**c**) we plot the mean value and variance of RR (removed) computed from the IBM. The dark green area contains 68% of the trajectories, and the light green area 95%. In figure (**d**) we plot the mean value and variance of CR (cumulated reported) computed from the IBM. The dark gray area contains 68% of the trajectories, and the light gray area 95%. We use 500 independent runs of the IBM simulations. The parameters we used for both computations are the following: $I_0 = 93$, $U_0 = 5$, $S_0 = 1.40005 \times 10^9 - (I_0 + U_0)$, $R_0 = RR_0 = CR_0 = 0$ and $f = 0.8$, $\tau_0 = 3.3655 \times 10^{-10}$, $N = 26$, $\mu = 0.148$, $\nu = \frac{1}{7}$, $\eta = \frac{1}{7}$, $t_0 = 13.3617$

Since the precise method of parameter inference was not central to our framework, we used a basic approach to estimate the values of pathogen transmission parameters here, namely least squares estimation. Many different methods are used

Table 7 Maximal standard deviation for the components I, R and U computed by stochastic simulations started at date t_1 with initial condition given by the solution to (1) with the parameters from Table 2. The ODE model (1) is solved up to $t = t_1$, and we take the solution to (1) at $t = t_1$ as initial condition for the stochastic simulations. $\sigma(t)$ is the maximum, at time t, of the standard deviations of the quantities $I(t)$, $R(t)$ and $U(t)$ in a sample of $n = 1000$ independent simulations started at $t = t_1$, and is expressed in number of individuals. We took $f = 0.8$ and other parameters are taken from Table 2

t_1	t_0	18	22	26	33	40	
Date	Jan. 14	Jan. 19	Jan. 23	Jan. 27	Feb. 3	Feb.10	
$\max_{t \geq t_1} \sigma(t)$	3717	1685	787	401	186	106	

to estimate transmission parameters in real-time during epidemics [1, 25], and our modelling framework could be extended to use these more sophisticated methods.

Despite the many simplifications in our modelling approach as presented here, we have provided an initial framework for predicting the ends of epidemic waves, and demonstrated the key principle that the end date of an epidemic wave depends sensitively on the proportion of infectious cases that report disease. Extending this framework to include additional epidemiological realism, so that ends of epidemic waves can be forecasted as accurately as possible, is an important target for future research. This will allow public health decision makers to plan control interventions effectively during infectious disease epidemics.

5 Supplementary Information

5.1 Formula to Compute the Probability Distribution of the Extinction Date

We use continuous-time Markov processes to compute the exact distribution of the date of end of the epidemic after the transmission rate is effectively taken as zero. We start on t_1 with initial values I_1, U_1, and R_1 for I-individuals, U-individuals and R-individuals, respectively. The evolution of each individual is guided by independent exponential processes, and we have the following:

(i) Each individual I will change state following an exponential clock of rate ν. When I changes its state, it will be transferred to the class of R-individuals with probability f and to the class of U-individuals with probability $(1 - f)$;
(ii) Each individual in the state U will change state following an exponential clock with rate η and become removed individual;
(iii) Each individual in the state R will change state following an exponential clock with rate η and become removed individual

Since the class I has only outgoing fluxes, the law of extinction for the I-individuals is

$$\mathbb{P}(I(t) = 0 \mid I(t_1) = I_1) = \left(\int_{t_1}^{t} \nu e^{-\nu(s-t_1)} \mathrm{d}s\right)^{I_1} = \left(1 - e^{-\nu(t-t_1)}\right)^{I_1},$$

and the probability to have some I-individual left at time t is

$$\mathbb{P}(I(t) = I \mid I(t_1) = I_1) = (1 - e^{-\nu(t-t_1)})^{I_1 - I} e^{-\nu I (t-t_1)}.$$

For the U-individuals and the R-individuals, the situation is more intricate. Indeed, the U-individuals and the R-individuals vanish at a constant rate η but new individuals appear from the I class at rate $(1-f)\nu$ and $f\nu$, respectively, depending on the remaining stock of I. Therefore the probability that U gets extinct before t also depends on the number of remaining I. It is actually easier to compute directly the extinction property for the sum $I + U$, which is our aim anyways.

When $\nu \neq \eta$, we obtain

$$\begin{aligned}
&\mathbb{P}(I(s) + U(s) = 0 \,\forall s \geq t \mid I(t_1) = I_1, U(t_1) = U_1) \\
&\quad = \left(1 - e^{-\eta(t-t_1)}\right)^{U_1} \\
&\qquad \times \left(\int_{t_1}^{t} \mathbb{P}(U \to RR \text{ before } t \mid I \to U \text{ at } s)\mathbb{P}(I \to U \text{ at } s) + \mathbb{P}(I \to R \text{ at } s)\mathrm{d}s\right)^{I_1} \\
&\quad = \left(1 - e^{-\eta(t-t_1)}\right)^{U_1} \\
&\qquad \times \left(\int_{t_1}^{t} \left(1 - e^{-\eta(t-s)}\right) \times (1-f)\nu e^{-\nu(s-t_1)} + f\nu e^{-\nu(s-t_1)} \mathrm{d}s\right)^{I_1} \\
&\quad = \left(1 - e^{-\eta(t-t_1)}\right)^{U_1} \\
&\qquad \times \left((1-f)\left(1 - e^{-\nu(t-t_1)} - \nu \frac{e^{-\nu(t-t_1)} - e^{-\eta(t-t_1)}}{\eta - \nu}\right) + f(1 - e^{-\nu(t-t_1)})\right)^{I_1} \\
&\quad = \left(1 - e^{-\eta(t-t_1)}\right)^{U_1} \times \left(1 - e^{-\nu(t-t_1)} - (1-f)\nu \frac{e^{-\nu(t-t_1)} - e^{-\eta(t-t_1)}}{\eta - \nu}\right)^{I_1},
\end{aligned}$$

where the RR-individuals are the removed individuals.

Similarly when $\eta = \nu$, we obtain

$$\begin{aligned}
&\mathbb{P}(I(s) + U(s) = 0 \,\forall s \geq t \mid I(t_1) = I_1, U(t_1) = U_1) \\
&\quad = \left(1 - e^{-\eta(t-t_1)}\right)^{U_1} \times \left(1 - e^{-\nu(t-t_1)} - (1-f)\nu(t-t_1)e^{-\eta(t-t_1)}\right)^{I_1}.
\end{aligned} \tag{17}$$

5.2 Cumulative Distribution of the Date of End of the Epidemic

The stochastic simulations introduced in Sect. 3.3 can be used, in particular, to precisely estimate the cumulative probability distribution of the date of end of the epidemic, defined as the last time at which the quantity $I + U$ is positive.

In order to get a measure of the precision we remark that the values taken by the cumulative probability distribution $f(t)$ can be estimated by the average of independent measures of the random variable

$$X = \mathbb{1}_{t_{ext} \leq t},$$

which follows an Bernouilli distribution of parameter $f(t)$. Consecutive runs of the individual-based simulations yield independent observations X_n of this distribution. By Hoeffding's inequality we have for all $\varepsilon > 0$ and $n \in \mathbb{N}$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_n - f(t)\right| \geq \varepsilon\right) \leq 2\exp\left(-2\varepsilon^2 n\right) =: \alpha,$$

and we achieved an error of at most $\varepsilon = 10^{-3}$ at risk $\alpha \leq 10^{-3}$ by running $n = -\frac{2}{\varepsilon^2}\ln\left(\frac{\alpha}{2}\right) \approx 15201805$ independent individual-based simulations to estimate the probability distribution of the extinction time (Fig. 5, $t_1 = 82$ i.e. March 23). Other curves are estimated on the basis of 152019 independent simulations, which amounts to an error of at most 10^{-2} at risk 10^{-3}.

Since the curves presented in Fig. 3 are so similar that it is difficult to see any difference between them, we computed the absolute error between each curve and the "reference" of $t_1 = 82$. We present the numerical values in Table 5. Notice that the error is actually below the estimated precision of the approximation.

5.3 Supplementary Figures

5.4 Supplementary Tables

We use cumulative reported data from the National Health Commission of the People's Republic of China and the Chinese CDC for mainland China. Before February 11, the data was based on laboratory confirmations. From February 11 to February 15, the data included cases that were not tested for the virus, but were clinically diagnosed based on medical imaging (patients that showed signs of pneumonia). There were 17,409 such cases from February 11 to February 15. The data from February 11 to February 15 specified both types of reported cases. From February 16, the data did not separate the two types of reporting, but reported the sum of both types. We therefore subtracted 17,409 cases from the cumulative

reported cases after February 15 to obtain approximate data for the cumulative numbers of reported cases based only on laboratory confirmations after February 15.The data is given in Table 4 with this adjustment.

Acknowledgments The numerical simulations presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr/).

Author Contributions All authors conceived and designed the study. Q.G. and P.M. analysed the data, carried out the analysis and performed the numerical simulations, Z.L. and P.M. conducted the literature review. All authors participated in writing and editing the manuscript.

Funding This research was funded by the National Natural Science Foundation of China (grant number: 11871007 (ZL)), NSFC and CNRS (Grant number: 11811530272 (ZL, PM)) and the Fundamental Research Funds for the Central Universities (ZL). This research was also funded by the Agence Nationale de la Recherche in France (Project name : MPCUII (QG, PM)), and Christ Church, Oxford, via a Junior Research Fellowship (RNT).

Conflicts of Interest The authors declare no conflict of interest.

References

1. T. Britton and E. Pardoux, *Stochastic Epidemic Models with Inference*, Springer (2019).
2. R.M. Cotta, C.P. Naveira-Cotta and P. Magal (2020), Modelling the COVID-19 epidemics in Brasil: Parametric identification and public health measures influence, Biology 2020, 9(8), 220.
3. N.G. Davies et al., Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study, *Lancet Public Health*, **5(7)** (2020), e375–e385.
4. J. Demongeot, Q. Griette and P. Magal (2020) SI epidemic model applied to COVID-19 data in mainland China Royal Society Open Science (2020), 7:201878.
5. N.M. Ferguson et al., Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, (2020), see: See www.imperial.ac.uk/mrc-global-infectious-disease-analysis/ covid-19/report-9-impact-of-npis-on-covid-19/.
6. L. Ferretti et al., Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, *Science*, **368(6491)** (2020), eabb6936.
7. C. Fraser et al., Factors that make an infectious disease outbreak controllable, *PNAS* **101(16)** (2004), 6146–6151.
8. T. Ganyani et al., Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020, *Eurosurveillance* **25** (2020), 2000257.
9. D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* **8** (1977), 2340–2361.
10. Q. Griette and P. Magal, Clarifying predictions for COVID-19 from testing data: the example of New-York State, (2021) Volume 6 (2021), 273–283.
11. Q. Griette, P. Magal and O. Seydi, Unreported cases for Age Dependent COVID-19 Outbreak in Japan, *Biology* **9** (2020), 132.
12. W. Guan et al., Clinical Characteristics of Coronavirus Disease 2019 in China, *New England Journal of Medicine*, (2020). Published on February 28, 2020, PMID: 32109013. https://doi.org/10.1056/NEJMoa2002032.
13. X. He et al., Temporal dynamics in viral shedding and transmissibility of COVID-19, *Nature Medicine*, **26** (2020), 672–675.

14. H. Lee and H. Nishiura, Sexual transmission and the probability of an end of the Ebola virus disease epidemic. *Journal of theoretical biology*, **471** (2019), 1–12.
15. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang and J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* (2020). https://doi.org/10.1126/science.abb3221
16. Z. Liu, P. Magal, O. Seydi and G. Webb, Understanding unreported cases in the 2019-nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions, *Biology*, **9(3), 50** (2020). https://doi.org/10.3390/biology9030050
17. Z. Liu, P. Magal, O. Seydi and G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data, *Mathematical Biosciences and Engineering* **17(4)** (2020), 3040–3051. https://doi.org/10.3934/mbe.2020172
18. Z. Liu, P. Magal, O. Seydi and G. Webb, A COVID-19 epidemic model with latency period, *Infectious Disease Modelling* **5** (2020), 323–337.
19. Z. Liu, P. Magal, O. Seydi and G. Webb, A model to predict COVID-19 epidemics with applications to South Korea, Italy, and Spain, *SIAM News* May 01 2020.
20. Z. Liu, P. Magal, G. Webb, Predicting the number of reported and unreported cases for the COVID-19 epidemics in China, South Korea, Italy, France, Germany and United Kingdom, *Journal of Theoretical Biology*, Volume 509, 21 (2021).
21. K. Mizumoto, K. Kagaya, A. Zarebski and G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25(10)** (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180
22. H. Nishiura, Y. Miyamatsu and K. Mizumoto, Objective determination of end of MERS outbreak, South Korea, 2015, *Emerging infectious diseases*, **22(1)** (2016), 146.
23. H. Nishiura, Methods to determine the end of an infectious disease epidemic: A short review, In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases* (eds. G. Chowell, J.M. Hyman), (2016), 291–301.
24. H. Nishiura et al., Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19), *International Journal of Infectious Diseases*, (2020). Published: March 13, https://doi.org/10.1016/j.ijid.2020.03.020.
25. P.D. O'Neill, Introduction and snapshot review: Relating infectious disease transmission models to data, *Statistics in Medicine*, **29** (2010), 2069–2077.
26. D. P. Oran, & E. J. Topol, (2020) Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. *Annals of internal medicine*, **173(5)**, 362–367.
27. A. Pan et al., Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China, *JAMA*, **323(19)** (2020), 1915–1923.
28. K. Prem 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*, **5** (2020), e261–e270.
29. J. Qiu, Covert coronavirus infections could be seeding new outbreaks, *Nature*, (2020). https://www.nature.com/articles/d41586-020-00822-x
30. C. Rothe et al., Transmission of 2019-nCoV infection from an asymptomatic contact in Germany, *New England Journal of Medicine*, (2020). https://doi.org/10.1056/NEJMc2001468
31. R. N. Thompson, O. W. Morgan and K. Jalava, Rigorous surveillance is necessary for high confidence in end-of-outbreak declarations for Ebola and other infectious diseases, *Philosophical Transactions of the Royal Society B*, **374(1776)** (2019), 20180431. https://doi.org/10.1098/rstb.2018.0431
32. R. N. Thompson, Epidemiological models are important tools for guiding COVID-19 interventions, *BMC Medicine*, **18** (2020), 152.
33. R. N. Thompson, F. A. Lovell-Read and U. Obolski, Time from Symptom Onset to Hospitalisation of Coronavirus Disease 2019 (COVID-19) Cases: Implications for the Proportion of Transmissions from Infectors with Few Symptoms. *Journal of Clinical Medicine*, **9(5)** (2020), 1297.

34. T. K. Tsang et al., Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health*, **5** (2020), e289–296.
35. C. Wang et al., Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China, *medRxiv*. https://doi.org/10.1101/2020.03.03.20030593
36. R. Wölfel et al., Virological assessment of hospitalized patients with COVID-2019, *Nature*, (2020). https://doi.org/10.1038/s41586-020-2196-x
37. Worldometer, Covid-19 Coronavirus Pandemic, (2020), see: See http://www.worldometers.info/coronavirus/.
38. The National Health Commission of the People's Republic of China http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml (accessed on 10 April 2020)
39. Chinese Center for Disease Control and Prevention. http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11809/ (accessed on 10 April 2020)

The Effect of Heterogeneity in Social Distancing on the Infection Peak for COVID-19

Connell McCluskey

1 Introduction

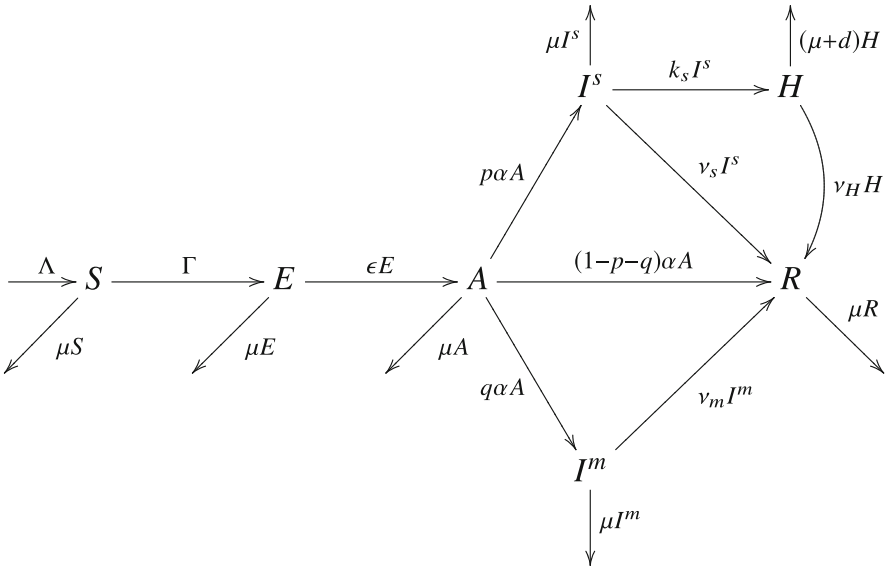
For many models, analytical calculations can give great insight into the beginning of an outbreak (the period of exponential growth) and also into the long-term behaviour (asymptotic stability). However, with COVID-19 it is extremely important to understand what happens between these extremes. Insight into questions about the timing and height of peaks is of fundamental importance, and is one focus of this work.

Another goal of this work is to provide insight into the effect of having different subgroups of a population that engage in social distancing (or other means of contact reduction) at different levels. Along with general analysis for multiple groups, we provide numerical simulations that focus on the case of two levels of social distancing.

The idea of social distancing as a policy to help control the spread of COVID-19 was introduced to Canada in March 2020; many other places adopted similar policies around the same time. Until then, Canadian mixing patterns had not substantially reacted to the novel virus. Before social distancing began, the following transfer diagram (which is a simplification of the model used in [6]) would have been appropriate.

C. McCluskey (✉)
Wilfrid Laurier University, Waterloo, ON, Canada
e-mail: cmccluskey@wlu.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_9



The compartments in the transfer diagram are:

S	Susceptible
E	Exposed (latently infected)
A	Asymptomatically Infectious
I^s	Infectious (and symptomatic) - Serious case
I^m	Infectious (and symptomatic) - Mild case
H	Hospitalized (and at risk of dying)
R	Recovered

We now split each of these compartments into n subgroups based on the level of engagement in social distancing. Thus, we have $S_1, S_2, \ldots, S_n, E_1, E_2, \ldots, E_n$, and so on for the other compartments that appear in the transfer diagram. Suppose the baseline contact rate of the population, before social distancing policies are introduced, is c. For $j = 1, \ldots, n$, the contact rate for group j, after social distancing policies are introduced, is $c\tau_j$, where we expect (but don't require) that $\tau_j \in [0, 1]$. Aside from social distancing, the n subgroups of an epidemiological compartment are assumed to be identical. For intuition, we assume[1] $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$. For the purposes of our analysis, the period before the introduction of social distancing is ignored; we assume social distancing was already in place at the beginning.

[1] Since subgroups are identical beyond the value of τ_j, there is no need for any of the τ's to be equal, but it does allow for discussion of the case where they are all equal. See Sect. 3.1.

Table 1 Parameters used in the incidence functions

Parameter	Value	Description
c	13.46	Number of contacts in the absence of social distancing
τ_j	varies	Contact multiplier for subgroup j
β_A	0.0249	Probability of infection per contact with A_j
β_s	0.0249	Probability of infection per contact with I_j^s
β_m	0.0249	Probability of infection per contact with I_j^m

Table 2 The remaining system parameters

Parameter	Value	Description
N_0	6,000,000	Total population size at the DFE
$1/\mu$	80 years	Life expectancy
Λ	$\mu N_0 \approx 205$	Birth rate (in people per day)
$1/\epsilon$	2.5 days	Duration of latency
$1/\alpha$	1 day	Duration of asymptotic infectiousness
p	0.07	Probability that a case is serious
q	$1 - p = .93$	Probability that a case is mild
$1/k_s$	6 days	Time for a serious case to lead to hospitalization
$1/\nu_s$	12 days	Time for serious cases to recover (outside of the hospital)
$1/\nu_m$	6 days	Time for mild cases to recover
$1/\nu_H$	18 days	Time for hospitalized cases to recover
p_d	0.078	Fraction of hospitalizations ending in death
$1/d$	$\nu_H \frac{p_d}{1-p_d}$	Time in hospital before death
f_j	varies	Fraction of the population in social distancing group j

We assume that a fraction $f_j \in (0, 1)$ of the population enters the j'th subgroup (with contact multiplier τ_j), through the birth/recruitment rate Λ, with $f_1 + \cdots + f_n = 1$. Furthermore, we assume that the initial distribution of the population is in accordance with these fractions.

The meaning of the system parameters, along with the values that are used in the numerical simulations can be found in Table 1 (for parameters related to the incidence functions) and Table 2 (for other parameters).

Let N_j be the number of non-hospitalized individuals in group j. All non-hospitalized individuals are assumed to engage in contacts. (Quarantining and self-isolation are not included in the model.) Then

$$N_j = S_j + E_j + A_j + I_j^s + I_j^m + R_j.$$

Let Γ_j be the incidence affecting S_j, calculated using proportional mixing. To calculate Γ_j, we determine the probability that a given contact is with an individual in a particular compartment. This is related to earlier work such as [1, 2]. The total contact rate of individuals in group A_k, for example, is $c\tau_k A_k$. Dividing by the total contact rate of the population $\sum_{i=1}^{n} c\tau_i N_i$, we see that the probability that a particular contact is with someone in group A_k is $\frac{\tau_k A_k}{\sum_{i=1}^{n} \tau_i N_i}$. The contact rate for an individual in group S_j is $c\tau_j$. Thus, the rate of contacts between groups S_j and A_k is $c\tau_j S_j \frac{\tau_k A_k}{\sum_{i=1}^{n} \tau_i N_i}$. Multiplying by the transmission probability β_A of the asymptomatic class gives the associated transmission rate. Calculating similar terms for I_k^s and I_k^m and summing over k gives

$$\Gamma_j = c\tau_j S_j \frac{\sum_{k=1}^{n} \tau_k \left(\beta_A A_k + \beta_s I_k^s + \beta_m I_k^m\right)}{\sum_{i=1}^{n} \tau_i N_i}.$$

We note that $\frac{\Gamma_j}{\Gamma_l} = \frac{\tau_j S_j}{\tau_l S_l}$, so that the force of infection for the various susceptible groups varies only by the change in the contact multiplier τ.

The differential equations for the system are

$$\begin{aligned}
\frac{dS}{dt} &= f_j \Lambda - \mu S - \Gamma_j \\
\frac{dE}{dt} &= \Gamma_j - (\epsilon + \mu)E \\
\frac{dA}{dt} &= \epsilon E - (\alpha + \mu)A \\
\frac{dI^s}{dt} &= p\alpha A - (\nu_s + k_s + \mu)I^s \\
\frac{dI^m}{dt} &= q\alpha A - (\nu_m + \mu)I^m \\
\frac{dH}{dt} &= k_s I^s - (\nu_H + \mu + d)H \\
\frac{dR}{dt} &= (1 - p - q)\alpha A + \nu_s I^s + \nu_m I^m + \nu_H H - \mu R,
\end{aligned}$$

where the variables $(S, E, \ldots, R)$ should all have a subscript j on them, but we have omitted the subscript in order to make the equations more legible.

In order to study the initial outbreak, it is appropriate to consider initial conditions near the disease-free equilibrium.

Parameters related to the incidence are described in Table 1, while the remaining parameters appear in Table 2. The population size and life expectancy were chosen with the Greater Toronto Area in mind. Other parameters came from [3, 5, 6] and [4], including Supplementary Table 5.

2 Disease-Free Equilibrium

The disease-free equilibrium (DFE) is given by

$$S_j^0 = f_j \frac{\Lambda}{\mu}, \quad \text{and} \quad E_j = A_j = I_j^s = I_j^m = H_j = R_j = 0,$$

for $j = 1, \ldots, n$. Note that at the DFE, $N_j = S_j^0$.

It is worth rewriting the incidence as

$$\Gamma_j = G_j B,$$

where

$$G_j = \frac{\tau_j S_j}{\sum_{i=1}^n \tau_i N_i} \quad \text{and} \quad B = c \sum_{k=1}^n \tau_k \left(\beta_A A_k + \beta_s I_k^s + \beta_m I_k^m \right).$$

Note that G_j is the probability that a given contact (by an infectious individual) is with an individual in class S_j. These probabilities are highly relevant for how the disease spreads and their values near the DFE affect the initial outbreak. At the DFE, we have

$$G_j = G_j^* = \frac{\tau_j f_j}{\sum_{i=1}^n \tau_i f_i}.$$

Note that $G_1^* + \cdots + G_n^* = 1$, and so the G_j can be used as coefficients in a weighted average. Let

$$\begin{aligned} \tau^* &= \sum_{j=1}^n \tau_j G_j^* \\ &= \frac{\sum_{j=1}^n \tau_j^2 f_j}{\sum_{i=1}^n \tau_i f_i}. \end{aligned} \tag{1}$$

This quantity τ^* will arise naturally when the basic reproduction number $\mathcal{R}_0$ is calculated. Due to the role that it will play, we refer to τ^* as *the effective contact multiplier* and the product $c\tau^*$ gives *the effective contact rate* of the population.

3 The Basic Reproduction Number $\mathcal{R}_0$

Using the next generation matrix method [7], with the infected variables ordered as

$$\left(E_1, \ldots, E_n, A_1, \ldots, A_n, I_1^s, \ldots, I_n^s, I_1^m, \ldots, I_n^m, H_1, \ldots, H_n, R_1, \ldots, R_n\right),$$

leads to rather large matrices for F and V, size $6n \times 6n$. Fortunately, they have a block form that can be used efficiently. First, $\mathcal{F} = (\Gamma_1, \ldots, \Gamma_n, 0, \ldots, 0)^T \in \mathbb{R}^{6n}$, with all other terms put in $\mathcal{V}$.

Differentiating Γ_j with respect to A_l (and similarly for I_l^s and I_l^m), and evaluating the results at the DFE, we obtain the matrix

$$\frac{c\beta_A}{\sum_{i=1}^n \tau_i f_i}\mathbb{T}, \quad \text{where} \quad \mathbb{T} = \begin{bmatrix} \tau_1^2 f_1 & \tau_1 f_1 \tau_2 & \cdots & \tau_1 f_1 \tau_n \\ \tau_2 f_2 \tau_1 & \tau_2^2 f_2 & \cdots & \tau_2 f_2 \tau_n \\ \vdots & \vdots & \ddots & \vdots \\ \tau_n f_n \tau_1 & \tau_n f_n \tau_2 & \cdots & \tau_n^2 f_n \end{bmatrix} = \begin{bmatrix} \tau_1 f_1 \\ \tau_2 f_2 \\ \vdots \\ \tau_n f_n \end{bmatrix} \begin{bmatrix} \tau_1 & \tau_2 & \cdots & \tau_n \end{bmatrix}.$$

The matrix $\mathbb{T}$ will appear as a block in the matrix F below. At this point it is worth noting that $\mathbb{T}$ has rank 1, and therefore it has $n-1$ eigenvalues that are 0, while the remaining eigenvalue is equal to the trace. Thus, the spectral radius ρ of $\mathbb{T}$ is given by

$$\rho(\mathbb{T}) = \text{trace}(\mathbb{T}) = \sum_{j=1}^n \tau_j^2 f_j.$$

Calculating F and V from $\mathcal{F}$ and $\mathcal{V}$ in the standard way (see [7]), using $\mathbf{0}$ to denote an $n \times n$ block of zeros and using $\mathbf{I}$ to denote the $n \times n$ identity, we obtain

$$F = \frac{c}{\sum_{i=1}^n \tau_i f_i} \begin{bmatrix} \mathbf{0} & \beta_A\mathbb{T} & \beta_s\mathbb{T} & \beta_m\mathbb{T} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

and

$$V = \begin{bmatrix} (\epsilon+\mu)\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\epsilon\mathbf{I} & (\alpha+\mu)\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -p\alpha\mathbf{I} & (\nu_s+k_s+\mu)\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -q\alpha\mathbf{I} & \mathbf{0} & (\nu_m+\mu)\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -k_s\mathbf{I} & \mathbf{0} & (\nu_H+\mu+d)\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -(1-p-q)\alpha\mathbf{I} & -\nu_s\mathbf{I} & -\nu_m\mathbf{I} & -\nu_H\mathbf{I} & \mu\mathbf{I} \end{bmatrix}.$$

(If the infected variables are reordered as $(E_1, \ldots, R_1, E_2, \ldots, R_2, \ldots, E_n, \ldots, R_n)$, then V becomes block diagonal, where the diagonal blocks are all identical and can be obtained by replacing each block of the matrix given here for V, with the scalar appearing on that block's diagonal.)

In order to write $\mathcal{R}_0$ concisely, we define T_A, T_s and T_m, the average times spent in classes A, I^s and I^m, respectively. They are given by

$$T_A = \frac{\epsilon}{(\epsilon+\mu)(\alpha+\mu)}, \qquad T_s = T_A \frac{p\alpha}{k_s+\nu_s+\mu} \quad \text{and} \quad T_m = T_A \frac{q\alpha}{\nu_m+\mu},$$

and appear in V^{-1}.

Due to the many rows of zeroes in F, the next generation matrix FV^{-1} also has many rows of zeros. In terms of $n \times n$ blocks we get

$$FV^{-1} = \frac{c}{\sum_{i=1}^{n} \tau_i f_i} \begin{bmatrix} M_1 & M_2 & M_3 & M_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where

$$M_1 = (\beta_A T_A + \beta_s T_s + \beta_m T_m)\, \mathbb{T}.$$

We can now write

$$\begin{aligned} \mathcal{R}_0 &= \rho\left(FV^{-1}\right) \\ &= \frac{c}{\sum_{i=1}^{n} \tau_i f_i} \rho\left(M_1\right) \\ &= \frac{c}{\sum_{i=1}^{n} \tau_i f_i} \left(\beta_A T_A + \beta_s T_s + \beta_m T_m\right) \rho\left(\mathbb{T}\right) \\ &= c \frac{\sum_{j=1}^{n} \tau_j^2 f_j}{\sum_{i=1}^{n} \tau_i f_i} \left(\beta_A T_A + \beta_s T_s + \beta_m T_m\right), \end{aligned}$$

and so

$$\mathcal{R}_0 = c\, \tau^* \left(\beta_A T_A + \beta_s T_s + \beta_m T_m\right). \tag{2}$$

3.1 Interpreting τ^*

From Eq. (2), we see that the manner in which the heterogeneity in contact rates comes into the expression for $\mathcal{R}_0$ is in $c\tau^*$. The individual values of $\tau_1, \ldots, \tau_n$ do not matter, except in how they affect τ^*. Two different sets of values of $(\tau_1, \ldots, \tau_n)$ that produce the same τ^* lead to the same value for the basic reproduction number.

Consider the case where all of the τ_j are equal. Suppose $\tau_1 = \cdots = \tau_n = \tau$. Then Eq. (1) gives $\tau^* = \tau$. This is equivalent to having a single group (i.e. $n = 1$), with everyone at the same contact level.

Thus, the value of $\mathcal{R}_0$ that arises from a population with heterogeneity in the contact rates, given by $(c\tau_1, \ldots, c\tau_n)$, would also be obtained if everyone in the population adopted a contact rate of $c\tau^*$.

This is the reason why τ^* can reasonably be called the effective contact multiplier of the population and $c\tau^*$ can be called the effective contact rate.

3.2 A Different Average τ

Averaging τ over the non-hospitalized population gives

$$\bar{\tau} = \frac{\sum_{j=1}^{n} \tau_j N_j}{\sum_{i=1}^{n} N_i}.$$

At the DFE, the hospitalized populations are zero, and $N_i = f_i \frac{\Lambda}{\mu}$. Since $\sum f_i = 1$, this average $\bar{\tau}$ becomes

$$\bar{\tau}_{\text{DFE}} = \sum_{j=1}^{n} \tau_j f_j. \tag{3}$$

The relationship between τ^* and $\bar{\tau}_{\text{DFE}}$ will be discussed in Sects. 4 and 6. However, comparing Eqs. (1) and (3) makes it immediately apparent that τ^* and $\bar{\tau}_{\text{DFE}}$ are generally different.

3.3 How $\mathcal{R}_0$ and τ^* Depend on the τ's

If all other parameters are held fixed while τ_l is varied, then the ratio of $\mathcal{R}_0$ to τ^* remains fixed. Thus, it is sufficient to determine how τ^* depends on τ_l, and the same correlation will apply to $\mathcal{R}_0$. With a bit of calculation, we obtain

$$\frac{\partial \tau^*}{\partial \tau_l} = \frac{2 f_l}{\bar{\tau}_{\text{DFE}}} \left(\tau_l - \frac{1}{2} \tau^* \right).$$

Recall that $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$ and also that τ^* (defined in Eq. (1)) is a weighted average of the τ's. This means that some of the τ's (including τ_n) are greater than $\frac{1}{2}\tau^*$, making the corresponding derivative positive.

On the other hand, it is possible that τ_1 (and some other τ's) are less than $\frac{1}{2}\tau^*$, making the corresponding derivative negative—a surprising result. We have the following:

- If $\tau_l > \frac{1}{2}\tau^*$, then an increase in τ_l leads to an increase in $\mathcal{R}_0$. This is certainly the case for τ_n.
- If $\tau_l < \frac{1}{2}\tau^*$, then an increase in τ_l leads to a decrease in $\mathcal{R}_0$. This may or may not be the case for τ_1.

4 Numerical Work

In this section we present the results of numerical simulations[2] for the simplified case of $n = 2$. This means that each epidemiological group from the original transfer diagram is divided into two subgroups, which differ only in the contact multipliers τ_1 and τ_2.

4.1 Parameters

System parameters that are associated with the original transfer diagram—except for the population size, the contact rate c and the transmission probabilities β_A, β_s and β_m—are taken from [6], as the model studied here is a modified version of the model studied there. The population size is taken to be 6 million, which is roughly the size of the Greater Toronto Area. We use $c = 13.5$, which comes from [4, Supplementary Table 5], work that provides data on contact rates for infectious respiratory diseases in several European countries. In the absence of more detailed information, we use $\beta_A = \beta_s = \beta_m$. The value[3] used is chosen so that in the absence of contact heterogeneity we obtain $\mathcal{R}_0 = 2.3$, taken from [3, 6]. All system parameters are given in Tables 1 and 2, which appear in Sect. 1.

With two subgroups, we have $f_2 = 1 - f_1$, and so there are three input parameters (f_1, τ_1 and τ_2), which can be varied. We choose to keep τ_2 fixed at 1, so that individuals in sub-group 2 can be thought of as maintaining their original contact level. This gives two input parameters (f_1 and τ_1), which can be varied. We make the further restriction that $\tau^* = 0.7$. This essentially sets f_1 to be a function of τ_1.

[2] Numerical solutions were calculated in Matlab using ode45.

[3] We experimented with distinct values, but there did not seem to be a significant difference in the results.

Rearranging Eq. (1) gives

$$f = f(\tau_1) = \frac{\tau_2(\tau_2 - \tau^*)}{(\tau_2 - \tau_1)(\tau_2 + \tau_1 - \tau^*)} = \frac{0.3}{(1 - \tau_1)(0.3 + \tau_1)}. \tag{4}$$

We now have a single input parameter τ_1 to be varied.

4.2 The Simulations

The goal is to investigate the significance of having different groups social distancing at different levels and to see how this affects the time-course of the outbreak. In particular, numerical simulations will allow us to see how an outbreak develops both during and after the initial period of exponential growth.

The choice of $\tau^* = 0.7$ results in a new basic reproduction number of 1.61 (coming from 70% of 2.3, the original value of $\mathcal{R}_0$). Thus, we still have a period of exponential growth. By varying the value of τ_1 (while choosing f_1 so that τ^* stays fixed), we are able to study:

- how heterogeneity affects the doubling time during the period of exponential growth,
- how heterogeneity affects the height and timing of the infection peak,
- the difference between predictions that explicitly include heterogeneity versus those that use the average value of $\bar{\tau}_{\text{DFE}}$ or τ^*.

We note that since τ_2 is held fixed at 1, lower values of τ_1 represent greater heterogeneity.

4.2.1 Doubling Time

The doubling time during the period of exponential growth was studied by finding the dominant eigenvalue $\widetilde{\lambda}$ of the Jacobian matrix at the DFE for each value of τ_1, and then determining the doubling time T_{double} as

$$T_{\text{double}} = \frac{\ln(2)}{\widetilde{\lambda}}$$

so that $e^{\widetilde{\lambda} T_{\text{double}}} = 2$. Calculations were completed numerically in Matlab and again independently in Maple. The observed result was that the doubling time remained constant at 10.85 days, as τ_1 was varied (while τ^* remained constant).

4.2.2 The Infection Peak

In Fig. 1, the blue curves are the most important as they represent the total population (i.e. groups 1 and 2 combined). The red curves give the quantities in the social distancing group (group 1); the green curves give the quantities for the non-distancing group. In the left-hand panel of Fig. 1, $f_1 = 1$ and so everyone is in the social distancing group. This is why the green curve is at 0 and the red curve is covered by the blue curve.

The top portion of each panel shows the total number of current infections ($E + A + I^s + I^m + H$) as a function of time. The second portion gives the number of hospitalized cases. The third portion gives the cumulative deaths. The fourth portion gives the number of susceptibles. More or less, it is better when the curves in the top three portions are lower and the curves in the fourth portion are higher—especially the blue curves. Figure 1 suggests that (for fixed τ^*) it is better to have τ_1 small.

As τ_1 is varied, the timing of the peak remains constant (see the left-hand panel of Fig. 2). The height of the peak decreases from 588,000 to 175,000 as τ_1 decreases from 0.7 to 0.1 (middle panel). That's a decrease of just over 70%.

Note that when $\tau_1 = 0.7$ (at the right-hand side of each panel of Fig. 2), we have $f_1 = 1$, as shown in the right-hand panel, and the entire population is in group 1. As τ_1 decreases from 0.7 (i.e. moving to the left within the panel), the fraction f_1 decreases so that less of the population is in group 1 so that τ^* remains fixed, before reaching a minimum at $\tau_1 = \frac{1}{2}\tau^* = 0.35$.

Throughout the simulations, $f_1 \geq 0.7$ (right-hand panel of Fig. 2). This means that the social distancing group is always more than 70% of the total population, explaining why the red curves in Fig. 1 are generally higher than the green curves. (This changes a bit, but not dramatically, if the value of τ^* is changed from 0.7.)

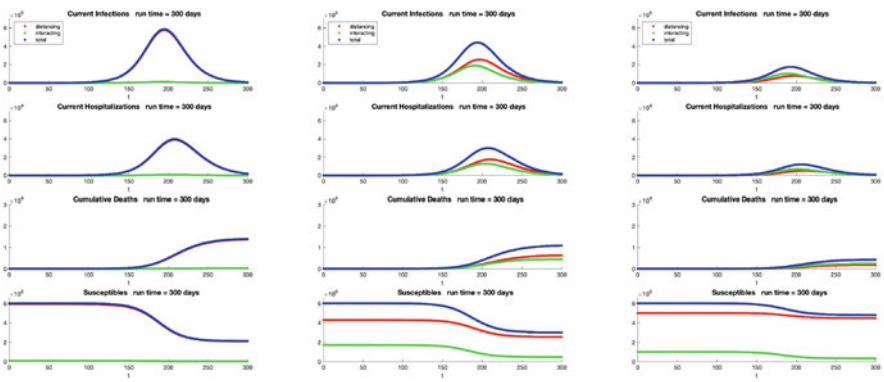


Fig. 1 Numerical plots. The left-hand panel has $\tau_1 = 0.7$ and $f_1 = 1$, so that everyone is in group 1 and everyone has a contact level that is 70% of the original level. The middle panel has $\tau_1 = 0.4$ and $f_1 = \frac{5}{7}$ so most of the population is at 40% of the original contact level and the remainder is at 100%. The right-hand panel has $\tau_1 = 0.1$ and $f_1 = \frac{5}{6}$

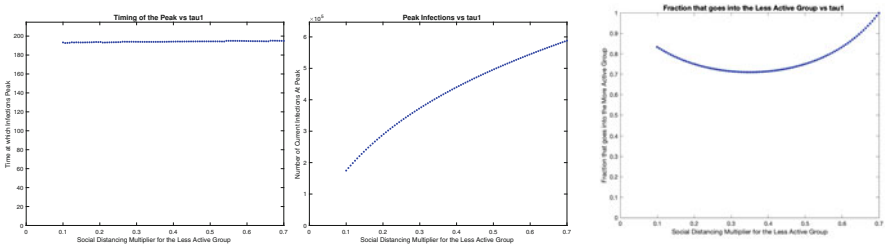


Fig. 2 The left-hand panel indicates that heterogeneity in social distancing does not appreciably affect the timing of the epidemic peak. The middle panel shows that the height of the epidemic peak has a strong dependence on the level of heterogeneity in social distancing (as represented by τ_1). The right-hand panel shows the fraction f_1 of the population in group 1 as a function of τ_1 (given that τ_2 and τ^* are held fixed)

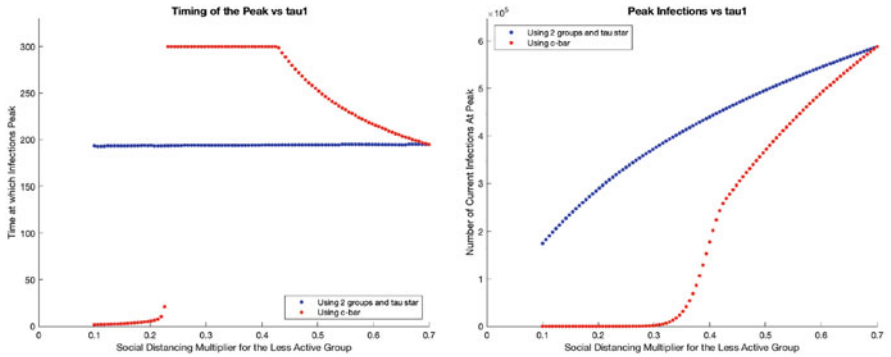


Fig. 3 The left-hand panel shows the predicted timing of the infection peak, while the right-hand panel shows the height of the peak. In each panel, the blue curve comes from explicitly using two subgroups with contact multipliers of τ_1 and τ_2, whereas the red curve comes from not using subgroups and using a single contact rate of $c\bar{\tau}_{\text{DFE}}$

The right-hand panel of Fig. 2 shows that even when the majority of the population is in a strongly social distancing group (about 83% of the population when $\tau_1 = 0.1$) a comparatively small group of non-distancers (about 17% of the population) are able to maintain the outbreak, with $\mathcal{R}_0$ fixed at 1.61.

4.2.3 Including Heterogeneity Versus Using $\bar{\tau}_{\text{DFE}}$

In Fig. 3, the blue curve in each panel is the same as in the first two panels of Fig. 2, giving the timing and height of the infection peak versus τ_1, with τ_2 and τ^* held fixed.

Separate simulations were run using a single contact level (i.e. 7 variables S, E, A, I^s, I^m, H and R with no subgroups) given by the average contact level of the population: $c\bar{\tau}_{\text{DFE}}$. (Note that this approach is natural, but ignores that clusters are

more likely to occur in the highly active subgroups of the population that interact primarily with each other.) The red curves in Fig. 3 give the resulting predictions for the timing and the height of the infection peak. All other parameters are the same as in the main simulations that led to the blue curves.

By using a contact rate of $c\bar{\tau}_{\text{DFE}}$, rather than $c\tau^*$, the value of $\mathcal{R}_0$ changes by a factor of $\frac{\bar{\tau}_{\text{DFE}}}{\tau^*}$. For low values of τ_1 (recalling that f_1 remains above 0.7—see the right-hand panel of Fig. 2), a contact rate of $c\bar{\tau}_{\text{DFE}}$ results in a value of $\mathcal{R}_0$ less than 1. This can be seen in the right-hand panel of Fig. 3, where the height of the peak is essentially zero for $\tau_1 < 0.3$. As τ_1 increases past 0.35, the height of the peak climbs, while remaining significantly lower than the blue curve. Only when τ_1 gets to 0.7 (at which point $\bar{\tau}_{\text{DFE}} = \tau^*$) do the red and blue curves converge.

This means that in the presence of heterogeneity, treating the contact level as being homogeneously at the population average underestimates the height of the peak. Additionally, when an outbreak occurs $\tau_1 > 0.35$) using $c\bar{\tau}_{\text{DFE}}$ leads to overestimating the time until the peak occurs. Such a prediction would encourage preparation for a lower peak that would happen later, leading to a state of underpreparedness.

It is important to address the left-hand portion of the left-hand panel of Fig. 3. The red curve is very low in this portion of the figure. However, by looking at the right-hand panel, we see that there is no outbreak for these low values of τ_1 (since heterogeneity has been inappropriately omitted). Thus, the "timing of the peak" is really coming from the small number of initial infections that are used to seed the numerical simulations.

The flatness of the upper portion of the red curve in the left-hand panel comes from using a maximum time of 300 days in the numerical simulations. Cases were still climbing when the simulation terminated.

For contrast, if heterogeneity is not explicitly accounted for by including subgroups, but is acknowledged by setting the single contact level to be $c\tau^*$, then it is equivalent to the result obtained for $\tau_1 = 0.7$ (since that gives $f_1 = 1$). In the right-hand panel of Fig. 3 this would give a third curve that is constant at a height equal to the intersection point of the blue and red curves.

Thus, a single contact level $c\bar{\tau}_{\text{DFE}}$ underestimates the height of the peak, possibly dramatically, while a single contact level $c\tau^*$ overestimates the height of the peak, also possibly dramatically.

5 A Summary of Key Observations

As τ_1 decreases from 0.7 to 0.1 (and f_1 correspondingly changes so that τ^* stays fixed) we have:

- $\mathcal{R}_0$ remains constant,
- the doubling time for the period of exponential growth remains constant,
- the timing of the peak (in the number of infections) remains constant,

- BUT the height of the peak decreases by as much as 70%,
- a small group of non-distancers can keep $\mathcal{R}_0$ elevated.

Ignoring heterogeneity in the contact levels can profoundly affect the accuracy of predictions about the height and timing of the infection peak.

6 Discussion

Many early models predicted peak levels of infection (and the resulting death toll) to be much higher than were seen in the first wave of COVID-19. Clearly, policies that were introduced to limit social interactions played a significant role in the reduction of cases. However, an additional factor that is less obvious, is that not everybody had the same contact level. This would have been the case even without interventions.

The exact doubling time found here is likely not the real-world value. (The basic compartmental model and parameters are not precise enough for that.) However, the fact that the doubling time depends on τ^* (but not directly on τ_j, f_j or $\bar{\tau}_{\mathrm{DFE}}$) is noteworthy. Two major population level quantities of interest are $\mathcal{R}_0$ and the doubling time. It is apparent that these depend strongly on τ^* rather than on $\bar{\tau}_{\mathrm{DFE}}$.

Often $\mathcal{R}_0$ and/or the doubling time is known through other means. It is then used to calculate the contact rate c or the transmission probability β or the product $c\beta$. The current work suggests that this should be done carefully due to heterogeneity in the contact levels.

The numerical simulations were done in the limited case of only 2 groups (represented by τ_1 and τ_2). However, it is clear that similar behaviour also happens with more than 2 groups—although likely at a reduced level.

Certainly, the most extreme case studied here (two subgroups with associated values of $\tau_1 = 0.1$ and $\tau_2 = 1$) is unlikely to be representative of a real population. Nevertheless, it highlights the fact that system parameters that give the same value for τ^* (and hence for $\mathcal{R}_0$) can yield wildly different predictions for the height of the infection peak (70% in this case). It follows that if heterogeneity is ignored, then quantitative predictions may suffer substantially.

If a subgroup's contact levels are very small compared to the full population (as represented by τ^*), then it becomes similar to that group being absent and the population being a little smaller. In that case, a new higher value of τ^* would be calculated based on the remainder of the population. If the subgroup with the low contact level were to engage in more contacts, this would (in some sense) shield the higher contact subgroups from infection, provided that it did not result in an increase of contacts for those higher contact subgroups. Nevetheless, the author would not recommend intentionally increasing any of the τ's. Instead, this observation can be seen as shining a light on the fact that higher activity groups appear to be able to sustain the epidemic on their own by maintaining high contact levels with others in the high contact population.

Combining Eqs. (3) and (4), we see that in our simulations

$$\bar{\tau}_{\mathrm{DFE}} = \frac{\tau_1 \tau_2}{\tau_2 + \tau_1 - \tau^*} = \frac{\tau_1}{\tau_1 + 0.3}.$$

Thus, for $\tau_1 = 0.7$ (the beginning of our simulations) we get $\bar{\tau}_{\mathrm{DFE}} = 0.7$, whereas for $\tau_1 = 0.1$ (the end of our simulations), we get $\bar{\tau}_{\mathrm{DFE}} = 0.25$ (a 64% reduction). As $\bar{\tau}_{\mathrm{DFE}}$ changes dramatically throughout the simulation, we have $\tau^* \equiv 0.7$ and so $\mathcal{R}_0$ remains fixed at 1.61.

Many models are (quite reasonably) focused on issues other than heterogeneity in contact levels. Nevertheless, those models often encapsulate the population's contact level in a single parameter c. In estimating a value of c from contact data $(c_1, \ldots, c_n)$ with associated frequencies $(f_1, \ldots f_n)$, it would be more appropriate to use an effective contact level associated with τ^*, rather than a standard average contact level associated with $\bar{\tau}_{\mathrm{DFE}}$. That is, it is better to use

$$c^* = \frac{\sum_{j=1}^{n} c_j^2 f_j}{\sum_{i=1}^{n} c_i f_i}, \qquad \text{rather than} \qquad \bar{c} = \sum_{i=1}^{n} c_i f_i.$$

Either way, though, predictions about the height of a peak can be profoundly affected by not explicitly including subgroups with different contact levels.

As societies and economies continue to navigate the pandemic, there will be people that continue to keep their contacts levels reduced and others that won't. Accounting for the impact of heterogeneity in contact levels will continue to be important.

Acknowledgments This work was submitted for publication on November 1, 2020.

References

1. Anderson, R. M. May, R. M. Medley, G. F., Johnson, A.: A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. IMA J. Math. Appl. Med. Biol., **3**, 229–263 (1986)
2. Jacquez, J. A., Simon, C. P., Koopman, J., Sattenspiel, L., and Perry, T.: Modeling and analyzing HIV transmission: The effect of contact patterns. Math. Biosci., **92**, 119–199 (1988)
3. Li, Q., Guan, X. et al: Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med., **382**, 1199–1207 (2020)
4. Mossong, J., Hens, N., Jit, M. et al: Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med, **5**, (3) (2008). https://doi.org/10.1371/journal.pmed.0050074
5. Tuite, A. R., Fisman, D. N., Greer, A. L.: Appendix 1 (as supplied by the authors): Mathematical modeling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada: Technical appendix. CMAJ, **192**, (19) (2020). https://www.cmaj.ca/content/cmaj/suppl/2020/04/08/cmaj.200476.DC1/200476-res-1-at.pdf Cited 1 Nov 2020

6. Tuite, A. R., Fisman, D. N., Greer, A. L.: Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. CMAJ, **192**, (19)E497–E505 (2020). https://doi.org/10.1503/cmaj.200476
7. van den Driessche, P., and Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci., **180**, 29–48 (2002)

Forecasting Demand for Personal Protective Equipment for Ontario Acute Care Hospitals During the First Wave of the COVID-19 Pandemic

Kali Barrett, Yoshiko Nakamachi, Terra Ierasts, Yasin A. Khan, Stephen Mac, David Naimark, Nathan M. Stall, Raphael Ximenes, Andrew M. Morris, and Beate Sander

1 Introduction

Novel Coronavirus Disease 2019 (COVID-19), caused by Severe Acute Respiratory Syndrome-related Coronavirus-2 (SARS-CoV-2), has placed tremendous strain on healthcare systems. Personal protective equipment (PPE)—designed to protect healthcare workers and patients, and promote infection prevention and control—proved to be one of the first vulnerabilities of healthcare systems when COVID-19

K. Barrett
Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

University Health Network, Toronto, ON, Canada

Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, ON, Canada

Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, ON, Canada

Department of Medicine, University of Toronto, Toronto, ON, Canada

Y. Nakamachi
University Health Network, Toronto, ON, Canada

Sinai Health, Toronto, ON, Canada

T. Ierasts
University Health Network, Toronto, ON, Canada

Y. A. Khan
Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

University Health Network, Toronto, ON, Canada

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_10

cases surged [1]. PPE shortages in Canada led governments, health regions, and hospitals to resort to extraordinary measures to source and procure supplies in a highly competitive and uncertain global marketplace [2]. Actual and feared PPE shortages contributed to burnout among healthcare workers, resulted in legal challenges over the right to access PPE, and were considered as a contributing factor in COVID-19 outbreaks in congregate settings [3–5].

As the pandemic evolved, recommendations for PPE use diverged from conventional infection control policies (e.g., contact, droplet) [6] to policies based on risk mitigation given a changing understanding of disease transmission. PPE shortages

Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, ON, Canada

Department of Medicine, University of Toronto, Toronto, ON, Canada

S. Mac
Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, ON, Canada

D. Naimark
Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, ON, Canada

Sunnybrook Hospital, Toronto, ON, Canada

N. M. Stall
Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Department of Medicine, University of Toronto, Toronto, ON, Canada

Sinai Health, Toronto, ON, Canada

Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada

R. Ximenes
Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, ON, Canada

Escola de Matemática Aplicada, Fundação Getúlio Vargas, Rio de Janeiro, Brazil

A. M. Morris
University Health Network, Toronto, ON, Canada

Department of Medicine, University of Toronto, Toronto, ON, Canada

Sinai Health, Toronto, ON, Canada

B. Sander (✉)
Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
e-mail: beate.sander@uhnresearch.ca

led some institutions to require staff to reuse PPE to conserve supplies. Increased community spread with subsequent concern for nosocomial spread of COVID-19 led many institutions to adopt universal masking policies. It was unclear how these new PPE policies would impact overall demand for PPE.

Accurate predictions of PPE demand during a pandemic are critical to ensure adequate supplies and prevent morbidity and mortality among patients and healthcare workers. Forecasting PPE demand informs estimates of remaining days of stock on hand, supports procurement decisions, aids in supply chain management, and assists with planning for the resumption of clinical activities such as non-urgent surgeries. Estimates of PPE demand must be based on empirical data that incorporates epidemic trajectories, clinical volumes, healthcare worker practice patterns, and other human factors. We used health system modelling, informed by data on clinical practice patterns to forecast provincial PPE demand in the acute care setting in Ontario, Canada, and to estimate the effect that changes to PPE use policies, such as universal masking and conservation strategies, have on demand.

2 Methods

We forecasted near-term (up to 60 days) PPE demand during the COVID-19 pandemic, taking into account COVID-19 epidemic trajectories and the acute care pathway for COVID-19 patients, acute care patient mix (COVID-19 confirmed and suspected patients, and non-COVID-19 patients), healthcare workers' practice patterns, and PPE required by type of activity. We considered two PPE strategies based on: (1) provincial public health recommendations for PPE use—*provincial-level study*; and (2) institutional policies of universal masking and conservation strategies- *hospital-level case study*.

We obtained research ethics board approval from the University of Toronto to access provincial datasets. Sinai Health waived research ethics board approval for the case study.

2.1 Provincial-Level Study

2.1.1 Forecasting COVID-19 Cases and Their Care Trajectory: COVID-19 Resource Estimator (CORE) Model

To forecast emergency department (ED) visits, hospital admissions and care pathways for COVID-19 patients in Ontario (population 14.6 million), we updated through refinement and calibration, our previously described COVID-19 resource estimator (CORE) health system model [7], a discrete time, dynamic, open, parallel,

individual-level health state transition model. We used the CORE model to forecast the daily number of COVID-19 cases presenting to the ED and being admitted to the hospital ward or intensive care unit (ICU), with or without invasive mechanical ventilation. To account for PPE use to care for suspected COVID-19 cases, i.e., persons under investigation (PUI) until infection is ruled out, we applied the reported ratio of PUI per confirmed COVID-19 patient in acute care [8–10]. We chose to base our case numbers on the CORE model results as opposed to daily reported numbers available in provincial data sets for two reasons. First, we developed this model in March 2020, to estimate future 60-day demand for PPE for provincial decision-makers. Second, simply extrapolating case numbers from existing epidemic curves would not provide data on the number of patients with COVID-19 admitted to hospital and at a given stage in their admission (day 1 in ICU, day 2 in ICU, etc.), which are required to allocate the appropriate touchpoint.

2.1.2 Determining PPE Use

PPE is used during episodes of patient contact. We therefore estimated the number of times patient-contact takes place per patient within a 24-h period. We defined these episodes of "patient touchpoints" as any time a health care worker (HCW) enters a patient room or is required to physically interact with a patient, thereby requiring PPE. We stratified the number of touchpoints per patient per 24-h period by in-patient hospital setting (ED, ICU, and ward), type of healthcare worker (physician, nurse, respiratory therapist, allied health professional), COVID-19 status of the patient (confirmed, suspected or negative), and type of contact (COVID−, COVID+, low-risk aerosol generating medical procedure [AGMP], high-risk AGMP). Examples of touchpoints for patients in the ICU are provided in Table 1.

In order to quantify the type and frequency of patient touchpoints in a 24-h period, we observed healthcare workers on various hospital units while they provided care to patients, and conducted in-the-moment empathy interviews for each group of HCWs. Observations and interviews occurred at two points in time: April 2, 2020 and April 17, 2020. The rationale for repeating this process was twofold: PPE policies were evolving as new information became available regarding COVID-19 infection prevention and control practices, and HCWs' behaviours around PPE use also changed as the pandemic evolved. Specifically, HCWs were adapting to their new work environment and were modifying the manner and frequency of patient care/touchpoints. For example, the use of cell phones, baby monitors, and intercoms to communicate with patients/staff in the room reduced the need to use PPE. Extending patient IV tubing so medication pumps could be placed outside the patient's room for medication management, coordinating several patient care tasks to occur at the same time, or remaining in the patient room for an extended period of time (1–2 consecutive hours)—particularly during the first 24–48 h of a newly admitted critical care patient—are just a few of the observed changes. Thus,

Table 1 Daily touchpoints for suspected or confirmed COVID-19 patients in the ICU

Mechanical ventilation status	Day of admission	Touchpoints from contact	Touchpoints from intubation and extubation	Touchpoints from proning
Receiving invasive mechanical ventilation	Day 1	57	8	31
	Day 2	53		31
	Day 3 onward	39.25		31
Not receiving invasive mechanical ventilation	Day 1	57		
	Day 2	53		
	Day 3 onward	39.25		
	Ward post ICU	40.25		

re-validation of touchpoints was essential to ensure we captured these changes in practice and PPE use.

Permission was obtained from the unit manager upon entry to each of the units and prior to any observations or interviews with HCWs. HCWs were informed that the purpose was to understand the frequency with which they had contact with their patient, the purpose for contact/entering the patient room, and the type of PPE worn during each of those touchpoints. HCWs were observed while they worked in their native environment and clarifying questions were asked while they were still in-context to elicit further information and insights to quantify touchpoints. Access to types of PPE and the distribution of PPE to HCWs on each of the units was also captured. HCW observations and interviews were conducted with physicians, nurses, respiratory therapists, allied health professionals, hospital support staff, diagnostic technicians, unit managers, managers of environmental services (housekeeping, transportation, and linen & laundry) and nutritional services. Observations occurred in 3 Emergency Departments, 4 Intensive Care Units (COVID+ve patient ICUs, COVID−ve patient ICUs and COVID mixed-patient ICUs), 7 General Internal Medicine Units (COVID+ve and COVID−ve units), and 1 Surgical Unit (COVID−ve).

Based on local clinical experience, we assumed that 15% of patients with COVID-19 in the ICU not requiring invasive mechanical ventilation would require non-invasive ventilation or oxygen via high-flow nasal cannula, and that touchpoints for these patients would require the low-risk AGMP PPE bundle. Although not considered an AGMP in provincial guidelines, local critical care clinical guidelines considered patient care associated with patient proning during invasive mechanical ventilation to require the use of N95 masks and face shields given the risk of ventilator circuit disconnection and aerosolization of respiratory droplets, and we therefore considered these touchpoints to require the low-risk AGMP PPE bundle.

Table 2 Personal Protective Equipment (PPE) Bundles

	Bundle per touchpoint			
	High Risk AGMP	Low Risk AGMP	COVID+/PUI	COVID−
Surgical mask			1	
N95 mask	1	1		
Gloves		2	2	2
Gloves (extended)	2			
Face shield		1	1	
Face shield with drape	1			
Gown	1	1	1	
Level 1 mask				

AGMP Aerosol-generating Medical Procedure, *COVID+/−* Coronavirus Disease positive, Coronavirus Disease negative, *ED* Emergency Department, *ICU* Intensive Care unit, *PPE* Personal Protective Equipment, *PUI* Person Under Investigation

In the absence of high-quality published data, we assumed that 60% of mechanically ventilated patients with COVID-19 were being proned. This assumption was based on the proportion of COVID-19 patients reported in UK case series data with a PaO_2/FiO_2 <150 mmHg, the clinical threshold below which proning is indicated [11, 12]. All intubations and extubations were considered high-risk AGMPs by our institution's infection prevention and control department, necessitating extended length gloves and a face-shield with drape for touchpoints associated with those procedures.

We determined PPE bundles based on provincial PPE recommendations [13], combining PPE elements—masks (level 1, surgical/level 3, N95), face shields (with/without drape), gloves (regular, extended) and gowns—for each patient touchpoint depending patient COVID-19 status (COVID-19 positive, PUI, COVID-19 negative) and risk of AGMP (none, low, high risk). PPE bundles are shown in Table 2. Provincial PPE recommendations [13] provide no guidance for reuse of PPE, and we therefore assumed PPE was discarded after a single touchpoint for this policy.

2.1.3 Analysis

We calculated the total number of daily touchpoints for all patients estimated from our model to be admitted to hospital from March 6–May 5, 2020. We allocated corresponding bundles of PPE to each touchpoint based on patient characteristics.

PPE was therefore calculated as:

$$Daily\ sum\ of: PPE\ bundles/touchpoint \times touchpoint/patient \times patients.$$

We reported total PPE demand per day, based on provincial recommendations for PPE use to care for confirmed and suspected cases of COVID-19 in Ontario,

for the period from March 6–May 5, 2020. This time period was chosen as our aim was to estimate projected demand for PPE during the peak of new COVID-19 cases hospitalized in Ontario. Our model estimated a peak number of daily new cases in the second week of April, which corresponded to the observed peak of daily new cases.

Our CORE model was validated and calibrated against reported numbers of confirmed cases and patients admitted to hospital and ICUs with COVID-19. The CORE model results used to inform this study were based on the post-calibration model [7]. We did not attempt to validate or calibrate our predictions of PPE demand as, at the time of model development, there was no publicly available data informing on PPE use by Ontario acute care hospitals. Furthermore, our a priori assumption was that historical patterns of PPE use, or "burn rates", would likely be influenced by conservation strategies and thus may underestimate demand for PPE.

2.2 Hospital-Level Case Study

2.2.1 Determining Patient Volume and Staffing at the Institution Level

Institutional PPE policies often differ from provincial recommendations. PPE policies implemented by some Ontario hospitals include: (1) universal wearing of surgical masks, (2) PPE conservation efforts requiring the reuse or extended use of surgical masks, face shields, and N95 masks during a shift, and (3) individual point of care risk assessment allowing the healthcare worker to request an N95 facemask instead of a surgical facemask [13]. We conducted a case study to explore the effect these different PPE use policies have on demand, using real world data for observed case numbers and staffing from the University Health Network (UHN), a tertiary academic hospital in Toronto for the period of April 7, 2020 to April 16, 2020. This time period was chosen as it represented a period of high and sustained numbers of new COVID-19 cases presenting to hospital.

We identified the total daily number of patients assessed in the ED, admitted to acute and ICU beds, and requiring mechanical ventilation at UHN from institutional and provincial data sources. We identified the daily number of confirmed COVID-19 cases and PUIs among UHN admitted patients from provincial databases and reports. Institutional Human Resource data was used to estimate the number of clinical and non-clinical employees working each day. Attending physician staffing data was obtained from hospital on-call duty lists. UHN physician trainee (residents and clinical fellows) staffing data was provided by the University of Toronto Post-Graduate Medical Education office. Medical and other interdisciplinary health students were not working at UHN during the period under consideration (Tables 5 and 6).

Determining PPE Use

Because universal PPE policies are implemented across the entire institution, we considered PPE issued to all clinical and non-clinical staff, e.g., administrative, porters, etc., as well as all staff caring for patients assessed in the ED or admitted to acute care hospital beds. However, we did not consider PPE required for touchpoints associated with surgical procedures, endoscopy, bronchoscopy or other medical procedures deemed AGMPs, cardiac arrests, ambulatory clinics, or non-COVID-19 related contact precautions given the variability in policy between institutions and data limitations. As part of UHN's conservation strategy, PPE bundles were amended to reflect three levels of staff PPE allocation: non-clinical staff allocated one level 1 mask per shift; clinical staff working on wards allocated two surgical masks and one face shield per shift; and clinical staff working in ICUs and EDs with high risk for AGMPs allocated two surgical masks, one face shield, and one N95 mask per shift. We amended PPE bundles per touchpoint to reflect PPE allocated to healthcare workers per 12 hour shift, and reduced PPE consumed at each patient touchpoint (Table 3).

2.2.2 Analysis

Total PPE demand was calculated based on PPE required for patient touchpoints, and PPE allocated to staff. Daily PPE demand was calculated as:

$$(PPE\ allocated\ to\ staff \times daily\ staff\ numbers)$$
$$+ (PPE\ bundles/touchpoint \times touchpoint/patient \times patients)$$

We calculated PPE demand for the entire institution, and also performed separate analyses for the ED/ICU (where AGMP procedures take place), and acute care wards. The unit level analyses were conducted to determine PPE consumed for direct patient care, without the effect of universal masking for non-clinical staff. We performed several scenario analyses for: (1) COVID-19 and PUI patient volume; (2) proportions of patient touchpoints deemed to need an N95 based on point of care risk assessment; and (3) ratios of healthcare workers to patients on acute care wards (Table 4).

For each scenario we estimated the difference in the amount of PPE that would be required between the PPE use policy based on provincial recommendations versus the institutional policy for PPE use that involved universal masking and conservation.

We could not validate our results against real world PPE consumption as institutional data corresponding to the PPE consumed at the point of patient care for this time period did not exist. We did have access to the amount of PPE issued to the corresponding units within the hospital, but issued volumes were not felt to be an accurate surrogate of consumption. Furthermore, the purpose of this model was

Table 3 Personal Protective Equipment (PPE) Bundles by PPE policy

	Provincial PPE use recommendation				Institutional PPE policy						
	Bundle per touchpoint				Bundle per touchpoint				Bundle per shift		
	High risk AGMP	Low risk AGMP	COVID+/PUI	COVID−	High risk AGMP	Low risk AGMP	COVID+/PUI	COVID−	ICU/ED staff	Clinical staff	Non-clinical staff
Surgical mask			1						2	2	
N95 mask	1	1			1				1		
Gloves		2	2	2		2	2	2			
Gloves (extended)	2				2						
Face shield		1	1						1	1	
Face shield with drape	1				1						
Gown	1	1	1		1	1	1				
Level 1 mask									1	1	1

AGMP Aerosol-generating Medical Procedure, *COVID+/−* Coronavirus Disease positive, Coronavirus Disease negative, *ED* Emergency Department, *ICU* Intensive Care unit, *PPE* Personal Protective Equipment, *PUI* Person Under Investigation

Table 4 Description of provincial-level and hospital-level studies

	Provincial-level study	Hospital-level case study
Patients included in analysis	• Patients with COVID-19 and persons under investigation (PUIs) for COVID-19 assessed in ED and admitted to acute care hospitals in Ontario	• All patients presenting to the ED and admitted to intensive care and acute care at University Health Network, Toronto Canada.
Time horizon	• 60 days (March 6–May 5, 2020)	• 10 days (April 7–April 16, 2020)
Sources for patient numbers	• Daily number of new cases of COVID-19 and PUIs presenting to Ontario hospital EDs and admitted to acute and intensive care beds (with/without mechanical ventilation) estimated by CORE model • Number of PUIs calculated based on observed ratios of PUIs: confirmed COVID-19 cases for Ontario over 60-day period: – 1:1 for patients admitted to ICU (source: Provincial and Regional CCSO COVID-10 Dashboard Reports, Critical Care Services Ontario) – 2:1 for patients admitted to ward (source: Daily Bed Census, Ontario Ministry of Health) – 4:1 for patients assessed in ED (source: Electronic Canadian Triage and Acuity Scale (eCTAS), Ontario Ministry of Health)	• Institutional and provincial daily data for number of patients who were PUI, COVID-19+ and COVID-19- who presented to the ED and were admitted to acute and intensive care at University Health Network, Toronto, Canada
Data to inform patient location within hospital	• CORE model [7]	• Institutional and provincial databases on acute care and intensive care bed occupancy
PPE policy considered	• Ontario Provincial Recommendation for the use of PPE	• Ontario Provincial Recommendation for the use of PPE • Institutional COVID-19 specific PPE use policy: conservation of masks and face shields and universal masking of health care workers

(continued)

Table 4 (continued)

	Provincial-level study	Hospital-level case study
PPE Consumption	• Patient touchpoints	• Patient touchpoints
Hospital staffing data	• Not applicable	• Total number of hospital employees: Institutional data of number of full-time equivalents • Resident and clinical fellow physician numbers: University of Toronto Post-Graduate Medical Education Office • Staff physician numbers: Institutional daily on-call schedules • ICU nurse:patient ratio assumed at 1:1/12 h shift • Ward nurse:patient ratio assumed at 1:4/12 h shift
Levels of analysis	• Ontario	• Hospital • ICU/ED • Ward
Scenario analyses	• None conducted	• Proportion of patients in hospital confirmed to have COVID-19 or be PUI (+/− 20% of basecase % PUI) • Ward nurse staffing ratios (1:6, 1:8) • Percentage of touchpoints for non-PUI patients deemed to be risk of AGMP through staff point of care assessments requiring upgrading in PPE: N95 mask and face shield (1%, 5%, 10%; base case 0%)

AGMP Aerosol Generating Medical Procedure, *COVID-19* Coronavirus Disease, *ED* Emergency Department, *ICU* Intensive Care Unit, *PPE* Personal Protective Equipment, *PUI* Person Under Investigation

to estimate the magnitude in difference for demand for PPE between a strategy of single use PPE to strategies employing conservation and universal masking, not to accurately predict the amount of PPE that this specific institution would require.

3 Results

For the *provincial-level study*, we estimated that 2198 confirmed COVID-19 patients were admitted to Ontario hospitals over the 60-day period from March 6–May 5, 2020. We estimated that there were an additional 48,356 COVID-19 or PUI

Table 5 Results for Province-level study: Total PPE demand for the care of suspected and confirmed COVID-19 patients in Ontario's acute care system using Ontario Provincial Recommendations for the use of PPE

	10-day demand (April 3–April 12)[a]	30-day demand (March 6–April 6)	45-day demand (March 6–April 20)	60-day demand (March 6–May 5)
Surgical masks	1,299,338	2,165,288	3,751,435	4,564,801
N95 masks	45,441	65,962	130,464	166,136
Gloves	3,688,383	5,815,652	10,726,711	13,572,199
Gloves (extended)	6384	10,296	17,148	20,311
Face shields	1,341,587	2,226,102	3,873,325	4,720,782
Face shields with drape	3192	5148	8574	10,156
Gowns	1,344,779	2,231,250	3,881,899	4,730,938

[a]Approximate epidemic peak

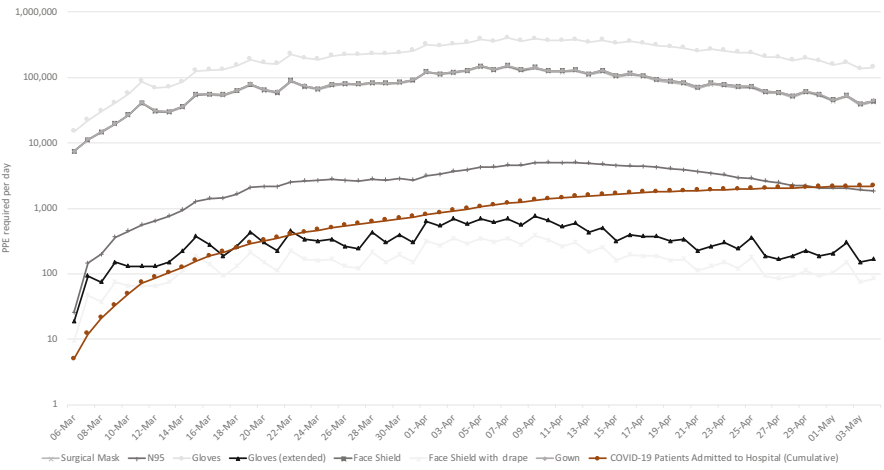


Fig. 1 Results for Province-level study: Total PPE demand for the care of suspected and confirmed COVID-19 patients in Ontario's acute care hospital system, March 6–May 5, 2020. Note: Y-Axis is on the logarithmic scale. *COVID-19* Coronavirus-19 disease, *PPE* Personal Protective Equipment

patients assessed in EDs who did not require admission. The total PPE required was estimated to be 13.5 million gloves, 4.7 million face shields and gowns, 4.5 million surgical masks, and 152,174 N95 masks over the 60-day period near the height of the pandemic (Table 5). PPE demand varied according to the epidemic pattern and resulting hospitalizations (Fig. 1).

Over the 10-day period of our *hospital-level case study*, daily total occupancy averaged: 166 patients assessed in the ED—of whom 15% were admitted, 440 patients occupying acute care beds, and 97 patients occupying ICU beds. The number of confirmed cases of COVID-19 admitted at UHN increased from 29 on April 7 to 66 on April 16, 2020. There was an average of 23 confirmed COVID-19 cases in the ICU per day.

Case study results are presented in Tables 6 and 7, and Fig. 2. Universal masking significantly increases demand for surgical masks at the institutional level given the large number of staff but is conserving at the ward and ICU level where PPE demand is driven by patient touchpoints. Our scenario analyses show that healthcare worker point of care risk assessments have a substantial impact on the demand for N95 masks. For example, the ward level analyses using provincial recommendations for PPE use, with 1% of touchpoints upgraded, resulted in an additional 2259 N95 masks used, but if N95 masks are re-used during one shift, the number of additional N95 masks needed is only 23. The same pattern was identified for analyses at the ICU/ED level.

4 Interpretation

Our results demonstrate that the PPE requirements to care for COVID-19 patients are substantial. We further demonstrate that policies aimed at reducing spread within institutions, such as universal masking, may result in substantial increases in PPE demand for surgical masks compared to provincial recommendation, and are unlikely to be offset by reduced PPE per patient touchpoint and conservation strategies. Institutional policies, without compensatory procurement, may therefore contribute to the depletion of supplies. In settings where the risk of community or nosocomial spread is low, policies of universal masking may unnecessarily deplete scarce resources; however, universal masking and conservation strategies may represent the best PPE policy in settings with high rates of community spread to help preserve supplies and protect healthcare workers.

A key strength of our model is the dynamic projection of COVID-19 cases and real-world patient contact "touchpoint" data. Simple consumption calculators, such as the CDC PPE calculator [14], are based on static historic trends of PPE use, and are not informed by changing epidemiology and clinical practice patterns. We caution against estimating PPE demand based on historical burn rates, as this may substantially underestimate PPE needs—especially if case volumes increase. Institutions at risk of depleting PPE stocks may implement conservation strategies such as reuse of masks or restricted distribution. It is imperative for health systems to plan PPE demand based on non-restrictive estimates of PPE use to prevent patient and healthcare worker morbidity and mortality. Our base-case analysis was therefore based on non-restrictive PPE utilization policies.

Our study is limited by the fact that we did not account for PPE consumed by touchpoints in some settings in acute care hospitals such as operating rooms or ambulatory clinics, and we do not therefore estimate the total PPE demand across the acute care sector. However, during the period studied, most elective surgeries were cancelled, and most ambulatory clinics were conducted virtually, and likely had little impact on demand for PPE. Touchpoint estimates and PPE policies used in our hospital-level analysis may not be generalizable to hospitals with different practices. We were unable to fully validate our results with actual PPE consumption.

Table 6 Results for hospital-level case study, base case analysis: Total PPE demand using institutional policies and Ontario provincial recommendations for the use of PPE over 10-day period (April 7–April 16, 2020)

	Institutional level analysis		ICU/ED level analysis		Ward level analysis	
	Institutional policy	Provincial recommendation	Institutional policy	Provincial recommendation	Institutional policy	Provincial recommendation
Surgical masks	157,720	52,195	8820	29,924	4500	22,271
N95 masks	4925	8603	4925	8603	0	0
Gloves	663,055	663,055	211,159	211,159	451,896	451,896
Gloves (extended)	1029	1029	1029	1029	0	0
Face shields	78,860	60,283	4410	38,012	2250	22,271
Face shields with drape	515	515	515	515	0	0
Gowns	60,797	60,797	38,526	38,526	22,271	22,271
Level 1 mask	83,250	0	4410	0	2250	0

ED Emergency Department, *ICU* Intensive Care Unit, *PPE* Personal Protective Equipment

Table 7 Results for hospital-level case study, scenario analyses

		Surgical mask (Level 3)	N95 facemask	Gloves	Gloves extended length	Face shield	Face shield with drape	Gown	Face mask (Level 1)
Institutional level analyses									
Institutional base case:									
0% touchpoints upgraded	Institutional Policy	15,7720	4925	663,055	1029	78860	515	60,797	83,250
	Provincial Policy	52,195	8603	663,055	1029	60,283	515	60,797	0
	Difference	*10,5525*	*−3678*	*0*	*0*	*18,577*	*0*	*0*	*83,250*
Institutional scenario analysis 1:									
1% touchpoints upgraded	Institutional Policy	15,7720	5669	663,059	1029	78,860	515	63,511	83,250
	Provincial Policy	51,673	11,839	663,059	1029	62,997	515	63,511	0
	Difference	*106,047*	*−6170*	*0*	*0*	*15,863*	*0*	*0*	*83,250*
Institutional scenario analysis 2:									
5% touchpoints upgraded	Institutional Policy	157,720	8647	663,072	1029	78,860	515	74,368	83,250
	Provincial Policy	49,585	24,783	663,072	1029	73,853	515	74,368	0
	Difference	*108,135*	*−16,136*	*0*	*0*	*5007*	*0*	*0*	*83,250*
Institutional scenario analysis 3:									
10% touchpoints upgraded	Institutional Policy	15,7720	12,370	663,088	1029	78,860	515	87,938	83,250
	Provincial Policy	46,975	40,963	663,088	1029	87,424	515	87,938	0
	Difference	*110,745*	*−28,594*	*0*	*0*	*−8564*	*0*	*0*	*83,250*
ICU/ED level analyses									
ICU/ED base case:									
13% of ED patients assessed PUI/COVID+	Institutional Policy	8820	4925	211,159	1029	4410	515	38,526	4410
42% ICU patients PUI/COVID+	Provincial Policy	29,924	8603	211,159	1029	38,012	515	38,526	0
0% touchpoints upgraded.	*Difference*	*−21104*	*−3678*	*0*	*0*	*−33,602*	*0*	*0*	*4410*

(continued)

Table 7 (continued)

		Surgical mask (Level 3)	N95 facemask	Gloves	Gloves extended length	Face shield	Face shield with drape	Gown	Face mask (Level 1)
Scenario analysis 1:									
1% touchpoints upgraded	Institutional Policy	8820	4925	211,162	1029	4410	515	39,204	4410
	Provincial Policy	29,624	9579	211,162	1029	38,689	515	39,204	0
	Difference	*−20,804*	*−4655*	*0*	*0*	*−34,279*	*0*	*0*	*4410*
Scenario analysis 2:									
5% touchpoints upgraded	Institutional Policy	8820	4925	211,176	1029	4410	515	41,913	4410
	Provincial Policy	28,427	13,486	211,176	1029	41,398	515	41,913	0
	Difference	*−19,607*	*−8561*	*0*	*0*	*−36,988*	*0*	*0*	*4410*
Scenario analysis 3:									
10% touchpoints upgraded	Institutional Policy	8820	4925	211,192	1029	4410	515	45,300	4410
	Provincial Policy	26,931	18,368	211,192	1029	44,785	515	45,300	0
	Difference	*−18,111*	*−13,444*	*0*	*0*	*−40,375*	*0*	*0*	*4410*
Scenario analysis 4									
Proportion of PUI/COVID+ 20% lower	Institutional Policy	8820	4891	206,760	963	4410	481	30,821	4410
(10% in ED, 34% in ICU)	Provincial Policy	23,875	6945	206,760	963	30,339	481	30,821	0
	Difference	*−15,055*	*−2054*	*0*	*0*	*−25,929*	*0*	*0*	*4410*
Scenario analysis 5:									
Proportion of PUI/COVID+ 20% higher	Institutional Policy	8820	4960	215,372	1099	4410	550	46,057	4410
(15% ED, 50% ICU)	Provincial Policy	35,813	10,244	215,372	1099	45,507	550	46,057	0
	Difference	*−26,993*	*−5285*	*0*	*0*	*−41,097*	*0*	*0*	*4410*

Scenario analysis 6:									
Proportion of PUI/COVID+ 20% higher	Institutional Policy	8820	4960	215,391	1099	4410	550	49,176	4410
(15% ED, 50% ICU)	Provincial Policy	34,022	15,154	215,391	1099	48,626	550	49,176	0
5% touchpoints upgraded	*Difference*	*−25,202*	*−10,194*	*0*	*0*	*−44,216*	*0*	*0*	*4410*
Scenario analysis 7:									
Proportion of PUI/COVID+ 20% higher	Institutional Policy	8820	4960	215,411	1099	4410	550	52,294	4410
(15% ED, 50% ICU)	Provincial Policy	32,232	20,063	215,411	1099	51,745	550	52,294	0
10% touchpoints upgraded	*Difference*	*−23,412*	*−15,103*	*0*	*0*	*−47,335*	*0*	*0*	*4410*
Ward level analyses									
Ward base case:									
10% ward patients PUI/COVID+	Institutional Policy	4500	0	451,896	0	2250	0	22,271	2250
0% touchpoints upgraded	Provincial Policy	22,271	0	451,896	0	22,271	0	22,271	0
225 staff/24 h (1:4 ratio)	*Difference*	*−17,771*	*0*	*0*	*0*	*−20,021*	*0*	*0*	*2250*
Scenario analysis 1:									
1% touchpoints upgraded	Institutional Policy	4500	23	45,1896	0	2250	0	24,308	2250
	Provincial Policy	22,048	2259	451,896	0	24,308	0	24,308	0
	Difference	*−17,548*	*−2237*	*0*	*0*	*−22,058*	*0*	*0*	*2250*
Scenario analysis 2:									
5% touchpoints upgraded	Institutional Policy	4500	113	451,896	0	2250	0	32,455	2250
	Provincial Policy	21,157	11,297	451,896	0	32,455	0	32,455	0
	Difference	*−16,657*	*−11,185*	*0*	*0*	*−30,205*	*0*	*0*	*2250*

(continued)

Table 7 (continued)

		Surgical mask (Level 3)	N95 facemask	Gloves	Gloves extended length	Face shield	Face shield with drape	Gown	Face mask (Level 1)
Scenario analysis 3:									
10% touchpoints upgraded	Institutional Policy	4500	225	451,896	0	2250	0	42,639	2250
	Provincial Policy	20,044	22,595	451,896	0	42,639	0	42,639	0
	Difference	*−15,544*	*−22,370*	*0*	*0*	*−40,389*	*0*	*0*	*2250*
Scenario analysis 4:									
Proportion of PUI/COVID+ 20% lower (8%)	Institutional Policy	4500	0	452,666	0	2250	0	17,417	2250
	Provincial Policy	17,417	0	452,666	0	17,417	0	17,417	0
	Difference	*−12,917*	*0*	*0*	*0*	*−15,167*	*0*	*0*	*2250*
Scenario analysis 5:									
Proportion of PUI/COVID+ 20% higher (12%)	Institutional Policy	4500	0	451,917	0	2250	0	26,126	2250
	Provincial Policy	26,126	0	451,917	0	26,126	0	26,126	0
	Difference	*−21,626*	*0*	*0*	*0*	*−23,876*	*0*	*0*	*2250*
Scenario analysis 6:									
Staffing Ratio 1:6 (150/24 h)	Institutional Policy	3000	0	451,896	0	1500	0	22,271	1500
	Provincial Policy	22,271	0	451,896	0	22,271	0	22,271	0
	Difference	*−19,271*	*0*	*0*	*0*	*−20,771*	*0*	*0*	*1500*

Scenario analysis 7:									
Staffing Ratio 1:8 (112/24 h)	Institutional Policy	2240	0	451,896	0	1120	0	22,271	1120
	Provincial Policy	22,271	0	451,896	0	22,271	0	22,271	0
	Difference	*−20,031*	*0*	*0*	*0*	*−21,151*	*0*	*0*	*1120*
Scenario analysis 8:									
Proportion of PUI/COVID+ 20% higher (12%)	Institutional Policy	4500	113	451,917	0	2250	0	36,117	2250
5% touchpoints upgraded	Provincial Policy	24,820	11,298	451,917	0	36,117	0	36,117	0
	Difference	*−20,320*	*−11,185*	*0*	*0*	*−33,867*	*0*	*0*	*2250*
Scenario analysis 9:									
Proportion of PUI/COVID+ 20% higher (12%)	Institutional Policy	4500	225	451,917	0	2250	0	46,109	2250
10% touchpoints upgraded	Provincial Policy	23,513	22,596	451,917	0	46,109	0	46,109	0
	Difference	*−19,013*	*−22,371*	*0*	*0*	*−43,859*	*0*	*0*	*2250*

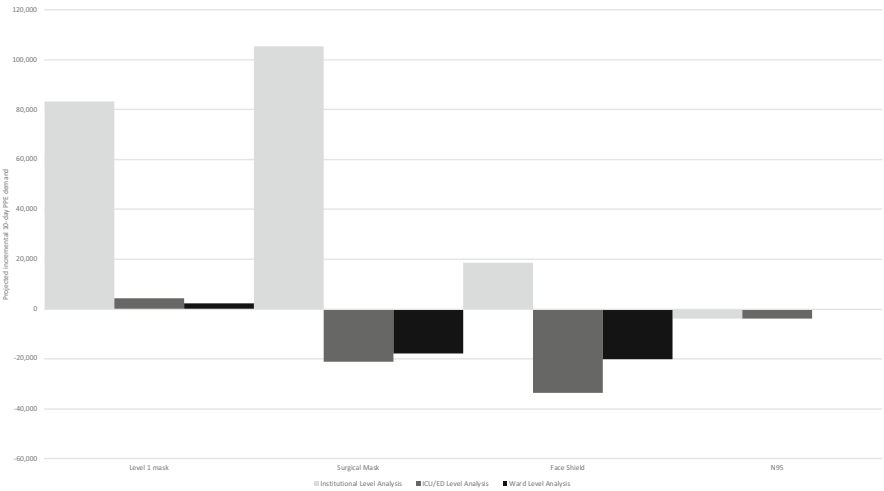


Fig. 2 Results for hospital-level case-study: Incremental PPE demand over 10-day period institutional PPE policy compared to provincial recommendations. Note: Positive numbers indicate institutional policy resulted in higher PPE demand. Negative numbers indicate institutional policy resulted in lower PPE demand. There was no difference in PPE demand for gloves, gowns, extended gloves, and face shield with drape. *ED* Emergency Department, *ICU* Intensive Care Unit, *PPE* Personal Protective Equipment

Further, PPE use may be influenced by other social factors (e.g. fear among staff), which may influence the proportion of point of care assessments that are upgraded and consume additional PPE. We did not explore the human factors that influence point of care risk assessments.

Our methodology to determine PPE demand based on case trajectories, patient touchpoints, and policy or site-specific PPE bundles can be applied to other settings, and can be adapted for use across the healthcare sector. To facilitate the adoption of our methods, we have published our model as a web-based, interactive tool, (www.covid-19-mc.ca) that allows users to customize model parameter inputs and generate results applicable to their local institution.

We demonstrate that the estimated volume of PPE required to care for confirmed and suspected COVID-19 cases in the Ontario acute care sector is substantial. Our case study further shows that institutional PPE policies that differ from provincial recommendations, such as universal masking, may result in a considerable increase in PPE demand, which needs to be considered in procurement. Decision-makers at all levels of government need to consider local epidemiology, practice patterns, and local policies that differ from provincial recommendations to procure adequate supplies of PPE over the coming months, particularly in the context of a subsequent waves of COVID-19.

References

1. 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(18):e41.
2. Artenstein AW. In Pursuit of PPE. N Engl J Med. Massachusetts Medical Society; 2020;382(18):e46.
3. Carter A. What we know about the Ontario long-term care homes in the military's scathing report. CBC News [Internet]. 2020. https://www.cbc.ca/news/canada/toronto/long-term-care-covid-19-caf-report-1.5587015
4. Vogel L. Canada's PPE crisis isn't over yet, say doctors [Internet]. CMAJ News. 2020 [cited 2020 Jun 5]. http://cmajnews.com/2020/04/29/covid-shortage-1095868/
5. Watts M, Wells A, Putyra M. Court places decision making for use of PPE in long-term care homes solely in the hands of nurses [Internet]. 2020 [cited 2020 Jun 5]. https://www.osler.com/en/resources/regulations/2020/court-places-decision-making-for-use-of-ppe-in-long-term-care-homes-solely-in-the-hands-of-nurses
6. Ontario Agency for Health Protection, Promotion PIDAC. Routine Practices and Additional Precautions in All Health Care Settings. 3rd Edition. Toronto: Queen's Printer for Ontario; 2012. 113 p.
7. Barrett KA, Khan YA, Mac S, Ximenes R, Naimark DMJ, Sander B. Estimation of COVID-19-induced depletion of hospital resources in Ontario, Canada, CMAJ. 2020; 192(24):E640-E646.
8. Ontario Ministry of Health. Electronic Canadian Triage and Acuity Scale (eCTAS).
9. Ontario Ministry of Health. Daily Bed Census.
10. Critical Care Services Ontario. Provincial and Regional CCSO COVID-19 Dashboard Reports.
11. Fan E, Del Sorbo L, Goligher EC, Hodgson CL, Munshi L, Walkey AJ, et al. An Official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine Clinical Practice Guideline: Mechanical Ventilation in Adult Patients with Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med. 2017;195(9):1253–63.
12. Guérin C, Reignier J, Richard J-C, Beuret P, Gacouin A, Boulain T, et al. Prone Positioning in Severe Acute Respiratory Distress Syndrome. N Engl J Med. 2013;368(23):2159–68.
13. Ontario Agency for Health Protection, Ontario PPH. IPAC recommendations for use of personal protective equipment for care of individuals with suspect or confirmed COVID-19. Toronto: Queen's Printer for Ontario; 2020.
14. Centers for Disease Control, Prevention. Personal Protective Equipment (PPE) Burn Rate Calculator [Internet]. [cited 2020 Jun 5]. https://www.cdc.gov/coronavirus/2019-ncov/hcp/ppe-strategy/burn-calculator.html

Don't Wait, Re-escalate: Delayed Action Results in Longer Duration of COVID-19 Restrictions

Amy Hurford and James Watmough

1 Introduction

Non-pharmaceutical public health interventions are an essential tool during a pandemic. Indeed, we can expect they will often be the only available tools at the start of a pandemic when little is known about a novel pathogen. However, these interventions have significant economic and social costs, and minimizing the duration and impact of interventions is paramount. In Canada, and in particular in Atlantic Canada and the Territories, contact tracing and testing combined with travel restrictions, quarantine of incoming travellers, and social distancing, including business closures and limits on larger gatherings are the main tools in use during the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. These have proven very effective in the smaller centres. At the time of writing border restrictions and traveller quarantine in conjunction with testing and contact tracing have allowed most businesses and social activity to resume a new, albeit slightly restricted, normal.

Current modelling efforts to understand the relative efficiency of the interventions include a full range of models from simple phenomenological forecasts [12] to complex network and agent-based models explicitly including households and healthcare systems [1, 6, 15]. While these details are necessary to predict healthcare needs and monitor capacity, we argue that the strategy for timing of implementing non-pharmaceutical community interventions can be based on much simpler linear deterministic models.

A. Hurford (✉)
Memorial University, St. John's, NL, Canada
e-mail: ahurford@mun.ca

J. Watmough
University of New Brunswick, Fredericton, NB, Canada
e-mail: watmough@unb.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_11

Specifically, we use a simple linear SIR model with case importation to determine the relationship between the timing of restrictions, duration of measures necessary to return the incidence to a set point, and the final size of the outbreak. We conclude from our analysis that delaying re-escalation of restrictions leads to increased duration of control measures and larger outbreaks. Conversely, earlier re-escalation results in shorter disruptions, smaller outbreaks, and consequently, lower economic and social costs.

2 Linear SIR Model

Consider a linear Susceptible-Infected-Recovered [8] model,

$$\frac{dI(t)}{dt} = ac_i S\frac{I(t)}{N} - \gamma I(t) + m_i \tag{1a}$$

$$\frac{dC(t)}{dt} = ac_i S\frac{I(t)}{N}, \tag{1b}$$

where $I(t)$ is the number of infected individuals, and $C(t)$ is the cumulative cases at time, t. The contact rate, c_i, is the rate that a susceptible individual contacts another individual in the population, $\frac{I(t)}{N}$ is the probability that the contacted individual is infected, where N is the total population size, and a is the probability of infection given a contact. The rate that infected individuals recover is γ infected individuals per day and the rate that cases are imported is m_i cases per day. The appropriate initial values are $I(0) = I_0$ and $C(0) = 0$.

The number of susceptible individuals in the population is S, and as an approximation, S is assumed to be unchanging, such that S/N is approximately equal to 1. This assumption is justified when the number of susceptible individuals is very large, and only a small fraction of these susceptible individuals have been infected during the study period. This is the case in many regions, for example, in Canada, as of October 29, 2020, there have been 228,542 clinical coronavirus disease 2019 (COVID-19) cases [9], which represents 0.6% of the Canadian population [10]. If 40% of cases are asymptomatic, then only around 1% of the Canadian population has been infected to date.

We consider the epidemiological dynamics subject to restrictions that affect the contact rate and the importation rate, and are implemented at $t = t_1$. Let c_i and m_i denote the contact and importation rates where $i = 1$ indicates prior to restrictions ($0 \leq t \leq t_1$), and $i = 2$ indicates after restrictions ($t_1 < t \leq t_2$). To simplify the notation, let $\beta_i = ac_i S/N = ac_i$ and $\lambda_i = \beta_i - \gamma$.

We consider a restricted parameter space, such that the number of infected individuals increases until the restrictions are implemented, and then decreases. This scenario represents the epidemiological dynamics for outbreaks that have subsequently been eradicated, for example outbreaks in New Zealand, Taiwan [17], and the provinces of Atlantic Canada [9]. The parameter space that we consider

is such that prior to restrictions the outbreak grows. There are two parameter combinations that allow for this possibility: (1) importations occur $m_1 > 0$; or (2) if importations do not occur, $m_1 = 0$ and $\lambda_1 > 0$. After restrictions, we assume that $\lambda_2 < 0$ and $m_2 = 0$, such that the outbreak dissipates, and infection is eventually eradicated.

Under these assumptions, the outbreak exponentially increases and then exponentially decreases, and we can solve Eq. (1a), such that,

$$I(t) = \begin{cases} \left(I_0 + \frac{m_1}{\lambda_1}\right) e^{\lambda_1 t} - \frac{m_1}{\lambda_1}, & \text{when } 0 \le t \le t_1, \text{ and} \\ I_1 e^{\lambda_2 (t - t_1)}, & \text{when } t_1 < t \le t_2 \end{cases} \tag{2}$$

where I_1 is the number of infected individuals when restrictions are implemented:

$$I_1 = \left(I_0 + \frac{m_1}{\lambda_1}\right) e^{\lambda_1 t_1} - \frac{m_1}{\lambda_1}. \tag{3}$$

Note that after the restrictions are enacted ($t > t_1$), the number of infections, $I(t)$, is decreasing (since $\lambda_2 < 0$), and when the number of infected individuals decreases below a value I_2, we assume that the restrictions might be lifted, and then the duration of the restrictions is calculated as,

$$\tau = \frac{1}{\lambda_2} \log\left(\frac{I_2}{I_1}\right), \tag{4}$$

where we have used log to denote the natural logarithm. If $\lambda_1 > 0$, the duration of restrictions, τ, is increasing with respect to the timing of restrictions, t_1:

$$\frac{d\tau}{dt_1} = -\frac{\lambda_1}{\lambda_2}\left(1 + \frac{m_1}{\lambda_1 I_1}\right). \tag{5}$$

Hence, delaying the implementation of the restrictions results in a longer period that the restrictions must be in place to reduce the active number of cases, $I(t)$, below the target I_2. If $\lambda_2 < \lambda_1 < 0$, but $\lambda_1 > -m_1/I_1$, the same result holds. However, if $\lambda_2 < \lambda_1 < -m_1/I_1$, then τ is decreasing with respect to t_1. Since in this last scenario, the number of active cases, $I(t)$ is decreasing for $0 < t < t_1$, it is not likely a scenario that would trigger re-escalation.

Next, we calculate the total number of infections (also referred to as the final size [4]). The cumulative number of infections at the time when the restrictions are implemented is,

$$C(t_1) - C(0) = \int_0^{t_1} \frac{dC(t)}{dt}\, dt,$$

$$C(t_1) - 0 = \int_0^{t_1} \beta_1 I(t)\, dt,$$

$$
\begin{aligned}
C(t_1) &= \beta_1 \int_0^{t_1} \left(I_0 + \frac{m_1}{\lambda_1} \right) e^{\lambda_1 t} - \frac{m}{\lambda_1}\, dt, \\
&= \beta_1 \left[\frac{1}{\lambda_1} \left(I_0 + \frac{m_1}{\lambda_1} \right) e^{\lambda_1 t} - \frac{m}{\lambda_1} t \right]_0^{t_1}, \\
&= \frac{\beta_1}{\lambda_1} \left[\left(I_0 + \frac{m_1}{\lambda_1} \right) e^{\lambda_1 t_1} - m_1 t_1 - \left(I_0 + \frac{m_1}{\lambda_1} \right) \right], \\
&= \frac{\beta_1}{\lambda_1} (I_1 - I_0 - m_1 t_1). \qquad (6)
\end{aligned}
$$

The total number of cases occurring after restrictions, which we denote $C(t_2)$, is found by a simple integration:

$$
\begin{aligned}
C(t_2) - C(t_1) &= \int_{t_1}^{t_2} \beta_2 I(t)\, dt, \\
&= \beta_2 \int_{t_1}^{t_2} I_1 e^{\lambda_2 (t - t_1)}\, dt, \\
&= \frac{\beta_2}{\lambda_2} I_1 e^{\lambda_2 (t - t_1)} \Big|_{t_1}^{t_2}, \\
&= \frac{\beta_2}{\lambda_2} I_1 \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right). \qquad (7)
\end{aligned}
$$

Therefore, adding Eqs. (6) and (7), the total number of infections in the outbreak is,

$$
C(t_2) = \frac{\beta_1}{\lambda_1} (I_1 - I_0 - m_1 t_1) + \frac{\beta_2}{\lambda_2} I_1 \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right). \qquad (8)
$$

Returning to Eq. (7), and rearranging Eq. (6) such that $I_1 = \frac{\lambda_1}{\beta_1} C(t_1) + I_0 + m_1 t_1$, we have that,

$$
\begin{aligned}
C(t_2) &= \frac{\beta_2}{\lambda_2} I_1 \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right) + C(t_1), \\
&= \frac{\beta_2}{\lambda_2} \left(\frac{\lambda_1}{\beta_1} C(t_1) + I_0 + m_1 t_1 \right) \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right) + C(t_1), \\
&= C(t_1) \left(\frac{\beta_2}{\lambda_2} \frac{\lambda_1}{\beta_1} \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right) + 1 \right) + \frac{\beta_2}{\lambda_2} (I_0 + m_1 t_1) \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right), \qquad (9)
\end{aligned}
$$

such that the total number of cases in the outbreak is a linearly increasing function of the total number of cases up until the restrictions are enacted, with a slope of $\left(\frac{\beta_2}{\lambda_2} \frac{\lambda_1}{\beta_1} \left(e^{\lambda_2 (t_2 - t_1)} - 1 \right) + 1 \right)$.

3 Comparing the Linear SIR Model Predictions to Data

To test the predictions of the linear SIR model, we fit Eq. (2) to data describing the number of active infections in Newfoundland and Labrador (NL) and New Brunswick (NB) for the March–May 2020 outbreaks that occurred in these Canadian provinces (Fig. 1; where these data are from [7]). We defined $t = 0$ as the day that the number of active cases first exceeded 10. We then use the estimated values for λ_1, λ_2 and m_1 to predict the number of active cases at the peak of the outbreak (Eq. (3)), the required duration of restrictions (Eq. (4)), and the total number of cases in the outbreak (Eq. (8) and Table 1). We assumed that restrictions were enacted on the day, t_1, corresponding to the peak number of active infections in the data. Fitting was performed using maximum likelihood and assuming normally distributed residuals.

We found a close agreement between the predictions of the linear SIR model and these same quantities calculated from data (Table 1), suggesting that the linear SIR model can effectively predict the necessary duration of restrictions.

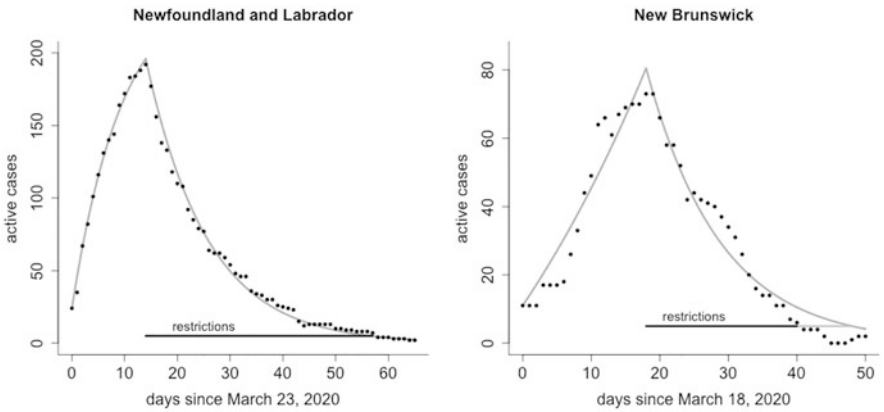


Fig. 1 The linear SIR model fitted to data describing the active number of COVID-19 cases in Newfoundland and Labrador and New Brunswick from March-May, 2020. Equation (1a), describing the number of active infections for the linear SIR model (grey line), is fit to data describing the number of active COVID-19 cases (black dots). We assume that restrictions are enacted at the time of the peak number of active infections in the data and remain until there are fewer than 5 active cases (black horizontal line). The duration of restrictions, τ, is predicted by the linear SIR model (Eq. (4); grey horizontal line). The fitted model parameters and the comparison of the model predictions with the data are reported in Table 1

Table 1 Parameter estimates and validation of the linear SIR model fitted to active case data (see Fig. 1)

		Newfoundland and Labrador		New Brunswick	
	Units	Fixed parameters			
Date corresponding to $t = 0$	Date	March 23, 2020		March 18, 2020	
Recovery rate, γ	Individuals per day	1/14		1/14	
Active cases when restrictions lifted, I_2	Individuals	5		5	
Initial number of infected individuals, $I(0)$	Individuals	10		10	
Day restrictions are enacted, t_1	Day	14		18	
End time, t_2	Day	65		50	
		Fitted parameters			
Exponential coefficient prior to restrictions, λ_1	Per day	-0.104		0.027	
Exponential coefficient after restrictions, λ_2	Per day	-0.086		-0.093	
Importation rate, m_1	Individuals per day	25.9		2.70	
		Data	Predicted	Data	Predicted
Active infected individuals at peak, I_1	Individuals	192	197	73	81
Duration of restrictions, τ	Days	43	43	22	30
Total number of cases, $C(t_2)$	Individuals	260	293	120	117

4 Comparison of the Linear SIR Model to a Nonlinear COVID-19 Model

The linear SIR model (Eqs. (1)) assumes that the number of susceptible individuals is not changing, and does not distinguish between different types of infected individuals, while deterministic mathematical models specifically developed for COVID-19 typically include these features. In this section, we examine the agreement between the predictions of the linear SIR model and a nonlinear COVID-19 model with different types of infected individuals, where the comparison between these models is illustrated in Fig. 2. The nonlinear COVID-19 model that we consider is based on [14], but without any age or spatial structure. The COVID-19 model equations are as follows:

$$\frac{dS(t)}{dt} = -\beta_i S(t)\frac{I_P(t) + b_C I_C(t) + b_A I_A(t)}{N}, \tag{10a}$$

$$\frac{dL(t)}{dt} = \beta_i S(t)\frac{I_P(t) + b_C I_C(t) + b_A I_A(t)}{N} - \delta_L(t), \tag{10b}$$

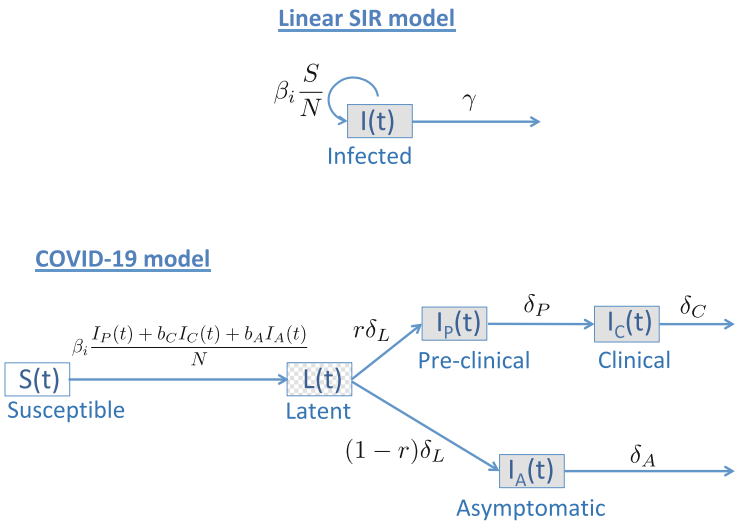


Fig. 2 Diagram comparing the Linear SIR model and the COVID-19 model. The linear SIR model has just one dynamic state, the number of infected individuals, $I(t)$, and the number of susceptible individuals, S, is assumed to be unchanging. The COVID-19 model (Eqs. (10)) has five dynamic states: susceptible individuals, $S(t)$, latently infected individuals, $L(t)$, and individuals with: pre-clinical infections (not yet showing symptoms), $I_P(t)$; clinical infections (showing symptoms), $I_C(t)$; and asymptomatic infections, $I_A(t)$. States representing infected individuals have some grey fill, infectious states are have solid grey fill, and since latently infected individuals are not infectious this compartment has a checkered grey fill. All model parameters are defined in Table 2

$$\frac{dI_P(t)}{dt} = r\delta_L L(t) - \delta_P I_P(t), \tag{10c}$$

$$\frac{dI_C(t)}{dt} = \delta_P I_P(t) - \delta_C I_C(t), \tag{10d}$$

$$\frac{dI_A(t)}{dt} = (1-r)\delta_L L(t) - \delta_A I_A(t), \tag{10e}$$

where $S(t)$ is the number of susceptible individuals, $L(t)$ is the number of latently infected individuals (infected, but not yet infectious), $I_P(t)$ is the number of individuals that have pre-clinical infections (infectious, but yet to be symptomatic), $I_C(t)$ is the number of individuals that have clinical infections (infectious and symptomatic), and $I_A(t)$ is the number of individuals that have asymptomatic infections (infectious, but without symptoms). All model parameters are defined and estimated in Table 2.

For these more complex COVID-19 mathematical models, analytic solutions such as the results in the *Linear SIR model* section are not possible. This model (Eqs. (10)) is a variation on a class of SLIAR models previously studied for influenza and SARS-CoV-2 [3, 5]. Several relevant results for this class of models, including general formulae for the basic reproduction number and the final size of the epidemic were established earlier [4].

Many models for SARS and influenza pandemics, including those for the current COVID-19 pandemic, fall into the following general form:

$$\frac{dx(t)}{dt} = \beta S(t)\Pi b x(t) - V x(t), \tag{11a}$$

$$\frac{dS(t)}{dt} = -\beta S(t) b x(t), \tag{11b}$$

$$\frac{dz(t)}{dt} = W x(t), \tag{11c}$$

where $x(t)$ is a vector of numbers of infected individuals in each stage, $S(t)$ is again the dynamically changing number of susceptible individuals, and $z(t)$ is a vector of numbers of individuals recovered through various routes of progression (typically either recovered from asymptomatic infection, recovered with only mild symptoms, recovered following hospitalization, or deceased). The parameter β is as defined previously, the vectors Π and b indicate the initial stage(s) of infection and relative transmission rates of the infectious stages, respectively, such that Πb is a rank one matrix. The matrices V and W contain rates of progression through and out of the stages of infection and recovery. This is a simplification of the Arino et al model which included a structured susceptible population [4]. The COVID-19 model (Eqs. (10)) is the special case of the general SLIAR model with $x(t) = (L(t), I_P(t), I_C(t), I_A(t))$, $\Pi = (1, 0, 0, 0)^T$, $b = (0, 1, b_C, b_A)$, and V a 4×4 matrix containing the remaining parameters. The total numbers of observed cases, $C(t)$, can be included by setting $z(t) = C(t)$ and $W = (0, \delta_P, 0, 0)$.

Table 2 Parameters and initial values for the linear SIR and the COVID-19 model

Description	Symbol	Units	Value	Reference
COVID-19 model				
Transmission rate prior to restrictions	β_1	Per time	0.571	To achieve 7 day doubling time
Transmission rate after to restrictions	β_2	Per time	0.185	To achieve 10 day halving time
Infectiousness of clinical relative to pre-clinical cases	b_C	Unitless	0.5	[11, 14]
Infectiousness of	b_A	Unitless	0.5	[11, 14]
Rate of become infectious after exposure	δ_L	Per time	1/3	[11, 14]
Rate of developing symptoms after becoming infectious	δ_P	Per time	1/2.1	[11, 14]
Rate of recovery for a clinical infection	δ_C	Individuals per time	1/2.9	[11, 14]
Rate of recovery for an asymptomatic infection	δ_A	Individuals per time	1/5	[11, 14]
Proportion of cases that are symptomatic	r	Unitless	0.6	Estimates vary, but see [11]
Initial values				
Initial susceptible	$S(0)$	Individuals	99,990	
Initial with preclinical infection	$I_P(0)$	Individuals	1.5	
Initial with clinical infection	$I_C(0)$	Individuals	2.5	
Initial with asymptomatic infection	$I_A(0)$	Individuals	3.0	
Initial latently infected	$L(0)$	Individuals	3.0	

(continued)

Table 2 (continued)

Description	Symbol	Units	Value	Reference
Linear SIR model				
Transmission rate prior to restrictions	β_1	Per time	0.224	To achieve 7 day doubling time
Transmission rate after to restrictions	β_2	Per time	0.057	To achieve 10 day halving time
Rate of importations prior to restrictions	m_1	Cases per time	0	To be consistent with the COVID-19 model
Rate of recovery	γ	Per time	1/8	$= 1/\delta_L + 1/\delta_A$ To be consistent with the COVID-19 model
Initial values				
Initial infected	$I(0)$	Individuals	10	
Initial cumulative infections	$C(0)$	Individuals	0	
Both models				
Total population size	Individuals	N	100,000	
Number of active cases when restrictions end	Cases	I_2	10	
Day when restrictions are enacted	Day	t_1	1–45 days	

Following the same analysis as we have done with the linear SIR model, we extend the Eqs. (11a) to include importation of infected individuals at rate m and linearize assuming $S(t)$ remains near S_0:

$$\frac{dx}{dt} = \beta S_0 \Pi b x - V x + m. \tag{12}$$

The solution can be expressed as $x(t) = \int_0^\infty e^{As} m(t-s)\, ds$ where for simplicity we have set $A = \beta S_0 \Pi b - V$. As before, supposing restrictions are put in place at time t_1, and that $\beta = \beta_1$ before restrictions and m is constant on $(0, t_1)$ and zero otherwise, and that after time t_1 we have $\beta = \beta_2$ and m identically zero, leads to solutions for $C(t_1)$ and $C(t_2)$. However, the resulting transcendental relations do not obviously lead to the simple expressions for the duration of restrictions or the total number of cases as with the linear SIR model. Hence, we instead proceed to a numerical comparison of the simple linear SIR model (Eqs. (1)) and the COVID-19 model (Eqs. (10)).

We assumed that the initial number of individuals in each infected state for the COVID-19 model (Eqs. (10)) was close to the right eigenvector (Table 2). We numerically solved Eqs. (10) in R using the deSolve package [16]. We compared the number of infected individuals under each model formulation ($I(t)$ for the linear SIR model and $L(t) + I_P(t) + I_C(t) + I_A(t)$ for the COVID-19 model). Figure 3 shows a close agreement between the dynamics of the linear SIR model (Eq. (1a)) and the nonlinear COVID-19 model (Eqs. (10)). All data and code for this Chapter are archived at https://github.com/ahurford/reescalation-chapter.

5 Discussion

We derive some simple equations to calculate the required duration of restrictions necessary to meet a target number of active infections (Eq. (4)) and the total number of cases in an outbreak (Eq. (8)), and while the model itself (Eq. (1a)) lacks complexity, when parameterized properly (Table 2), the dynamics are nearly identical to those of more complex nonlinear coupled ordinary differential equations (Eqs. (10), i.e. [14]). We also show that it is simple to estimate the parameters for the linear SIR model from epidemic data (Fig. 1), which then allows for the duration of restrictions and the total number of cases to be predicted using Eqs. (4) and (8). From the expression for the necessary duration of restrictions (Eq. (4)), we are able to show that restrictions always need to be in place longer when actions is delayed (Eq. (5)).

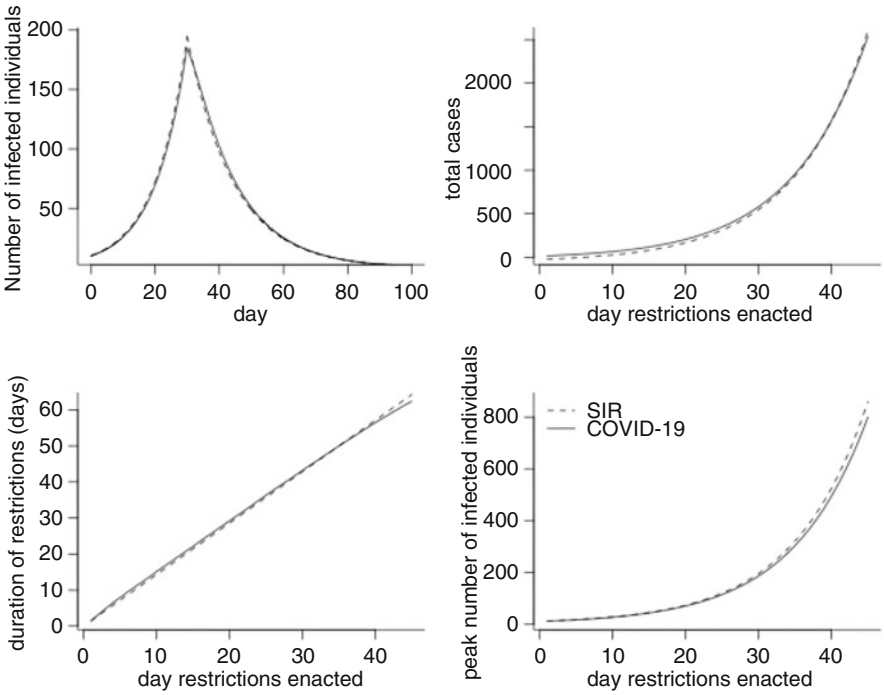


Fig. 3 The dynamics and predictions of the linear SIR model are in close agreement with the nonlinear COVID-19 model. We find a close agreement between the linear SIR model (dashed lines; Eq. (1a)) and the COVID-19 model (solid lines; Eqs. (10)) for the number of infected individuals over time (top left), the total number of cases during an outbreak (top right), the duration of restrictions (bottom left), and the peak number of infected individuals (bottom right). For the COVID-19 model, the number of infected individuals is calculated as $I_P(t) + I_C(t) + I_A(t) + L(t)$. All parameters values are listed in Table 2

5.1 Public Health Implications

Whether public health measures will be enacted to mitigate the spread of COVID-19 is a decision that public health officials should make prior to the discovery of exponentially growing cases numbers in the local community. Regarding SARS-CoV-2, the novel coronavirus, whereby the majority of the population is susceptible to infection, waiting to enact the restrictions only delays the inevitable, whereby restrictions with undesirable impacts will still have to occur, just at a later date. However, while action is delayed, the number of cases will continue to increase exponentially, and an unintuitive consequence of delayed restrictions, is that the restrictions, and their undesirable impacts, will need to be in place longer, to reduce the number of cases to the same level (Fig. 4).

While delayed restrictions might be justified to protect the economy and keep children in school, this is short sighted: delayed action will yield these benefits *now*,

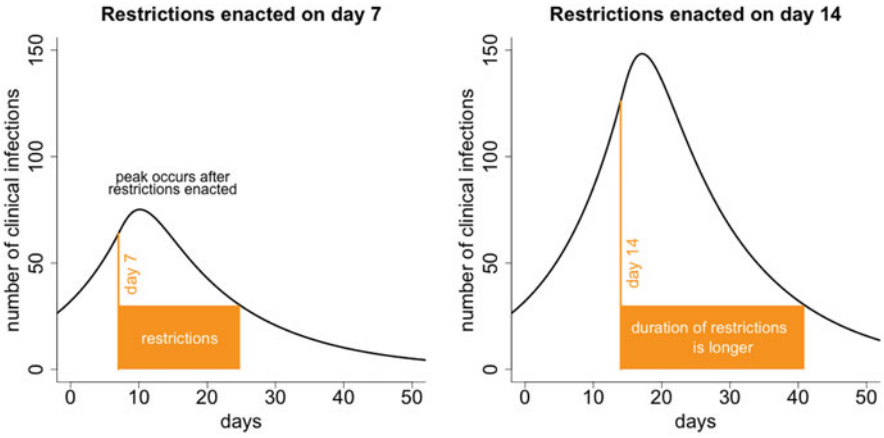


Fig. 4 Delayed action results in a longer duration of COVID-19 restrictions. We calculated the number of active clinical infections, $I_C(t)$, (black line) for the COVID-19 model (Eqs. (10)) with restrictions enacted on either day 7 (left) or day 14 (right) and remaining in place until there are 30 or fewer active clinical cases of COVID-19 (orange boxes). Delaying action results in a longer duration of restrictions, because a constant decay rate beginning from a higher peak will require longer to reduce the number of cases to the same level. The peak in the number of clinical cases occurs after the enactment of the restrictions because some individuals that were infected prior to the restrictions have yet to show symptoms. Elsewhere in this chapter (Eq. (5) and Fig. 3), we have shown that delaying restrictions results in a longer period of restrictions generally, not just for restrictions at 7 and 14 days as illustrated here. All parameter values are listed in Table 2

while ultimately these impacts will still be felt, just later and for longer. The only reasons to delay restrictions are if (1) a vaccine is likely in the near future, (2) conditions are likely to change in the near future, for example, a school break or a low tourism season, (3) there was never any intention to implement any restrictions, or (4) infection prevalence in the community is very low, such that exponential growth has not been established, and the outbreak may go extinct without any intervention.

Frequently, the question has been posed 'what should be a trigger for lockdowns?', where the trigger is in terms of the number of active cases [13]. We suggest that, providing that exponential growth is established, there is no benefit to delaying action. Such early action does not prioritize epidemiology over other considerations, because early action also enables the restrictions to be lifted after a shorter time, allowing for unrestricted economic, social, and education activities to resume sooner, and potentially allowing for a longer period of unrestricted activities until restrictions need to be enacted again.

Finally, while the effective reproduction number, R_t, has become a popular method for communicating epidemic trends, this metric has limitations, including that R_t is a lagged measure since the number of secondary infections generated per infected person can only be known after that person recovers, or by making assumptions via the method of nowcasting [2]. By contrast, the exponential growth

rates, λ_1, and λ_2 are instantaneous measures, and can be easily communicated as doubling times. As an alternative to R_t, we suggest estimating λ_1, λ_2 and m_1, as shown in Fig. 1 and using Eq. (1a) to predict the future time course of the epidemic, and using Eq. (4) to communicate the likely duration of restrictions shortly after they have been enacted.

Our key message is succinctly summarized by the phrase 'Don't wait, re-escalate' as coined by Christina Bancej of the Public Health Agency of Canada. We find that delaying re-escalation of restrictions to prevent the spread of COVID-19 results in not only in more infections, but also longer periods of restrictions. As such, we recommend not waiting to enact restrictions.

References

1. Abdollahi, E., Haworth-Brockman, M., Keynan, Y., Langley, J.M., Moghadas, S.M.: Simulating the effect of school closure during COVID-19 outbreaks in Ontario, Canada **18**(1), 230. https://doi.org/10.1186/s12916-020-01705-8.
2. Adam, D.: A guide to R: the pandemics misunderstood metric **583**(7816), 346–348. https://doi.org/10.1038/d41586-020-02009-w. http://www.nature.com/articles/d41586-020-02009-w. Number: 7816 Publisher: Nature Publishing Group
3. Arino, J., Brauer, F., van den Driessche, P., Watmough, J., Wu, J.: Simple models for containment of a pandemic. **3**(8), 453–457. https://doi.org/10.1098/rsif.2006.0112. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1578753&tool=pmcentrez&rendertype=abstract. Publisher: The Royal Society London
4. Arino, J., Brauer, F., Van Den Driessche, P., Watmough, J., Wu, J.: A final size relation for epidemic models **4**(2), 159. Publisher: American Institute of Mathematical Sciences
5. Arino, J., Portet, S.: A simple model for COVID-19 Publisher: Elsevier
6. Barrett, K., Khan, Y.A., Mac, S., Ximenes, R., Naimark, D.M.J., Sander, B.: Estimation of COVID-19–induced depletion of hospital resources in Ontario, Canada **192**(24), E640–E646. https://doi.org/10.1503/cmaj.200715. http://www.cmaj.ca/content/192/24/E640. Publisher: CMAJ Section: Research
7. Berry, I., Soucy, J.P.R., Tuite, A., Fisman, D.: Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada **192**(15), E420–E420. https://doi.org/10.1503/cmaj.75262. http://www.cmaj.ca/content/192/15/E420. Publisher: CMAJ Section: Letters
8. Brauer, F.: Compartmental models in epidemiology **1945**, 19–79. https://doi.org/10.1007/978-3-540-78911-6_2
9. Canada, P.H.A.O.: Coronavirus disease (COVID-19): Outbreak update. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html. Last Modified: 2020-10-28
10. Government of Canada, S.C.: Canada's population clock (real-time model). https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2018005-eng.htm. Last Modified: 2020-09-29
11. Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M., Eggo, R.M.: Age-dependent effects in the transmission and control of COVID-19 epidemics **26**(8), 1205–1211. https://doi.org/10.1038/s41591-020-0962-9. http://www.nature.com/articles/s41591-020-0962-9. Number: 8 Publisher: Nature Publishing Group
12. Fisman, D.N., Hauck, T.S., Tuite, A.R., Greer, A.L.: An IDEA for short term outbreak projection: Nearcasting using the basic reproduction number **8**(12). https://doi.org/10.1371/journal.pone.0083622. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3877403/

13. Karatayev, V.A., Anand, M., Bauch, C.T.: Local lockdowns outperform global lockdown on the far side of the COVID-19 epidemic curve **117**(39), 24575–24580. https://doi.org/10.1073/pnas.2014385117. https://www.pnas.org/content/117/39/24575. Publisher: National Academy of Sciences Section: Biological Sciences
14. Miller, I.F., Becker, A.D., Grenfell, B.T., Metcalf, C.J.E.: Disease and healthcare burden of COVID-19 in the united states **26**(8), 1212–1217. https://doi.org/10.1038/s41591-020-0952-y. https://www.nature.com/articles/s41591-020-0952-y. Number: 8 Publisher: Nature Publishing Group
15. Ng, V., Fazil, A., Waddell, L.A., Bancej, C., Turgeon, P., Otten, A., Atchessi, N., Ogden, N.H.: Projected effects of nonpharmaceutical public health interventions to prevent resurgence of SARS-CoV-2 transmission in Canada **192**(37), E1053–E1064. https://doi.org/10.1503/cmaj.200990. http://www.cmaj.ca/content/192/37/E1053. Publisher: CMAJ Section: Research
16. Soetaert, K., Petzoldt, T., Setzer, R.W., ddaspk.f, P.N.B.f., dvode.f, zvode.f), dvode.f, G.D.B.f., zvode.f), radau5.f, E.H.f., radau5a), ddaspk.f, A.C.H.f., dlsode.f, dvode.f, zvode.f, opdkmain.f, opdka1.f), dlinpck.f), C.M.f., ddaspk.f, L.R.P.f., dlsoda.f), dsparsk.f), Y.S.f., ddaspk.f), C.W.U.f.: deSolve: Solvers for initial value problems of differential equations ('ODE', 'DAE', 'DDE'). https://CRAN.R-project.org/package=deSolve
17. Summers, D.J., Cheng, D.H.Y., Lin, P.H.H., Barnard, D.L.T., Kvalsvig, D.A., Wilson, P.N., Baker, P.M.G.: Potential lessons from the Taiwan and New Zealand health responses to the COVID-19 pandemic **0**(0). https://doi.org/10.1016/j.lanwpc.2020.100044. https://www.thelancet.com/journals/lanwpc/article/PIIS2666-6065(20)30044-4/abstract. Publisher: Elsevier

Learning COVID-19 Mitigation Strategies Using Reinforcement Learning

Nicholas Denis, Alexander El-Hajj, Blair Drummond, Yamina Abiza, and Krishna Chaitanya Gopaluni

1 Introduction

The novel coronavirus 2019 (COVID-19) pandemic prompted by severe acute respiratory syndrome (SARS-CoV-2) has introduced many challenges for public health officials and global economies. Since its onset, many modelling efforts related to COVID-19 have been published, ranging from estimating epidemiological parameters [1–3] to forecasting hospitalizations, deaths and infection events [4–6]. Public health officials generally use two modelling approaches: forecasting models and dynamic models [7, 8]. Forecasting models estimate the future number of a specified metric (e.g., deaths) in a population [9]. In contrast, dynamic models illustrate the virus's potential implications over a specified period, depending on how the virus functions and the likely outcomes of public health interventions [10]. Indeed, there is a demand for reliable models that can guide policy development to mitigate the virus, effectively detail the virus's perceived spread, and yield future predictions.

An effective mechanism for fighting an epidemic of such nature is through mitigation strategies. Many well-established epidemiological models are used to determine which non-pharmaceutical interventions (NPIs) will be most effective for examining the outcomes of preventive simulation strategies. These models generally fall within two categories: equation-based and agent-based [11]. The compartmental model, the most common type of equation-based model, contains a population usually assumed to be a homogenous mixture while consisting of

Nicholas Denis and Alexander El-Hajj authors contributed equally.

N. Denis (✉), A. El-Hajj, B. Drummond, Y. Abiza, K. C. Gopaluni
Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, ON, Canada
e-mail: nicholas.denis2@canada.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_12

compartments derived from health status [12]. In contrast, agent-based models (ABM) are computer simulations designed to include interactions between agents and their environment. Despite their differences, both approaches typically assume uniform dynamics applied across the modelled population. These approaches can capture metric means for large-populations; however, they are unfit to perform more granular analysis, for instance, conditioned on a distinct demographic. Moreover, human populations are not uniform in their dynamics, and individuals may behave quite differently depending on various attributes such as age and employment.

For this reason, we propose to utilize Reinforcement Learning (RL) with an agent-based modelling environment. The agent population and their state representation include pertinent demographic attributes. This approach allows agents to learn granular behaviours, called *policies*, conditioned on demographic data. The RL environment is encoded to minimize infection events within a population. By providing an array of actions related to everyday activities (e.g., going to work), the agents are able to learn behaviours that dramatically reduce certain metrics, including infections and hospitalizations, compared to different baselines. Rather than modelling specific scenarios such as "school closures", by providing the RL agents with the freedom to select and learn from their own actions, this approach allows for a more expressive range of scenarios to be considered, and thus a larger set of scenarios to be optimized over. The learned policies can then be analyzed, conditioned on state or agent attributes, to provide a more granular analysis on the types of mitigation strategies that could lead to a dramatic reduction in the spread of SARS-CoV-2, which can inform policy for real-life mitigation strategies and restrictions.

The main contributions of this work include:

- A novel ABM-RL-based modelling approach for learning and analyzing pandemic mitigation strategies using Ontario, Canada epidemiological, socioeconomic, health, and social data.
- We model two scenarios involving non-compliant agent subpopulations and show their effect on the spread of SARS-CoV-2 within the population.
- We provide an analysis and insights on the importance of mitigation strategy compliance within schools and the asymptomatic agent subpopulation, while highlighting the potential importance of recommendations and guidelines for individuals self-isolating in the same household as a SARS-CoV-2-positive person.
- We provide a Kibana dashboard for the public to analyze and explore the results of our experiments, including supplementary tables and figures (see Ref. [13] for link).

The rest of the paper is organized as follows: Sect. 2 provides a review of reinforcement learning and its mathematical framework. Section 3 details the methodology and implementation details. Section 4 describes the experimental results and Sect. 5 provides a discussion of the work.

2 Reinforcement Learning

RL is a field of machine learning wherein an agent interacts with an environment at successive time steps by observing its current state, selecting an action, subsequently resulting in the environment transitioning to a new successor state and returns a scalar reward to the agent [14]. The goal is to learn a state-dependant action selection strategy that maximizes the expected discounted sum of cumulative rewards over an infinite horizon. We introduce the mathematical formalism for representing and solving RL problems, then describe how it is applied to learning mitigation strategies for reducing COVID-19 infections within an agent-based simulation environment.

2.1 Markov Decision Processes

A Markov Decision Process (MDP) is represented as a tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \gamma \rangle$. $\mathcal{S}$ is the set of states of the environment, $|\mathcal{S}| < \infty$. $\mathcal{A}$ is the set of actions which the agent has access to, $|\mathcal{A}| < \infty$. $P := \{P^a\}_{a \in \mathcal{A}}$, is the dynamics of the environment. $\forall a \in \mathcal{A}$, P^a is the state transition kernel associated to action a. Under the above assumptions, $P^a \in [0, 1]^{\mathcal{S} \times \mathcal{S}}$ is a stochastic $|\mathcal{S}| \times |\mathcal{S}|$ transition matrix, where $P^a(i, j)$ is the probability of the environment transitioning to state j when action a is taken from state i. $R : \mathcal{S} \times \mathcal{A} \to \Delta(\mathbb{R})$, is the reward model of the environment that is a function mapping state-action pairs to distributions over $\mathbb{R}$. R is assumed to be almost surely bounded (i.e. $\exists M < \infty$, such that $|R(s, a)| \leq M, \forall (s, a) \in \mathcal{S} \times \mathcal{A}$, almost surely). $\gamma \in (0, 1)$ is called the discount-factor.

The MDP framework receives its name from the assumption that the environment dynamics are Markovian. That is,

$$\mathbb{P}(s_{t+1}|a_t, s_t, a_{t-1}, s_{t-1}, ..., a_1, s_1) = \mathbb{P}(s_{t+1}|a_t, s_t),$$

where s_k, a_k denotes the state $s \in \mathcal{S}$ and action $a \in \mathcal{A}$ taken at time k. This property is assumed to be true $\forall s \in \mathcal{S}, a \in \mathcal{A}$ and $\forall t \in \mathbb{N}$.

2.2 Policies and Value Functions

Agent action selection strategies (behaviours) are represented through a *policy*, which is a deterministic map from states to action: $\pi : \mathcal{S} \to \mathcal{A}$. Let R_n represent the random reward received at time point $n \in \mathbb{N}$. The *infinite horizon return* indexed at time $t \in \mathbb{N}$,

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k}, \tag{1}$$

is the random variable representing the random sum of discounted rewards over the infinite future. Under the above assumptions G_t is bounded almost surely $\forall s \in \mathcal{S}, a \in \mathcal{A}, t \in \mathbb{N}$. With policies and returns defined, one can introduce the definition of a value function. Let π be a fixed policy, then the *state value function*, $V^\pi : \mathcal{S} \to \mathbb{R}$,

$$V^\pi(s) := \mathbb{E}\big[G_t | s_t = s\big], \tag{2}$$

where the expectation is taken with respect to the environment dynamics, P, R, and with respect to actions taken while following policy π. The *state-action value function*, $Q^\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, is defined similarly:

$$Q^\pi(s, a) := \mathbb{E}\big[G_t | s_t = s, a_t = a\big]. \tag{3}$$

The objective in RL is to learn an optimal policy π^*, and optimal value functions, V^{π^*}, Q^{π^*}. A policy, π^*, is optimal if $\forall s \in \mathcal{S}$, $\pi^* \in \underset{\pi \in \mathcal{A}^{\mathcal{S}}}{arg\,max}\ V^\pi(s)$. It can be shown that under π^*, a recurrent structure exists for the value functions [15], relating the value of the current state to the expected immediate reward under π^* and the expected discounted value of the successor state:

$$\begin{aligned} V^{\pi^*}(s) &= Q^{\pi^*}(s, \pi^*(s)) = \max_{a \in \mathcal{A}}\ Q^{\pi^*}(s, a) \\ &= \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} P(s'|s, a)\Big[\mathbb{E}[R(s, a)] + \gamma Q^{\pi^*}(s', a)\Big] = \max_{a \in \mathcal{A}}\ \mathbb{E}\big[R(s, a) + \gamma V^{\pi^*}(s')\big]. \end{aligned}$$

Furthermore, the Bellman Optimality Operator [15] is a contraction mapping on the Banach space of value functions, which allows for the use of the value iteration algorithm [15] to solve for the optimal value functions, and hence optimal policies. However, this operator requires full knowledge of the environment dynamics, P, R, which is typically unavailable, as is the case for this study. For this reason we employ a model based approach and implement a variant of Q-learning [16], called Dyna-Q learning [17].

2.3 Model Based RL and Dyna-Q Learning

Value function based solution methods in RL typically utilize *temporal difference learning* algorithms [14]. Briefly, given an estimate of $Q(s, a)$, and a transition event $\langle s, a, r, s' \rangle$, one can use the immediate return as an unbiased estimate of the expected immediate reward. For a learning rate, $\alpha \in (0, 1)$, the Q-learning update is given as:

$$Q(s, a) \leftarrow (1 - \alpha) Q(s, a) + \alpha\big(r + \gamma \max_{b \in \mathcal{A}}\ Q(s', b)\big), \tag{4}$$

where $Q(s', b)$ is evaluated using the current Q-function. Tabular Q functions are implemented as $|\mathcal{S}| \times |\mathcal{A}|$ matrices, and the current policy can be *greedily* followed by simply taking the action that maximizes the row vector $Q(s, .)$. Under mild assumptions, as in this experiment, Q-learning is known to converge to Q^{π^*} [16]. To facilitate faster learning and convergence rates, we introduce a variant of the Dyna-Q learning algorithm. Dyna-Q [17] learning involves building a model of the environment, $\hat{p} \approx P, \hat{r} \approx R$, which we now describe. Let $\mathcal{D}$ be a database which holds transition events realized by agents acting in the environment of the form $\langle s, a, r, s' \rangle$. Model-based learning is employed in conjunction with Q-learning. Specifically, each transition event, $\langle s, a, r, s' \rangle$, is used to update a tabular model. A simple count-based model on empirical transitions is used to approximate $P(s_{t+1} = s' | a_t = a, s_t = s)$:

$$\hat{p}(s_{t+1} = s' | a_t = a, s_t = s) = \frac{\#\Big\{ \langle x, \alpha, \rho, x' \rangle \in \mathcal{D} | x = s, \alpha = a, x' = s' \Big\}}{\#\Big\{ \langle x, \alpha, \rho, x' \rangle \in \mathcal{D} | x = s, \alpha = a \Big\}}, \tag{5}$$

where #{ .} denotes set cardinality. Similarly, in order to approximate $R(s, a)$, let $R_{s,a} = \{r | \langle x, \alpha, r, x' \rangle \in \mathcal{D}, x = s, \alpha = a\}$ be the set of recorded rewards received from agents taking action $a \in \mathcal{A}$ from state $s \in \mathcal{S}$, then $\hat{r}(s, a) = \frac{\sum_{r \in R_{s,a}}}{\# R_{s,a}}$.

Every transition event of every agent is added to $\mathcal{D}$ and included in the model, and Q-learning updates are performed exclusively using simulated transition events sampled from the model. Specifically, $\forall s \in \mathcal{S}$, $\forall a \in \mathcal{A}$, counters of occurrences for $s_t = s$, and for $(a_t = a | s_t = s)$ (conditional action selection) are maintained. In this way, a simulated transition event $\langle \tilde{s}, \tilde{a}, \tilde{r}, \tilde{s}' \rangle$ is sampled from the model via:

$$\tilde{s} \sim \hat{p}(s), \tag{6}$$

$$\tilde{a} \sim \hat{p}(a | \tilde{s}) \tag{7}$$

$$\tilde{r} = \hat{r}(\tilde{s}, \tilde{a}) \tag{8}$$

$$\tilde{s}' \sim \hat{p}(s_{t+1} = \tilde{s}' | a_t = \tilde{a}, s_t = \tilde{s}) \tag{9}$$

Using the simulated transition event, $\langle \tilde{s}, \tilde{a}, \tilde{r}, \tilde{s}' \rangle$, a Q-learning update is performed, with learning rate $\alpha \in (0, 1)$, where:

$$Q(\tilde{s}, \tilde{a}) \leftarrow (1 - \alpha) Q(\tilde{s}, \tilde{a}) + \alpha (\tilde{r} + \gamma * \max_{b \in \mathcal{A}} Q(\tilde{s}', b)). \tag{10}$$

2.4 Framing Learning Mitigation Strategies as an MDP

We now describe how the MDP framework is applied to learning mitigation strategies. Agent states comprise different categorical values relating to facets, including age and current location (Table 1). Though a total of slightly under 100,000 states are feasible within our environment, each agent may only experience a small subset, as some state attributes are specific to them (e.g., number of comorbidities) and remain constant throughout a simulation. In this way, we can represent entire populations' policies using a single tabular Q-table. Agent actions are related to work, school, home, social and economic activities (Table 2). Rewards are sparse, in that only transition events resulting in an agent becoming infected, or infecting another agent, results in a small negative reward; all other transition events result in zero rewards. Agents "sleep" between 11 pm–5 am, inclusive, taking no actions, and each hour between 6 am–10 pm, inclusive, each agent selects an action (when possible), actions are resolved, states are updated, rewards received, and each agents transition event is appended to a database $\mathcal{D}$, the model is updated, and the model is sampled to perform a number of Dyna-Q learning updates.

Table 1 Agent states

Category	Values
Age	{0-5, 6-10, 11-15, ..., 95+}
Employment status	{True, False, Student}
Employment type	{NE, E}
Live in senior center	{True, False}
# Comorbdities	{0, 1, 2+}
Work/school hour	{True, False}
Node location	{Home, E Food, ..., School}
Symptomatic	{True, False}
Susceptible	{True, False}

Note: E = Essential; NE = Nonessential

Table 2 Action categories

Work	Social	Economic	School	Home
Go to work	Social Visit	NE Other	School	Home
Go to work + SD	Social Visit + SD	NE Other + SD	School + SD	Home + SD
Work from home		E Food		Home + ISO
Work from home + SD		E Food + SD		
		E Health		
		E Health + SD		
		E Other		
		E Other + SD		

Note: SD = Social distancing; ISO = Self-isolation; E = Essential; NE = Nonessential

3 Methods

In this section we describe the methodology pertaining to the simulation environment, epidemiological parameters, reinforcement learning and the wildcard and attrition experiments.

3.1 Simulation Environment

Below we describe how the simulation environment is generated. We address generating the agent population, employment assignment and building the agent community.

3.1.1 Agent Population

Given the user-specified population size, a population is generated with agents belonging to households or senior centers (i.e., senior centers or long-term care homes). Households and their occupants are sampled from empirical distributions obtained from the 2016 Canadian Census of Population, Statistics Canada surveys, and other institutional sources [13, 18–28], with agent attributes relating to age and employment status sampled accordingly. A portion of the population is assigned to senior center nodes using data from the Canadian Institute for Health Information [26].

3.1.2 Assigning Agent Employment

Employment node types were categorized as: essential health, essential food, essential other, nonessential school, or nonessential other. The classification of essentiality was referenced from an Ontario news releases [29], which outlined the businesses allowed to remain open during the pandemic. Health, food, school and the other categories represent high-level aggregates corresponding to sector level (i.e., two-digit) North American Industry Classification System (NAICS) codes [13, 30]. The employment statistics based on the sector-level NAICS codes were aggregated to categories, including business size by employment node [13, 19, 21]. These aggregates were used in the sampling process to assign employment nodes to an employed agent and assign the maximum number of employees to their given employment node. The hours of operation are generalized for each employment node [13]. Each employed agent assigned to a work node is designated a schedule of 5 days, with 8 h of scheduled employment consistent with the hours of operation for that employment node.

3.1.3 Agent Community

As described below (Sect. 3.3.1), agents can take actions involving visiting a particular node type or social visit. To represent preference or neighbourhood structure, each agent was assigned a subset of each node type [13]. Moreover, each agent is assigned a subset of the agent population with sample size as a function of age, representing the social circle in which they can visit [13, 22–24]. These node subsets act as constraints imposed on each agent regarding the ability to visit when taking an appropriate action. For instance, when a given agent selects an action for visiting an essential food node, that agent samples from its essential food node subset, representing its preference over all essential food nodes within the environment.

3.2 Epidemiological Parameters

3.2.1 Infection Events

Agents exist in an SEIR model-like setting [31], where all agents begin as susceptible to infection, and upon being exposed to an infected agent, one may become infected. The severity of infection is categorized as either: asymptomatic, mild, severe, or critical, each associated with their respective sampling distributions anatomized by age (Table 3). Depending on the severity level, an agent may transition through the following successive stages: latent period (non-infectious, asymptomatic), asymptomatic, symptomatic, require hospitalization, require Intensive Care Unit (ICU), and finally either death or recovery (Table 4), with durations sampled from Table 5.

Except for mortality rates (Table 4) conditioned on the presence of comorbidities, all epidemiological parameters were provided by the Public Health Agency of Canada (PHAC) [personal communications, Victoria Ng]. Aside from the latent period, a constant value of 3.931058% was used to represent the probability of infection given an infected and susceptible agent interacted. Infection events are sampled based on the presence of infected agents at a particular node. At each time

Table 3 Probability of infection severity

Age	Mild	Severe	Critical
20–43	0.752	0.039	0.009
44–53	0.686	0.085	0.029
54–63	0.654	0.114	0.032
64–73	0.571	0.167	0.061
74–83	0.342	0.234	0.224
84+	0.157	0.357	0.286

Note: Agents age 0–19 have probabilities similar to agents age 20–43.

Table 4 Probability of death[a] by number of comorbidites

Age	0	1	2	3+
20–43	0.005	0.005	0.007	0.010
44–53	0.025	0.027	0.035	0.050
54–63	0.069	0.076	0.097	0.139
64–73	0.131	0.144	0.148	0.263
74–83	0.200	0.220	0.279	0.399
84+	0.489	0.537	0.684	0.977

[a] $\mathbb{P}$(death|critical severity)
Note: Agents age 0–19 have zero probability of death.

Table 5 Sampling distributions for duration of infection stages

Incubation period	Latent period	Symptomatic	Hospital (severe/critial)	ICU (critical)
$\mathcal{N}$(5.1, 1.0)	$\mathcal{N}$(3.7, 1.0)	$\mathcal{N}$(5.0, 1.0)	$\mathcal{N}$(8.0, 1.0)/$\mathcal{N}$(6.0, 1.0)	$\mathcal{N}$(10.0, 1.0)
	C(2.0, 5.0)	C(0, 20.0)	C(1.0, 20.0)/C(1.0, 10.0)	C(1.0, 10.0)

Note: $\mathcal{N}$: Normal Distribution; mean, standard deviation, C: Clip (limit) the values; min, max.

step (i.e., 1 h) during the simulation, several contact events are sampled for each infected agent at a given node. This is done by sampling a maximum of five agents present at the node location (ten for students while at school) for interaction. Each infectious-susceptible interaction event leads to a new infection with the probability given above.

3.2.2 Contact Tracing

Testing for SARS-CoV-2 is implemented when an agent shows symptoms and takes an essential health action, and if an agent requires hospitalization due to their infection's severity, we assume the agent is tested upon arrival at the hospital.

Each agent maintains a list of every other agent they have been in contact with over the previous 7 day period. Agent A' can only be added to the contact tracing list of Agent A if A' belongs to the same social circle or is employed at the same employment node as A. This heuristic aims to represent that someone should only be able to share contact information of people they know. Upon testing positive, the agent, their cohabitants (household or senior center node), and all agents belonging to the positive agents' contact tracing list are sequestered to their household nodes for 14 days of self-isolation. If an agent requires hospitalization during the self-isolation period, they transition to a hospital node. Agents in self-isolation can still interact with other agents that share the same household node and can receive a social visit from agents but cannot take any actions themselves. Agents self-isolating receive a discrete state associated with self-isolation, and no actions are permitted. Likewise, agents admitted to the hospital, ICU or after death, transition to distinct states respective of each, and no actions are permitted.

3.3 Reinforcement Learning

3.3.1 Actions

A critical factor in expanding the space of representable behaviours is the inclusion of relevant actions endowed to the agents. Actions should represent abstractions of conventional everyday decisions and activities that a population may encounter (Table 2). Each action has a Socially Distanced (+SD) equivalent, whereby the infection rate is reduced by a factor of 0.6 [10].

Depending on the state, some actions may not be available. This requirement ensures common-sense rules to the environment. For example, an unemployed agent should not have access to actions for to going to work. During school, asymptomatic agents must take school related actions, while symptomatic agents can also get tested (take essential health action), or stay home. During work hours, an employee working at an essential node must take a work related action, while agents employed at nonessential nodes may stay home (work from home). When symptomatic, all employed agents may stay home, go to work, or go to an essential health node.

3.3.2 States and Rewards

States represent the information agents have about their environment and are depended on by agents for action selection. As seen in Table 1, the state is represented by the concatenation of various information-relevant categories, either binary or categorical. The environment provides very sparse rewards, with a reward of -1 whenever an agent becomes infected and for each other agent they infect. All other rewards are 0. The Q-values of all state-action pairs are initialized to 0.

3.3.3 Model Learning

Dyna-Q learning updates use a learning rate of $\alpha = 0.1$. During RL training epochs, after each "hour" of the simulation each agent's transition events are added to $\mathcal{D}$, the model is sampled N times, and each sample is used to perform a Q-learning update. We set N to the size of the agent population.

3.3.4 ϵ-greedy

ϵ-greedy is an action-selection strategy that balances exploration and exploitation [14]. With probability ϵ, legal actions are taken uniformly at random, and with probability $1 - \epsilon$, the policy is followed. This mechanism allows agents to *explore* all actions with some non-zero probabilty. ϵ-greedy is only implemented during training epochs. We initialize $\epsilon := 1$, then linearly anneal ϵ to 0.1 over the first 40

training epochs, and keep it fixed at 0.1 for the remaining training epochs. Testing epochs can be viewed as implementing ϵ greedy with $\epsilon := 0$.

4 Results

We first describe the results of the RL experiments, then explore the learned agent behaviours. Next, we compare the learned behaviours to various baselines. Finally, we provide results for two non-compliant subpopulation scenarios.

4.1 Agent Learning

This experiment's primary intention is to allow the agent population to learn optimal behaviours, π, that maximize $V^{\pi}(s), \forall s \in \mathcal{S}$. Based on the formulation of the MDP (Sect. 2.1), this is equivalent to minimizing infection events within the population. The subsequent experiments within this section aim to compare and evaluate the learned behaviours, compared to various baselines, in terms of their ability to minimize infections, hospitalizations, ICU occupancy and deaths.

An agent-based simulation environment was used to allow a population of agents to learn behaviours that minimize infection events (Sect. 3.1). Each simulation generated 50,000 agents, and in each simulation of 120 days, an initially infected subset of 500 agents was randomly sampled. 100 simulation epochs were run, alternating between training (using epsilon greedy, Sect. 3.3.4) and testing learned behaviours. Q-learning updates occur throughout the training simulations, whereas the policies are fixed and followed deterministically during the testing epochs.

Figure 1 plots the daily new infections (top) and cumulative new infections (bottom) over each epoch. The agents progressively learn behaviours that reduce the number of infection events occurring within the population. Moreover, as the agents' policies improve from epoch to epoch, as expected, the testing epochs demonstrate more dramatic advancements and show improved reductions in the number of infection events over the training epochs. Similar plots for other recorded metrics (e.g., deaths, ICU) are available to view on our Kibana dashboard [13]. The policy resulting in the lowest number of infections was saved and used for further experiments.

4.2 Learned Agent Behaviour Analysis

The learned policies from Sect. 4.1 can then be analyzed and treated as a database that maps states to its actions' values. In viewing the Q-table as a database, the policy associated with a specific state (i.e., 35 years old, employed, one comorbidity,

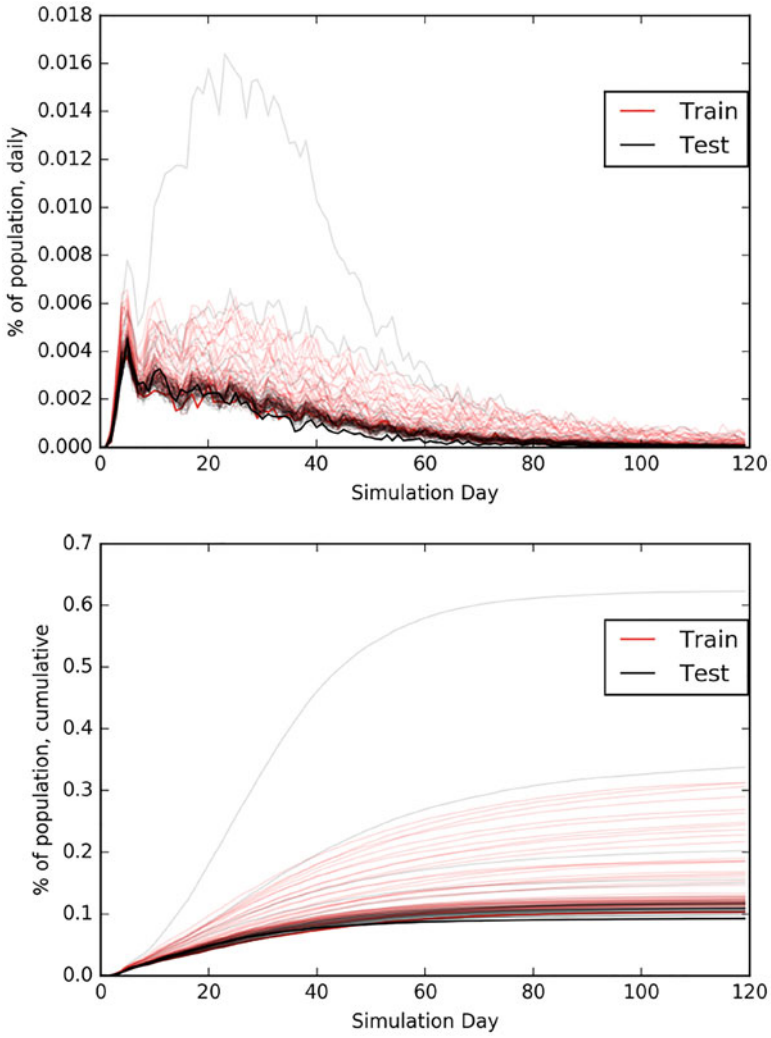


Fig. 1 Agents learn to reduce infections over time. Daily (top) and cumulative (bottom) infections as a percentage of the population size over each RL training and testing epoch. Each epoch is plotted slightly darker in colour, leaving the darkest plot for the best performing epoch

outside of work hours, currently at home, no symptoms) can be determined and utilized to produce summary statistics marginalized over state attributes. The states included in the Q-table provide a multitude of dynamics to explore. For instance, one can condition on subpopulations based on various attributes (i.e., by age or the number of comorbidities).

Figure 2 provides a visualization of behaviours related to essential employed workers during work hours. When conditioned on work hours of asymptomatic

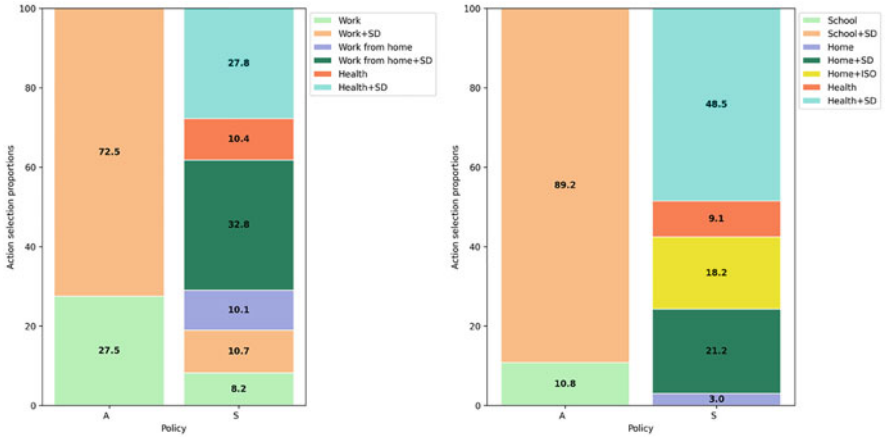


Fig. 2 Analysis of learned agent behavarious for work and school related states. Learned policies for essential workers (left) and for students (right) differ depending on presence of symptoms

agents, 72.5% of agent states resulted in going to work and practicing Social Distancing (+SD), compared to 27.5% for going to work without practicing Social Distancing (-SD). Note that the environment logic prevents essential workers from getting tested (visiting essential health nodes) or working from home when asymptomatic. However, when symptomatic, those actions become available to the essential worker agents. When symptomatic, agents chose to forgo work and get tested for 38.2% of the agent states (27.8% +SD, 10.4% -SD), work from home for 42.9% of the agent states (32.8% +SD, 10.1% -SD) and go to work for 18.9% of the agent states (10.7% +SD, 8.2% -SD). Symptomatic agents employed at nonessential work nodes chose to forgo work and get tested in 47.3% of the agent states (36.8% +SD, 10.5% -SD), work from home for 47.3% of the agent states (36.8% +SD, 10.5% -SD) and go to work +SD for remaining 5.4% of the agent states (0% -SD).

The proportion of action selections for students during school hours, as shown in Fig. 2 reveals a similar trend in preference. When asymptomatic, students at school are 8.2x more likely to use +SD than -SD school actions. Notably, during school hours, 0% of symptomatic students went to school, as these agents prefer to get tested (57.6%; 48.5% +SD, 9.1%) -SD or stay home (42.4%; 18.2% +ISO, 21.2% +SD, 3% -SD). Interestingly, outside of school hours, asymptomatic student agents spent 10.5% (7.9% +SD, 2.6% -SD) of their free time making social visits, whereas symptomatic student agents completely stopped all social visits.

We analyze the learned agent behaviours outside of work and school hours (Fig. 3). We break down the analysis by aggregating agent states by employment status outside of work/school hours and further decompose the analysis by the presence or absence of symptoms. For all categories, the +SD action variant is significantly preferred to the -SD variant. Asymptomatic agents learn to spend most of their time at home (25.6%, 43.4%, 26.2% for the employed, student and unemployed agents, respectively), while symptomatic, the agents strongly prefer to

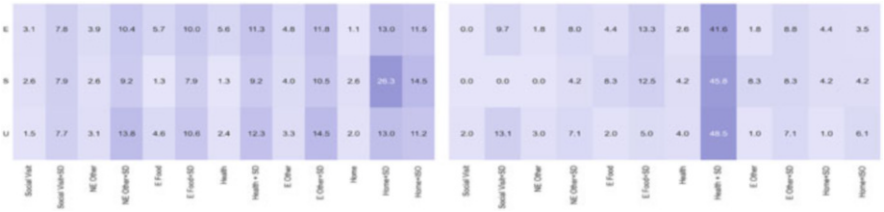


Fig. 3 Heatmap representing the distribution over actions selected by policy for states outside of school or work for asymptomatic (left) and symptomatic agents (right). E: employed agent, S: student agent, U: unemployed agent

get tested (44.2%, 50.0%, 52.5% for the employed, student and unemployed agents, respectively). Clearly, getting tested and partaking in the self-isolation and contact tracing mechanisms is invaluable for minimizing infection events, as agents learn to do this without explicitly being provided with any positive reward.

4.3 Learned Agent Behaviour Performance

The performance of the learned behaviours in terms of relevant metrics (infections, hospitalization, ICU, death) was assessed by comparing three baselines: (1) Baseline 1 (B1) involves agents taking actions uniformly at random, constrained to -SD actions; (2) Baseline 2 (B2) involves agents taking actions uniformly at random, including +SD actions; (3) Baseline 3 (B3) extends B2 by including testing, contact tracing and self-isolation; and (4) Learned Behaviours (LB) denotes actions selected using the policy learned using RL (Sect. 4.1) and employs testing, contact tracing and self-isolation. Each experiment comprises 50 epochs. All results, including mean fold reductions and metric plots, can be found in the Kibana dashboard [13].

Figure 4 demonstrates an expected result, in that B1 results in the most infection events, B2 improves upon B1, B3 improves upon B2, and LB outperforms all the baselines with roughly 97%, 89%, and 30% of the population becoming infected in B1, B2, and B3, respectively, whereas, in LB, the proportion that gets infected is 4.4%. It is observed that the RL behaviours (LB) lead to a 22.3, 20.4, and 6.79 mean fold reduction in total cumulative infection events over B1, B2, and B3, respectively. Introducing testing, contact tracing and self-isolation in B3 contributes to a noteworthy decrease in the mean fold reduction compared to B1 and B2. Improving upon B3, LB leads to a further reduction in the spread of COVID-19 due to an optimized state-dependent action selection strategy. Of note, in comparison to B3, the agents in LB learn to be responsible by taking certain precautions such as reducing ineffective actions (i.e., social visits), while increasing social distancing, getting tested, and staying at home. When comparing other metrics, it was found

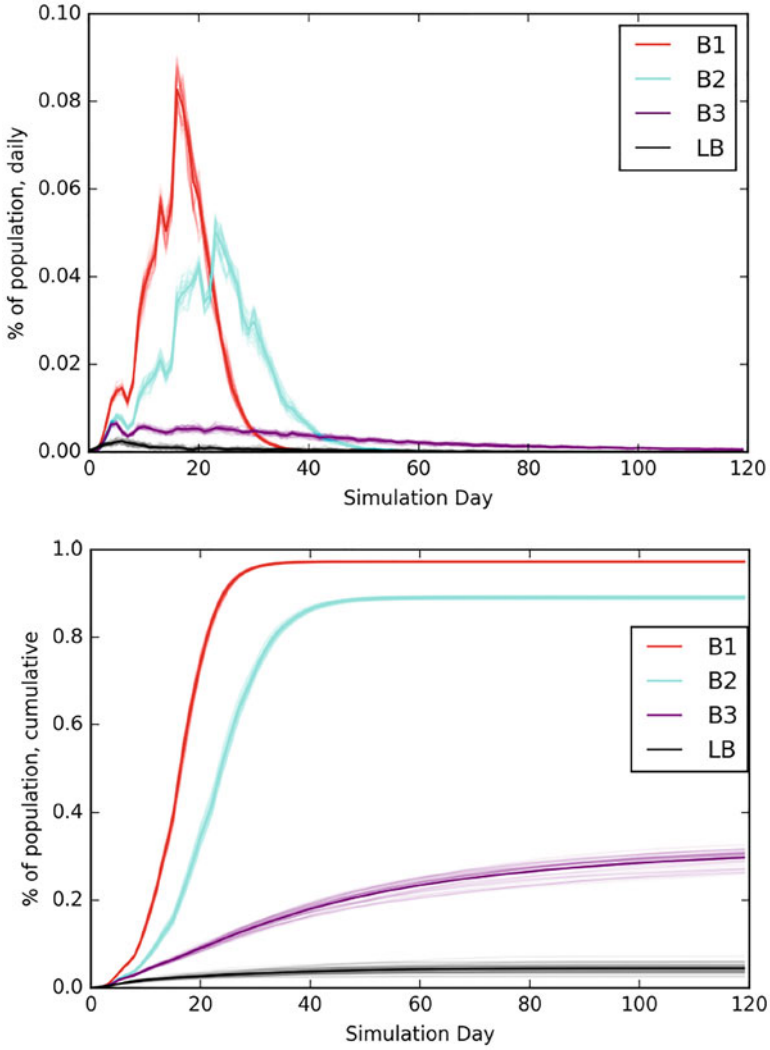


Fig. 4 Comparison of daily (top) and cumulative (bottom) infection events between B1, B2, B3, and LB. Each experiment is run with 50 repeats and plotted in light colour, with the mean plotted in dark color

that LB resulted in a 20.6, 18.1, and 8.2 mean fold reduction in total deaths over B1, B2, and B3, respectively, an 11.4, 10.3, and 3.9 mean fold reduction in total hospitalizations over B1, B2, and B3, respectively and finally a 16.7, 14.9, and 5.9 mean fold reduction in the number of ICU admittances over B1, B2, and B3, respectively.

4.4 Wildcard and Attrition Experiments

We address two modelling scenarios wherein agents may prefer a non-social distancing default behaviour to the learned RL mitigation behaviour. The first scenario involves Wildcard (WC) agents, and the second models behaviour attrition (ATT). While generating the population, each agent samples a value, $\delta \in [0, 1]$, representing the probability that the agent follows the learned RL policy or with probability $1 - \delta$, in which the agent follows a predefined default behaviour (See Supplementary Kibana Tables [13] for details). This scenario represents a population with a distribution over agents related to how often they are willing to comply with NPI strategies. Attrition is represented similarly; each agent in the population is assigned a value δ, initialized to 1. We represent attrition by sampling each day with probability 0.5, whether the agent's δ value exponentially decays by a factor of 0.95.

Figure 5 compares the daily new infections (top) and cumulative new infections (bottom) for LB, LB + WC, and LB + ATT. The results confirm that attrition leads to a larger number of infections throughout the simulation than wildcard agents, with a 3.8 and 2.4 fold increase in the number of infection events over LB for LB + ATT and LB + WC, respectively. Figure 5 (top) shows that LB + WC eventually results in daily infection events decreasing to zero, concomitant with a plateau in the cumulative number of infections (Fig. 5, bottom). However, in the presence of attrition, the population sees a monotonic increase in cumulative infection events as the daily number of infections fails to converge down to zero, plateauing roughly half-way through the simulation. As attrition represents the progression of a population from uniformly following behaviours that minimize infection events to uniformly following default behaviours (e.g., a relaxation of population restrictions and return to normal life, pre-COVID-19), it is vital to find an attrition/relaxation strategy that does not lead to large increases (e.g., waves) in infection events. Though our results are inconclusive, we conjecture that running the LB + ATT experiments for longer (e.g., 180 days) could likely lead to a second wave of infection events.

Of interest, Fig. 5 (top) demonstrates periodic trends in the infections throughout the simulation runs. Vertical bars are plotted for Saturdays and Sundays within the simulations to help visualize the periodicity. In the presence of wildcard agents (LB + WC, LB + ATT), one can see that the weekends correspond to valleys (local minima), with the mid-week typically associated with peaks (local maxima). Interestingly, the opposite trend is seen for LB, with mid-week days corresponding to valleys (local minima) and weekends with peaks (local maxima). It is worth noting that with the exception of schools, all employment nodes are open 7 days a week and employed agents sample five employment days uniformly over the week. Hence, it is expected that no periodicity related to such employment nodes can be observed. The only facet of the simulation environment that could encode a weekday/weekend periodicity must come from the schools. This result demonstrates that in the absence of wildcard agents, when the agents present in schools are uniformly following the

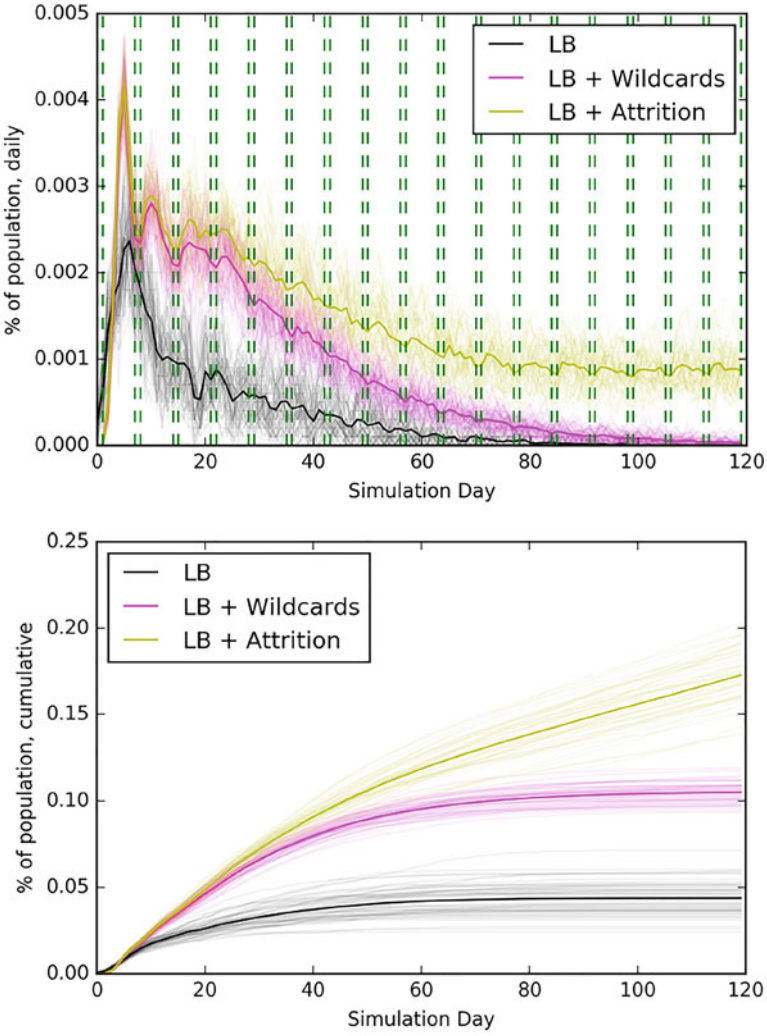


Fig. 5 Comparison of daily (top) and cumulative (bottom) infection events between LB, LB + WC and LB + attrition. Saturdays and Sundays within the 50 simulation runs are marked in green

learned RL behaviours, this leads to decreases in infection events during the week, while in the presence of wildcard agents, schools can become hotbeds for infection events. We believe this result provides strong supporting evidence on the importance of uniform compliance of recommended NPI strategies within schools.

4.5 Infection Event Analysis

Infection events are recorded for all simulations, allowing for a statistical analysis conditioned on state and attributes of infecting and infected agents, as well as the node type. Table 6 presents several of these infection events. For baselines B1, B2, and B3, the proportion of all infection events that occur when agents are taking social visit actions are 14.3%, 9.9% and 10.0%, respectively, whereas, for LB, only 0.2% of all infection events result from social visit actions. The LB + WC and LB + ATT saw these numbers increase to baseline values at 8.6% and 13.5%, respectively. Similarly, 9.6%, 7.5% and 2.5% of all infection events occur within schools for B1, B2, and B3, while only 1.1% for LB. The LB + WC and LB + ATT saw these numbers increase to baseline values of 6.6% and 8.6%, respectively. This, in addition to results from Section 4.2, demonstrate that RL allows the agents to learn to minimize riskier actions, leading to a reduction in overall infection events and a reduction in infection events related to particular states (i.e., during school hours).

Oppositely, the proportion of all infection events occurring at home are 15.4%, 14.2%, 44.5%, and 45.4% for B1, B2, B3 and LB, respectively. Again, the LB + WC and LB + ATT saw these numbers return to baseline values at 10.6% and 15.3%, respectively. The high proportion of infection events occurring in the household of LB agents is explained by the fact that the LB agents spend significantly less time taking actions that lead them outside their households due to the agent's improved mitigation strategy. In addition, B3 and LB both include self-isolation mechanisms leading agents to be sequestered to their place of residence. Further analysis revealed that for LB, 60.57% of all infection events occurred when two members of the same household were self-isolating together after one member had tested positive for COVID-19. We believe this result strongly suggests that as a population improves its NPI strategies, infection events become more concentrated in self-isolating agents' households. Thus, guidelines and strategies on effectively self-isolating with an infected cohabitant become increasingly crucial for further reducing infections.

Finally, we analyzed the proportion of infection events based on whether the infecting agent was symptomatic or not. For B1, B2 and B3, it was observed that 27.4%, 24.0%, and 43.4% of infections involved a symptomatic infecting agent, while LB and LB + WC resulted in 47.0% and 17.5%, respectively. In raw numbers, LB resulted in a 30.5, 29.2 and 7.3 fold reduction in infections by asymptomatic

Table 6 Infection events (%)

Infection event	B1	B2	B3	LB	LB +WC	LB + ATT
Being social	14.3	9.9	10.0	0.2	8.6	13.5
At school	9.6	7.4	2.5	1.1	6.6	8.6
At home	15.4	14.2	44.5	45.4	10.6	15.3
Symptomatic	27.4	24.0	43.4	47.0	17.5	22.9

Note: ATT = Attrition

infecters over B1, B2, and B3, respectively, as well as a 3.7 and 5.5 fold reduction in such infections for LB + WC and LB + ATT, respectively. These results stress the importance of maintaining NPI strategies such as testing, contact tracing, and self-isolation when asymptomatic, as a large proportion of all infection events may occur from asymptomatic individuals.

5 Discussion

In this paper, we demonstrate the potential of RL combined with agent-based modelling in the context of learning optimal behaviours for controlling the COVID-19 pandemic. The general nature of the RL simulation environment allows it to be easily customizable for different nations, provinces or communities, and different pandemics. Using mainly Ontario epidemiological, health, social, and socioeconomic data, we showed that the RL model could be trained to dramatically reduce the number of infections, hospitalizations, ICU admittance and deaths over time.

This work introduces RL as a novel and expressive modelling tool with the capabilities to allow researchers and policymakers to consider policy recommendations related to novel mitigation strategies with increased granularity over standard scenario-based approaches [4, 5]. By conditioning the analysis on subsets of the agent population and states, policies and guidelines can focus and target at-risk subpopulations. For instance, our results suggest that as a population increasingly practices social distancing strategies, the importance of recommendations and education about safely self-isolating at home with a SARS-CoV-2 positive cohabitant may become increasingly important. Additionally, Sects. 4.4–4.5 suggest the importance of uniform compliance within schools to reduce infections within the population.

Our RL simulation environment allows agents to learn behaviours that dramatically reduce infection, hospitalizations, ICU admittance and deaths compared to various baselines. Section 4.2 reveals that this can be attributed to an increase in the use of SD actions and an increase in agents staying at home, while also unveiling the importance of these decisions when agents do not show signs of symptoms. When symptomatic, the agents learned to decrease risky behaviour such as social visits and learned to opt to get tested and subsequently stay at home under self-isolation. However, when wildcard agents are present, agents following these mitigation behaviours have a reduced impact on the population. Further studies on this interaction between compliant versus non-compliant subpopulations and varying the attrition rates and mechanisms may be informative for policy surrounding re-opening strategies and how to inform how and when to return to *normal* life. Another exciting avenue for future work includes investigating different self-isolating strategies and their impact on further mitigating infection events within the population.

Acknowledgments The authors would like to Acknowledge Victoria Ng, Aamir Fazil and Nick Ogden from Public Health Agency of Canada for their support and feedback on experiments conducted, Deirdre Hennessey for support and Saeid Molladavoudi, Shirin Roshanafshar, Manolo Malaver-Vojvodic and Bryan Paget for reviewing the manuscript.

References

1. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... & Xing, X. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*
2. Banka, P., & Comiskey, C. M. (2020). The incubation period of COVID-19: A scoping review and meta-analysis to aid modelling and planning. *medRxiv*
3. Mahsin, M., Lee, S., Vickers, D., Guigue, A., Williamson, T., Quan, H., ... & Ravani, P. (2020). The Coronavirus 2019 pandemic in Canada: the impact of public health interventions on the course of the outbreak in Alberta and other provinces. *medRxiv*
4. Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., ... & Monod, M. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820), 257–261
5. COVID, I. Modeling COVID-19 scenarios for the United States. *Nature medicine*
6. Jung, S. M., Akhmetzhanov, A. R., Hayashi, K., Linton, N. M., Yang, Y., Yuan, B., ... & Nishiura, H. (2020). Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *Journal of clinical medicine*, 9(2), 523
7. Ogden, N. H., Fazil, A., Arino, J., Berthiaume, P., Fisman, D. N., Greer, A. L., ... & Waddell, L. A. (2020). Predictive modelling of COVID-19 in Canada. *CCDR*, 46(6)
8. Public Health Agency of Canada. (2020). Mathematical modelling and COVID-19. https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/epidemiological-economic-research-data/mathematical-modelling.html
9. Zhang, W., Zhao, W. W., Wu, D., & Yang, Y. (2020). Predicting COVID-19 trends in Canada: a tale of four models. *Cognitive Computation and Systems*, 2(3), 112–118.
10. Tuite, A. R., Fisman, D. N., & Greer, A. L. (2020). Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. *CMAJ*, 192(19), E497–E505
11. Adam, D. (2020). Special report: The simulations driving the world's response to COVID-19. *Nature*, 580(7803), 316
12. Hunter, E., Mac Namee, B., & Kelleher, J. D. (2018, December). A Comparison of Agent-Based Models and Equation Based Models for Infectious Disease Epidemiology. *AICS*, (pp. 33–44)
13. Statistics Canada. (2020). COVID-19 Reinforcement Learning Dashboard. https://kibana.covid.cloud.statcan.ca/s/covid19rl/goto/a1261b3dbc4f17c6b52b6d65bbd6dac4
14. Sutton, R. S., & Barto, A. G. (2011). Reinforcement learning: An introduction. Note: a recurrent structure exists for the value functions
15. Martin L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, 1994.
16. P. Dayan C.J.C.H. Watkins. Q-learning. Machine Learning, 8:279–292, 1992.
17. Sutton, R. S. Dyna, an Integrated Architecture for Learning, Planning and Reacting. AAAI Spring Symposium, 151–155, 1991.
18. Statistics Canada. (2019a). Table 14-10-0327-01 Labour force characteristics by sex and detailed age group, annual. https://doi.org/10.25318/1410032701-eng
19. Statistics Canada. (2019b). Table 33-10-0222-01 Canadian Business Counts, with employees, December 2019. https://doi.org/10.25318/3310022201-eng

20. Statistics Canada. (2018). Table 13-10-0096-01 Health characteristics, annual estimates. https://doi.org/10.25318/1310009601-eng
21. Statistics Canada. (2017a). Census Profile, 2016 Census: Ontario (Province). Catalogue no. 98-316-X2016001. Ottawa. http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E
22. Statistics Canada. (2014a). Table 45-10-0020-01 Number of close relatives, by sex (x 1,000). https://doi.org/10.25318/4510002001-eng
23. Statistics Canada. (2014b). Table 45-10-0021-01 Number of close friends, by sex (x 1,000). https://doi.org/10.25318/4510002101-eng
24. Statistics Canada. (2014c). Table 45-10-0022-01 Number of acquaintances, by sex (x 1,000). https://doi.org/10.25318/4510002201-eng
25. Milan, Anne, Nora Bohnert, Sandrine Levasseur and Franois Pag. (2012). Living arrangements of seniors. Census in Brief. September. Statistics Canada Catalogue no. 98–312–2011003.
26. Canadian Institute for Health Information. (2019). Continuing Care Reporting System (CCRS) 2018–2019 Profile of Residents in Residential and Hospital-Based Continuing Care. Ottawa: CIHI; 2019
27. Cancer Care Ontario. Ontario Cancer Statistics 2018 Report. Cancer Care Ontario 2018. https://www.cancercareontario.ca/en/statistical-reports/ontario-cancer-statistics-2018-report
28. Public Health Agency of Canada (2016). Chronic Disease and Injury Indicator Framework. Edition 2016, using data from the Canadian Community Health Survey 2014
29. Order Under Subsection 7.0.2 (4)—Closure of Places of Non-Essential Businesses. (2020). O. Reg. 119/20.
30. Statistics Canada. (2017b). North American Industry Classification System (NAICS) Canada 2017 Version 3.0. https://www.statcan.gc.ca/eng/subjects/standard/naics/2017/v3/index
31. Shil, P. (2016). Mathematical Modeling of Viral Epidemics: A Review. *Biomedical Research Journal*, 3(2), 195–215

Joint Modeling of Hospitalization and Mortality of Ontario Covid-19 Cases

Dexen D. Z. Xi, C. B. Dean, and E. M. Renouf

1 Introduction

In epidemiology, various empirical methods have been developed to quantify the outbreak of infectious diseases. One approach models public health data as time series processes (Zeger et al. 2006), which is typically suitable when an outcome is observed for a long period of time. Time series models generally assume that the observation today is linearly related to the observations lagged several days prior, with additive error terms independently and identically distributed (i.e. i.i.d.) from a normal distribution with a mean of zero and an unknown variance. The average, the trend, and the seasonality of the outcome process can then be specified in the model.

Time series models have previously been used in public health studies of infectious diseases. Examples of the diseases and study regions where time series models have been applied are Campylobacter and measles in Montreal, Canada (Allard 1998), diarrhoea in Peru (Checkley et al. 2000), and Covid-19 in Italy [1]. Times series models are prominently studied in field outside of public health, such as econometrics, where a technique, termed cointegration analysis, can further assess whether if there is correlation in the long run between two processes [2]. For example, cointegration analysis was applied to various processes of stock prices to examine if the SARS outbreak in 2003 had an impact on them (e.g. Chen et al. 2018). As hospitalization data regarding Covid-19 are collected and become

D. D. Z. Xi (✉)
National Research Council Canada, Ottawa, ON, Canada
e-mail: dazhong.xi@nrc-cnrc.gc.ca

C. B. Dean · E. M. Renouf
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada
e-mail: cdean@uwaterloo.ca; elizabeth.renouf@uwaterloo.ca

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_13

available, several authors have indeed studied the relationship between the daily number of cases and stock prices (e.g. Zeren and Hizarci 2020; [3]).

Another potential approach for studying the relationship between two outcomes is through *so-called* joint-outcome modeling (Dunson 2000; Henderson and Shimakura (2003)). This approach links the outcomes through a latent variable, a shared error term that is incorporated in the models for each of the outcomes, which then induces an underlying correlation between the outcomes. The method has been utilized in linking, for example, various outcomes that are count data (Feng and Dean 2012; Juarez-Colunga et al. 2017), survival data [4], and presence/absence data (Lundy and Dean 2018), where the latent variable is shared among the outcomes. Such a methodology has not been considered in linking time series data, and it may provide a novel perspective for understanding the long-run relationship between two time series processes.

In this study, we analyze the daily number of new hospitalizations and the daily number of new deaths from Covid-19 in Ontario as autoregressive processes. In infectious disease studies, these two processes are key indicators in an outbreak (e.g. Trivedi et al. 2012). We chose to model hospitalized cases instead of the number of new infections because testing was initially limited to the sickest patients or those recently returned from travel, so that case counts did not reflect the true progression of transmission. Section 2 outlines two frameworks for assessing the relationship between hospitalizations and deaths, where a cointegration analysis and a joint modeling framework are used to understand and model the long-run relationship between these two outcomes. Section 3 presents results of the analysis on the Ontario data using each framework, identifying the unique perspective that each framework provides. Section 4 closes with a discussion of the utility of each of the frameworks and potential ways that the models can be extended.

2 Models and Methods

2.1 Cointegration Analysis

We assume that the time series process, y_t, $t = p + 1, \ldots, n$ follows an autoregressive model with lag p, termed an AR(p) model, defined as

$$y_t = \mu + \theta_1 (y_{t-1} - \mu) + \cdots + \theta_p \left(y_{t-p} - \mu\right) + \varepsilon_t,$$

where μ is the intercept; $\theta_s = \sigma_s/\sigma_0$, $s = 1, \ldots, p$ such that $\sigma_s = \text{COV}(y_t, y_{t+s})$, the covariance between y_t and y_{t+s}, is the autocorrelation coefficient associated with lag s; ε_t is the random error assumed to be distributed as i.i.d. $N(0, \sigma^2)$, $t = p + 1, \ldots, n$. Inference on the model is straight forward when the time series process is stationary, that is, if the intercept and the autocorrelation are both fixed and do not depend on t. This is equivalent to stating that the process has a *unit root*, or, $|\theta_s| < 1$.

A non-stationary process can often become stationary by differencing y_t with respect to time d times, and such process is denoted as $y_t \sim I(d)$. For example, if $y_t \sim I(1)$, then $y_t^* = \Delta y_t = y_t - y_{t-1}$ is stationary; whereas if $y_t \sim I(2)$, then $y_t^{**} = \Delta y_t^* = y_t^* - y_{t-1}^*$ is stationary. The value d is often referred as the *order of integration*.

Two outcome processes y_{kt}, $k = 1, 2$ are cointegrated with other, if there exists an integer constant b such that $y_{1t} \sim I(d)$, $y_{2t} \sim I(d)$ and $z_t \sim I(d - b)$, where z_t is a linear combination of y_{1t} and y_{2t}. In other words, given two processes that are stationary after differencing d times, if their residuals z_t are stationary by differencing less than d times, the two processes are related in a unique long-run relationship and they are termed cointegrated. Heuristically, variables that are cointegrated do not deviate from each other in the longer term.

We first need to determine the value of d that supports stationarity in the two outcome processes. Several tests can be used to determine if a process, y_t is stationary. For example, an AR(1) process with intercept zero can be written as

$$y_t = \theta y_{t-1} + \varepsilon_t.$$

For testing that y_t is non-stationary, the Dickey-Fuller test [5] tests that $\pi = 0$, $\pi = \theta - 1$, in the rearranged model framework

$$y_t - y_{t-1} = \theta y_{t-1} - y_{t-1} + \varepsilon_t$$

$$\Delta y_t = (\theta - 1)\, y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \pi y_{t-1} + \varepsilon_t.$$

When the null hypothesis is true, the stochastic error term accumulates over time and hence the process is unstable. Hence the null and the alternative hypotheses can be written as

$$H_0 : \pi = 0$$

$$H_1 : \pi < 0.$$

If there is significant evidence to reject H_0, we conclude that y_t is stationary. More generally, assuming an AR(p) process analogously yields the augmented Dickey-Fuller (ADF) test, utilizing the same null and alternative hypotheses, here using the modeling framework

$$\Delta y_t = \pi y_{t-1} + \sum\nolimits_{s=1}^{p-1} \gamma_s \Delta y_{t-s} + \varepsilon_t.$$

The parameters can be estimated using least squares. The test statistic π follows a Dickey-Fuller distribution whose p-value is computed through Monte Carlo methods (i.e. [6, 7]; Chang et al. 2017). Although the Dickey-Fuller test is a standard in the literature, we note that alternative tests, such as the Phillips-Perron (PP) test, the Elliott-Rothenberg-Stock (ERS) test, and the Schmidt-Phillips (SP) test may also be used; see Pfaff [2] for a description of these tests.

2.2 *Joint Modelling*

Let $y_{kt}, k = 1, 2, t = p + 1, \ldots n$ be two time series processes, each process with lag p_k and $p = \max(p_1, p_2)$. This model assumes that the processes quantify outcomes measured at the same values of t. The model is defined by

$$y_{kt} = \mu_k + \theta_{k1}\left(y_{k,t-1} - \mu_k\right) + \cdots + \theta_{kp_k}\left(y_{k,t-p_k} - \mu_k\right) + b_{kt} + \varepsilon_{kt},$$

where, associated with outcome k: μ_k is the intercept; $\theta_{ks} = \sigma_{ks}/\sigma_{k0}$, $s = 1, \ldots, p_k$ such that $\sigma_{ks} = \mathrm{COV}(y_{kt}, y_{k,t+s})$, the covariance between y_{kt} and $y_{k,t+s}$, is the autocorrelation coefficient with lag s; ε_{kt} are random errors distributed as i.i.d. $N\left(0, \sigma_k^2\right)$; and $\boldsymbol{b}_t = (b_{1t}, b_{2t})^T$ is a 2×1 vector of random effects, independent from ε_{kt}, used to model the shared variability between the outcomes. The distribution of $\boldsymbol{b}_t$, $t = p + 1, \ldots, n$, is i.i.d. $Q(\boldsymbol{b}|\boldsymbol{D}) = N_2(\boldsymbol{0}, \boldsymbol{D})$, with a 2×1 mean vector $\boldsymbol{0}$ and a 2×2 symmetric and positive definite variance-covariance $\boldsymbol{D}$. Each outcome is of an order of integration d_k. In other words, the outcomes will need to be differenced d_k times before model development in order to achieve stationarity in the transformed outcomes.

It is convenient to represent the framework in matrix notation. The response $\boldsymbol{Y}_k$, the design matrix $\boldsymbol{X}_k$, and the associated vectors of parameters and random effects are specified as follows

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\theta}_k + \mu_k\left(1 - \sum_{s=1}^{p_k}\theta_{k,s}\right) + \boldsymbol{B}_k + \boldsymbol{\epsilon}_k,$$

where

$$\boldsymbol{Y}_k = \begin{bmatrix} y_{k,p+1} \\ \vdots \\ y_{k,n} \end{bmatrix}, \boldsymbol{X}_k = \begin{bmatrix} y_{k,p} & \cdots & y_{k,p+1-p_k} \\ \vdots & \ddots & \vdots \\ y_{k,n-1} & \cdots & y_{k,n-p_k} \end{bmatrix}, \boldsymbol{\theta}_k = \begin{bmatrix} \theta_{k,1} \\ \vdots \\ \theta_{k,p_k} \end{bmatrix}$$

$$\boldsymbol{B}_k = \begin{bmatrix} b_{k,p+1} \\ \vdots \\ b_{k,n} \end{bmatrix}, \boldsymbol{\epsilon}_k = \begin{bmatrix} \varepsilon_{k,p+1} \\ \vdots \\ \varepsilon_{k,n} \end{bmatrix}.$$

The joint posterior distribution is expressed as

$$p\left(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{D} | \boldsymbol{y}\right) \propto p\left(\boldsymbol{y} | \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}\right) Q\left(\boldsymbol{b} | \boldsymbol{D}\right) p\left(\boldsymbol{\mu}\right) p\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{D}\right) p\left(\boldsymbol{\sigma}\right),$$

where $\boldsymbol{y} = (\boldsymbol{y}_{p+1}, \ldots, \boldsymbol{y}_n)$, $\boldsymbol{y}_t = (y_{1t}, y_{2t})$, $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, $\boldsymbol{b} = (\boldsymbol{b}_{p+1}, \ldots, \boldsymbol{b}_n)$, and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$. The first term on the right-hand side is the conditional likelihood

$$p\left(\boldsymbol{y} | \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}\right) \propto \prod_{t=p+1}^{n} f\left(\boldsymbol{y}_t | \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}_t\right),$$

where $f(\boldsymbol{y}_t | \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}_t)$ is the joint density function of $\boldsymbol{y}_t$. We impose different constraints on the term b_{kt} and ε_{kt} to create four joint outcome models and these are shown in Table 1. For instance, Model B defines the vectors of error terms as $(b_{1t} + \varepsilon_{1t}, b_{2t} + \varepsilon_{2t})^T = (\gamma b_t, b_t + \varepsilon_{2t})^T$. This model assumes that all the variability in y_{1t} is explained by the term b_t which follows i.i.d.$N\left(0, \sigma_b^2\right)$ and is scaled by the factor loading parameter γ; as well, that all the variability in y_{2t} is explained by the sum of b_t and the additive error term ε_{2t}, where ε_{2t} is $N\left(0, \sigma_{\varepsilon 2}^2\right)$. Since the outcomes are independent, given the shared random effect, we have the joint density expressed as

$$\prod_{t=p+1}^{n} f\left(\boldsymbol{y}_t | \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}_t\right) = \prod_{t=p+1}^{n} \prod_{k=1}^{2} f_k\left(y_{kt} | \mu_k, \boldsymbol{\theta}_k, b_{kt}\right),$$

where $f_k(y_{kt} | \mu_k, \boldsymbol{\theta}_k, b_{kt})$ is the marginal density function of y_{kt}. Finally, the product of the prior distributions is given by

$$p\left(\boldsymbol{\mu}\right) p\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{D}\right) p\left(\boldsymbol{\sigma}\right) = \prod_{k=1}^{2} \left[p\left(\mu_k\right) p\left(\theta_{k1}\right) \ldots p\left(\theta_{kp_k}\right) p\left(\sigma_k\right)\right] p\left(\gamma\right) p\left(\sigma_b\right).$$

Choices of the distributions of the priors will be discussed more fully in the next section.

Table 1 Parameterization of the joint models

Model form	$\begin{bmatrix} b_{1t}+\varepsilon_{1t} \\ b_{2t}+\varepsilon_{2t} \end{bmatrix}$	$f_k(y_{kt}\mid \mu_k, \boldsymbol{\theta}_k, b_{kt})$	$f(\boldsymbol{y}_t \mid \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{b}_t)$
A	$\begin{bmatrix} \gamma b_t+\varepsilon_{1t} \\ b_t \end{bmatrix}$	$y_{1t} \sim N\left(u_1, \gamma^2\sigma_b^2+\sigma_{\varepsilon 1}^2\right) y_{2t} \sim N\left(u_2, \sigma_b^2\right)$	$\boldsymbol{y}_t \sim N_2\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} \gamma^2\sigma_b^2+\sigma_{\varepsilon 1}^2 & \gamma\sigma_b^2 \\ \gamma\sigma_b^2 & \sigma_b^2 \end{bmatrix}\right)$
B	$\begin{bmatrix} \gamma b_t \\ b_t+\varepsilon_{2t} \end{bmatrix}$	$y_{1t} \sim N\left(u_1, \gamma^2\sigma_b^2\right) y_{2t} \sim N\left(u_2, \sigma_b^2+\sigma_{\varepsilon 2}^2\right)$	$\boldsymbol{y}_t \sim N_2\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} \gamma^2\sigma_b^2 & \gamma\sigma_b^2 \\ \gamma\sigma_b^2 & \sigma_b^2+\sigma_{\varepsilon 2}^2 \end{bmatrix}\right)$
C	$\begin{bmatrix} b_t+\varepsilon_{1t} \\ \gamma b_t \end{bmatrix}$	$y_{1t} \sim N\left(u_1, \sigma_b^2+\sigma_{\varepsilon 1}^2\right) y_{2t} \sim N\left(u_2, \gamma^2\sigma_b^2\right)$	$\boldsymbol{y}_t \sim N_2\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} \sigma_b^2+\sigma_{\varepsilon 1}^2 & \gamma\sigma_b^2 \\ \gamma\sigma_b^2 & \gamma^2\sigma_b^2 \end{bmatrix}\right)$
D	$\begin{bmatrix} b_t \\ \gamma b_t+\varepsilon_{2t} \end{bmatrix}$	$y_{1t} \sim N\left(u_1, \sigma_b^2\right) y_{2t} \sim N\left(u_2, \gamma^2\sigma_b^2+\sigma_{\varepsilon 2}^2\right)$	$\boldsymbol{y}_t \sim N_2\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} \sigma_b^2 & \gamma\sigma_b^2 \\ \gamma\sigma_b^2 & \gamma^2\sigma_b^2+\sigma_{\varepsilon 2}^2 \end{bmatrix}\right)$

where $u_k = \mu_k + \theta_{k1}\left(y_{kt-1} - \mu_k\right) + \cdots + \theta_{kp_k}\left(y_{kt-p_k} - \mu_k\right)$

3 Result and Analysis

3.1 Ontario Data

We obtain data from the daily epidemiological summaries provided by Public Health Ontario.[1] To study the delayed effect of hospitalization on mortality, the daily number of new hospitalizations 6 days prior and the daily number of new deaths are defined as the outcomes of interest. We shift the time between these two outcomes by 6 days because recent research shows that a reasonable estimate of the median time from hospitalization to death for Covid-19 varies between 4 days (i.e. [15]; The Ontario Public Health surveillance report[2]) to 7.5 days (i.e. Zhao et al. 2020). Additional evidence for using a 6-day lag is that a basic generalized additive model examining the relationship between hospitalizations at various lags and deaths gives the highest deviance explained at a 6-day lag period. There are $n = 78$ observations from March 29 to June 14. Figure 1 provides an illustration of the data. The left panel is the base 10 logarithm of the cumulative number of hospitalizations (black solid lines) and deaths (red dashed lines) and the right panel provides the daily number of these outcomes. On the right panel, both processes demonstrate a downward trend starting in May, while their residual, defined as their difference (blue dotted line), appears stationary. We define the daily number of new hospitalizations and new deaths by y_{1t} and y_{2t} and identify here the potential long-run relationship between them, if any.

3.2 Cointegration Analysis of Ontario Data

To assess if y_{1t} and y_{2t} are cointegrated, we first need to identify an appropriate model for each process, respectively denoted as AR(p_k), $k = 1, 2$. For $p_k = 1, \ldots, 10$, the Akaike information criterion (AIC) of the models yield a minimum at $p_1 = 5$ for y_{1t} and at $p_2 = 3$ for y_{2t}. We apply the ADF test to y_{1t} and y_{2t} to determine if they are non-stationary under the models selected with the minimum AIC. The p-values for the tests are 0.258 and 0.193, respectively, suggesting that y_{1t} and y_{2t} are not stationary. Taking the first order difference of each process and reapplying the above procedure on $y^*_{1t} = \Delta y_{1t}$ and $y^*_{2t} = \Delta y_{2t}$ yields a minimum AIC for each model at $p_1 = 9$ for y^*_{1t} and $p_2 = 8$ for y^*_{2t}. Under the models with the minimum AIC, the p-values for the tests are 0.027 and 0.052, respectively, suggesting that y^*_{1t} and y^*_{2t} are stationary. Identifying an appropriate model for $z_t = y_{1t} - y_{2t}$ as an AR(p) process, yields a minimum over $p = 1, \ldots, 10$ of the AIC at $p = 3$.

[1] https://covid-19.ontario.ca/covid-19-epidemiologic-summaries-public-health-ontario#daily

[2] https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-severe-outcomes-ontario-epi-summary.pdf?la=en

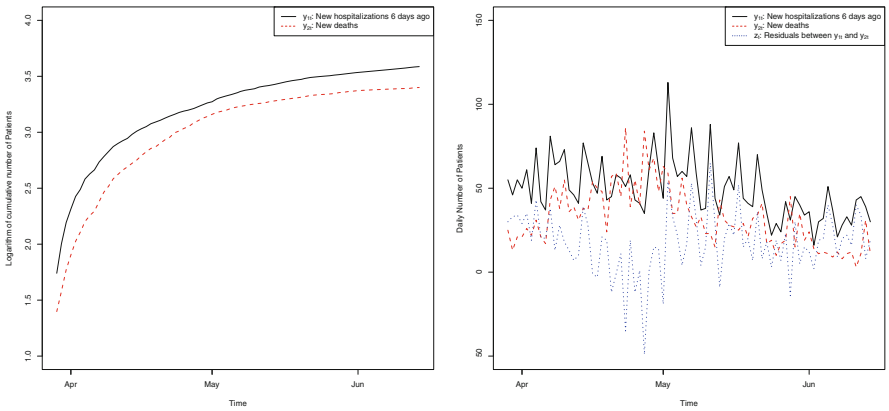


Fig. 1 The left panel illustrates the logarithm of the cumulative number of hospitalizations 6 days prior (black) and the cumulative number of deaths (red). Hospitalizations and deaths grow with a decreasing rate over time. The right panel plots the daily number of these quantities and their residuals (blue) against time. The processes are identified as having a long-term correlation through the cointegration analysis described in the text

The corresponding p-value for testing non-stationarity of z_t, distributed as AR(3), is 0.019, suggesting that z_t is stationary. This evidence indicates that there is long-run correlation between y_{1t} and y_{2t}; that y_{1t} and y_{2t} are cointegrated such that y_{1t}~$I(1)$, y_{2t}~$I(1)$ and z_t~$I(0)$.

3.3 Joint Modeling of the Ontario Data

The joint model is fitted by the adaptive Markov Chain Monte Carlo (MCMC) method described in Xi et al. (2020). We assume vague priors commonly used in the literature: for $k = 1, 2$ and $s = 1, \ldots, p_k$, $p(u_k)$ and $p(\theta_{ks})$ follow i.i.d. $N(0, 10000)$; $p(\gamma)$, $p(\sigma_b)$ and $p(\sigma_{\varepsilon k})$ follow i.i.d. half. $N(0, 10000)$. Credible intervals are obtained as the lower and upper 2.5% quantiles of the posterior density. The goodness of fit of the models are assessed by their deviance information criteria (DIC) with models having low DIC considered to offer a good fit to the data (Spiegelhalter et al., 2002).

We consider model parameterization in three ways. Four forms of the joint model as provided in Table 1 are considered; four choices of order of integration based on the result of the cointegration analysis above and additionally exploring the use of the responses themselves as well as first differences: $(d_1, d_2) = (0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$; a hundred combinations of the number of lags: $p_k = 1, \ldots, 10$ for each of $k = 1, 2$. Hence a total of 1600 models are estimated. We first select the models with the optimal number of lags using the DIC criterion under each of the joint models and the forms of the order of integration, then choose an overall model that provides the best fit.

Table 2 Statistics assessing model fits for the candidate models

Form	d1	d2	p1	p2	DIC
A	0	0	4	10	518
A	0	1	4	10	518
A	1	0	5	10	516
A	1	1	3	10	520
B	0	0	10	4	493
B	0	1	10	3	487
B	1	0	10	2	485
B	1	1	10	3	482
C	0	0	5	10	521
C	0	1	4	10	517
C	1	0	5	10	517
C	1	1	3	10	518
D	0	0	10	4	494
D	0	1	10	3	486
D	1	0	10	2	486
D	1	1	10	3	484

Table 2 lists the 16 optimal models along with their DIC. Including an outcome-specific variability term, ε_{kt}, in modeling the outcome death (i.e. as in models B and D) yields a much better fit than incorporating such a term in modeling the outcome hospitalization (i.e. as in models A and C). Models with a factor loading on hospitalization (i.e. B) have slightly better fit than those with a factor loading on death (i.e. D). Both of the outcomes hospitalization and death are best fitted with an order of integration $d_k = 1$. This is consistent with the results from our cointegration analysis. We note that for all models omitting the additive error term, ε_{kt}, yields that the maximum number of lagged terms needs to be considered. We note that a better fit may be produced by incorporating an even higher number of lagged terms in the model, but a model with high number of lagged terms is not conducive to model parsimony.

The parameter estimates along with their 95% credible intervals for model B, with the lowest DIC, are presented in Table 3. The intercepts of the model are non-significant with estimates of μ_k respectively as 1.560 $(-0.171, 3.350)$ and -1.631 $(-0.280, 1.080)$ for $k = 1, 2$; recall the responses here are first order differences since $d_k = 1, k = 1, 2$. Although model B incorporated 10 lagged terms for the outcome hospitalization, only the credible intervals for the coefficients of the first two lagged terms do not include zero. The values of the coefficients of the leading lagged term, θ_{k1}, $k = 1, 2$, are estimated respectively as -0.295 $(-0.410, -0.165)$ and -0.741 $(-0.977, -0.506)$, suggesting that the correlation of death ($k = 2$) with observations on the previous day is stronger than the corresponding correlation for hospitalization ($k = 1$). The standard deviation of the shared variability, σ_b, has an estimate of 1.416 $(1.151, 1.841)$. The factor loading parameter, γ, and the standard deviation of the outcome-specific variability in modelling death, σ_2, have estimates of 13.644 $(11.129, 14.876)$ and 12.533 $(10.610, 15.131)$, respectively. The variance

Table 3 Posterior estimates of the model parameters

	$Q_{.025}$	$Q_{.500}$	$Q_{.975}$
μ_1	−0.171	1.560	3.350
μ_2	−1.631	−0.280	1.080
θ_{11}	−0.410	−0.295	−0.165
θ_{12}	−0.341	−0.172	−0.046
θ_{13}	−0.053	0.020	0.094
θ_{14}	0.056	0.143	0.283
θ_{15}	−0.042	0.056	0.225
θ_{16}	−0.006	0.058	0.161
θ_{17}	−0.170	−0.077	0.010
θ_{18}	−0.101	−0.038	0.044
θ_{19}	−0.093	0.054	0.161
θ_{110}	−0.071	0.023	0.116
θ_{21}	−0.977	−0.741	−0.506
θ_{22}	−0.598	−0.318	−0.035
θ_{23}	−0.489	−0.254	−0.017
γ	11.129	13.644	14.876
σ_b	1.151	1.416	1.841
$\sigma_{\varepsilon 2}$	10.610	12.533	15.131

of the outcomes is parameterized as $\gamma^2\sigma_b^2$ for y_{1t} and $\sigma_b^2+\sigma_{\varepsilon 2}^2$ for y_{2t}, with estimates of 373.06 (258.57, 537.15) and 159.08 (114.48, 229.82). These estimates suggest that although there is dependence between hospitalizations and deaths, much of the variability in these outcomes is unexplained as outcome specific random error.

The left panel of Fig. 2 illustrates the posterior estimates of the shared random effect, b_t, along with their 95% credible intervals, plotted against time. The peak value of 5.38 on May 2 reflects the peak of hospitalization 6 days prior in Fig. 1. The right panel of Fig. 2 provides the estimated joint distribution of the outcomes, reflecting the positive, weak correlation in these outcomes as discussed earlier.

4 Discussion

The co-integration analysis identified a long-run relationship between hospitalizations and deaths subsequently modeled through a joint outcome autoregressive model with a shared latent random effect. The first order differences of hospitalizations 6 days prior, and deaths, in the joint outcome model are autoregressively correlated with the observations 2 and 3 days ago respectively. The autocorrelation could be a result of the reporting schedule by public health units as many of them do not report on weekends. The weak dependence between the outcomes may be due in part to reporting lags in both hospitalizations and deaths in the Ontario data. The data are reported to Public Health Ontario by 34 different public health units, and while the reporting lag is not currently quantified, it likely varies by health unit

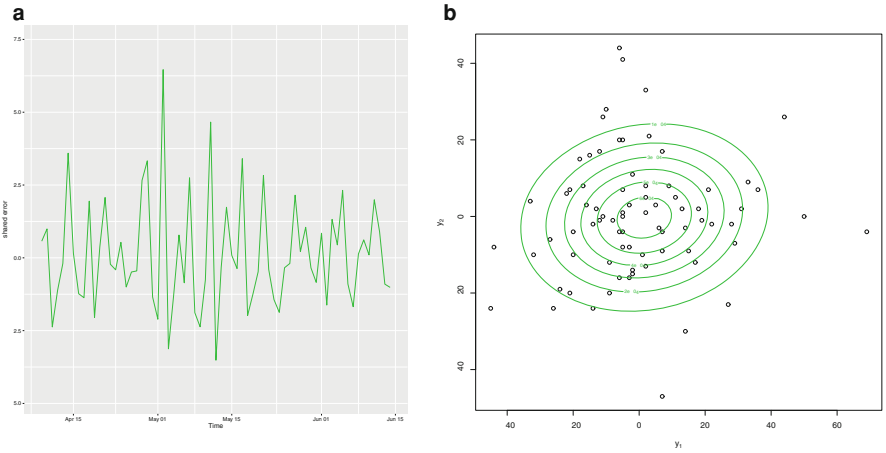


Fig. 2 Posterior estimates of the shared random effect (left panel) and the estimated joint distribution (right panel) of the outcomes, y_1 and y_2 being the first order difference of hospitalizations 6 days prior and deaths, respectively. The posterior estimates of the shared random effect have a peak on May 2, reflecting the peak in daily hospitalizations. The estimated joint distribution of the outcomes demonstrates a weak dependence between the outcomes

and by outcome. In future work we hope to be able to adjust for the lags in both outcomes.

The framework can be extended in several ways to reduce the unexplained variability, enhance predictability, and sharpen linkages across the outcomes. An ARIMA model that has moving averaging error terms may better describe the structure of the variability, and it may also be useful to incorporate autoregressive structures in the latent random effect. Comparisons with multivariate time series frameworks may help identify the benefits of using shared random effect for modeling joint outcomes beyond ease of interpretation. Environmental data associated with each day, such as temperature and humidity (i.e. [13, 16]), as well as geographical information, if available, may be included into the model as explanatory covariates. As the uncertainty in the model is reduced and with stronger linkages evidenced across the outcomes, given any current increment in hospitalization, more accurate predictions of future mortality may be obtained through the estimated joint distribution of the outcomes.

Acknowledgments This paper was first presented at the Fields Institute Seminar Series, 2020-2021 COVID-19 Math Modelling Seminar, on October 6, 2020. The authors thank the participants for their comments and helpful suggestions.

References

1. Zeger SL, Irizarry R, and Peng RD. 2006. On time series analysis of public health and biomedical data. Annu. Rev. Public Health, 27: 57–79
2. Allard R. 1998. Use of time-series analysis in infectious disease surveillance. Bulletin of the World Health Organization, 76(4): 327
3. Checkley W, Epstein LD, Gilman RH, Figueroa D, Cama RI, Patz JA and Black RE. 2000. Effects of EI Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. The Lancet, 355(9202): 442–450
4. Ding G, Li X, Shen Y and Fan J. 2020. Brief Analysis of the ARIMA model on the COVID-19 in Italy. medRxiv
5. Pfaff B. 2008. Analysis of integrated and cointegrated time series with R. Springer Science & Business Media
6. Chen MP, Lee CC, Lin YH and Chen WY. 2018. Did the SARS epidemic weaken the integration of Asian stock markets? Evidence from smooth time-varying cointegration analysis. Economic research-Ekonomska istraživanja, 31(1): 908–926
7. Zeren F and Hizarci A. 2020. The impact of COVID-19 coronavirus on stock markets: evidence from selected countries. Muhasebe ve Finans İncelemeleri Dergisi, 3(1): 78–84
8. Şenol Z and Zeren F. 2020. Coronavirus (COVID-19) and stock markets: The effects of the pandemic on the global economy. Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi, 7(4): 1–16
9. Dunson DB. 2000. Bayesian latent variable models for clustered mixed outcomes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(2): 355–366
10. Henderson R and Shimakura S. 2003. A serially correlated gamma frailty model for longitudinal count data. Biometrika, 90(2): 355–366
11. Feng CX and Dean CB. 2012. Joint analysis of multivariate spatial count and zero-heavy count outcomes using common spatial factor models. Environmetrics, 23(6): 493–508
12. Juarez-Colunga E, Silva GL and Dean CB. 2017. Joint modeling of zero-inflated panel count and severity outcomes. Biometrics, 73(4): 1413–1423
13. Tsiatis AA and Davidian M. 2004. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica, 809–834
14. Lundy ER and Dean CB. 2021. Analyzing heaped counts versus longitudinal presence/absence data in joint zero-inflated discrete regression models. Sociological Methods & Research, 50(2): 567–596
15. Trivedi TK, DeSalvo T, Lee L, Palumbo A, Moll M, Curns A, Hall AJ, Patel M, Parashar UD and Lopman BA. 2012. Hospitalizations and mortality associated with norovirus outbreaks in nursing homes, 2009–2010. Jama, 308(16): 1668–1675
16. Dickey DA and Fuller WA. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a): 427–431
17. Park JY. 2003. Bootstrap unit root tests. Econometrica 71: 1845–95
18. Wei J. 2014. On Bootstrap Evaluation of Tests for Unit Root and Cointegration. Acta Universitatis Upsaliensis
19. Chang Y, Sickles RC and Song W. 2001. Bootstrapping unit root tests with covariates. mimeographed, Department of Economics, Rice University.
20. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, ..., Cookingham J. 2020. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 323(20): 2052–2059
21. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D and Wang MH. 2020. 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. International journal of infectious diseases, 92: 214–217

22. Chan K-H, Peiris JM, Lam S, Poon L, Yuen K, Seto WH. 2011. The effects of temperature and relative humidity on the viability of the SARS coronavirus. Advances in virology 2011:734690
23. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. 2020. Temperature, Humidity, and Latitude Analysis to Estimate Potential Spread and Seasonality of Coronavirus Disease 2019 (COVID-19). JAMA Network Open 3: e2011834

Evaluating the Risk of Reopening the Border: A Case Study of Ontario (Canada) to New York (USA) Using Mathematical Modeling

Pei Yuan, Elena Aruffo, Qi Li, Juan Li, Yi Tan, Tingting Zheng, Jummy David, Nick Ogden, Evgenia Gatov, Effie Gournis, Sarah Collier, Beate Sander, Guihong Fan, Jane M. Heffernan, Jun Li, Jude Dzevela Kong, Julien Arino, Jacques Bélair, James Watmough, and Huaiping Zhu

1 Introduction

SARS-CoV-2 can rapidly spread within and between populations. Travel restrictions and border closures have been among the first control measures to be implemented during the COVID-19 pandemic. After several months of lockdown, regional economic outcomes have suffered greatly [1]. Borders have therefore begun to be

Supplementary Information The online version contains supplementary material available at (https://doi.org/10.1007/978-3-030-85053-1_14).

Pei Yuan, Elena Aruffo, Qi Li and Juan Li contributed equally.

P. Yuan · E. Aruffo · Y. Tan · J. David · J. M. Heffernan · J. D. Kong · H. Zhu (✉)
Funding: This research was supported by the Canadian Institutes of Health Research, Canadian COVID-19 Math Modelling Task Force and the Natural Sciences and Engineering Research Council of Canada, Toronto, ON, Canada
e-mail: yuanpei@yorku.ca; aruffoe@yorku.ca; lbxty@yorku.ca; jummy30@yorku.ca; jmheffer@yorku.ca; jdkong@yorku.ca; julien.arino@umanitoba.ca; belair@dms.umontreal.ca; watmough@unb.ca; huaiping@yorku.ca

Q. Li
Shanghai Normal University, Shanghai, China
e-mail: 1000443603@smail.shnu.edu.cn

J. Li
Shanxi University, Taiyuan, China
e-mail: lijuanjuan0120@163.com

T. Zheng
Xingjiang University,Ürümqi, China
e-mail: ztt0711@163.com

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_14

reopened to allow the growth of local economies, but to do so, changes to safety measures, such as COVID-19 tests, health certificates, and quarantine requirements [2–5], to keep the importation of cases and possible spread of infection in destination regions under control.

While reopening borders will help relaunch local and global economies, a complete and uncontrolled opening can result in new waves of outbreaks. More locally, Prince Edward Island (Canada) experienced a new cluster of COVID-19 cases stemmed from a traveler from the USA [6]. Infected travelers are key to the global spread of COVID-19 [7–10]. Therefore, the critical question is if there is a way to reopen borders but ensuring minimal COVID-19 spread risk.

There has been an enormous amount of trade and transportation across the Canada-USA land border. On average, more than 70,000 trucks and 700,000 people crossed the Buffalo-Niagara Falls border each month in 2019 [11]. Between January to June 2020 there were only 744,489 overnight visitors into Ontario from the USA (4136 per day), a 73.7% decrease compared to the same period in 2019 (15,698 per day). Overnight overseas visitors also dropped [12] (by 71.3%). Simultaneously, Ontario's hotel occupancy rate dropped by 31.1% (from 65.4% to 34.3%), and

N. Ogden
Public Health Agency of Canada, Ottawa, ON, Canada
e-mail: nicholas.ogden@canada.ca

E. Gatov · E. Gournis · S. Collier
Toronto Public Health, Toronto, ON, Canada
e-mail: evgenia.gatov@toronto.ca; jenny.gatov@mail.utoronto.ca; effie.gournis@toronto.ca; sarah.collier@toronto.ca

B. Sander
Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
e-mail: beate.sander@uhnresearch.ca

G. Fan
Columbus State University, Columbus, GA, USA
e-mail: fan_guihong@columbusstate.edu

J. Li
Xidian University, Xi'an, China
e-mail: lijun@xidian.edu.cn

J. Arino
Department of Mathematics, University of Manitoba, Winnipeg, MB, Canada
e-mail: Julien.Arino@umanitoba.ca

J. Bélair
Université de Montréal, Montreal, QC, Canada
e-mail: belair@dms.umontreal.ca

J. Watmough
University of New Brunswick, Fredericton, NB, Canada
e-mail: watmough@unb.ca

revenue per available room dropped by 55.6% [12]. Although no direct data show the specific losses caused by the border closure, open borders play a crucial role in both economic and social lifelines. However, the USA's situation is very worrying, with 6,656,799 cumulative cases on September 18, 2020 and currently about 40,000 new cases per day [13]. Thus, a reopening of the Canada-USA border requires careful consideration. Now the border is still closed [14, 15], if and when to reopen made frequent headlines of both countries.

Some studies have confirmed that border closure has played a crucial role in delaying and controlling the advancement of COVID-19 [16–25] Dickens et al. [16] showed that effective testing and a mandatory 14-day quarantine of all travelers into Canada are needed to curb COVID-19 if the border is reopened. Linka et al. [17], however, suggested a complete travel ban.

Another COVID-19 restrictive measure that has been implemented is the stay-at-home policy (SAHP). SAHP is effective in controlling the spread of the infection [26, 27] in concert with border closures. Ontario entered phase three with community reopening at the end of July, and while transmission has been kept under control, recently, there is evidence that the epidemic is resurging [28]. If borders are reopened, it may increase epidemic resurgence in Canada beyond our capacity to control.

Given the extensive land border between Canada and the USA, it is essential to investigate the impact of a full/partial reopening of the border on the spread of COVID-19 in Canada. Herewith we present a compartmental mathematical SEAIR model (following the Susceptible-Exposed-Asymptomatic-Infectious (prodromal phase)-Infectious (with symptoms)-Recovered) that is employed to quantify different border reopening strategies. Using the household structure model for Toronto, we will incorporate daily travelers, traveler quarantine policy, and border testing or required health certificates. We define an average risk index to forecast the intensity of COVID-19. Investigating the effective reproduction number and forecasting simulations is also carried out to inform public health decision-makers on the best border reopening strategies that will contain the spread of the epidemic.

2 Method

2.1 Data and Materials

Canada and the USA share the longest land border in the world. However, we focus only on investigating visitors from USA to Canada from the border of Ontario and New York State given the large number of vehicles crossing the border [29], and consider the extreme case that visitors all enter Toronto as a worst scenario of increasing incoming travelers after border reopens. We obtained daily new confirmed case data, by episode date and reporting date in Toronto from February 24, 2020, to August 31, 2020 [28]. This data is used to validate our model during

escalation (February 24–mid-March), mitigation (mid-March–mid-May), and de-escalation (stage 1: mid-May–mid-June, stage 2: mid-June–mid-July, stage 3: mid-July to border reopening) stages of the epidemic before border reopening, and to evaluate the risk of transmission in Toronto over time. The model validation results are presented in Figure ESMC1 in the electronic supplementary material (ESM C).

2.2 Compartmental Model: Description and Assumptions

The federal government of Canada determines entry into Canada. Therefore, the federal government decides border control measures, whereas what happens after an individual has been granted access into the country is determined by provincial governments. To control case-importation, and consequently, a rapid spread of COVID-19 in Ontario, we assume that the Ontario will implement safety measures for border reopening, applicable at the border upon arrival once travelers are granted entry into the country. We assume that upon arrival at the frontier, traveler health certificates (with an issue date that does not exceed 2 days prior arrival) will be required and that COVID-19 symptom detection checks (i.e., temperature testing) will be implemented. We also assume that all eligible travelers entering Canada will need to provide proof of domicile and a 14-day quarantine plan. Moreover, they must follow all local control policies, i.e., Nonpharmaceutical Interventions (NPIs), personal protective equipment (PPE) requirements, and social distancing restrictions. Modeling assumptions are summarized in Table 1 (Appendix).

We extend our household-based transmission model with a Susceptible-Exposed-Asymptomatic-Infectious (prodromal phase)-Infectious (with symptoms)-Recovered framework proposed in [27] by assuming that a proportion of S, E, A and I (prodromal stage) (m_S, m_E, m_A, m_{I_1}, respectively) travelers will be allowed to cross the border. We note that individuals with positive tests will not cross the border. We do, however, allow for faulty testing, and assumed that the effective detection rate of asymptomatic infections via testing is η.

The model is structured over three periods: before and after implementing SAHP, and after border reopening. In the latter two periods, the population is divided into SAHP (home quarantine) compliant and non-compliant subpopulations, with subscript q and g, respectively.

For travelers, they are divided into quarantined and community. A proportion ω of travelers will be considered part of the community, since not all travelers will strictly respect the mandatory quarantine/isolation policies. Consequently, a proportion $1 - \omega$ will be in the quarantined household sub-group.

Figure 1 shows the flow diagram for modeling [27]. Details of model derivation are in electronic supplementary materials (ESM A).

Given the current epidemics in Canada [30], the planned reopening scheduled on October 21 is improbable. Using available data, we simulate the impact of reopening on Toronto if the border reopened on September 21, 2020 [31]. We will calculate and present useful information to public health on safe reopening once the daily

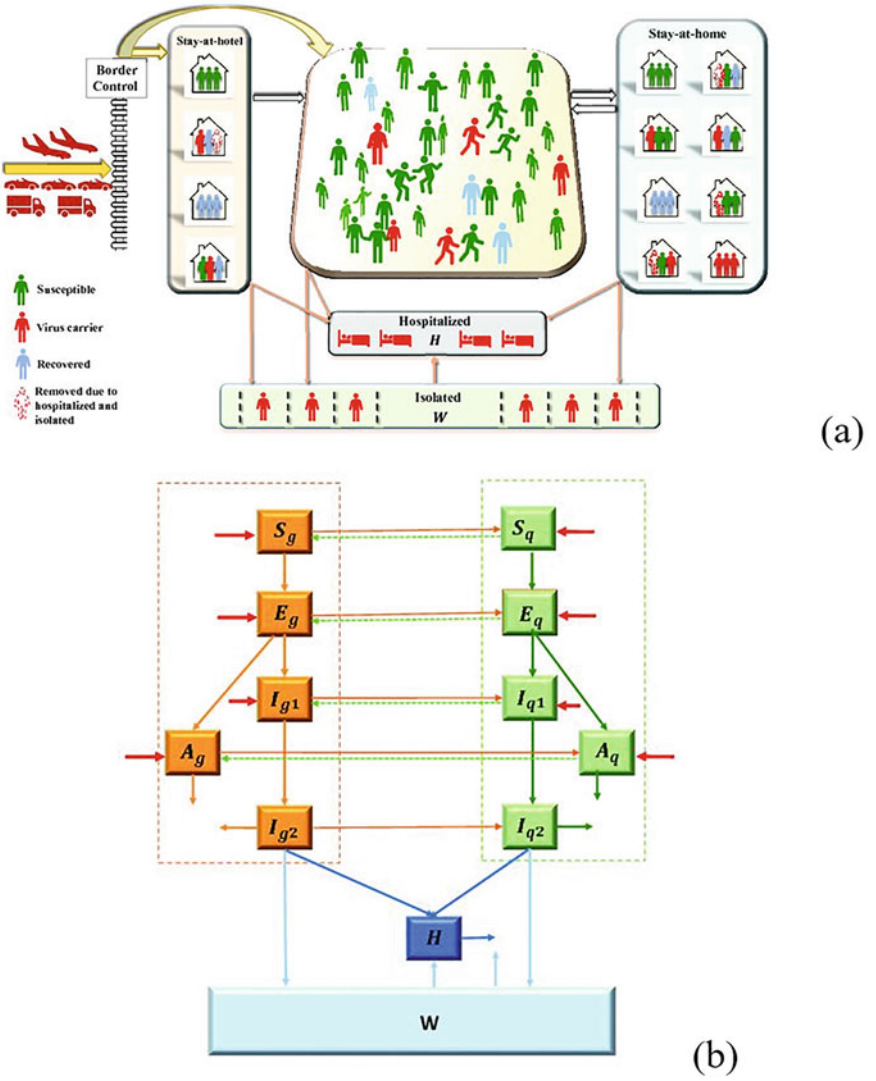


Fig. 1 Modeling with household structure. (**a**) shows the activity and response of different groups. (**b**) Schematic diagram after border reopening, red solid arrow indicates of importation of travelers into quarantined (q) or non-quarantine into Ontario (assuming that they all go to Toronto). Solid lines indicate movement between classes. Dashed lines represent the virus transmission routes

cases in Ontario drop below 100. We will also forecast the infection trend for 1, 12, and 24 weeks after this date. The degree of infection is compared between different strategies, considering variations in the number of daily travelers crossing the border, and variations in parameter for the effectiveness of border control, the ratio of individuals who will go to the non-quarantined group after arrival.

2.3 Risk Indicators of the Border Opening

2.3.1 The Instantaneous Reproduction Number

The instantaneous reproduction number, R_t,is defined as

$$R_t = \frac{I_t}{\sum_{j=1}^{t} p_j I_{t-j}}, \tag{1}$$

which can be estimated by the statistical approach [32] using episode date [28]. Here I_t is the new cases on day t and p_j is the discretized distribution of the serial intervals, assuming a Gamma distributed serial interval of 7.5 days with a standard deviation of 3.4 days [33].

2.3.2 Risk Indicator

We define a risk indicator at time t as

$$Risk_t = PercentRank\left(new\ infection_t\right) * 100 \tag{2}$$

indicating the risk of COVID-19 infection on a given day t. Estimation of $Risk_t$ is computed by calculating the rank of the percentile of daily new infections in R [34], indicating the value below which a given percentage of observations is contained. The risk is measured by a function PercentRank, given the relative rank of the number of new infections on a given day in the historical data. A percentage representing those is less than or equal to the value. A non-parametric approach is used in which data do not follow a particular distribution, and the highest and lowest values are excluded. This indicator reflects the degree of risk compared to the current epidemic period since the first wave of the Toronto epidemic peaked in April. A value close to 100 indicates that a new peak of the epidemic has appeared. Also, we define low risk if $0<Risk_t \leq 30$, moderate risk if $30<Risk_t \leq 50$, high risk if $Risk_t>50$. The risk indicator has also been used in other fields, like in microbiological control levels [35]. It is applicable to inform the public about the risk of COVID using this risk indicator.

3 Results

3.1 Risk in Toronto

Figure 2 plots the risk indicator $Risk_t$ for different stages of escalation, mitigation, and de-escalation. We observe that in the 2 weeks before stage 1 reopening was

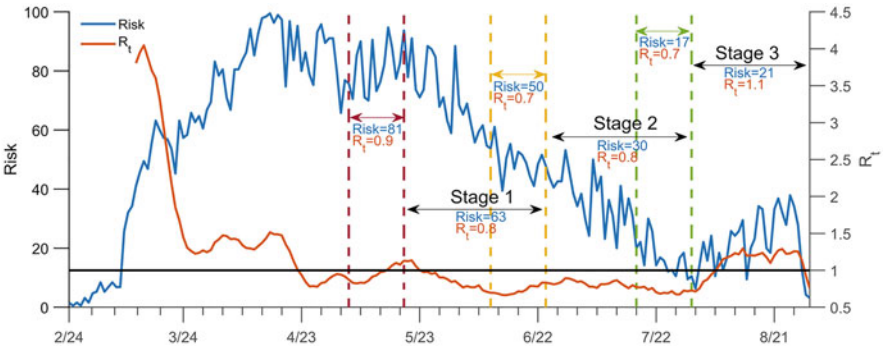


Fig. 2 Risk of COVID-19 in Toronto. Risk indicator in Toronto from February 24 to August 31. The average risk and R_t 2 weeks before and in the period of reopening stage 1, 2 and 3 are presented. The dark solid line indicates the critical threshold $R_t = 1$ of the instantaneous reproduction number. All dates are in 2020

implemented, the risk indicator is very high (Risk = 81) but the average R_t has declined to 0.9. Within the 2 weeks before stage 2 reopening, the average risk has decreased to 50, with $R_t = 0.7$. Finally, before Toronto entered stage 3, the risk dropped to 17 with $R_t = 0.7$. Although the effective reproduction numbers before each reopening stage are below 1, the epidemic risk is entirely different. We also note that the R_t increases to values greater than 1 in stage 3, but the risk value is not high [21].

3.2 Effect of Border Control Measures

Border control will affect the epidemic in two ways: managing the number of people allowed to cross the border and controlling what travelers do after crossing. Figure 3 shows that 1500 daily travelers will generate a sufficient number of infections to cause a new transmission wave, which will be more severe than the first wave. Observe that if the quarantine policy is strictly followed by all travelers (Fig. 3a, b), the daily new cases increase until December 31 and slightly decrease by April 30, 2021 (panel A). We also see that the number of new infections that enter the region does not differ significantly between a 50 or 100% detection level. By April 30 there is only a 5000 person difference in the cumulative number of imported cases (panel B). If 100% of all travelers quarantined upon entering, the level of government resources invested in health certificate checking and temperature testing is not over burdensome. However, we observe a similar outcome when the detection rate is

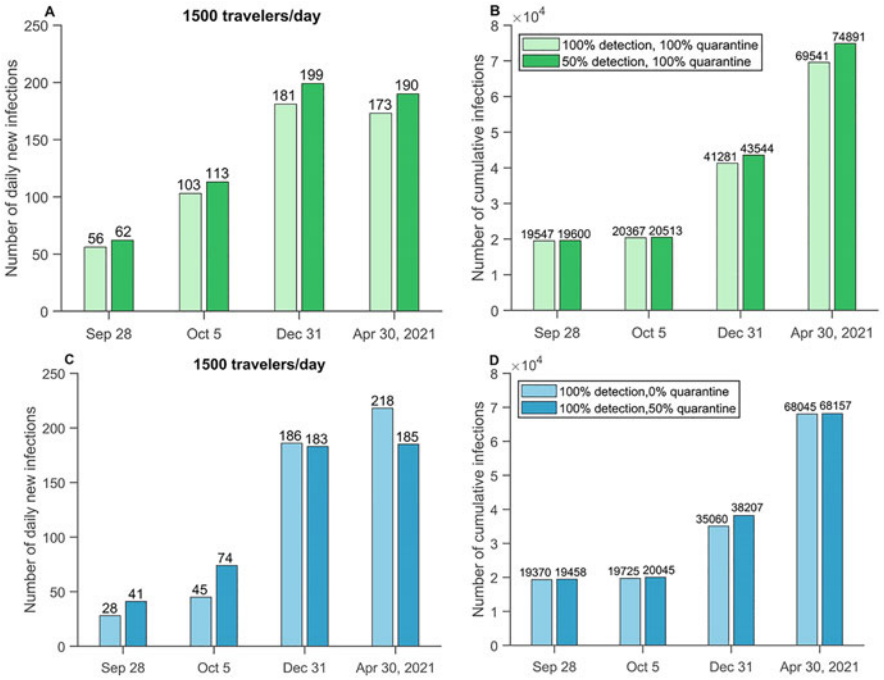


Fig. 3 Effect of border control measures. (top row) Number of daily new infections (**a**) and cumulative infection (**b**) in Toronto 1 week after border reopening (September 28, 2020) to the end of April 2021 for 1500 travelers, when $\omega = 0$ and $\eta = 1, 0.5$. (bottom row) Number of daily new infections (**c**) and cumulative infection (**d**) in Toronto 1 week after border reopening (September 28, 2020) to the end of April 2021 for 1500 travelers daily, when $\eta = 1$ and $\omega = 1, 0.5$. ω= proportion of travelers not following home quarantine orders. η= effective detection rate of asymptomatic infections

100% in detecting possible infection importation (panels C and D)—the difference between the cumulative number of infections given a 0–50% quarantine uptake rate is miniscule (panel D). We note that the results shown in Fig. 3 may not reflect government investment optimization and may instead solely indicated that daily 1500 travelers might overwhelm the system. Both scenarios have been investigated with fewer daily travelers (500 or 1000) in electronic supplementary material (ESM C) (see Fig. ESMC2-C3). While the number of daily and the cumulative number of infections are reduced, we find a similar outcome when comparing detection levels and quarantine rates.

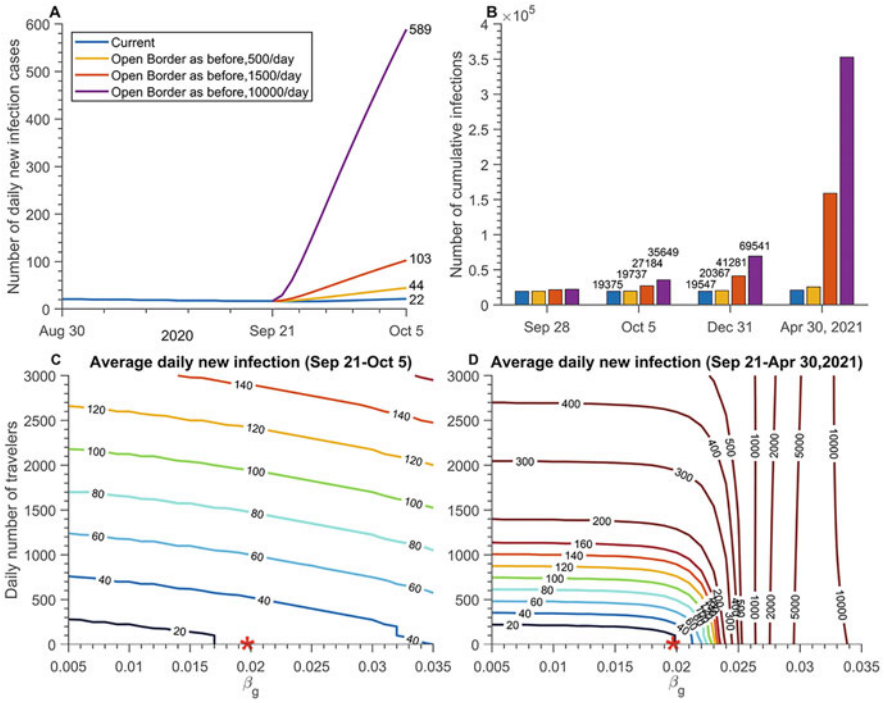


Fig. 4 Effect of importation of travelers. Number of daily new infections (**a**) and number of cumulative infections with different daily number of travelers (0 (border closed), 500, 1500, 10,000 (the situation before pandemic)) under the best border control measures, $\omega = 0$, $\eta = 1$. Contour plot of average daily new infections after border reopening (**c**) within 2 weeks (September 21–October 5) and (**d**) in the long run (September 21, 2020–April 30, 2021) with different daily number of travelers and β_g. The red star indicates the current state in Toronto ($\beta_g = 0.019$). ω= proportion of travelers not following home quarantine orders. η= effective detection rate of asymptomatic infections. β_g= probability of transmission per contact outside household

3.3 *Effect of Importation of Travelers*

The number of infection cases will increase over time when 500, 1500 or 10,000 (86% of 2019) travelers cross the border daily under perfect conditions that all are quarantined, and the test efficacy is 100% (Fig.4a, b). It is visible that if more than 1500 travelers cross the border daily, the number of infections increases sharply between December 2020 and April 2021. Moreover, if 10,000 travelers are entering Ontario daily, the epidemic will become much severer with daily reported cases over 600 in October that keeps increasing after that. We also observe that opening the border to 500 travelers daily (yellow) does not result in a large outbreak, but the cases are still rising if compared to the current situation.

The number of daily new cases decreases as the number of daily travelers and the probability of transmission outside household (β_g) decreases in the short and long term (Fig. 4c, d). With the current NPIs policy (β_g= 1.9%), an average of 100 or less daily new infections 2 weeks after border reopening are possible if a maximum of 2000 travelers enter Ontario. It will also happen if the current restrictions are lifted (increasing the value of β_g). However, in the long run (until the end of April 2021, Fig. 4d), the daily number of travelers should be restricted to 1000 to keep the average daily new infection below 100 under current NPIs interventions. And if NPIs are relaxed a little (β_g larger than 2.3%), daily cases will exceed 100 if daily number of travelers is above 200. Moreover, Toronto might experience 10,000 daily newly reported cases with relaxed NPIs if borders are opened for an extended period.

3.4 Tradeoff Between Border Reopening and Local Risk

If we open the border for more to enter Canada (represented by Ontario here), we will face increasing local transmission (Fig. 5). Currently, the risk in Toronto (blue star, Fig. 5) is high, we find that to mitigate the risk (in the interval (0–30)), we can only allow about 100 travelers to enter Ontario each day. If more than 300 people cross the border, the average risk will become medium (between 30 and 50), and if more than 1200 travelers to enter, the risk will become high (above 50). We also observe that if the risk in Toronto increases above 30, the average risk will always be medium or high, even just a few travelers to cross the border. On the other hand, if the local risk is relatively low, the average risk remains low if we allow a maximum of daily 500 travelers to cross the border.

4 Discussion

Our findings suggest that the border may be reopened with the restricted number of travelers, under the strictest border control measures when the daily cases in Ontario is roughly 100. Despite that, the risk of local transmission will rise. Hence reopening is imprudent given the current arising situation. Effective detection of infectious visitors at the border and quarantine of passengers after entry can reduce the risk of reopening to a certain extent.

However, we also observe that, in the absence of efficient border control measures and quarantine of travelers, reopening the border might induce a new local outbreak even with a low number of visitors. Moreover, the current local risk is still at a critical point of resurgence and, a slight relaxation of the current

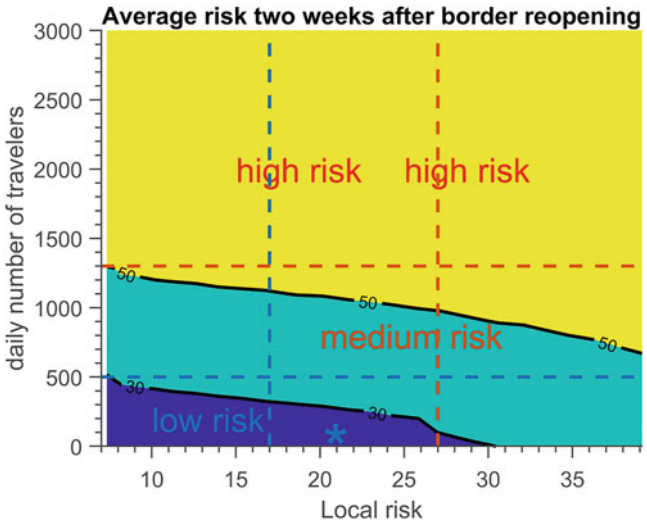


Fig. 5 The tradeoff between border reopening and local risk. Contour plot of average risk in the period of 2 weeks after border reopening on September 21 with different daily number of travelers and local risk. The blue star is the current state.

control measures may result in a new wave of outbreak. Without strengthening local prevention and control measures, it would be even more impractical to reopen the border. Also, even with perfect control measures (efficient testing process at the border and strict implementation of quarantine), if a larger number of passengers, for example, 10,000 per day, to enter Ontario, it will result in a second wave at least three times larger than the original COVID-19 wave in Toronto.

Reasonable and effective risk measurement indicators are crucial for short-term risk forecasting and timely adjustment of staggered measures, especially when it is foreseen that SARS-CoV-2 may persist for a long time. When the instantaneous reproduction number R_t becomes less than 1, we will consider that the local epidemic is well mitigated or controlled. However, when considering border reopening, it is not enough to judge the situation of the epidemic from R_t alone, since we noticed that during the first phase of Toronto's reopening, R_t was already less than 1, but the local infection risk was still high. To ensure more accurate short-term predictions, we consider both R_t and risk indicators as a measurement of infection levels.

In the short run, our new indicator of risk of infection has been useful to establish different levels of risk: low, medium and high. We identified that the current risk in Toronto is low and, in case of reopening, this level can be maintained only if the local risk is below 30 and at most 500 travelers can cross the border. Allowing more

people to enter Ontario will result in an increase of the general risk of infection in the province. Hence, we recommend keeping the level of risk low in Ontario, using NPIs, and reopening the border to a minimal number of travelers.

We mainly discuss the number of passengers allowed to enter if the border reopens and local risk. Our results should be interpreted as the best-case scenario indicating the maximum possible number allowed after border reopening. However, neither the effectiveness of border detection nor compliance of individuals coming from abroad with quarantine can be fully guaranteed. People may get upset with social distancing and other control measures, and children have started returning to school in early September, which makes local epidemic prevention and controls more difficult. Therefore, keeping the border closed might be a more appropriate and safer choice. Reopening the border might be a feasible plan when a vaccine is available.

Our work provides an essential reference for public health, it has some limitations. Scenario analyses are conducted under the current epidemic situation in USA. However, it is not very likely that the epidemic in USA will mitigate sooner. If the USA epidemic were to become controlled, our results would need to be re-examined. Moreover, our findings may be too pessimistic for lifting travel restrictions towards countries where the epidemic is well controlled, such as South Korea and China.

In conclusion, reopening the borders to the USA is possible only if the mandatory quarantine, high efficiency of testing at the frontier and a maximum daily number of 500 travelers to cross, if the use of NPIs is enforced strictly, or strengthened further and if the daily cases in Canada drop drastically about 100.

Appendix

Table A.1 Model assumptions

General setting	a.	No birth, death or immigration
	b.	We divide the population into two groups: one consisting of individuals who follows SAHP (indicated by subscript q) and another consisting of individuals who do not opt for this intervention (subscript g). Due to influences of self-protection consciousness and severity of the epidemic, people are assumed to move from one group to another with stay-at-home rate (denoted by $q(t)$) or going out rate (denoted by $g(t)$)
	c.	Each subpopulation is further the divided into Susceptible ($S_i(t)$), Exposed ($E_i(t)$), Asymptomatic (subclinical) infection ($A_i(t)$), Infectious pre-symptomatic (will eventually show symptoms) ($I_{i1}(t)$) and Infectious symptomatic ($I_{i2}(t)$)
	d.	Both $A_i(t)$ and $I_{i1}(t)$ are infectious virus carriers. Individuals in $A_i(t)$ never show symptoms, while individuals in $I_{i1}(t)$ develop into symptomatic classes ($I_{i2}(t)$) after a specified period of time

(continued)

Table A.1 (continued)

	e.	Mild symptomatic infections ($I_{i2}(t)$), may choose to either isolate themselves at home (or other places). If the quarantine is respected well enough, these infections are fully isolated and, consequently, do not contribute to the spread of the virus. Otherwise, they are still a source of infection until recovery
	f.	Two further compartments encode for severe infections: the fully isolated $W(t)$, and the hospitalized H(t) who are all severely affected. Neither of these compartments contribute to infection transmission
Household structure setting	a.	All households contain n ($n = 3$) individuals and family members are homogeneously mixing i.e., contacting each other randomly
	b.	The infection rate of asymptomatic and symptomatic infectious individuals to the susceptible is the same among the household
	c.	Two members in a family cannot be infected by one household member at the same time t
	d.	Every family except for those with symptomatic members has an equal opportunity to be released from quarantine after the SAHP is relaxed
	e.	Households with infected symptomatic individuals will continue to be quarantined after the SAHP is relaxed
	f.	For family members following SAHP, susceptible $S_q(t)$ are only infected by infectious individuals in the home $A_q(t)$, $I_{q1}(t)$ *or* $I_{q2}(t)$
	g.	When there are no infections in a household, the family is safe and is no longer be involved in the transmission of COVID-19
Assumed border control measures	a.	Travelers are required to provide certificate of health to border control officials. Travelers indicate all the people entering The test must report an issued date which does not exceed 2 days prior arrival. ONLY travelers with negative results are allowed to enter Canada. (Note that this is specific to the Canada-USA border.)
	b.	Further tests, such as rapid test (POC) or temperature check, are implemented to all travelers eligible to enter Canada. ONLY travelers with negative tests will be then allowed to enter Canada. (Note that this is specific to the Canada-USA border.)
	c.	The daily number of travelers is restricted
Assumed traveler restrictions	a.	Travelers eligible to enter Ontario/Canada need to provide proof of domicile. Visitors are only allowed to stay in hotels or isolated in camping areas
	b.	Travelers must follow the local control policies, such as mandatory NPI's, PPE's, and social distancing, etc.

Note: See electronic supplementary material for model details and derivation process

Electronic Supplementary Material

Data 1 (DOCX 468 kb)

References

1. Statistics Canada. Recent development in the Canadian economy, 2020: COVID-19, third edition. 2020. https://www150.statcan.gc.ca/n1/pub/11-626-x/11-626-x2020013-eng.htm (3 August 2020, date last accessed).
2. Etias.Info. The EU Is Reopening Its Borders for Internal Travel. 2020. https://www.etias.info/eu-reopening-borders-internal-travel/ (30 July 2020, date last accessed).
3. McClanahan P. Europe's Patchwork Reopening. The New York Times. 2020. https://www.nytimes.com/2020/06/05/travel/europe-reopening-tourism-covid.html (30 July 2020, date last accessed).
4. Chiew CJ, Li Z, Lee VJ. Reducing onward spread of COVID-19 from imported cases: quarantine and 'stay at home' measures for travelers and returning residents to Singapore. J Travel Med 2020; 27: taaa049.
5. Li X, Liu M, Zhou R, et al. Centralized medical quarantine for imported COVID-19 in Shanghai, China, J Travel Med 2020; 27: taaa109.
6. Silberman A. Canada adds health officials at U.S. border crossings to screen for COVID-19. CBC News. 2020. https://www.cbc.ca/news/canada/new-brunswick/canada-adds-health-officials-us-border-1.5645990 (15 July 2020, date last accessed).
7. Liu JY, Chen TJ, Hwang SJ. Analysis of Imported Cases of COVID-19 in Taiwan: A Nationwide Study. Int J Environ Res Public Health 2020; 17: 3311.
8. Ministry of Health (New Zealand) 3 new imported cases of COVID-19. 2020. https://www.health.govt.nz/news-media/media-releases/3-new-imported-cases-covid-19-1 (10 October 2020, date last accessed).
9. Luft A. Travelers who sat close to COVID-19-infected passenger on Air Canada flight will be alerted: Quebec public health. CTV News. 2020. https://montreal.ctvnews.ca/travellers-who-sat-close-to-covid-19-infected-passenger-on-air-canada-flight-will-be-alerted-quebec-public-health-1.4825506 (1 September 2020, date last accessed).
10. Davidson S. More planes land in Canada with COVID-19-infected passengers. Here's where the flights came from. CTV News. 2020. https://toronto.ctvnews.ca/more-planes-land-in-canada-with-covid-19-infected-passengers-here-s-where-the-flights-came-from-1.5076714 (1 September 2020, date last accessed).
11. Tableau Server. Border Crossing Entry Data. 2020. https://explore.dot.gov/views/BorderCrossingData/Monthly?:isGuestRedirectFromVizportal=y&:embed=y (30 August 2020, date last accessed).
12. Government of Ontario. Tourism research. 2020. http://www.mtc-currentperformance.com (18 Sep 2020, date last accessed).
13. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19). 2020. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html (30 July 2020, date last accessed).
14. National Post Staff. Hundreds of thousands of U.S. visitors are still crossing border into Canada each week. National Post News. 2020. https://nationalpost.com/news/hundreds-of-thousands-of-u-s-visitors-are-still-crossing-border-into-canada-each-week (22 July 2020, date last accessed).
15. Government of Canada. Coronavirus disease (COVID-19): Travel restrictions, exemptions and advice. 2020. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/latest-travel-health-advice.html (29 July 2020, date last accessed).
16. Dickens BL, Koo JR, Lim JT, et al. Strategies at points of entry to reduce importation risk of COVID-19 cases and reopen travel. J Travel Med 2020; taaa141.
17. Linka K, Rahman P, Goriely A, Kuhl E. Is it safe to lift COVID-19 travel bans? The Newfoundland story. Comput Mech 2020; 66: 1081-1092.
18. Hu Z, Cui Q, Han J, et al. 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: 231-240.

19. Huang Q, Kang YS. Mathematical Modeling of COVID-19 Control and Prevention Based on Immigration Population Data in China: Model Development and Validation. JMIR Public Health Surveill 2020; 6: e18638.
20. Sun T, Weng D. Estimating the effects of asymptomatic and imported patients on COVID-19 epidemic using mathematical modeling. J Med Virol 2020; 92: 1995-2003.
21. Hossain MP, Junus A, Zhu X, et al. The effects of border control and quarantine measures on the spread of COVID-19. Epidemics 2020; 32: 100397.
22. Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020; 368: 395-400.
23. Kraemer MG, Yang CY, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020; 368: 493-497.
24. Linka K, Peirlink M, Costabal FS, Ellen K. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. Comput Methods Biomech Biomed Engin 2020; 23: 710-717.
25. Song W, Zang P, Ding Z, et al. Massive migration promotes the early spread of COVID-19 in China: a study based on a scale-free network. Infect Dis Poverty 2020; 9:109.
26. Gao S, Rao J, Kang Y, et al. Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US. JAMA Netw Open 2020; 3: e2020485.
27. Yuan P, Li J, Aruffo E, et al. Efficacy of 'Stay-at-Home' Policy and Transmission of COVID-19 in Toronto, Canada: A Mathematical Modeling Study. 2020. Available at SSRN: https://doi.org/10.2139/ssrn.3678581
28. City of Toronto. COVID-19. 2020. https://www.toronto.ca/home/covid-19/ (31 Aug 2020, date last accessed).
29. Statistics Canada. Number of vehicles travelling between Canada and the United States. 2020. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2410000201 (22 July 22 2020, date last accessed).
30. Government of Canada. Coronavirus disease (COVID-19). 2020. https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html (3 August 2020, date last accessed).
31. Government of Canada. Travel restriction to Canada remain in place for September long weekend. 2020. https://www.canada.ca/en/border-services-agency/news/2020/08/travel-restrictions-to-canada-remain-in-place-for-september-long-weekend.html (30 Aug 2020, date last accessed).
32. Cori A, Ferguson NM, Fraser C, et al. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 2013; 178: 1505-1512.
33. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020; 382: 1199–207.
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. https://www.R-project.org/ (30 July 2020, date last accessed).
35. Gordon O, Goverde M, Pazdan J, et al. Comparison of different calculation approaches for defining microbiological control levels based on historical data. PDA J Pharm Sci Technol 2015; 69: 383-398.

Optimal Staged Reopening Schedule Based on ICU Capacity: A Model-Informed Strategy

Kyeongah Nah, Michael Chen, Ali Asgary, Zachary McCarthy, Francesca Scarabel, Yanyu Xiao, Nicola Luigi Bragazzi, Jane M. Heffernan, Nicholas H. Ogden, and Jianhong Wu

1 Introduction and Background

Reopening plans and strategies have been developed and revised in every country around the world as part of the COVID-19 pandemic recovery process. These plans and strategies are based on some basic principles and golden rules focusing on the importance of existing measures such as increasing access to diagnostic testing, contact tracing, and access to PPE (personal protective equipment), among others [1]. While the COVID-19 reopening will be a staged process with staggered reopening of activities in order to minimize risk, reopening details vary from one country/region to another, depending on their unique demographic, economic,

K. Nah · Z. McCarthy · N. L. Bragazzi
Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

M. Chen · J. Wu (✉)
Department of Mathematics and Statistics, York University, Toronto, ON, Canada
e-mail: wujh@yorku.ca

A. Asgary
Disaster and Emergency Management, School of Administrative Studies and Advanced Disaster and Emergency Rapid-Response Simulation (ADERSIM), York University, Toronto, ON, Canada

F. Scarabel
Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

CDLab—Computational Dynamics Laboratory, Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy

Y. Xiao
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, USA

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_15

health care system, and dynamics of disease transmission. Reopening has been occurring under much uncertainty about many factors including the availability and effectiveness of vaccines/treatment drugs [2, 3]. Thus, it is very challenging for decision makers to determine the timing and level of relaxation which should be considered for their jurisdictional circumstances [4, 5]. Therefore, mathematical modelling and statistical and optimization techniques are needed to reduce this uncertainty and assist decision makers. To address these challenges, we propose a modelling framework that can be used to examine and assess the impacts of different reopening strategies, and find optimum, calculated, and feasible solutions [6, 7]. The outcomes are based on the expected access and arrival time of an efficacious vaccine/treatment drug for SARS-CoV-2.

Ontario, with more than 14 million people, is the largest Canadian province and was among the first provinces to have confirmed COVID-19 cases, starting in late January 2020, from international travelers. While monitoring the situation during the month of February of 2020, the province started to implement a number of pandemic response measures, with public school closures in March 12 of 2020 (escalation phase 1), and a declaration of the state of emergency in March 17 of 2020 including cancelation of public events, places and establishments (escalation phase 2), followed by closure of non-essential workplaces on March 24 of 2020, and by more strict measures including the prohibition of gatherings greater than five people on March 25 of 2020 (escalation phase 3).

On April 27 of 2020, the government of Ontario issued its framework for reopening the province after going through several escalation phases for nearly 2 months [8]. The reopening framework envisioned three incremental and gradual stages in which moving from one stage to the next will depend on the success and outcomes of the current stage, which can be evaluated based on the impacts of the reopening measures against the number of predicted cases admitted in ICU. This reopening framework was based on six key principles of responsibility, evidence-informed, resourced, monitored, responsive and effective, and clear communication. In particular, the framework emphasized the adequacy of health system capacity to respond to resurgence in COVID-19 infections and efficiency of implemented measures. Accordingly, easement of public health measures in response to the COVID-19 were assessed based on four criteria, including virus spread and containment, health system capacity, public health system capacity, and incidence tracking capacity. The main goal of this study was to assess Ontario's reopening

J. M. Heffernan
Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

Modelling Infection and Immunity Lab, Centre for Disease Modelling, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

N. H. Ogden
Public Health Risk Sciences Division, National Microbiology Laboratory, St-Hyacinthe/Guelph Public Health Agency of Canada, Guelph, Canada

plans by taking into consideration the capacity of the health system based on the province's disease transmission model that incorporates these reopening criteria. More specifically, we developed and applied an optimization technique to the disease transmission model which maximizes the relaxation of social contacts while remaining within the health system's capacity, considering an expected time for SARS-CoV-2 vaccine/treatment drug availability.

2 Materials and Methods

2.1 The Transmission Dynamics Model

We use the parameterized model describing COVID-19 transmission dynamics in Ontario, Canada used in previous studies [9, 10]. The population is divided into susceptible (S), exposed (E), asymptomatic infectious (A), infectious with symptoms (I), and recovered (R) compartments according to the epidemiological status of individuals. We also include diagnosed and isolated (D), isolated susceptible (S_q), and isolated exposed (E_q) compartments based on control interventions. Within the modelling framework, we also account for contact tracing, where a proportion, q, of individuals exposed to the virus are traced and isolated (we will also refer to q as 'quarantine proportion' or 'quarantine fraction'). The quarantined individuals can either move to the compartment E_q or S_q, depending on whether transmission occurred (with probability β), while the other proportion, $1 - q$, consists of individuals exposed to the virus who are missed from contact tracing and, therefore, move to the exposed compartment E once infected, or stay in the compartment S otherwise. The transmission dynamics model is

$$S' = -\left(\beta c + cq\left(1-\beta\right)\right) S\left(I + \theta A\right)/N + \lambda S_q,$$

$$E' = \beta c\left(1-q\right) S\left(I + \theta A\right)/N - \sigma E,$$

$$I' = \sigma \varrho E - \left(\delta_I + \alpha + \gamma_I\right) I,$$

$$A' = \sigma\left(1-\varrho\right) E - \gamma_A A,$$

$${S_q}' = \left(1-\beta\right) cq S\left(I + \theta A\right)/N - \lambda S_q,$$

$${E_q}' = \beta cq S\left(I + \theta A\right)/N - \delta_q E_q,$$

$$D' = \delta_I I + \delta_q E_q - \left(\alpha + \gamma_D\right) D,$$

$$R' = \gamma_I I + \gamma_A A + \gamma_D D,$$

where N denotes the total population in Ontario, Canada. In order to better capture the different phases of the escalation of intervention measures, the contact rate and quarantine proportion are assumed to be time-dependent. Specifically, after March 2 of 2020, the contact rate $c(t)$ and quarantine proportion $q(t)$ are exponentially decreasing or increasing with exponential rates r_1 and r_2, respectively. That is, by introducing T_0, T_1 and T_s corresponding to March 14, March 18 and March 24 of 2020, respectively, we have

$$c(t) = \begin{cases} c_0, & t < T_0, \\ c_1, & t < T_1, \\ c_2, & t < T_s, \\ (c_2 - c_b)\, e^{-r_1(t-T_s)} + c_b, & t \geq T_s, \end{cases}$$

and

$$q(t) = \begin{cases} q_0, & t < T_s \\ (q_0 - q_b)\, e^{-r_2(t-T_s)} + q_b, & t \geq T_s, \end{cases}$$

where c_0, c_1, c_2 are the constant contact rates and q_0 is the quarantine proportion before March 24 of 2020. The parameters c_b and q_b denote the minimum contact rate and the maximum quarantine proportion, respectively, estimated after March 24 of 2020. A full parameterization of the transmission model for each phase of the escalation process in Ontario, Canada has been proposed in [11]. Table 1 contains the parameter definitions and estimated values from [11] (Fig. 1).

2.2 Division of the Diagnosed Class

We further divide the class $D(t)$, the number of diagnosed but not-yet-recovered population, as follows:

$$D = D_{mild} + D_{icu} + D_{ward},$$

distinguishing the confirmed cases into those who show only mild symptoms and are not hospitalized (D_{mild}),are hospitalized in ICU with severe symptoms (D_{ICU}),and are hospitalized in non-ICU units (D_{ward}). We describe each of the populations by

$$\begin{aligned} D'_{mild} &= (1-h)\left(\delta_I I + \delta_q E_q\right) - (\alpha + \gamma_{mild})\, D_{mild}, \\ D'_{icu} &= h\,(1-w)\left(\delta_I I + \delta_q E_q\right) + b_{ward} D_{ward} - (\alpha + \gamma_{icu})\, D_{icu}, \\ D'_{ward} &= hw\left(\delta_I I + \delta_q E_q\right) - (\alpha + \gamma_{ward} + b_{ward})\, D_{ward}, \end{aligned}$$

where h is the proportion of hospitalized cases among newly confirmed cases and w is the proportion which are hospitalized in non-ICU units. The values of h and w are estimated to be 0.16 and 0.77, respectively [12]. The recovery rates for individuals in each states are $\gamma_{mild} = 1/5$, $\gamma_{icu} = 1/14$ and $\gamma_{ward} = 1/12$ [13–15]. For the following simulations, the initial conditions given as $D_{mild}(0) = 4$, $D_{icu}(0) = 0$ and $D_{ward}(0) = 1$. The value of b_{ward} is estimated to be 0.26*1/3 which is the product of

Table 1 Parameter estimates for COVID-19 in Ontario, Canada, taken from [11]

Parameter		**Definitions**	**Mean (Std)**
c_0		Contact rate before March 14	11.5801 (0.3456)
c_1		Contact rate between March 14 to March 18	10.1202 (0.9185)
c_2		Contact rate between March 18 to March 24	8.0495 (0.2787)
$c(t)$	c_2	Constant contact rate at March 24	8.0495 (0.2787)
	r_1	Exponential decrease of contact rate	0.0466 (0.0152)
	c_b	Minimum contact rate after March 24	2.1987 (0.2400)
β		Probability of transmission per contact	0.1469 (0.0023)
q_0		Fraction of quarantined exposed individuals before March 24	0.1145 (0.0114)
$q(t)$	q_0	Quarantine fraction at March 24	0.1145 (0.0114)
	r_2	Exponential increase of quarantine fraction	0.1230 (0.0123)
	q_b	The maximum quarantine fraction	0.3721 (0.0371)
σ		Transition rate of exposed individuals to the infected class	1/5
λ		Rate at which the quarantined uninfected contacts were released into the wider community	1/14
ϱ		Probability of having symptoms among infected individuals	0.7036 (0.0261)
δ_I		Transition rate of symptomatic infected individuals to the quarantined infected class	0.1344 (0.0134)
δ_q		Transition rate of quarantined exposed individuals to the quarantined infected class	0.1237 (0.0086)
γ_I		Recovery rate of symptomatic infected individuals	0.1957 (0.0111)
γ_A		Recovery rate of asymptomatic infected individuals	0.139
γ_D		Recovery rate of quarantined diagnosed individuals	0.2
α		Disease-induced death rate	0.008
θ		Modification factor of asymptomatic infectiousness	0.0275 (0.0128)
Initial values		**Definitions**	**Mean (Std)**
$S(0)$		Initial susceptible population	1.471×10^7
$E(0)$		Initial exposed population	8.9743 (0.6558)
$I(0)$		Initial symptomatic infected population	5.3887 (0.9442)
$A(0)$		Initial asymptomatic infected population	19.4186 (3.9406)
$S_q(0)$		Initial quarantined susceptible population	0
$E_q(0)$		Initial quarantined exposed population	0
$D(0)$		Initial quarantined diagnosed population	5
$R(0)$		Initial recovered population	0

the two quantities: proportion of patients who are moving to ICU beds among those who are originally admitted to non-ICU units, 0.26 [16], and the average duration of pre-ICU period of 3 days [16].

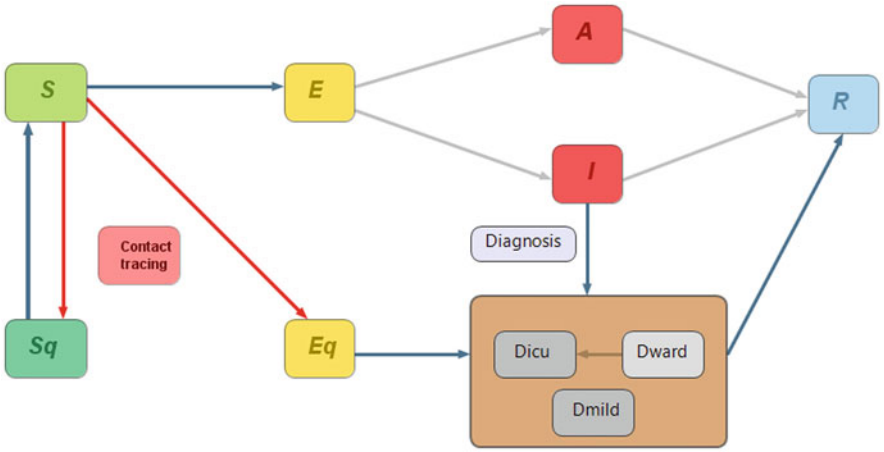


Fig. 1 The flowchart of the transmission dynamics model, where the population is stratified by susceptible (S), exposed (E), asymptomatic infectious (A), symptomatic infectious (I), and recovered status (R), and by quarantine and isolation status (i.e. Sq and Eq). The compartment of diagnosed but not yet resolved cases is further stratified by the need for the use of hospital wards (Dward) and/or ICU beds (Dicu), and mild (Dmild)

2.3 De-escalation Considerations

We investigate de-escalation using the same model. A simple de-escalation plan is to simply 'reverse' the escalation steps, as studied in [11]. There, three main de-escalation phases were considered:

De-escalation-Phase 1: Opening of workplaces;

De-escalation-Phase 2: Resumption of public events and activities;

De-escalation-Phase 3: School opening.

In [11], authors focused on the effect of different parameters on the control reproduction number R_c, defined as the average number of secondary infections that one primary infected individual would produce in a fully susceptible population under control measures, with the goal of investigating which parameter combinations, and public health intervention measures, would allow to avoid a rebound of the epidemic (by achieving $R_c < 1$).

Here we relax this requirement by allowing the reproduction number to be above 1 during the de-escalation phase, hence allowing for a second wave, but constrained so that the critical cases will not exceed the health system capacity. More precisely, we study the optimal initiation time of each de-escalation phase in order to keep the number of diagnosed cases below a certain threshold while we gradually retrieve the pre-intervention contact rate.

Herein, a de-escalation strategy refers to a set of initiation timings corresponding to each de-escalation phase. An acceptable de-escalation strategy refers to the strategy which makes the number of diagnosed cases below a certain threshold. This threshold is set by the health system capacity in the considered area (which

is Ontario, Canada, in the present study). We compute the cost of a de-escalation strategy by quantifying the reduction in contacts until we return to the pre-pandemic contact level. We then find the optimal acceptable de-escalation strategy, depending on various public health intervention scenarios (in terms of the improvement of contact tracing, diagnosis, and recommendations for PPE usage) put in place after re-opening, and on different time-windows of cost-evaluation. We start by evaluating the three-phase de-escalation process outlined above, with contact rates estimated in the escalation phases [11]. We then evaluate a more gradual de-escalation process of the workplace resumption divided into five subphases.

2.4 *Feasible De-escalation Plans Based on the Health System Capacity*

We consider four time-points, t_0, t_1, t_2 and t_3, each of which corresponds to the beginning of cost-evaluation (May 20, 2020), de-escalation-phase 1, de-escalation-phase 2 and de-escalation-phase 3. The three latter time-points are to be optimally selected.

Each de-escalation phase is characterized by a contact rate, with $c_{r,0}$ denoting the contact rate before reopening, and $c_{r,1}$,$c_{r,2}$ and $c_{r,3}$ denoting the (increasing) contact rates in each de-escalation phase (see Table 2). The contact rate $c_{r,3}$ is assumed to be at the value estimated before interventions, hence it corresponds to a full resumption of social contacts and activities.

We consider the acceptable de-escalation strategies which meet the constraints $D_{icu}(t) \leq \bar{D}_{icu}$ for $t_1 < t < t_0 + T$, where, $t_0 + T$ is the time corresponding to the end of cost-evaluation and $\bar{D}_{icu}$ is a capacity of the ICU beds available for COVID-19 patients. The total number of existing ICU and acute beds in Ontario, Canada as of April 16 is 3504 and 20,354, respectively [17]. We assume that only a portion of these resources is available for COVID-19 patients.

Table 2 Parameters considered in the reopening plans

Parameter	Definitions	Lower bounds	Upper bounds
T	Days of cost-evaluation (counting from May 20, 2020)	163 (Nov 1, 2020)	28 (Mar 1, 2021)
β_r	Transmission probability after $t = t_1$	0.13	0.1469
q_r	Quarantine proportion after $t = t_1$	0.3721	0.5
$\delta_{I,r}$	Detection rate after $t = t_1$	0.1344	0.25
$c_{r,0}$	Contact rate before de-escalation	3.759	3.759
$c_{r,1}$	Contact rate during de-escalation-phase 1	8.0495	8.0495
$c_{r,2}$	Contact rate during de-escalation-phase 2	10.1202	10.1202
$c_{r,3}$	Contact rate during de-escalation-phase 3	11.5801	11.5801

A de-escalation strategy refers to the vector $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ determined by

$$\varepsilon_i = t_i - t_{i-1},$$

for $i = 1, 2, 3$. Here, ε_1 represents the length of pre-reopening period starting from t_0 (the beginning of cost-evaluation), ε_2 and $\varepsilon_3$are the duration of the de-escalation phases 1 and 2, respectively. The initiation time of de-escalation phase 3 is at the time $t_3 = t_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$.

We solve the following optimization problem of minimizing the intensity of reduced contacts (cost)

$$\sum_{i=1}^{3} \left(c_{r,3} - c_{r,i-1}\right) \varepsilon_i,$$

with constraints

$$\max_{t_1 < t < t_0 + T} D_{icu}(t) \leq \underline{D}_{icu},$$

$$\varepsilon_1 + \varepsilon_2 + \varepsilon_3 \leq T,$$

and

$$\varepsilon_i \geq 0$$

for $i = 1, 2, 3$. We further restrict the values of ε_1, ε_2, ε_3 to be integer multiples of 14 days (including 0 day) considering the minimum time unit required for the implementation of each de-escalation strategy.

The range of each parameter used for scenario analysis is presented in Table 2. The days of cost-evaluation are set so that the end of the evaluation time corresponds to the beginning, middle and end of the flu season (i.e., November 2020, January 2021, March of 2021). The underlying assumption is that, after re-opening, we would enforce other control measures while we relax the regulations resulting in the increase of the contact rate. Therefore, the upper bound of the transmission probability and lower bounds of the quarantine proportion and detection rates are set to be at the level of pre-reopening time presented in Table 1. Lower bound of the transmission probability is obtained from the result of the similar modeling study [18] in China, setting this level of PPE as a maximum level we can achieve. The lower bound for the quarantine proportion is set at the level of q_b=0.3721, the estimated value of the maximum fraction of the quarantine in Table 1. The upper bound for the detection rate is set to be 0.25, assuming that 4 days is the minimum length of time required for testing and case-confirmation.

The minimum level of contact rate before reopening is set to be 3.759, calculated using weighted contact matrices from the POLYMOD project with no contacts in schools, workplaces and communities while the contacts in households increased 20% due to provincial emergency orders since March 17 of 2020 in Ontario, Canada [19–21]. The minimum level of contact rate is achieved prior to the beginning of cost-evaluation (earlier than May 20) and therefore remained to be constant at the level of 3.759 till the first reopening date.

3 Results

The optimal and acceptable de-escalation strategy depends on various public health intervention scenarios in terms of improvement of contact tracing (increase of the quarantine proportion, q), diagnosis (increase of the diagnosis rate, δ_I), and recommendations for PPE usage (decrease of the transmission probability, β) put in place after re-opening ($t \geq t_1$), as well as the length of cost-evaluation (T).

3.1 Optimal De-escalation Strategy Under Different Cost Evaluation Periods

Figure 2 shows the optimal de-escalation strategies and the resulting numbers of cases in ICU with three distinguished lengths of cost-evaluation (T) assuming that 50% of ICU beds are available for COVID-19 patients and that public health measures (such as contact tracing, diagnosis and personal protection) are kept at high levels. In the three different scenarios, we set the end of the evaluation period at March 2021, January 2021 and November 2020.

As we can see from the top panel of Fig. 2, the timing of acceptable reopening date is delayed when the duration of cost-evaluation (T) is prolonged, indicating that the plan for the early reopening should be in line with the early release of vaccines/drugs. In all cases, the optimal strategies allow the immediate initiation of de-escalation phase 1 and then transition to de-escalation phase 3 without the implementation of de-escalation phase 2.

In the following section, we compare the optimal de-escalation strategies with the three specific scenarios on the time of vaccine availability, at the beginning, middle, and at the end of influenza season.

Scenario 1: With the early release of vaccines in November 2020 (T = 163), the optimal strategy is to initiate the de-escalation phase 1 immediately and then transition to phase 3 in early October. With this strategy, the rebound will lead to exceeding the ICU beds around early November.

Scenario 2: When T = 224 (January 2021), the optimal strategy is to initiate de-escalation phase 1 immediately, and then transition to de-escalation phase 3 in mid December. With scenario 2, the rebound will lead to exceeding the ICU beds around early January, 2021.

Scenario 3: When T = 283 (March 2021), the optimal strategy is to initiate de-escalation phase 1 immediately, and then transition to de-escalation phase 3 in late February. With scenario 3, the rebound will lead to exceeding the ICU beds around March, 2021.

The bottom-right panel shows the percentage of cases in ward beds. While the number of cases in ICU reaches near to its limit at the end of cost-evaluation (time of vaccine availability), the number of cases in ward beds will remain at less than 5% of the total ward beds by the end of cost evaluation.

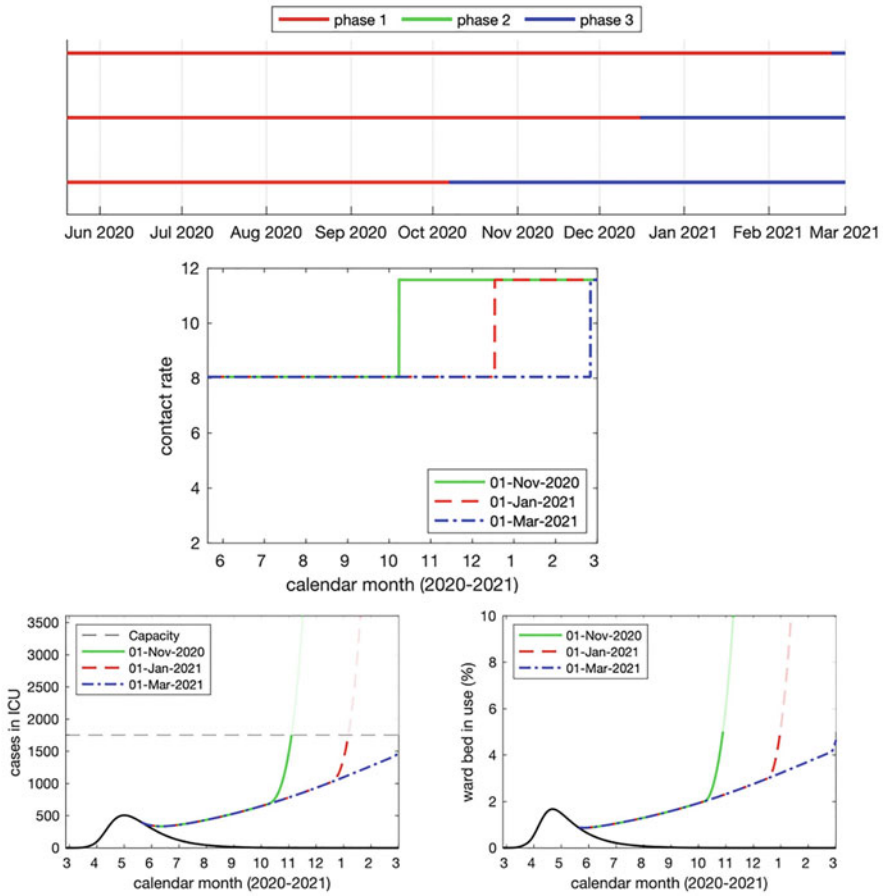


Fig. 2 Optimal de-escalation strategies under three different cost evaluation periods. Top panel shows the timing of optimal de-escalation phases with vaccine availability in March 2021 (top line), January 2021 (middle line) and November 2020 (bottom line). Middle panel shows the contact rate at the optimal strategy. Bottom panels show the predicted number of confirmed cases in ICU, $D_{ICU}(t)$, (left) and the predicted percentage of ward beds occupied by COVID-19 patients, $D_{ward}(t)$, (right). Here, D_{icu}=1752, assuming that 50% of the existing ICU beds are available for COVID-19 patients. The results are obtained at the minimum level of quarantine proportion($q = 0.3721$), maximum level of diagnosis rate ($\delta_I = 0.25$), and the minimum level of transmission probability ($\beta = 0.13$) within their respective ranges in Table 2. Note that in the case of vaccine availability (mass vaccination) in March of 2021, Ontario should remain in stage 2 almost all the time until late February 2021

3.2 *Optimal De-escalation Strategy Under Different Hospital Bed Capacity*

Given the same parameters and the three cost-evaluation periods used in the previous section, Fig. 3 presents optimal de-escalation strategies with different ICU bed capacities available for COVID-19 patients. We can see that the plan for the early reopening and de-escalation should be in line with the improved numbers of ICU beds as well as the early release of vaccines.

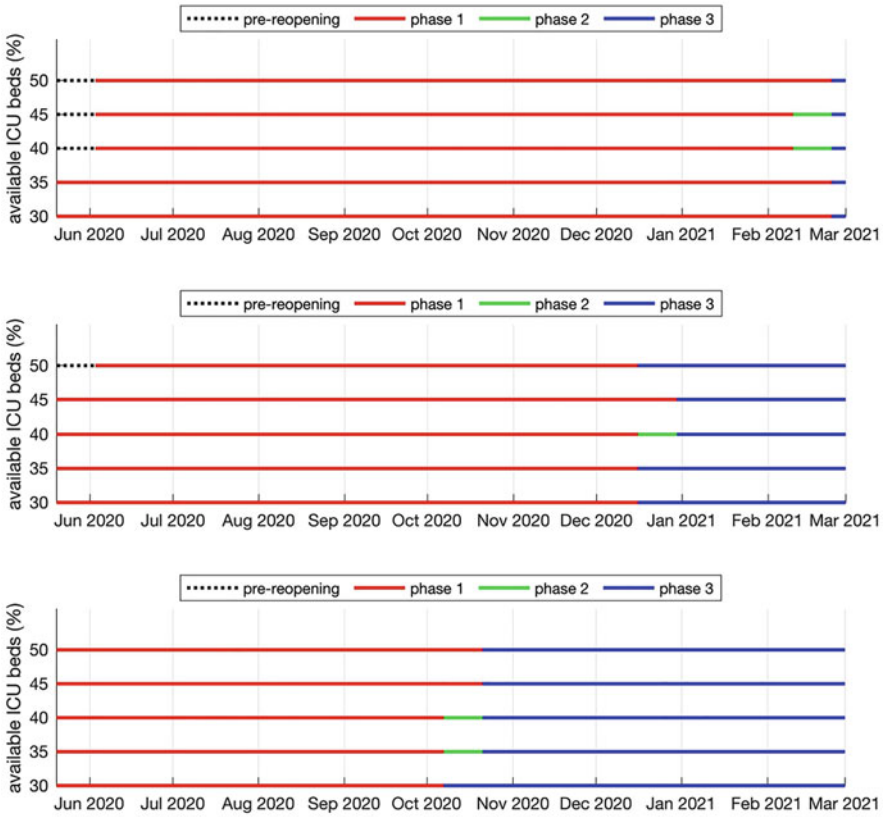


Fig. 3 Optimal de-escalation strategies under different ICU bed capacity. Each of the lines shows the timing of optimal de-escalation phases under different % of ICU beds available for COVID-19 patients. Each panel is with vaccine availability in March 2021 (top panel), January 2021 (middle panel) and November 2020 (bottom panel). The results are obtained at the minimum level of quarantine proportion($q = 0.3721$), maximum level of diagnosis rate ($\delta_I = 0.25$), and the minimum level of transmission probability ($\beta = 0.13$) within the range in Table 2

3.3 *Optimal De-escalation Strategy Under Various Intervention Scenarios*

In Table 3 and Figs. 4 and 5, we present the optimal de-escalation strategies obtained at different levels of quarantine proportion(q), diagnosis rate (δ_I), by fixing the

Table 3 Optimal de-escalation strategies depending on the percentage of contacts traced and quarantined (q) and the percentage of symptomatic diagnosed, ($\delta_I/(\delta_I+\alpha+\gamma_I)$)

	% contacts traced and quarantined		
% symptomatic diagnosed	37%	44%	50%
40% (δ_I= 0.1344)	Oct 7–Oct 21–Nov 4	Aug 12–Dec 2–Dec 16	Jul 1–Dec 2–Dec 16
49% (δ_I= 0.1922)	Aug 12–Nov 18–Dec 2	Jun 3–NA–Dec16	May 20–Oct 21–Nov 4
55% (δ_I=0.25)	May 20–NA–Dec 16	May 20–Oct 7–Nov 4	May 20–Jul 15–Sep 23

Note: The red and yellow cells correspond to the high and low cost. Corresponding initiation timing of each phase are presented in the cells. The transmission probability (β) is fixed to be 0.13. Here, it is assumed that the vaccines are available on Jan 1, 2021 (T = 235) andD_{icu}=1752. For optimal de-escalation strategies which skip a phase, NA appears to identify this case

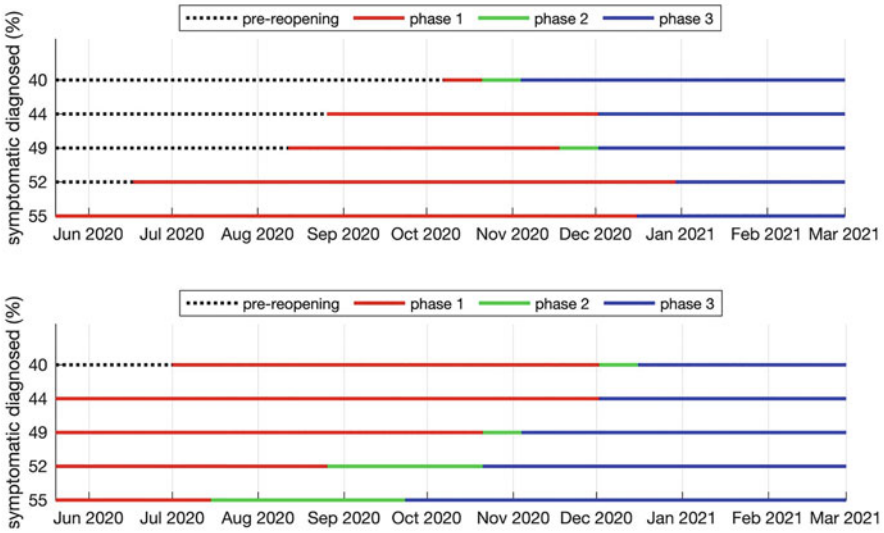


Fig. 4 Optimal de-escalation strategies under different percentages of symptomatic diagnosed with minimum level of quarantine proportion, $q = 0.3721$, (upper panel) and maximum level of quarantine proportion, $q = 0.5$(lower panel). Here, $\delta_I = 0.1344$, $\beta = 0.13$, T = 235 (Jan 01, 2021) and D_{icu}=1752

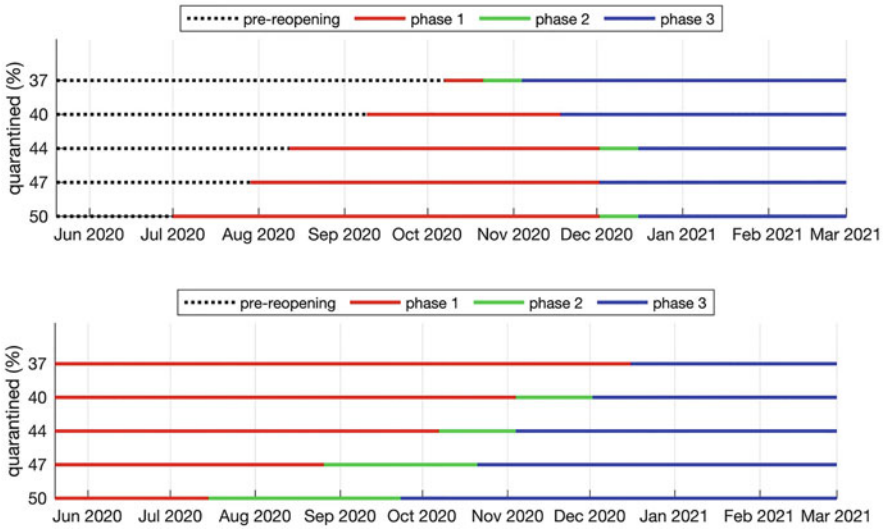


Fig. 5 Optimal de-escalation strategies under different percentages of contacts traced (q) with minimum level of diagnosis rate, $\delta_I = 0.1344$, (upper panel) and maximum level of diagnosis rate, $\delta_I = 0.25$(lower panel). Here,$\beta = 0.13$, T = 235 (January 01, 2021) andD_{icu}=1752

transmission probability (β), end of the cost evaluation period and the fraction of ICU beds available.

Sensitivity to diagnosis capacity (see Table 3, Fig. 4): while the maximum diagnosis rate allows for an immediate reopening (May 20 of 2020), a lower (and likely more realistic) diagnosis rates of 0.134 (corresponding to a time to diagnosis of about 7.5 days) at the minimal level of contact tracing (37%), requires to delay the reopening by almost 5 months, to the beginning of October, 2020. Re-opening on May 20, 2020 is allowed only by almost unrealistic values of diagnosis rate (corresponding to the average time from symptom onset to diagnosis of 4–5 days). In the lower panel of Fig. 4, we can see that the higher diagnosis capacity is related with the earlier reopening (the date of initiating de-escalation phase 1) when the quarantine proportion is high. With low quarantine proportion, the higher diagnosis capacity is related.

Sensitivity to quarantine proportion (see Fig. 5): With maximum level of diagnosis rate ($\delta_I = 0.25$), the higher quarantine proportion is related with the earlier initiation of phase 2 and 3. When the quarantine proportion is also at the maximal level (50%), de-escalation phase 3 can follow immediately after the end of de-escalation phase 1. With minimum level of diagnosis rate ($\delta_I = 0.1344$), the higher quarantine proportion is related to the earlier reopening.

3.4 Optimal De-escalation Strategy with Gradual Implementation of De-escalation Phase 1

In the following, we consider the optimization problem with five sub-stages for the de-escalation phase 1 (workplace resumption):

$$\min \sum_{i=1}^{7} \left(c_{r,7} - c_{r,i-1}\right)\ \varepsilon_i,$$

with constraints

$$\max_{t_1<t<t_0+T} D_{icu}(t) \leq \underline{D}_{icu},$$

$$\sum_{i=1}^{7} \varepsilon_i \leq T,$$

$$\varepsilon_i \geq 0 \qquad \text{(and)}$$

for $i = 1,...7$, where $D_{ICU}(t)$ is the solution of the transmission dynamics model with initial conditions in Table 1 and $D_{mild}(0) = 4$, $D_{icu}(0) = 0$ and $D_{ward}(0) = 1$. The contact rates corresponding to each of the de-escalation sub-stages for phase 1 and other phases are set to be $c_{r,1} = 4.6171$, $c_{r,2} = 5.4752$ $c_{r,3} = 6.3333$, $c_{r,4} = 7.1914$, $c_{r,5} = 8.0495$, $c_{r,6} = 10.1202$ and $c_{r,7} = 11.5801$. The contact rates are obtained by dividing the increase in contact rate due to workforce into five substages with equal increase of contacts per day. Using weighted contact matrices for the workplaces, we calculated that the first four contact levels corresponding roughly to 30%, 57%, 72%, 87% and 98% workforce resumption in phase 1; the last two contact rates are estimated for phase 2 (resumption of public events and activities) and phase 3 (school reopening).

Figure 6 presents the results. In all of the cases, the optimal strategies allow the immediate relaxation of the contact level to be at 4.6171 (first sub-stage of de-escalation phase 1), which is obtained by ~30% of workforce resumption. In the following, we compare optimal de-escalation strategies with three distinguished scenarios regarding the time of vaccine availability or improved number of ICU beds.

Scenario 1: With the early release of vaccines or improved numbers of ICU beds in November of 2020 (T = 163), the optimal strategy is to initiate the fourth sub-stage of de-escalation phase 1, corresponding to 86–87% resumption of workforces, in mid July of 2020. The initiation of de-escalation phase 3 is at the end of July of 2020.

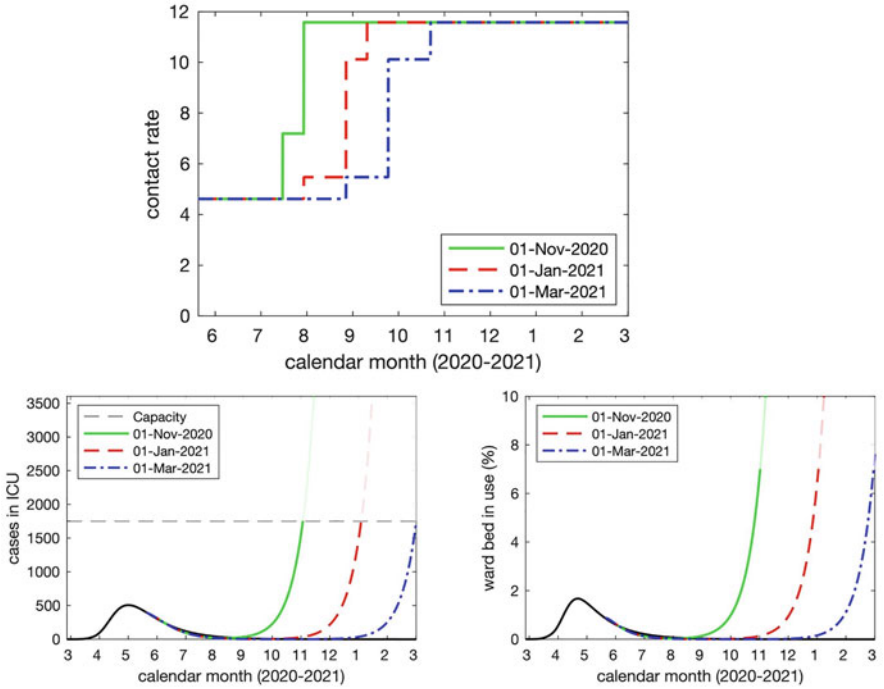


Fig. 6 Optimal de-escalation strategies under three different cost evaluation periods. The upper panel shows the contact rate at the optimal strategy with vaccine availability in March of 2021, January of 2021 and November of 2020. Bottom panels show the predicted number of confirmed cases in ICU, $D_{icu}(t)$, (left) and the predicted number of ward beds occupied by Covid-19 patients, $D_{ward}(t)$, (right). Here, D_{icu}=1752, $q = 0.3721$, $\delta_I = 0.25$and $\beta = 0.13$

Scenario 2: With the early release of vaccines or improved numbers of ICU beds in January of 2021 (T = 224), the optimal strategy is to initiate the second sub-stage of de-escalation phase 1 at the end of July of 2020, corresponding to 56–57% resumption of workforces. The initiation of de-escalation phase 2 is at the end of August of 2020 and the initiation of de-escalation phase 3 should be at the beginning of September of 2020.

Scenario 3: With the early release of vaccines or improved numbers of ICU beds in March of 2021 (T = 283), the optimal strategy is to initiate the second sub-stage of de-escalation phase 1 at the end of August of 2020, corresponding to 56–57% resumption of workforces. The initiation of de-escalation phase 2 is at the end of September of 2020 and the initiation of de-escalation phase 3 should be at the end of October of 2020.

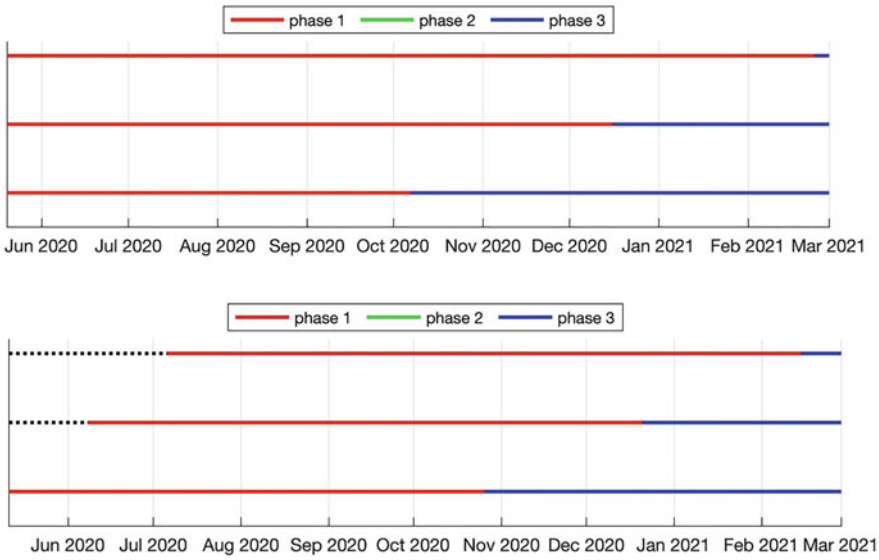


Fig. 7 Optimal de-escalation strategies under two different pre-reopening contact rates. Each of the panels shows the timing of optimal de-escalation phases with pre-reopening contact rates $c_{r,0} = 3.759$ (top) and $c_{r,0} = 4.6171$ (bottom), respectively. The three lines in the panels correspond with vaccine availability in March of 2021 (top line), January of 2021 (middle line) and November of 2020 (bottom line). Here, D_{icu}=1752, $q = 0.3721$, $\delta_I = 0.25$and $\beta = 0.13$

3.5 Sensitivity of the Optimal De-escalation Strategy to the Pre-reopening Contact Rate

As of May 19, 2020, the government of Ontario announced reopening of some businesses. This may have lifted the contact rate of pre-reopening stages. In Fig. 7, we observe that the optimal de-escalation timings are not highly sensitive to the different levels of pre-reopening contact rates, although higher contact rate during the pre-reopening stage may cause a slightly delayed initiation of phase 1 and phase 3.

4 Discussion and Conclusion

Using a transmission dynamics model parameterized through model fitting to cumulative incidence data during different social distancing escalation phases [11], we identified the optimal timing for social-distancing de-escalation for Ontario, Canada as a case study. The timing of acceptable reopening date is delayed when the expected date of vaccine/drug availability is prolonged. While ICU beds in

use reach its limit by the end of cost-evaluations, less than 10% of ward beds are needed. Increasing the capacity of ICU beds available for COVID-19 patients and enforcing other measures (diagnosis, PPE, contact tracing) would allow earlier reopening dates and therefore reduce the number of social contacts lost during the social-distancing period. Specifically, our simulations show that improving contact tracing and diagnosis capacity has a significant effect on the reopening date (Figs. 4 and 5), whereas increasing ICU beds capacity has only a minor effect (Fig. 3).

While using a different disease transmission and optimization method, our findings are in line with other optimization and de-escalation studies such as those conducted by Rawson et al. [22] in the UK and German et al. [23] in Germany. These studies also emphasize the importance of gradual reopening and implementation of various public health measures and constraints such as social distancing, antibody testing, isolation of infectious individuals and contact tracing.

In our study, we have only considered the de-escalation scenario of reversing steps implemented during the escalation phases. However, the actual reopening strategy may be more gradual. The Ontario government was taking a gradual reopening approach for the reopening of workplaces and public places. Starting from May 4 of 2020, businesses which could meet public health guidelines started to reopen, including seasonal business and construction projects. Although we have considered the scenario of gradual workplace reopening by having sub-phases of de-escalation phase 1, the contact rates set up for our sub-phases may not match with the contact rates during the actual gradual reopening of businesses. Once we obtain more accurate estimates of the contact rates during those small steps, the optimal reopening plans onwards can be re-evaluated. In Sect. 3.4, we explored the optimal control problem with fixed and equal increase of contacts per day for five sub-stages of de-escalation phase 1. Based on our algorithm, we can also conduct the optimization with different sub-stages assigned by various contact rates corresponding to increased levels of workforce resumption, such 25%, 50%, 75%, 90% and 100%.

Note that in our cost function, the number of contacts lost during three de-escalation phases have equal weights and we did not consider the difference in economic impact the number of social contacts lost have during workplace closure/public event closure/school closure. Also, we assumed that the contact rates are constant during de-escalation phases. Therefore, the resulting optimal timing should be carefully interpreted considering the delay between the approval of reopening and actual implementation.

We expect that the reopening plan would take place with constant monitoring of the risk indicators (e.g. incidence), and therefore the reopening plans may change over time responding to the situations. As a next step, we need to build a model which incorporates the adaptive reopening strategies by identifying the risk indicators and criteria which form the adaptive strategy.

In the study, we establish a functional and flexible framework for determining optimal de-escalation strategies capable of performing key scenario analyses to inform decision-making. To accomplish this, we have utilized our previously established modelling studies in Ontario, Canada [9, 11]. We note that the estab-

lished framework is amenable to expansion and modification to address specific research questions at hand. Further, the systematic process presented of model-fitting, scenario analysis and interpretation may be used as a reference to inform optimal de-escalation strategies in other regions. More broadly, this study highlights the opportunity for mathematical model-based techniques to play a pivotal role in the management, risk assessment, and control of COVID-19 globally.

As a final remark added when the paper is being processed for publication in January of 2021, we note that the Province of Ontario has unfortunately experienced a second outbreak started in the fall of 2020, and the Province entered a full lockdown on December 26 of 2020, although a gradual mass vaccination is indeed expected to start in March of 2021. This scenario of vaccine availability in March of 2021 was considered in our simulations, and our model-based simulation suggested optimal timing for the Province to have advanced to phase 2 and phase 3 social distance de-escalation should be August of 2020 and October of 2020, respectively. The Province, however, decided to start its region (public health unit)-specific staged reopening process much earlier: the Toronto Public Health Unit, for example, started its phase 2 (officially called Stage 2) and phase 3 on June 24 and July 31, 2020 respectively. This pre-mature advancing to staged reopening process was halted indeed due to the ICU capacity limitation. As a result, the Province reverted the reopening process and the Toronto Public Health Unit moved back to the modified Stage -2 on October 10, 2020, and further back to the Red Zone (Control) on November 7, 2020, and then Grey Zone (Lockdown) on November 24, 2020. The issue of optimal timing for reopening unfortunately arises again with an evolving vaccine availability timeline, so our methodology could potentially be tested.

Acknowledgements This project has been funded by the CIHR 2019 Novel Coronavirus (COVID-19) rapid research program.

References

1. Crespo Márquez, A., & SIM Research Group. (2020). Understanding R-based planning decisions and control for COVID-19 recovery.
2. Kopecky, K., & Zha, T. (2020). *Impacts of COVID-19: Mitigation Efforts versus Herd Immunity* (No. 2020-3). Federal Reserve Bank of Atlanta.
3. Omer SB, Malani P, del Rio C. The COVID-19 Pandemic in the US: A Clinical Update. *JAMA*. 2020;323(18):1767–1768. doi:https://doi.org/10.1001/jama.2020.5788
4. Agarwal, A., Alomar, A., Sarker, A., Shah, D., Shen, D., & Yang, C. (2020). Two Burning Questions on COVID-19: Did shutting down the economy help? Can we (partially) reopen the economy without risking the second wave?. *arXiv preprint arXiv:2005.00072*.
5. Dingel, Jonathan I. and Patterson, Christina and Vavra, Joseph, Childcare Obligations Will Constrain Many Workers When Reopening the US Economy (April 18, 2020). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-46. Available at SSRN: https://ssrn.com/abstract=3579711 or https://doi.org/10.2139/ssrn.3579711

6. Osmar Pinto Neto, Josa Clark Reis, Ana Carolina Brisola Brizzi, Gustavo Jose Zambrano, Joabe Marcos de Souza, Wellington Pedroso E. Amorim, Rodrigo Cunha de Mello Pedreiro, Bruno de Matos Brizzi, Ellysson Oliveira Abinader, Deanna M. Kennedy, Renato A Zangaro, medRxiv 2020.04.26.20081208; doi: https://doi.org/10.1101/2020.04.26.20081208.
7. Kraay, A. N. M., Nelson, K., Zhao, C., Weitz, J. S., & Lopman, B. A. (2020). Modeling serological testing to inform relaxation of social distancing for COVID-19 control. *medRxiv*.
8. Government of Ontario, 2020, A Framework for Reopening our Province. https://files.ontario.ca/mof-framework-for-reopening-our-province-en-2020-04-27.pdf
9. Barrett, K., Khan, Y. A., Mac, S., Ximenes, R., Naimark, D. M., & Sander, B. (2020). Potential magnitude of COVID-19-induced healthcare resource depletion in Ontario, Canada. *Canada (3/25/2020)*.
10. Tang et al, Quantifying the Role of Social Distancing, Personal Protection and Case Detection in Mitigating COVID-19 Outbreak in Ontario, Canada
11. Tang et al, De-escalation by Reversing the Escalation with a Stronger Synergistic Package of Contact Tracing, Quarantine, Isolation and Personal protection: Feasibility of Preventing a COVDI-19 Rebound in Ontario, Canada, as a Case Study, Biology
12. PHAC Corona Virus Disease 2019 (Covid-19) Daily Epidemiology Update (Updated: 11 May 2020, 11:00 ET)
13. Sanche S, Lin YT, Xu C et al., The novel coronavirus, 2019-nCoV, is highly contagious and more infectious than initially estimated. *medRxiv* 2020 Feb 11 [Epub ahead of print]. https://doi.org/10.1101/2020.02.07.20021154.
14. Moghadas SM, Shoukat A, Fitzpatrick MC et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc Natl Acad Sci U S A*. 2020 Apr 3 [Epub ahead of print]. https://doi.org/10.1073/pnas.2004064117
15. Shoukat A, Wells CR, Langley JM, Singer BH, Galvani AP, Moghadas SM. Projecting demand for critical care beds during COVID-19 outbreaks in Canada. CMAJ. 2020 Jan 1.
16. Tuite, A. R., Fisman, D. N., & Greer, A. L. (2020). Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. *CMAJ*.
17. Government of Ontario, Ontario Significantly Expands Hospital Capacity to Prepare for Any COVID-19 Outbreak Scenario, April 16, 2020, https://news.ontario.ca/opo/en/2020/04/ontario-significantly-expands-hospital-capacity-to-prepare-for-any-covid-19-outbreak-scenario.html (last access: May 19, 2020)
18. B. Tang, F. Xia, N.L. Bragazzi, Z. McCarthy, X. Wang, S. He, X. Sun, S. Tang, Y. Xiao, J. Wu. Lessons drawn from China and South Korea for managing COVID-19 epidemic: insights from a comparative modeling study. Bull World Health Organ (2020) https://doi.org/10.2471/blt.20.257238.
19. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS computational biology. 2017;13(9):e1005697.
20. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine. 2008;5(3).
21. Government of Ontario, Ontario Enacts Declaration of Emergency to Protect the Public, https://news.ontario.ca/opo/en/2020/03/ontario-enacts-declaration-of-emergency-to-protect-the-public.html (Last access: May 23, 2020)
22. Rawson, T., Brewer,T., Veltcheva, D., Huntingford, C., and Bonsall, M.B., 2020. How and when to end the COVID-19 lockdown: an optimisation approach, medRxiv https://doi.org/10.1101/2020.04.29.20084517.
23. German, R., Djanatliev, A., Maile, L., Bazan, P. & Hackstein, H. Modeling Exit Strategies from COVID-19 Lockdown 431 with a Focus on Antibody Tests. medRxiv (2020).
24. Hethcote HW, Waltman P. Optimal vaccination schedules in a deterministic epidemic model. Mathematical Biosciences. 1973;18(3–4):365-81.

Mathematics of the Pandemic

M. Ram Murty and V. Kumar Murty

1 Introduction

Mathematics offers a framework for tackling the coronavirus pandemic that is now confronting the human race. We will show below how basic mathematics already gives us a viable method to control the spread of the virus. More advanced techniques can be used to make predictions and plot future trajectories. This, of course, does not preclude the essential aspect of pharmaceutical research needed to develop a vaccine. It however offers us a method of social behaviour essential to thwart the spread of the disease until a vaccine is developed.

This paper is largely expository and can be dubbed "basic epidemiology for (pure) mathematicians." Our main references are the book by Bailey [1] and Chapter 21 of [10]. Though this paper is a condensed exegesis of known results, we do however make some new remarks concerning the approximate solution by Kermack and McKendrick [16] of their "SIR" model to study epidemics. A study of this approximation and how it falls short of the reality has been discussed in several papers such as [15] and [12]. One of our main results is that a small correction term in the Kermack–McKendrick solution improves the error term in their model.

Daniel Bernoulli [5] was the first to propose a mathematical model to study the spread of smallpox back in 1760. But the idea of using mathematics to study

Research of both authors is partially supported by NSERC Discovery grants.

M. Ram Murty (✉)
Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada
e-mail: murty@queensu.ca

V. Kumar Murty
Fields Institute for Research in Mathematical Sciences, Toronto, ON, Canada
e-mail: director@fields.utoronto.ca; murty@math.toronto.edu

V. K. Murty, J. Wu (eds.), *Mathematics of Public Health*, Fields Institute Communications 85, https://doi.org/10.1007/978-3-030-85053-1_16

the spread of diseases did not quite catch on (pardon the pun). The simplest mathematical model to study the transmission of disease is called the SI model, where there is no recovery. It was proposed by Hamer [13] in 1906 and then expanded by Ross [21] in 1915. Later, Kermack and McKendrick [16] in 1927 proposed the celebrated SIR model which accomodates recovery from the infection. (Incidentally, Ross's work on malaria won him the 1902 Nobel Prize and a knighthood in 1911.) By the middle of the twentieth century, the mathematics of epidemics evolved into a serious subject and the book by Bailey [1] is considered a primary source.

In this paper, we will examine the SIR model. After surveying the paper by Kermack and McKendrick, we examine the question of error in the approximate solution in [16]. Then, we discuss the topic of "exact solutions" mainly from a combinatorial perspective via an application of the Lagrange inversion theorem. This leads to the introduction of the important role played by the Lambert W-function. We derive at the end an assortment of results that are also of number theoretic interest. In particular, we give two new integral expressions for the Lambert function for positive real arguments.

In the simplest model, we have a *branching process*. This means that each infected individual meets k other people and infects every individual he meets with probability p. Thus, the number of new cases generated by a single individual is $R_0 = pk$, which is called *basic reproductive number*. The infected people go on to infect other people in the same way and we can model this using graph theory by a *tree* structure. It is then clear that if we let q_n be the number of infected people at the n-th level of this tree, then $q_n = (pk)^n$. Since we want q_n to go to zero, we derive our first theorem.

Theorem 1 *The number of infected people will eventually tend to zero if* $pk < 1$.

This simple result already gives us a powerful method to deal with the pandemic. First, we need to know p. That is obtained by collecting data, which is very important and this means extensive testing of the population. The probability p is approximated by the ratio of the number of infected people by the total number of people tested. Of course, a total lockdown of the entire population will substantially diminish the spread of the disease. However, this is not practical in the long run and Theorem 1 tells us that we can relax this and curtail the number of people an individual can come into contact with. That number is $[1/p]$, where $[\cdot]$ is the greatest integer function. We record this as:

Corollary 2 *The number of people each individual can come into contact with should be less than* $[1/p]$ *in order to contain the pandemic.*

Our simple model, though expedient in giving us a quick understanding of the pandemic and its spread, is not an accurate reflection of the real state of things. For instance, there is no single value of p that can be ascribed to an entire nation since often, the value of p changes from region to region. Therefore, it is necessary to view p as a function of two co-ordinates x, y giving the geographical position. The

same can be said of the quantity k. Thus, if $\mathcal{R}$ is the region we are interested in, the number of people infected at the n-th level in the region $\mathcal{R}$ is

$$\iint_{\mathcal{R}} p(x, y)^n k(x, y)^n dxdy.$$

Suppose that

$$f(x, y) = \lim_{n\to\infty} p(x, y)^n k(x, y)^n$$

exists almost everywhere. This is a reasonable assumption since our functions are locally constant functions. As the sequence of functions

$$f_n(x, y) = p(x, y)^n k(x, y)^n$$

is bounded by the total population of the region, an application of Lebesgue's dominated convergence theorem (see page 26 of [22]) shows that if we want

$$\lim_{n\to\infty} \iint_{\mathcal{R}} p(x, y)^n k(x, y)^n dxdy = 0,$$

then $f(x, y) = 0$ almost everywhere. In other words, we must have $p(x, y)k(x, y) < 1$ almost everywhere. We thus arrive at our second theorem.

Theorem 3 *The pandemic is contained if $p(x, y)k(x, y) < 1$ almost everywhere in every part of the region.*

In order to formulate a practical public policy, it is then prudent to consider these localized probabilities and functions.

There are several models that describe the spread of epidemics. Broadly, they can be grouped into two categories: discrete and continuous with respect to the time parameter. Our theorems above deal with the discrete time model. Much of this paper is devoted to the continuous time model which uses the theory of differential equations. By contrast, the discrete model uses graph theory. In this model, each person of the population is represented by a vertex. A vertex can be in one of two states: susceptible or infected. A directed edge from node i to node j means that i can infect j. The rate of infection β is attached to each edge and the rate of recovery γ is attached to each infected node. This results in a graph G and one defines the epidemic threshold of the graph G as the value τ such that if $\beta/\gamma < \tau$ then the epidemic dies over time and if $\beta/\gamma > \tau$, the epidemic spreads over time. In [25], the authors show that if λ_1 is the largest eigenvalue of the adjacency matrix of G, then the epidemc threshold of G is $1/\lambda_1$. This gives an interesting connection between spectral graph theory and the study of epidemics. We refer the reader to [9, 25] and the survey [6] for further details.

2 The SI Model

The SI model to study epidemics is discussed in Chapter 5 of Bailey [1]. We give a brief exposition. At time $t = 0$, we consider a population of size $N + a$, with N "susceptible" and a persons "infected". Let $S(t)$ be the number of susceptible people and $I(t)$ the number of infected people at time t respectively. We assume that the infection rate is β and that

$$\frac{dS(t)}{dt} = -\beta S(t) I(t).$$

As we assume the population is constant throughout the epidemic, we have $S(t) + I(t) = N + a$ so that our differential equation reduces to

$$\frac{dS(t)}{dt} = -\beta S(t)(N + a - S(t)),$$

which is easily solved using basic calculus. We have

$$S(t) = \frac{N(N + a)}{N + ae^{(N+a)\beta t}}.$$

Thus, as $S(t) + I(t) = N + a$, we have

$$\frac{dI(t)}{dt} = -\frac{dS(t)}{dt} = \frac{aN(N + a)^2 e^{(N+a)\beta t}}{(N + ae^{(N+a)\beta t})^2},$$

often called the *epidemic curve* since it gives the rate at which infections occur. It is not difficult to see that this attains its maximum at time

$$t = \frac{\log(N + a)}{\beta(N + a)}.$$

If β is very large, the peak will be reached very early in the time period. Though this model is simple, it leads to some interesting probability theory for which we refer the reader to Chapter 5 of [1].

3 The SIR Model

To understand the spread of a virus in a community, Kermack and McKendrick formulated in 1927 the so-called SIR model. (See in particular equation (29) of [16].) This is one of the basic models that mathematical biology uses to study epidemics. Other models are variations on this theme. There are three quantities that need to be studied as a function of time. The first is the number of "susceptible"

people who can contract the disease. This is denoted $S(t)$ as a function of the time parameter t. The next is the number of infected people at time t, denoted $I(t)$. The third is the number of "recovered" people $R(t)$. This explains the acronym 'SIR.'

These quantities are inter-related. We make the following assumptions.

$$S(t) + I(t) + R(t) = N, \tag{1}$$

where N is the total population of the region under consideration. In other words, we assume our population of the region is constant and does not increase with time t. We assume that a proportion γ of the infected people will recover. (Sadly, the word 'recover' may also euphemistically include deaths.) Thus,

$$\frac{dR(t)}{dt} = \gamma\, I(t). \tag{2}$$

The number of new infections is proportional to the number of interactions between susceptible people and the infected people and so, taking into account the recovery rate, we see that the differential equation encoding this fact is

$$\frac{dI(t)}{dt} = \beta\, S(t)I(t) - \gamma\, I(t), \tag{3}$$

where β is the infection rate. Finally, the number of susceptible people satisfies

$$\frac{dS(t)}{dt} = -\beta S(t)I(t). \tag{4}$$

These equations together with

$$\frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} = 0,$$

obtained by differentiating (1) give us a system of three differential equations govering the spread of the pandemic. This can also be deduced simply by adding up (4), (3) and (2). In essence, we actually have two differential equations. Let us record this in the following system describing the SIR model:

$$\begin{cases} \dfrac{dS(t)}{dt} = & -\beta S(t)I(t) \\ \dfrac{dI(t)}{dt} = \beta\, S(t)I(t) - & \gamma\, I(t) \\ \dfrac{dR(t)}{dt} = & \gamma\, I(t). \end{cases} \tag{5}$$

Fortunately, this system can be simplified as follows. Dividing (4) by (2) and using the chain rule, we get

$$\frac{dS}{dR} = -\frac{\beta}{\gamma} S,$$

so that

$$-\frac{\beta}{\gamma} R(t) = \log(S(t)/S(0))$$

since at time $t = 0$, we have $R(0) = 0$. In other words, $S(t) = S(0)\exp(-\frac{\beta}{\gamma} R(t))$. Inserting this back into (2) we obtain

$$\frac{dR}{dt} = \gamma \left(N - R(t) - S(0)\exp(-\frac{\beta}{\gamma} R(t)) \right),$$

which is the basic differential equation governing the behaviour of the epidemic.

The ratio β/γ is often denoted R_0 in the literature[1] and is called the *basic reproduction number* or *basic reproduction ratio* and it is not unrelated to the concept introduced in the earlier section. The reciprocal $\rho = 1/R_0$ is called the *removal rate* by Bailey [1]. The rate at which infections are increasing is given by (3) and we can re-write it as

$$\frac{dI(t)}{dt} = \gamma[R_0 S(t) - 1]I(t),$$

from which we note that as $S(t) \leq S(0)$, the function $I'(t)$ is decreasing if $R_0 S(0) < 1$. We therefore see that the epidemic can start only when $S(0) > 1/R_0 = \rho$. Thus, the reproduction ratio measures the spread of the epidemic.

Using the notation of R_0, our differential equation for R now becomes

$$\frac{dR}{dt} = \gamma \left(N - R - S(0)e^{-R_0 R} \right). \tag{6}$$

This differential equation has no closed form solution for R and one approximates the exponential function by its Taylor series. Thus, the equation studied becomes

$$\frac{dR}{dt} = \gamma \left(N - S(0) + (R_0 S(0) - 1)R - (R_0^2 S(0)/2)R^2 \right). \tag{7}$$

This is a special case of the generalized Riccati differrential equation.

[1] This is unfortunate notation since R_0 has nothing to do with the R function. Bailey [1] denotes this as $1/\rho$ and this is a better notation which we also use sporadically, whenever convenient.

We could have equivalently divided (4) by (3) and obtained a relation between $I(t)$ and $S(t)$ but this arrangement is simpler and follows [16]. For the sake of clarity, we rewrite (7) as

$$\frac{dR}{dt} = \gamma(A + BR + CR^2), \quad \text{with} \quad \begin{cases} A = N - S(0), \\ B = R_0 S(0) - 1, \\ C = -R_0^2 S(0)/2. \end{cases} \tag{8}$$

Let us observe that the discriminant of this quadratic is

$$(R_0 S(0) - 1)^2 + 2(N - S(0))R_0^2 S(0) > 0. \tag{9}$$

Now we insert the following lemma from first year calculus (which we leave as an exercise to the reader):

Lemma 4 *Let* $\Delta = B^2 - 4AC > 0$. *Then*

$$\int \frac{dR}{A + BR + CR^2} = -\frac{2}{\sqrt{\Delta}} \tanh^{-1}\left(\frac{2CR + B}{\sqrt{\Delta}}\right).$$

Thus, (8) is solved by

$$\frac{2}{\sqrt{\Delta}} \tanh^{-1}\left(\frac{2CR + B}{\sqrt{\Delta}}\right) = -\gamma t + \phi \tag{10}$$

where Δ is given by (9), and

$$\phi = \frac{2}{\sqrt{\Delta}} \tanh^{-1}\left(\frac{B}{\sqrt{\Delta}}\right),$$

is the integration constant determined by $R(0) = 0$. Hence, we deduce that

$$R(t) = \frac{-B + \sqrt{\Delta} \tanh\left(-\frac{\sqrt{\Delta}}{2}(\gamma t - \phi)\right)}{2C}.$$

As tanh is an odd function, we can simplify this to

$$R(t) = \frac{R_0 S(0) - 1 + \sqrt{\Delta} \tanh\left(\frac{\sqrt{\Delta}}{2}(\gamma t - \phi)\right)}{R_0^2 S(0)},$$

which agrees with formula (30) in the Kermack–McKendrick paper [16].

We record the solution of this in the following theorem and interpret the results.

Theorem 5 *The solution of (8) is given by*

$$R(t) = \frac{R_0 S(0) - 1 + \sqrt{\Delta}\tanh\left(\frac{\sqrt{\Delta}}{2}(\gamma t - \phi)\right)}{R_0^2 S(0)},$$

where

$$\phi = \frac{2}{\sqrt{\Delta}}\tanh^{-1}\left(\frac{R_0 S(0) - 1}{\sqrt{\Delta}}\right),$$

and

$$\Delta = (R_0 S(0) - 1)^2 + 2(N - S(0))R_0^2 N^2 S(0) > 0.$$

Since we are interested in the infection rate which is the derivative of $R(t)$, we find easily: (there is a typo in [16], $\sqrt{-q}$ in formula (31) should be $-q$. This seems to have been corrected in the book by Bailey [1] on page 83.)

Theorem 6

$$I(t) = \frac{\Delta}{2R_0^2 S(0)} sech^2\left(\frac{\sqrt{\Delta}}{2}(\gamma t - \phi)\right).$$

The function appearing in the above theorem is often called the *epidemic curve*. It is a symmetrical bell-shaped curve that reflects what is often seen in epidemics, namely that new cases continue to rise until they hit a certain peak point and then slowly reduce in number. Let us note that this model does not take into account variable rates of infection or removal. One can, for instance, drastically reduce the rate of infection by following the suggested guidelines of personal hygiene, mask wearing and social distancing with or without a vaccine.

To determine the maximum number of infected people, we can differentiate this function representing $I(t)$ and note that it attains its maximum value of

$$\frac{\Delta}{2R_0^2 S(0)} \quad \text{at} \quad t_0 = \frac{\phi}{\gamma}.$$

We also see from Theorem 5 that

$$R(\infty) = \frac{\rho^2}{S(0)}\left(S(0)R_0 - 1 + \sqrt{\Delta}\right).$$

At the beginning of the outbreak, we expect N and $S(0)$ to be very close, and so we can approximate $\sqrt{\Delta}$ by $S(0)R_0 - 1$. Thus,

$$R(\infty) \approx 2\rho\left(1 - \frac{\rho}{S(0)}\right).$$

How good is this approximation? Using Taylor's theorem with error term, it is possible to write down an approximation of the error incurred by using only the quadratic term. We can also analyse if the approximation has any serious repercussions with respect to making predictions. This we do in the next section.

4 Error Terms

We will now analyze the error terms incurred in the use of the quadratic approximation of the exponential function in the previous section. This will also give us insight into an understanding of using higher polynomial approximations and the errors that arise in that context.

Let us write

$$e^{-R_0 R} = 1 - R_0 R + \frac{1}{2}R_0^2 R^2 + E$$

so that in fact, the difference

$$\int_0^t \frac{dR}{(N - R - S(0)e^{-R_0 R})} - \int_0^t \frac{dR}{(N - S(0) + (R_0 S(0) - 1)R - \frac{1}{2}R_0^2 S(0)R^2)}$$

is equal to

$$\int_0^t \frac{S(0)E\,dR}{(N - R - S(0)e^{-R_0 R})(N - S(0) + (R_0 S(0) - 1)R - \frac{1}{2}R_0^2 S(0)R^2)}. \tag{11}$$

By Taylor's theorem with approximation (see for example, Theorem 1.18 on pages 36–37 of [17]), we have

$$|E| = R_0^3 e^{-R_0 \xi}\frac{R^3}{3!},$$

for some $\xi \in [0, t]$. As the function $e^{-R_0 r}$ is decreasing, we have

$$|E| \leq R_0^3 \frac{R^3}{3!}.$$

Thus, the error (11) is bounded by $O(\log t)$ where the implied constant depends on N and $S(0)$.

If on the other hand, we use a cubic polynomial approximation to $e^{-R_0 R}$, then the resulting error integral is of the form

$$\int_0^t \frac{E_3(R)dR}{f(R)}$$

where $E_3(R) = O(R^4)$ and $f(R)$ is a polynomial in R of degree 6. Thus, the integral is convergent and we can write it as

$$\int_0^\infty \frac{E_3(R)dR}{f(R)} - \int_t^\infty \frac{E_3(R)dR}{f(R)}$$

which is of the form

$$\text{constant} \quad + O\left(\frac{1}{t}\right).$$

One can also use a quartic approximation to $e^{-R_0 R}$ and this leads to a similar conclusion with the error being of the form

$$\text{constant} \quad + O\left(\frac{1}{t^2}\right).$$

In the last two cases, therefore, the error is "negligible" in the sense that as t goes to infinity, there is no significant increase in the term we would obtain if we use either a cubic or quartic approximation to the exponential function.

Since

$$\int_0^t \frac{dR}{A+BR+CR^2} - \int_0^t \frac{dR}{A+BR+CR^2+DR^3} =$$

$$\int_0^\infty \frac{DR^3 dR}{(A+BR+CR^2)(A+BR+CR^2+DR^3)} + O\left(\frac{1}{t}\right),$$

the integral on the right hand side is a convergent integral. In other words, (10) changes to

$$\frac{2}{\sqrt{\Delta}} \tanh^{-1}\left(\frac{2CR+B}{\sqrt{\Delta}}\right) = -\gamma t + \phi + \phi' + O\left(\frac{1}{t}\right),$$

where ϕ' is a suitable constant alluded to above. We state this formally as:

Theorem 7 *The exact solution for $R(t)$ in the SIR model (5) satisfies*

$$R(t) = \frac{R_0 S(0) - 1 + \sqrt{\Delta}\tanh\left(\frac{\sqrt{\Delta}}{2}(\gamma t - \phi - \phi' + O\left(\frac{1}{t}\right))\right)}{R_0^2 S(0)},$$

where

$$\phi = \frac{2}{\sqrt{\Delta}}\tanh^{-1}\left(\frac{R_0 S(0) - 1}{\sqrt{\Delta}}\right),$$

and

$$\Delta = (R_0 S(0) - 1)^2 + 2(N - S(0))R_0^2 N^2 S(0) > 0.$$

5 Kendall's Exact Solution

The implication of the deliberations of the previous section to the result of Kermack and McKendrick (sometimes called K and K in the literature) is that one need only adjust the constant ϕ to correct the error incurred in using the quadratic approximation to $e^{-R_0 R}$. We underline this observation in light of a paper written by Kendall [15] in 1956, where he shows that the use of the quadratic approximation "consistently underestimates the infection rate" (see page 151 of [15]). Referring to the approximation we have recorded in Theorem 6 above, he adds that "it is curious that the K and K approximation should have been accepted without comment for nearly 30 years; the exact solution is easily obtained and the difference between the two can be of practical significance."

Kendall uses a calculus of variations argument to show that the K and K approximation underestimates the rate of infection β. For the sake of clarity, we outline his argument following [15] and [1]. Suppose that β is a function of R. Then, proceeding as we did before we can solve for $S(t)$:

$$S(t) = S(0)\exp\left(-\frac{1}{\gamma}\int_0^R \beta(r)dr\right),$$

which leads to as earlier

$$\frac{dR(t)}{dt} = \gamma\left(N - R - S(0)\exp\left(-\frac{1}{\gamma}\int_0^R \beta(r)dr\right)\right).$$

The function

$$\beta(R) = \frac{2\beta}{(1 - R_0 R) + (1 - R_0 R)^{-1}}$$

leads to the K and K approximation. Thus, $\beta(0) = \beta$ and $\beta(R) < \beta$ for $0 < R < 1/R_0$ and this is the basis of Kendall's statement that the K and K model underestimates the infection rate β. He adds that for $R > 1/R_0$, the model gives a negative infection rate. Because of these objections, he suggests a renormalization using the exact solution.

The "exact" solution alluded to is simply the transcendental function obtained by integrating (6). It is surprising that it was (perhaps independently) re-discovered by the authors in [14] in 2014. Neither Kendall nor the classic text book by Bailey which discusses Kendall's work are mentioned in the references of [14].

Kendall's approach can be described as follows. Using basic calculus, it is easily seen

$$N - R - S(0)e^{-R_0 R} = 0$$

has exactly two real roots, one negative and one positive, denoted $-\eta_1$ and η_2 (using the notation of page 85 of [1]). Thus,

$$\gamma t = \int_0^R \frac{dr}{N - r - S(0)e^{-R_0 r}}, \qquad 0 < R < \eta_2. \tag{12}$$

The integral diverges for $R \to \eta_2$ and so, $R(\infty) = \eta_2$. But the integral also diverges if $S(0) = N$ which suggests that there is an infinite amount of time before the epidemic starts, which is absurd. This absurdity is resolved by changing the origin to the point where $S = 1/R_0$ which is referred to as the center of the epidemic. The peak of the epidemic occurs when $I(t)$ reaches its maximum, which is when $I'(t) = 0$. Since

$$\gamma I'(t) = \frac{d}{dt}\left(\frac{dR}{dt}\right) = \gamma^2 I(t)\,(R_0 S(t) - 1)\,,$$

we see that the peak of the epidemic curve occurs when $S(t) = 1/R_0$ which is the center of the epidemic defined above. From (5), the maximum number of infections also occurs at the same point. This leads to the parametric solution described in the following 1956 theorem due to Kendall [15].

Theorem 8 *The parametric solution of the SIR model (5) is:*

$$t = \frac{1}{\gamma}\int_0^R \frac{dr}{I(0) - r + R_0^{-1}(1 - e^{-rR_0})},$$

$$\frac{dR}{dt} = \gamma(I(0) - r + R_0^{-1}(1 - e^{-R_0 r})),$$

where $-\infty < t < \infty$ *and* $-\zeta_1 < r < \zeta_2$, *with* $-\zeta_1$ *and* ζ_2 *being the unique negative and positive roots of*

$$I(0) - \zeta + R_0^{-1}(1 - e^{-R_0\zeta}) = 0. \tag{13}$$

The meaning of the roots $-\zeta_1$ and ζ_2 is that $\zeta_1 + \zeta_2$ is the total number of recovered people during the entire pandemic. In fact, ζ_1 is the number before the peak and ζ_2 is the number after the peak. It is therefore of some interest to determine these roots with some accuracy. This we do in the next section using the Lagrange inversion formula and the Lambert W-function. We study roots of exponential equations such as (13) from a general perspective.

6 An Application of the Lagrange Inversion Formula

If we want to determine the extremal points of $R(t)$, we need only look at when the right hand side of Eq. (6) vanishes. In other words, we want to solve for R in the equation

$$N - R - S(0)e^{-R_0 R} = 0.$$

Lagrange's inversion formula states that if we have a functional relation $f(r) = t$ with f being analytic at 0 and $f'(0) \neq 0$, then we can invert and write r as a power series in t. The precise formula is

$$r(t) = r(0) + \sum_{n=1}^{\infty} c_n t^n$$

where

$$c_n = \frac{1}{n!} D^{n-1}\left(\phi(t)^n\right)\Big|_{t=0}, \tag{14}$$

where $D = d/dt$ and $\phi(t) = 1/r'(t)$.

We give a simple proof of the Lagrange inversion formula using the Cauchy residue theorem, and then apply it to derive the needed facts about the Lambert W-function.

Given a formal power series

$$f(t) = \sum_{n=-\infty}^{\infty} c_n t^n,$$

we use the notation

$$[t^n](f(t)) := c_n.$$

Thus, for instance $[t^n](f(t)) = [t^{n+m}](t^m f(t))$, a fact we will use later in the proof of the proposition below.

Proposition 9 *(Lagrange Inversion Formula) Suppose $f(z)$ is analytic in a neighborhood of $z = 0$ with $f(0) = 0$ and $f'(0) \neq 0$. Then f^{-1} is analytic in a neighborhood of $z = 0$ and*

$$[z^n](f^{-1}(z)) = [z^{n-1}]\left(\frac{z^n}{nf(z)^n}\right).$$

Proof Since $f'(0) \neq 0$, we have by the inverse function theorem (see page 62 of [17]) that $f^{-1}(z)$ is well-defined and analytic in a neighborhood of $f(0) = 0$. Consequently, it has a power series expansion in a neighborhood of zero. By the Cauchy residue theorem,

$$[z^n](f^{-1}(z)) = \frac{1}{2\pi i}\int_C \frac{f^{-1}(w)dw}{w^{n+1}},$$

with C being a sufficiently small circle centered at zero. We change variables in the integral by setting $w = f(v)$ which is a conformal map if C is of sufficiently small radius. Since $f^{-1}(f(v)) = v$, we have

$$[z^n](f^{-1}(z)) = \frac{1}{2\pi i}\int_{C'} \frac{vf'(v)dv}{f(v)^{n+1}},$$

where C' is the closed contour image of C under our conformal mapping. Our integral can be re-written as

$$-\frac{1}{2\pi in}\int_{C'} vd\left(\frac{1}{f(v)^n}\right) = \frac{1}{2\pi in}\int_{C'}\left(\frac{1}{f(v)^n}\right)dv,$$

on integrating by parts and noting that the residue of $d(v/f(v)^n)$ at $v = 0$ is zero. Thus, by the Cauchy residue theorem, we conclude that

$$[z^n](f^{-1}(z)) = [z^{-1}]\left(\frac{1}{nf(z)^n}\right) = [z^{n-1}]\left(\frac{z^n}{nf(z)^n}\right),$$

the last equality being clear by shifting the power series appropriately. □

Remark 10 This elegant proof is due to Whittaker and Watson [26]. The analyticity assumption in the proposition is a red herring and can be dispensed with using the theory of formal power series. A proof centered on these ideas can be found in Chapter 5 of [24]. It is straightforward to see that Proposition 9 gives (14).

It is not difficult to see that the above proof can be suitably modified to yield the following more general result.

Proposition 11 *Suppose $f(z)$ is analytic in a neighborhood of $z = 0$ with $f(0) = 0$ and $f'(0) \neq 0$. Then f^{-1} is analytic in a neighborhood of $z = 0$ and for each $1 \leq k \leq n$, we have*

$$[z^n](f^{-1}(z))^k = [z^{n-k}]\left(\frac{kz^n}{nf(z)^n}\right).$$

Another version of the same theorem is often useful in applications. To this end, we introduce the *Bell polynomials* which are defined as follows. Suppose that

$$f(z) = \sum_{n=1}^{\infty} c_n \frac{z^n}{n!}$$

Then

$$\exp(f(z)) = \sum_{n=0}^{\infty} B_n(c_1, c_2, ..., c_n)\frac{z^n}{n!}. \tag{15}$$

Thus, the so-called n-th *complete Bell polynomial* is given by

$$B_n(c_1, ..., c_n) = \frac{\partial^n}{\partial z^n} \exp\left(\sum_{n=1}^{\infty} c_n \frac{z^n}{n!}\right)\Big|_{z=0}.$$

The Lagrange inversion formula can also be stated in terms of the Bell polynomials (see page 151 of [7]).

One more variation of the Lagrange inversion formula is given in [19]. To present it, we let

$$a_1 z + a_2 z^2 + \cdots = w$$

be a power series which converges in a neighborhood of $z = 0$. If $a_1 \neq 0$, this gives a conformal mapping of a sufficiently small disc centered at $z = 0$ onto a neighborhood of $w = 0$. Thus, we can write

$$z = b_1 w + b_2 w^2 + \cdots .$$

If we let

$$\phi(z) = \left(a_1 + a_2 z + a_3 z^2 + \cdots\right)^{-1}$$

then $w = z/\phi(z)$.

Proposition 12 *Let f be any analytic function. With ϕ and w as above, we have*

$$\frac{f(z)}{1 - w\phi'(z)} = \sum_{n=0}^{\infty} \frac{w^n}{n!} \frac{d^n}{dx^n} \left(f(x)\phi(x)^n\right)\Big|_{x=0}.$$

Proof For C a circle enclosing $\zeta = 0$ and oriented counterclockwise, we have

$$\frac{1}{n!} \frac{d^n}{dx^n} \left(f(x)\phi(x)^n\right)\Big|_{x=0} = \frac{1}{2\pi i} \int_C \frac{f(\zeta)\phi(\zeta)^n d\zeta}{\zeta^{n+1}},$$

by Cauchy's formula. Thus, for $|w\phi(\zeta)/\zeta| < 1$, we have

$$\sum_{n=0}^{\infty} \frac{w^n}{n!} \frac{d^n}{dx^n} \left(f(x)\phi(x)^n\right)\Big|_{x=0} = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{\zeta} \left(\sum_{n=0}^{\infty} (w\phi(\zeta)/\zeta)^n\right) d\zeta$$

which is

$$\frac{1}{2\pi i} \int_C \frac{f(\zeta) d\zeta}{\zeta - w\phi(\zeta)}.$$

Since $|w\phi(\zeta)| < |\zeta|$, we deduce from Rouché's theorem that ζ and $\zeta - w\phi(\zeta)$ have the same number of zeros in C which means the zero is unique, simple and equal to z. Thus the integral is

$$\frac{f(z)}{1 - w\phi'(z)}.$$

This completes the proof. □

We apply the Lagrange inversion theorem to study the inverse function of the map $w \mapsto we^w$. This map sends zero to zero and satisfies the hypothesis of Proposition 9. Its inverse function, called the Lambert W-function is sometimes denoted as W_0 and

we can compute its power series easily using the Lagrange inversion formula. We find

$$W_0(z) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} z^n. \tag{16}$$

Using Stirling's formula, it is easy to see that this power series converges absolutely for $|z| < 1/e$.

It seems that the series (16) was independently re-discovered by Ramanujan where Question 738 on page 332 of [20] asks one to show that for $0 < x < 1$,

$$x = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} x^n e^{-nx}. \tag{17}$$

It is not clear what proof Ramanujan may have had, but it follows immediately from our derivation above. Indeed, from (16), we see that

$$W_0(-xe^{-x}) = -\sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} x^n e^{-nx},$$

so that the right hand side of (17) is $-W_0(-xe^{-x}) = x$, since W is the inverse function of the map $z \mapsto ze^z$. According to Berndt [4], Ramanujan had a more general result of which (17) is a special case and proofs of the related entries that appear in Ramanujan's notebooks are supplied on page 70 of [4]. It appears that Ramanujan re-discovered Lambert's function. Lambert himself introduced his function in 1758 to solve trinomial equations and Ramanujan's "Entry 14" in his famous notebooks does just that. Euler [11] extended Lambert's work but it seems to be E.M. Wright [27] who recognized the importance of the function in solving certain transcendental equations.

Proposition 11 gives immediately the Taylor series around $z = 0$ for all powers of $W(x)$. Indeed,

$$W_0(x)^k = \sum_{n=k}^{\infty} \frac{-k(-n)^{n-k-1}}{(n-k)!} x^n,$$

which can be re-written as

$$\left(\frac{W_0(x)}{x}\right)^k = \sum_{n=0}^{\infty} \frac{k(n+k)^{n-1}}{n!} (-x)^n. \tag{18}$$

This series is valid for $|x| < 1/e$.

In combinatorics, there is a cognate function called the Tree function, denoted $T(x)$ which is the inverse function of the map $w \mapsto we^{-w}$. Using the Lagrange inversion theorem, it is not difficult to see that

$$T(x) = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} x^n,$$

which again converges for $|x| < 1/e$. It is so-called because n^{n-1} is the number of labelled trees having n vertices. Clearly $T(x) = -W_0(-x)$. The analog of (18) is

$$T(x)^k = \sum_{n=0}^{\infty} \frac{k(n+k)^{n-1} x^{n+k}}{n!}.$$

If in Proposition 12, we let $\phi(z) = e^z$ and $f(z) = e^{az}$, we obtain

$$\sum_{n=0}^{\infty} \frac{(n+a)^n}{n!} w^n = \frac{e^{az}}{1 - we^z} = \frac{e^{az}}{1 - z}.$$

In particular, setting $a = 0$, we deduce

$$\frac{1}{1 - T(x)} = \sum_{n=0}^{\infty} \frac{n^n x^n}{n!}.$$

But W_0 given by (16) is only a (tiny) piece of the inverse function of the complex map $w \mapsto we^w$. We can plot the map $w \mapsto we^w$ for w real and see that its global minimum occurs at $w = -1$. From the graph, it is evident that the inverse function has two real branches, one branch whose range includes $(-1, \infty)$ and another whose range includes $(-\infty, -1]$. These branches are denoted W_0 and W_{-1} respectively. In fact, in analogy with the logarithm function, the inverse function of $w \mapsto we^w$ has countably many branches. More precisely, $W_0(z)$ extends to an analytic function on $\mathbb{C} \backslash (-\infty, -1]$ with -1 as a branch point. The countably many branches are denoted W_k, $k \in \mathbb{Z}$. We refer the reader to [8] for details. In the next section, we discuss the analytic continuation of $W_0(x)$ and give two integral formulas for positive real values of x (Fig. 1).

7 Integral Formulas for $W_0(x)$

This function $W_0(x)$ and its other branches are as ubiquitous in mathematics as the logarithm function and its range of wide applicability is discussed in [8]. Comtet (see pages 228–229 of [7]) derived the asymptotic behaviour of the principal branch

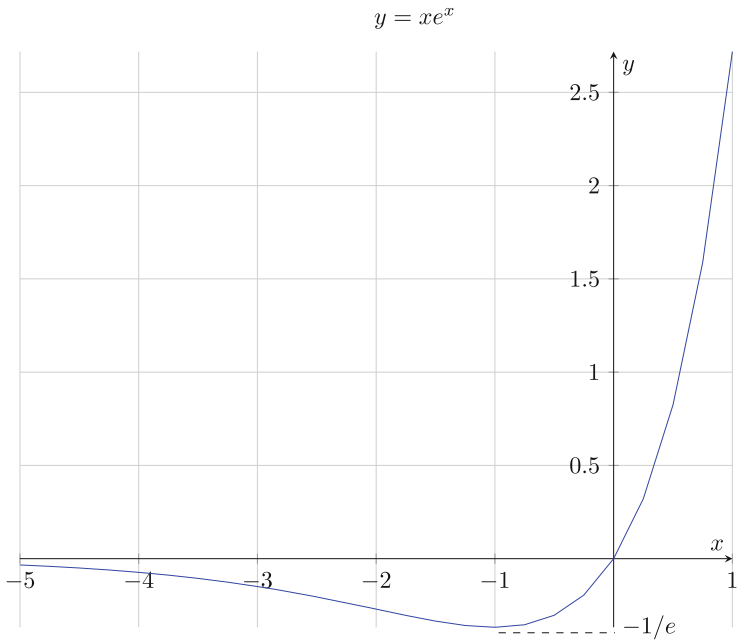


Fig. 1 $y = xe^x$

of $W_0(x)$. He proved that

$$W_0(x) = \log x - \log\log x - \sum_{n=1}^{\infty} \frac{(-1)^n}{(\log x)^n} \sum_{m=1}^{n} s(n, n-m+1) \frac{(\log\log x)^m}{m!},$$

where $s(n, k)$ is the Stirling number of the first kind. As $x \to \infty$, this shows that

$$W_0(x) = \log x - \log\log x + o(1).$$

It is possible to derive a similar expansion for the other branches as in [8]. Introducing the notation

$$\log_k z := \log z + 2\pi i k, \tag{19}$$

as in [8], we have

$$W_k(x) = \log_k x - \log\log_k x - \sum_{n=1}^{\infty} \frac{(-1)^n}{(\log_k x)^n} \sum_{m=1}^{n} s(n, n-m+1) \frac{(\log\log_k x)^m}{m!}.$$

(The "outer" log is the principal branch of the logarithm.) Here, we briefly discuss methods of extending the domain of the Lambert W-function. In this context, it may help the reader to recall how the logarithm function is analytically continued to the cut complex plane $\mathbb{C}\backslash(-\infty, 0]$. On this region, the principal branch of the logarithm is defined

$$\log z = \log|z| + i \arg z. \qquad -\pi < \arg z < \pi.$$

The branches of the logarithm are then given simply by (19). Following [8], we can give an analogous description of the branches of the W-function. But first, we review some basic information concerning the branches of W. Let us write $z = we^w$ with $z = x + iy$ and $w = u + iv$, and u, v, x, y real. Then

$$\begin{aligned} x &= e^u(u\cos v - v\sin v) \\ y &= e^u(v\cos v + u\sin v) \end{aligned}$$

Under this mapping, the image of the x-axis (that is, the curve $y = 0$) consists of the curves

$$v = 0 \quad \text{or} \quad u = -v\cot v$$

in the (u, v)-plane. This means that

$$x = -e^u\frac{v}{\sin v} < 0 \iff (2k-1)\pi < v < (2k+1)\pi, \qquad k \in \mathbb{Z}.$$

The curve which separates the principal branch W_0 from W_1 and W_{-1} is

$$-v\cot v + vi : \quad -\pi < v < \pi.$$

We give an analytic integral formula for $W_0(x)$ for $x > 0$ in the following theorem.

Theorem 13 *For $x > 0$,*

$$\frac{1}{x(1+W_0(1/x))} = \int_{-\infty}^{\infty} \frac{du}{(e^u - xu)^2 + \pi^2 x^2}$$

Proof Let R_T be the closed rectangular contour with vertices

$$(-T, -\pi), \quad (T, -\pi), \quad (T, \pi), \quad (-T, \pi)$$

oriented counterclockwise. Consider the integral

$$I_T := \int_{R_T} \frac{dw}{e^w + xw}.$$

Since

$$\frac{1}{|e^w + xw|} \leq \frac{1}{||e^w| - |x||w||} \to 0, \quad \text{as} \quad |w| \to \infty,$$

we see that the vertical integrals in the contour integral tend to zero as $T \to \infty$. Thus,

$$\lim_{T\to\infty} I_T = \int_{-\infty}^{\infty} \frac{du}{e^{u-\pi i} + x(u - \pi i)} - \int_{-\infty}^{\infty} \frac{dxu}{e^{u+\pi i} + x(u + \pi i)}.$$

Therefore,

$$\lim_{T\to\infty} I_T = \int_{-\infty}^{\infty} \frac{2\pi i x du}{(-e^u + xu)^2 + \pi^2 x^2}$$

Let us examine the integrand of I_T. Writing $w = u + iv$, and keeping in mind that $x > 0$, we see that $e^w + xw = 0$ if and only if

$$e^u \cos v = -xu, \qquad e^u \sin v = -xv.$$

In the region R_T, the second equation $e^u \sin v = -xv$ has a solution only when $v = 0$ in which case the first equation $e^u \cos v = -xu$ becomes $1/x = (-u)e^{-u}$ and we see the solution to this is $-u = W_0(1/x)$. This is the unique singularity of our integrand in the region, for T sufficiently large. If we denote by α this solution, we have

$$\lim_{T\to\infty} \frac{I_T}{2\pi i} = \frac{1}{e^\alpha + x}.$$

Since $e^\alpha = -x\alpha$, we find

$$\int_{-\infty}^{\infty} \frac{du}{(e^u - xu)^2 + \pi^2 x^2} = \frac{1}{x(1 + W_0(1/x))},$$

as claimed. □

The number $W_0(1)$ is often called the "omega constant" and denoted as Ω in the literature. Our integral formula in this special case yields the elegant formula

$$\frac{1}{1+\Omega} = \int_{-\infty}^{\infty} \frac{dx}{(e^x - x)^2 + \pi^2}$$

usually attributed to Adamchik, though there doesn't seem to be any published paper of his on the topic. Our formula can be viewed as a generalization of his result.

8 Zeros of Exponential Polynomials

Equation (13) in Kendall's parametric solution of the SIR model encoded in Theorem 8 is a special case of an exponential polynomial. In this section, we make some number-theoretic remarks concerning the arithmetic nature of roots of such polynomials. This will also lead to some interesting results about special values of the Lambert function.

An exponential polynomial has the general form

$$\sum_{i=1}^{n} p_i(z)e^{\alpha_i z},$$

where $p_i(z)$ is a polynomial for $1 \leq i \leq n$ and $\alpha_1, ..., \alpha_n$ are distinct complex numbers. We want to study zeros of such polynomials when $\alpha_1, ..., \alpha_n$ are algebraic numbers and $p_i(z)$ are non-zero polynomials with algebraic coefficients.

Theorem 14 *If $\alpha_1, ..., \alpha_n$ are distinct algebraic numbers, and $p_1(z), ..., p_n(z)$ are polynomials with algebraic coefficients, then any non-zero root z_0 of*

$$\sum_{i=1}^{n} p_i(z)e^{\alpha_i z} \tag{20}$$

is transcendental unless z_0 is a common root of all the polynomials $p_i(z)$ for $1 \leq z \leq n$.

Proof We recall the following version of a theorem of Lindemann and Weierstrass proved in 1885 (see Theorem 4.1 of [18] on page 15): if $\alpha_1, ..., \alpha_n$ are distinct algebraic numbers, then $e^{\alpha_1}, ..., e^{\alpha_n}$ are linearly independent over the field of algebraic numbers $\overline{\mathbb{Q}}$. So, if z_0 is a non-zero algebraic root of (20), then

$$\sum_{i=1}^{n} p_i(z_0)e^{\alpha_i z_0} = 0.$$

Since $\alpha_1 z_0, ..., \alpha_n z_0$ are then distinct algebraic numbers, and $p_i(z_0)$ are algebraic numbers, this would contradict the Lindemann-Weierstrass theorem, unless z_0 is a common root of all the polynomials $p_i(z)$, $1 \leq i \leq n$. □

Corollary 15 *Any root of (13) is transcendental.*

Proof We need only observe that in Kendall's parametric solution of the SIR model, Eq. (13) is an exponential equation and as all the constants appearing there are rational numbers, an immediate application of the theorem gives the result. □

Corollary 16 *If α is algebraic and lies in $\mathbb{C}\backslash(-\infty, 0]$, then $W_k(\alpha)$ is transcendental.*

Proof $W_k(\alpha)$ is the root of the exponential polynomial equation $ze^z - \alpha = 0$. Since $\alpha \neq 0$, we see that $W_k(\alpha) \neq 0$. The result is now evident. □

A minor variation of this result leads to a small generalization. The previous corollary is the case with $n = 1$ of the following.

Corollary 17 *If $\alpha_1, ..., \alpha_n$ are non-zero algebraic numbers and $c_1, ..., c_n$ are rational numbers, then at least one of the following is true:*

(a) $c_1 W(\alpha_1) + \cdots + c_n W(\alpha_n)$ is either zero or transcendental;

(b) $W(\alpha_1)^{c_1} \cdots W(\alpha_n)^{c_n}$ is non-zero and transcendental,

where W is any branch of the Lambert function.

Proof We have

$$W(\alpha_1)^{c_1} \cdots W(\alpha_n)^{c_n} \exp\left(c_1 W(\alpha_1) + \cdots + c_n W(\alpha_n)\right) = \alpha_1^{c_1} \cdots \alpha_n^{c_n},$$

Suppose that $W(\alpha_1)^{c_1} \cdots W(\alpha_n)^{c_n}$ is non-zero and algebraic. Then

$$\exp\left(c_1 W(\alpha_1) + \cdots c_n W(\alpha_n)\right)$$

is algebraic. But this contradicts the Lindemann-Weierstrass theorem unless $c_1 W(\alpha_1) + \cdots + c_n W(\alpha_n) = 0$. □

As a consequence of these comments on the transcendence of $W(z)$ for algebraic z, we deduce immediately from Theorem 13:

Corollary 18 *For any positive algebraic x, the integrals*

$$\int_{-\infty}^{\infty} \frac{du}{(e^u - xu)^2 + \pi^2 x^2}$$

are transcendental numbers.

A celebrated conjecture of Schanuel (see page 17 of [18]) predicts that if $x_1, ..., x_n$ are linearly independent over the rational number field, then the transcendence degree of

$$\mathbb{Q}(x_1, ..., x_n, e^{x_1}, ..., e^{x_n})$$

is at least n. Thus, if $\alpha_1, ..., \alpha_n$ are algebraic numbers such that $W(\alpha_1), ..., W(\alpha_n)$ are linearly independent over $\mathbb{Q}$, then applying the functional relation of the Lambert function leads to the prediction that the transcendence degree of

$$\mathbb{Q}(W(\alpha_1), ..., W(\alpha_n))$$

is at least n. In other words, Schanuel's conjecture predicts that if $\alpha_1, ..., \alpha_n$ are algebraic numbers such that $W(\alpha_1), ..., W(\alpha_n)$ are linearly independent over $\mathbb{Q}$, then they are algebraically independent. In particular, this would suggest that $W(\alpha_1), ..., W(\alpha_n)$ are linearly independent over $\overline{\mathbb{Q}}$. Perhaps this last implication can be proved using existing techniques from transcendental number theory.

9 The Mellin Transform of $W(x)$

The Mellin inversion formula allows us to derive another integral formula for $W(x)$ for $x > 0$. As noted in [8], the Mellin transform of $W(x)$ is

$$\int_0^\infty W(x)x^{s-1}dx = \int_0^\infty e^{-W(x)}x^s dx,$$

since $W(x)/x = e^{-W(x)}$. Setting $W(x) = u$ so that $x = W^{-1}(u) = ue^u$, and using the fact that

$$dx = e^W(1 + W)dW,$$

we get

$$\int_0^\infty W(x)x^{s-1}dx = \int_0^\infty (ue^u)^s(u + 1)du.$$

Putting $us = -t$, we obtain

$$\int_0^\infty W(x)x^{s-1}dx = -\int_0^\infty e^{-t}(-t/s)^s(-t/s + 1)dt/s.$$

The right hand side is

$$-(-s)^{-s}\int_0^\infty e^{-t}t^s(1 - t/s)dt/s = -(-s)^{-s}\left\{\Gamma(s) - \frac{\Gamma(s + 2)}{s^2}\right\}.$$

Using the functional relation $\Gamma(s + 2) = (s + 1)s\Gamma(s)$, this simplifies to

$$\int_0^\infty W(x)x^{s-1}dx = (-s)^{-s}\frac{\Gamma(s)}{s}.$$

Since the logarithm is analytic in $\mathbb{C}\backslash(-\infty, 0]$ and the Γ-function has simple poles only at negative integers, we see that the right hand side is analytic in the region $-1 < \mathrm{Re}(s) < 0$.

For ease of reference, we recall the Mellin inversion formula (see p. 273 of [23]):

Proposition 19 *Suppose that $F(s)$ is a function of the complex variable $\sigma+it$ which is regular in the infinite strip $S = \{s : a < \sigma < b\}$ and for arbitrary small positive number ϵ, $F(s)$ tends to zero uniformly as $|t| \to \infty$ in the strip $a+\epsilon \leq \sigma \leq b-\epsilon$. Then, if the integral*

$$\int_{-\infty}^{\infty} F(\sigma + it)dt$$

is absolutely convergent for each value of σ in the open interval (a, b) and if for positive real values of x and a fixed $c \in (a, b)$ we define

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}F(s)ds,$$

then in the strip S, we have

$$F(s) = \int_0^{\infty} x^{s-1}f(x)dx.$$

We can apply the Mellin inversion formula to deduce:

Theorem 20 *Let $-1 < c < 0$. For $x > 0$ we have*

$$W(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}(-s)^{-s}\frac{\Gamma(s)}{s}ds. \tag{21}$$

Proof To apply Proposition 19, we verify the growth conditions on $(-s)^{-s}\Gamma(s)/s$. To this end, let us recall Stirling's approximation: for $s = \sigma + it$ with $\sigma, t \in \mathbb{R}$, we have

$$|\Gamma(s)| = \sqrt{2\pi}\,|t|^{\sigma-1/2}e^{-\pi|t|/2}(1 + O(|t|^{-1}))$$

for $|t|$ sufficiently large. Moreover, writing $\log(-s) = \log|s| + i\arg(-s)$, we see that

$$|(-s)^{-s}\Gamma(s)/s| = |s|^{-\sigma}e^{t\arg(-s)}|t|^{\sigma-3/2}e^{-\pi|t|/2}$$

satisfies the condition of the theorem in that it tends to zero uniformly as $|t| \to \infty$ in the strip $-1+\epsilon \leq \sigma \leq -\epsilon$. We need only note that in this region $|\arg(-s)| \leq \pi/2$. □

Incidentally, the above theorem can be used to derive the power series representation of $W(x)$ for $|x| < 1/e$ obtained earlier using the Lagrange inversion theorem. To see this, we truncate the integral of the theorem and view it as

$$\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT} x^{-s}(-s)^{-s}\frac{\Gamma(s)}{s}ds.$$

Having done this, we move the contour to the left and note that the singularities of the integrand occur at negative integers. The residue is easily calculated. At $s = -n$, it is

$$-x^n n^{n-1}(-1)^n/n!$$

and both the horizontal and vertical integrals tend to zero since $|x| < 1/e$. Summing this over $n \geq 1$ gives (16). (We leave the details to the reader as an exercise in complex analysis.)

Our integral formulas are valid only for $x > 0$ and it would seem natural to inquire if there is an analytic power series that gives the analytic continuation of the principal branch of W. This was done recently by Beardon [2]. Using fairly elementary complex analysis, he shows that for $z \in \mathbb{C}\backslash(-\infty, -1/e]$ we have the following series for the principal branch:

$$W_0(z) = \sum_{m=1}^{\infty} a_m \left(\frac{\sqrt{ez+1}-1}{\sqrt{ez+1}+1} \right)^m, \qquad a_m = \sum_{n=1}^{m} \frac{(-n)^{n-1}}{n!} \left(\frac{4}{e}\right)^n \binom{m+n-1}{m-n}.$$

He also indicates how one may derive similar series representations of the other branches.

After this digression into the theory of the Lambert function, we describe how it relates to an exact solution of the SIR model in the next section.

10 Solving $A + Br + e^{Cr} = 0$ and Kendall's Arrangement

One can solve for roots of exponential equations using the W-function. Indeed, suppose we have an equation of the form

$$A + Br + e^{Cr} = 0, \tag{22}$$

with A, B, C given, and we want to solve for r. Thus, $(A + Br)e^{-Cr} = -1$ and we can re-write this as

$$\frac{B}{C}\left(-\frac{AC}{B} - Cr\right) e^{-Cr - AC/B} = e^{-AC/B}.$$

Thus, $-Cr - AC/B = W_k(Ce^{-AC/B}/B)$ so that all the roots are given by

$$-\frac{A}{B} - \frac{1}{C} W_k\left(\frac{C}{B} e^{-AC/B}\right).$$

For algebraic values A, B, C, the non-zero roots of (22) are transcendental numbers by the theorem of Hermite and Lindemann alluded to earlier. It is curious to note that, by contrast, $W_k\left(-\frac{AC}{B}e^{-AC/B}\right) = -AC/B$ is algebraic.

We apply this discussion to determine the roots of (13) appearing in Kendall's Theorem 8 which we re-write as

$$-[R_0 I(0) + 1] + R_0\zeta + e^{-R_0\zeta} = 0.$$

The positive root ζ_2 is

$$\zeta_2 = \frac{R_0 I(0) - 1}{R_0} + \frac{1}{R_0} W_0\left(-e^{1-R_0 I(0)}\right).$$

Since $R_0 > 1$ in the case of an epidemic, and we may suppose that there are initially at least two infected people, we see that $1 - R_0 I(0) < -1$ so that the value of this root can easily be determined using the power series (16) since the argument of the W-function lies in the domain of absolute convergence.

To find the negative root $-\zeta_1$, we have

$$-\zeta_1 = \frac{R_0 I(0) - 1}{R_0} + \frac{1}{R_0} W_{-1}\left(-e^{1-R_0 I(0)}\right).$$

Interestingly, by our earlier remark, both ζ_1 and ζ_2 are transcendental numbers. This should not be a cause for too much consternation since, after all, the SIR model is not an exact description of "reality" regarding the pandemic but only a mathematical tool to enable us to understand its behaviour.

According to Kendall's arrangement, the total number of people infected during the entire epidemic will be $\zeta_1 + \zeta_2$ which is

$$\frac{1}{R_0}\left(W_0\left(e^{1-R_0 I(0)}\right) - W_{-1}\left(e^{1-R_0 I(0)}\right)\right).$$

11 The Hadamard Product of $A + Bz + e^{Cz}$

The function $f(z) = A + Bz + e^{Cz}$ is of order 1 (in the sense of Hadamard) and as such, admits a factorization of the form

$$e^{c_0 z + c_1} \prod_{\omega} \left(1 - \frac{z}{\omega}\right) e^{z/\omega},$$

where the product is over the zeros of $f(z)$ and c_0, c_1 are appropriate constants. These are easily determined as follows. Setting $z = 0$ gives $e^{c_1} = A + 1$. We will

assume $A+1 \neq 0$. Taking the logarithmic derivative gives

$$\frac{f'(z)}{f(z)} = c_0 + \sum_{\omega} \Big[\frac{1}{z-\omega} + \frac{1}{\omega} \Big],$$

so by setting $z=0$ we get $c_0 = f'(0)/f(0) = (B+C)/(A+1)$. This proves:

Proposition 21

$$A + Bz + e^{Cz} = (A+1) e^{(B+C)z/(A+1)} \prod_{\omega} \left(1 - \frac{z}{\omega}\right) e^{z/\omega},$$

where the product is over the zeros of $A + Bz + e^{Cz}$.

As we observed in the previous section all the roots of $f(z)$ can be written as

$$\omega_k := -\frac{A}{B} - \frac{1}{C} W_k \left(\frac{C}{B} e^{-AC/B} \right).$$

We also noted that the Hermite-Lindemann theorem shows that ω_k is transcendental whenever A, B, C are algebraic. We can deduce the following curious result, reminiscent of Euler's explicit evalutaion of $\zeta(2n)$.

Proposition 22 *For* A, B, C *algebraic, the sums*

$$\frac{\ell_{n+1}}{n!} := \sum_{\omega} \frac{1}{\omega^{n+1}}$$

are algebraic for $n \geq 1$.

Proof By Taylor's theorem, we have

$$\frac{1}{n!} \frac{d^n}{dz^n} \left(\frac{f'(z)}{f(z)} \right) \Big|_{z=0} = -\sum_{\omega} \frac{1}{\omega^{n+1}}.$$

The left hand side is an algebraic number. □

12 Kendall's Integral and Lagrange Inversion

The exact solution to the SIR model provided by Kendall [15] amounts to writing the time function as a function of R via an integral of the form

$$\int_0^R \frac{dr}{A + Br + e^{cr}},$$

with suitable values of A, B, C. Though perhaps not practical, we want to show that it is theoretically possible to invert this and write R as a function of the time parameter t. To do this, we first write the above integral as a power series in R as follows. Using our Hadamard factoriztion we see that

$$A + Br + e^{Cr} = (A+1)\exp\left(\frac{(B+C)r}{A+1} + \sum_{\omega}\left[\log\left(1-\frac{r}{\omega}\right) + \frac{r}{\omega}\right]\right).$$

Thus,

$$\frac{1}{A + Br + e^{Cr}} = (A+1)^{-1}\exp\left(-\frac{B+C}{A+1}r + \sum_{n=2}^{\infty}\sum_{\omega}\omega^{-n}\frac{r^n}{n}\right).$$

It seems convenient to define $\ell_1 = -(B+C)/(A+1)$ so that applying (15) gives

$$\frac{1}{A + Br + e^{Cr}} = (A+1)^{-1}\sum_{n=0}^{\infty} B_n(\ell_1, ..., \ell_n)\frac{r^n}{n!}.$$

We can now integrate term by term and deduce that

$$\int_0^R \frac{dr}{A + Br + e^{Cr}} = (A+1)^{-1}\sum_{n=0}^{\infty} B_n(\ell_1, ..., \ell_n)\frac{R^{n+1}}{(n+1)!}.$$

The right hand side is of the form $Rg(R)$ with $g(0) \neq 0$, as required by Proposition 9. We can then invert this power series using the Lagrange inversion formula.

13 Further Remarks

The approximation of the exponential function with a quadratic polynomial leads to the above analysis. In particular, it implies that there is a single maximum for the number of infected people. In practice however, there seem to be waves of the epidemic pointing to a series of local maxima. So perhaps, this model may have to be modifed by allowing β to be a function of time and introducing an oscillation factor.

It is evident from the discussion of the previous sections that by using higher order polynomial approximations of the exponential function, one is led to the study of integrals of the form

$$\int \frac{dr}{f(r)}$$

where $f(r)$ is a polynomial. One can evaluate such integrals using the method of partial fractions and we immediately see that r is an algebraic function over the field generated by e^t.

In the particular cases of cubic and quartic approximations to the exponential function, we end up with explicit formulas for $r(t)$ in terms of the exponential function since we have formulas for roots of the general cubic and quartic polynomials. The general quintic can also be solved using elliptic functions. Thus, it may be fruitful to push this analysis to at least these three levels of approximation and derive further local maxima of the $I(t)$ function. This would be only of theoretical interest since our analysis shows that the Kermack and McKendrick solution is very accurate if we introduce the constant ϕ' mentioned in Theorem 7.

There are efficient numerical methods that can be used to solve the original differential equation without using the polynomial approximations of the exponential function. Most notable in this array of methods is the Runge-Kutta method. The idea here is to use the differential equation for r as the starting point of developing its Taylor series. We will not go into details but refer the reader to section 14 of Chapter 8 of [3]. Since we are interested in knowing before hand the peaks of the number of infected people at any given time, it is the derivative of the function r that we will need to focus on and determine its zeros. This can be done numerically through these methods.

Acknowledgments We thank Siddhi Pathak, Troy Day and Jianhong Wu for their careful reading and helpful remarks on an earlier version of this paper.

References

1. N.T.J. Bailey, The Mathematical Theory of Infectious Diseases and its Application, Charles Griffin and Company, London, 1975.
2. A.F. Beardon, The principal branch of the Lambert W-function, Computational Methods and Function Theory. https://doi.org/10.1007/s40315-020-00329-6, published online 13 August 2020.
3. G. Birkhoff and G.-C. Rota, Ordinary differential equations, Blaisdell Publishing Company, Waltham, Massachusetts, 2nd Edition, 1969.
4. B. Berndt, Ramanujan's Notebooks, Part I, Springer-Verlag, New York, 1985.
5. D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite verole et des avantages de l'inoculation pour la prevenir, *Mem. Phys. Acad. Roy. Soc., Paris.* I, 1760.
6. S. Cioaba, Some applications of eigenvalues of graphs, in Structural Analysis of Complex Networks, edited by M. Dehmer, 357–379, (2011), Springer.
7. L. Comtet, Advanced Combinatorics, D. Reidel Publishing Company, Dodrecht, Holland, 1974.
8. R.M. Corless, G.H. Gonner, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth, On the Lambert W-function, Advances in Computational Mathematics, Vol. 5, (1996), 329–359.
9. M. Draief, A. Ganesh, L. Massoulié, Threshold for virus spread on networks, *Annals of Applied Probability, 18* (2008), 359–378.
10. D. Easley and J. Kleinbert, Networks, Crowds and Markets: Reasoning about a highly connected world, Cambridge University Press, 2010.

11. L. Euler, De serie Lambertina Plurimisque eius insignibus proprietatibus, reprinted in Opera Omnia, Series Prima, Vol. 6: Commentationes Algebraicae. Leipzig, Germany: Teubner, pp. 350–369, 1921.
12. J. Gani and R.J. Swift, On the Kermack–McKendrick approximation for removals from an epidemic, *Math. Scientist*, ***35***, (2010), 29–36.
13. W.H. Hamer, Epidemic disease in England, *Lancet*, ***1*** (1906), 733–739.
14. T. Harko, F.S.N. Lobo, and M.K. Mak, Exact solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates, *Applied Mathematics and Computation*, ***236*** (2014), 184–194.
15. D.G. Kendall, Deterministic and stochastic epidemics in closed populations, *Proc. Third Berkeley Symp. Math. Statistics and Probability*, ***4*** (1956), 149–165, Berkeley and Los Angeles, University of California Press.
16. W.O. Kermack and A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. of the Royal Society A*, ***115*** (1927), 700–721.
17. M. Ram Murty, A Second Course in Analysis, Hindustan Book Agency, 2020.
18. M. Ram Murty and P. Rath, Transcendental Numbers, Springer, 2014.
19. G. Pólya and G. Szegö, Problems and Theorems in Analysis I, Springer-Verlag, New York, 1978.
20. S. Ramanujan, Collected Papers, edited by G.H. Hardy, P.V. Seshu Aiyar and B.M. Wilson, AMS Chelsea Publishing, American Mathematical Society, Providence, Rhode Island, 2013.
21. R. Ross, Some a priori pathometric equations, *British Medical Journal*, ***1*** (1915), 546–547.
22. W. Rudin, Real and Complex Analysis, Third Edition, McGraw-Hill, 1987.
23. I. Sneddon, The Use of Integral Transforms, McGraw-Hill, 1972.
24. R. Stanley, Enumerative Combinatorics, Volume 2, Cambridge University Press, 1999.
25. Y. Wang, D. Chakraborti, C. Wang, C. Faloutsos, Epidemic spreading in real networks: an eigenvalue viewpoint, in Proceedings of the Symposium on Reliable Distributed Computing, Florence, Italy, 25–34.
26. E.T. Whittaker and G.N. Watson, A Course in Modern Analysis, Cambridge University Press, 1927.
27. E.M. Wright, Solution of the equation $ze^z = a$, *Bulletin of the American Mathematical Society*, ***65*** (1959), 89–93.