Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Chapter 10

Mathematical modeling as a tool for policy decision making: Applications to the COVID-19 pandemic

J. Panovska-Griffiths[a,b,c,*], C.C. Kerr[d,e], W. Waites[f,g], and R.M. Stuart[h,i]

[a]*Department of Applied Health Research, University College London, London, United Kingdom*
[b]*Institute for Global Health, University College London, London, United Kingdom*
[c]*The Wolfson Centre for Mathematical Biology and The Queen's College, University of Oxford, Oxford, United Kingdom*
[d]*Institute for Disease Modeling, Global Health Division, Bill & Melinda Gates Foundation, Seattle, WA, United States*
[e]*School of Physics, University of Sydney, Sydney, NSW, Australia*
[f]*School of Informatics, University of Edinburgh, Edinburgh, United Kingdom*
[g]*Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom*
[h]*Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark*
[i]*Disease Elimination Program, Burnet Institute, Melbourne, VIC, Australia*
[*]*Corresponding author: e-mail: j.panovska-griffiths@ucl.ac.uk*

Abstract

The coronavirus disease 2019 (COVID-19) pandemic highlighted the importance of mathematical modeling in advising scientific bodies and informing public policy making. Modeling allows a flexible theoretical framework to be developed in which different scenarios around spread of diseases and strategies to prevent it can be explored. This work brings together perspectives on mathematical modeling of infectious diseases, highlights the different modeling frameworks that have been used for modeling COVID-19 and illustrates some of the models that our groups have developed and applied specifically for COVID-19. We discuss three models for COVID-19 spread: the modified Susceptible-Exposed-Infected-Recovered model that incorporates contact tracing (SEIR-TTI model) and describes the spread of COVID-19 among these population cohorts, the more detailed agent-based model called Covasim describing transmission between individuals, and the Rule-Based Model (RBM) which can be thought of as a combination of both. We showcase the key methodologies of these

Handbook of Statistics, Vol. 44. https://doi.org/10.1016/bs.host.2020.12.001

approaches, their differences as well as the ways in which they are interlinked. We illustrate their applicability to answer pertinent questions associated with the COVID-19 pandemic such as quantifying and forecasting the impacts of different test-trace-isolate (TTI) strategies.

Keywords: Epidemiological modeling, COVID-19, SEIR models, Agent-based models, Rule-based models

1 Introduction

1.1 Overview of mathematical modeling

As we write this in October of 2020, the world remains gripped by COVID-19 pandemic caused by the spread of a severe acute respiratory syndrome coronavirus (SARS-CoV-2). Since the emergence of this new virus, mathematical sciences—particularly modeling—have been at the forefront of policy decision making around it.

Mathematical and computational models are a way to understand the processes in complex systems that underlie empirical observations and to generate possible future trajectories of these systems. Strictly speaking, mathematical models refer to the actual framework of composing a set of equations or theoretical approaches, while computational models refer to the numerical and computational approaches used to solve the mathematical framework. In practice, these terms are used interchangeably since solving practical mathematical models analytically is rarely possible; hence, we often simply refer to both as models. The overarching purpose of models is to allow a flexible framework in which different scenarios can be tested, different questions related to future behavior can be posed and evaluated, and potential future behavior can be predicted. There is a difference between explanatory models that attempt to explain current behavior and predictive models that extend this into the future, and we will discuss this further in Section 1.2.

The overall aim of mathematical modeling is to generate answers to questions we can't get from observations. The answers are then used to understand, manage and predict future behavior of complex systems and processes, for example, to inform public policy and future decision making. The statistician George Box said, "all models are wrong; some models are useful." It is important to understand that modeling is like any other technology: it can be properly applied or not, it may produce output that admits a useful interpretation, or it may not. In Section 1.3, we discuss the notion of a "correct" model; it is important to understand this notion of correctness against the background of utility of mathematical modeling.

1.2 Modeling to explain or to predict

Mathematical modeling provides a framework that, given data, facilitates understanding of how changes within the framework can affect outcomes.

Modeling combined with data can explain past behavior, predict and forecast future behavior, and evaluate how changes may alter these predictions. Explaining past and current trends and predicting future trends are two different aspects of modeling. The clear distinction between the two concepts was highlighted by a number of scientists such as Forster and Sober (1994), Forster (2002), Hitchcock and Sober (2004), and Dowe et al. (2007).

Explanatory modeling combines theory with data to test hypotheses and explain behavior. Regression modeling, or curve-fitting to data, is a type of explanatory modeling that is widely used in statistical sciences. Explanatory modeling can answer questions like "Which population cohort is at greatest risk of infection"? by looking backwards and finding the key parameters (covariates) that help explain the observed patterns in the data. These types of models have a long history in helping to explain patterns in disease and public health. While continuing the trajectory of the curve that best explained historical trends could in theory be used to predict future trends, the statistical model best capable of explaining past trends may not be well suited to forecasting future trends.

In contrast, predictive modeling (Shmueli, 2010) consists of building a mechanistic framework that is explicitly designed to be able to explain both historical patterns and future states. In contrast to explanatory modeling, within this framework different cogs within the system can be built to resemble possible future behavior. Predictive modeling could explore different scenarios and answer questions of the general form: "What would happen to X if we did Y"? Specific examples of such questions include: "What would happen to the COVID-19 epidemic under strict social distancing for 2 months"? or "If things carried on like today, how would the epidemic look in 3, 6 or 12 months"? Finally, predictive modeling can also look at trade-offs and optimise outcomes by answering questions like: "What is the best strategy to take if we want to achieve a given outcome"?

Although both explanatory and predictive models are important and have been used widely to gain a better understanding of the COVID-19 pandemic, our focus in this chapter will be on predictive models. An overview of explanatory models can be found in Shmueli (2010).

1.3 What does it mean for models to be "right"?

Building a "correct" model is a more complex concept than may initially be thought. In the previous section, we addressed the question of why we build models: we do so in order to answer questions about things that we cannot observe. But since the questions we put to models are often complex, knowing whether the answer is "right" is also not straightforward. This is best illustrated with an example. In early 2020, the widely-publicised mathematical model produced by Imperial College London (Ferguson et al., 2020) suggested that mortality from COVID-19 was around 1% and an epidemic in a susceptible population of 67 million people could cause around 670,000

deaths in the absence of any interventions to reduce the spread of the virus (Ferguson et al., 2020). In reality, numerous non-pharmeceutical interventions were deployed, and the UK had recorded around 70,000 deaths by end of November 2020. Does this mean that the model "got it wrong"? How can we assess and validate the model under these circumstances?

This example illustrates a fundamental difficulty with evaluating predictive models. If we model a policy scenario which doesn't manifest, does that invalidate the projections produced by the model? We can try to answer this by going back to the model and asking whether it would have predicted the actual trajectory of the epidemic if given information on what policies ended up being enacted. Exercises of this nature have been attempted for the Imperial model, for example (Rice et al., 2020).

Rather than asking whether a model is *right*, it is more instructive to ask whether it is *useful*. It is for this reason that a better understanding of the processes of modeling and a greater awareness of how and when models can be reliably used are important. The COVID-19 pandemic illustrated that even the often-cited maxim among modelers that "the model is as good as the data it uses" may not always be the most useful framework, because even at the onset of the pandemic where data was scarce, modeling still proved useful.

1.4 Aims and purposes of this work

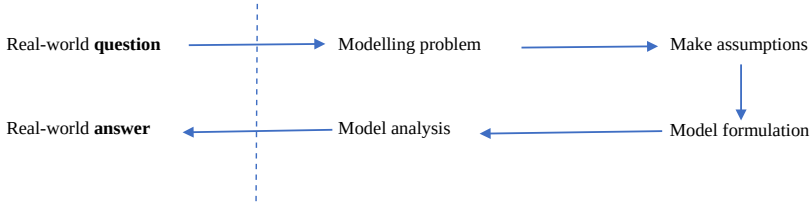
This chapter aims to bring together the current thinking about mathematical modeling of infectious diseases, zooming in on the modeling of the COVID-19 pandemic that was used to guide policy. We first give a short history of mathematical modeling (Section 2) before presenting a scoping overview of the key published models until September 2020 used to capture COVID-19 transmission dynamics (Section 3). Then in Section 4, we showcase three of the models that we have applied to understanding the COVID-19 epidemic, briefly describing their methodology and illustrating their application to answer a specific policy question. Our conclusion then brings it all together, summarising the usefulness of modeling.

2 Modeling of infectious disease

2.1 The essence of modeling infectious diseases

The essence of mathematical modeling involves building a framework, often based on a system of equations, that simplifies and mimics reality in order to derive answers to real-life questions, as shown schematically in Fig. 1A. Although these systems of equations are simplifications of reality, they can nevertheless be used to investigate real-world questions, either by solving the equations directly if possible, or by simulating outputs from the system for given values of the parameters within the equations (Fig. 1B).

A



B

	Empirical	Mechanistic
Deterministic	Predicting virus spread from a **regression analysis** over time	Predicting virus spread based on known historic behaviour and based on **ordinary differential equations**
Stochastic	Analysis of **variance in regression model** of virus	Predicting virus spread based on different historic behaviour and based on **probabilistic differential equations**

FIG. 1 (A) Schematic of mathematical modeling describing broadly the steps of modeling that allow answering real world problems. (B) Broad division of mathematical models.

The parameters of the mathematical model can be refined when we compare the outputs of the model to information that we already know, for example, available data on the reported number of infections or the confirmed number of deaths due to the infection. This process of model refinement (or calibration) can be done iteratively a number of times sweeping through the parameter space, manually or in an automated way (Taylor et al., 2010), until the output of the model agrees with what we already know, e.g., about the virus spread. While simple models with only a small number of parameters can be calibrated with simple parameter sweep and sensitivity analysis, calibration of complex models require extensive complex computational techniques. A multitude of statistical approaches have been applied to this problem, some of which rely on a likelihood function and some which do not (Andrianakis et al., 2015; Kennedy and O'Hagan, 2001). In essence, these methods attempt to minimise the difference between the model and the data. More formally, this means minimising a specified objective function defined on the model by starting with an initial model parametrisation and seeing how the difference changes as the parameters change. A subset of the methods discussed in Andrianakis et al. (2015) use Bayesian approaches, in which a prior distribution is set over the parameter space, which is then updated to create a posterior distribution once the information from the data has been taken into account.

Once the model is calibrated it can be used to tell us more about future behavior of the virus spread, i.e., make predictions or forecasts. For example, in epidemiological modeling, it can predict the epidemic curve, i.e., the curve that gives the number of infections caused by the virus over time. Changes to the model parameters can mimic possible future interventions and this

would allow the calibrated model to be useful in making predictions of what the epidemic curve may look like in the future.

Model discrepancy and measurement bias are two aspects that complicate model calibration. While model discrepancy explains the difference between the mathematical model and the reality as described by the data sources, the measurements bias is related to describing the measurement error, which can be caused by both the data and the calibration method. Separating the measurement bias from model discrepancy is important in calibrating models and when defining "goodness of fit." Accounting for both is important in predictive modeling.

Ideally, the system of equations within the modeling framework would be solved analytically, deriving exact solutions using historic behavior (initial or boundary conditions). But for non-trivial models it is rarely possible to obtain such analytical solutions; the complexity of modeling real-life scenarios quickly produces non-trivial models. The alternative is numerical solution of the system. Broadly speaking, the modeler has a choice when conducting a numerical simulation of the model: to formulate the model as an initial-value problem of ordinary differential equations describing entire population groups or to consider each change to the state of individuals in the population as a discrete stochastic event. For differential equations, simulating the system always gives the same result. For stochastic simulations, the inherent randomness means that each simulation results in a different evolution of the system and ascertaining how the system will behave on average requires running many such simulations and reporting the mean or the median. These two strategies are related: with suitable assumptions, the time evolution of the system described with differential equations is approximated by the average time-evolution of the stochastic system. There are important differences: the more fine-grained the model is, the less feasible it becomes to formulate differential equations for it, and the only choice is stochastic simulation. Furthermore, only some choices of distribution for the timing of events are compatible with the assumptions required for the mean trajectory of the stochastic system to coincide with the solution to the differential equations. These differences mean that the practice is that some classes of model are typically simulated using one or the other technique and this leads to two categories of widely used models: compartmental models and agent-based models (ABMs).

2.2 History of modeling infectious diseases

Mathematical modeling has a long history of being used for understanding how a virus can spread in a population.

The simplest kind of disease model is based on the concept of splitting the population in compartments, i.e., compartmental modeling and was introduced in the seminal paper by Daniel Bernoulli in 1776 (Bernoulli, 1766). This model described the spread and vaccination against smallpox and was

revisited and discussed in detail in Dietz and Heesterbeek (2002). This was extended by Ross (1911) who, in work for which he was awarded the Nobel Prize in Medicine, used what we today call dynamic transmission modeling. When he published his first dynamic malaria model, Ross introduced the phrase "*a priori pathometry*" to describe the scientific process of modeling transmission dynamics (Ross, 1911). He expanded on this in his work from 1911, presenting a new set of equations for the demonstration of the dynamics of the transmission of malaria between mosquitoes and humans (Ross, 1916). The importance of Ross' work was that he believed that explaining epidemics and epidemic control quantitatively were extremely challenging and had to be combined with predictive quantitative measures; hence, giving rise to many of the processes of mathematical modeling of infectious diseases as we know them now. In his description of the work he called this "*a priori notions of observational data and exploration of patterns*" that emerge when we use observational data combined with statistical analysis; as outlined in Ross (1916). The concept of combining data with a system of theoretical equations forms the fundamental framework of mathematical modeling of infectious diseases. Further advances were made possible by Ross' collaborations with the mathematician Hilda Hudson, whose technical expertise opened the door for an understanding of the patterns that different models could produce (Ross and Hudson, 1917a,b). These developments formed the seed for the work of Kermack and McKendrick, widely considered as the conceptors of the present-day mathematical epidemiology. Their first paper from 1927 (Kermack and McKendrick, 1927), acknowledged Ross and Hudson's work and cemented the notion of "mathematical epidemiology" or "mathematical theory of epidemics" as they called it. Subsequent work by Kermack and McKendrick (1932, 1933, 1937, 1939) expanded this theory and also defined the basic reproduction number in terms of model parameters; we discuss their model in more detail in Section 2.2.2. Ross, Kermack and McKendrick influentially recognized the importance of mathematical epidemiology, and their ideas motivated the work by Macdonald (1950, 1952, 1955, 1956) and Anderson et al. (1992). This is the basis of compartmental models that are presently used in modeling infectious diseases across a number of diseases, including COVID-19; more details on this are presented in Section 3. In the next section, we give a brief description of the simplest compartmental modeling framework, while in Section 4.1 we show how we have extended this to apply it to modeling the spread of COVID-19.

2.2.1 Lotka-Volterra equations, SIR models and reproduction number

Every university mathematical epidemiology course generally starts with introducing the modeling framework that tracks the temporal evolution of populations of Susceptible to the infection (S) cohort, the subset of these that

become infected (I) and the subset of these that recover from the infection (R); this describes the model of Kermack and McKendrick (1927).

But before we discuss in more details this classic Kermack-McKendrick SIR model we will introduce the classic Lotka-Volterra system of equations as a pair of first order, non-linear, differential equations that describe the dynamics of biological systems in which two species, a predator and a prey, interact (Murray, 1989). They were proposed independently by Lotka in 1925 and Volterra in 1926 just before Kermack-McKendrick SIR model was introduced in 1927. The classic Lotka-Volterra system comprises Eqs. (1)–(2) and taught modeling courses generally start with introducing this system of equations before subsequently studying them using steady-state, perturbation and bifurcation analysis exploring the existence and stability of the long-term solutions of the system; for details, see for example, Chapters 3 and 4 of Murray (1989):

$$\frac{dx}{dt} = x(a - by) \tag{1}$$

$$\frac{dy}{dt} = y(cx - d) \tag{2}$$

Assuming that $a,b,c,d > 0$, this system describes x as the predator and y as the prey with the parameters c and d representing the competition between them. The parameters a and d describe how quickly $x(t)$ and $y(t)$, respectively, grow and decay exponentially in absence of the other species; solving the system with $b = c = 0$ gives

$$x(t) = x_0 e^{at}, y(t) = y_0 e^{-dt}$$

where x_0 and y_0 are the initial values of x and y, i.e., $x_0 = x(t = 0)$, $y_o = y(t = 0)$.

If the parameters b and c are included, the system of Eqs. (1)–(2) can be solved in special cases as shown by Varma, Wilson or Burnside (Varma, 1977; Wilson, 1980). The numerical solutions of the system of equations, for a choice of model parameters and initial conditions, are shown in Fig. 2A and B.

The oscillatory solutions in Fig. 2A are a result of the cyclic behavior of the system confirmed in the phase plot diagram in Fig. 2B. Interestingly, using two different solvers in MATLAB to solve even a very simple system of equations such as (1)–(2) can give slightly different results; hence it is important that we discuss the robustness of numerical solutions in Section 2.3.

In fact, we can derive the cyclic solutions depicted in Fig. 2B, representing the phase trajectories, analytically by directly integrating the phase plane equation (brief introduction to phase plane analysis can be found in the Appendix A of Murray (1989))

$$\frac{dv}{du} = \alpha \frac{v}{u}\left(\frac{u - 1}{1 - v}\right) \tag{3}$$

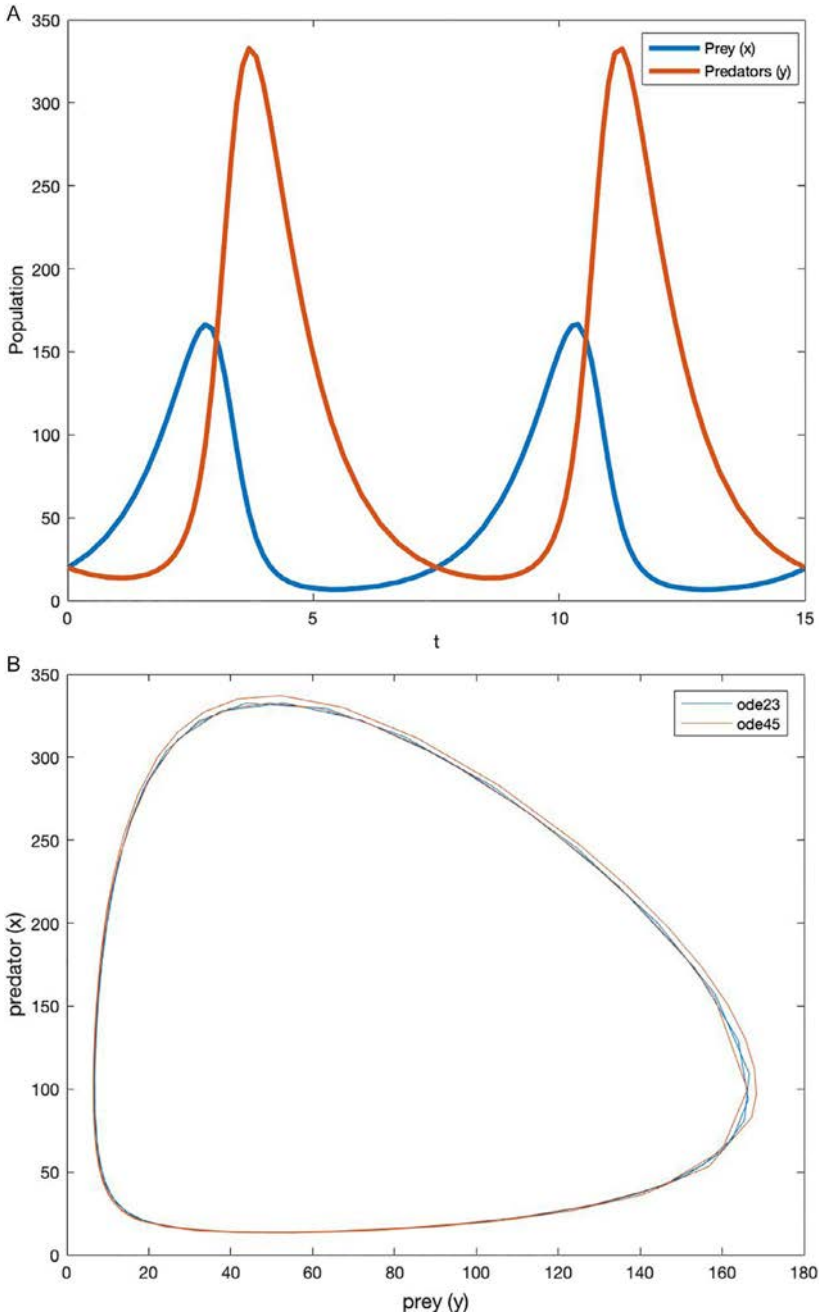


FIG. 2 Solutions to the Lotka-Volterra system of Eqs. (1)–(2) when a = 1 = d, b = 0.01 and c = 0.02. The initial conditions assume that $x_0 = y_0 = 20$ in this example and we run the system for t = [0,15]. (A) Predator/prey populations over time. (B) Phase plane diagram. In (A) we present the temporal profiles of x and y showing the periodic solutions of the system which are confirmed in the phase plane diagram in (B). When generating the phase plane diagram, we solve the system of equations using two different numerical solvers ode45 and ode23 illustrating the subtle differences present on even the simplest set of equations as (1)–(2).

Here, u and v represent the non-dimensionalised (or rescaled) versions of the variables x and y defined as $u = \frac{cx}{d}, v = \frac{by}{a}$ and $\alpha = d/a$. Eq. (3) has singular points at $u = v = 0$ and at $u = v = 1$ and integrating it directly we get $\alpha u + v - ln\, u^{\alpha} v = C$, where C is a constant with a minimum $1 + \alpha$ occurring when $u = v = 1$. It is a straightforward mathematical conjecture to show that, for a given constant $C > 1 + \alpha$, the trajectories in the phase plane are closed; illustrated for a given parameter regime in Fig. 2B.

An important aspect to note from Fig. 2A and B is that the asymptotic solutions, i.e., the solutions of the system that the system settles to as $t \to \infty$, depend on the model parameters. Even in the simplest case, setting $a = c = 1$; and the system only has two non-zero parameters that describe the strength of the competition between x and y, there are still three possible longtime solutions of the system: the coexistence steady state shown in Fig. 2A; or a limit case of this steady state where either the predator or the prey goes extinct in the longtime if the competition term from one species is much stronger (mathematically an order of magnitude larger) than the other. To determine these we need to set the derivatives in (1)–(2) to zero and solve the system of algebraic equations to determine the steady states and their stability; details on this method can be found in Appendix A of Murray (1989).

Why is the Lotka-Volterra system of equations so important when doing modeling? One of the reasons is that the classic SIR framework of Kermack-McKendrick is very similar to this system of equations.

Instead of having two competitor species, as the Lotka-Volterra system, the SIR framework splits the population cohort into three groups of Susceptible (S), Infected (I) and Recovered (R) populations. Analogous to the Lotka-Volterra system of equations, it is based on the principle of mass action that describes the rates of transition between these classes (Fig. 3A). In the simplest SIR model, the one that Kermack-McKendrick developed, births and deaths can be ignored and there is no waning of immunity, allowing only two possible transitions: spread of infection with transfer of individuals from the susceptible to the infected class (at a rate β) and recovery from the infection with transfer of individuals from the infected to the recovered class (at a rate γ). We note that different notations of the I and R compartments exist: I are sometimes called infected, infective or infectious while R are either recovered or removed; the original Kermack-McKendrick model referred to these as infected and recovered and comprised the system of Eqs. (4)–(6) below:

$$\frac{dS}{dt} = -\beta SI/N \tag{4}$$

$$\frac{dI}{dt} = \beta SI/N - \gamma I \tag{5}$$

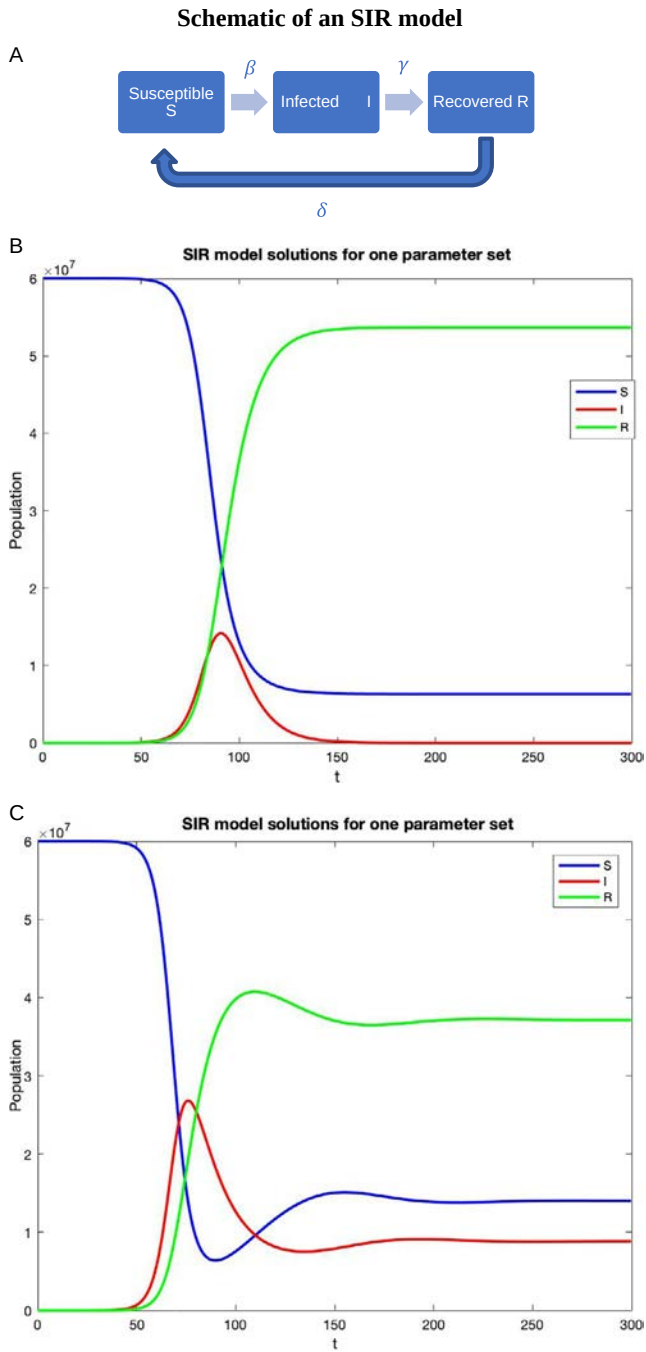


FIG. 3 (A) Schematic of SIR model. Schematic of a simple SIR model without and with waning of immunity (at a rate δ) describing the system of Eqs. (3)–(4) and numerical solutions for two set of parameters. (B and C) Model solutions for a parameter set. The model parameters are $\beta = 5 \times 10^{-9}$, $\gamma = 0.12$ in (B) and $\gamma = 0.07$ in (C) and $\delta = 0.0$ in (B) and $\delta = 1/60$ in (C). The initial conditions assume that $N(t=0) = 6 \times 10^7$, $I(t=0) = 10$, $R(t=0) = 0$, i.e., there are 10 infected people at the onset and that the basic reproduction number is $R0 = \frac{\beta}{\gamma} N(t=0) = 2.5$. We simulate the system for t = [0,300]. In (B) we present the temporal profiles showing the system settles to a steady state that clears the virus while in (C) the long-time solution shows the coexistence of the infected, susceptible and recovered populations (this is endemic epidemic).

$$\frac{dR}{dt} = \gamma I \tag{6}$$

Here, the parameter β represents the infectiousness rate, i.e., the rate at which the susceptible population transfers to the infected population, and γ is the rate of transfer from the infected to recovered population, with $1/\gamma$ describing the average length of the infectiousness period. It is important to note that this model assumes homogeneous mixing in the population which means that everyone interacts with equal probability with everyone else and discards situations where there is a heterogenous mixing across ages or settings. Details on the model adaptations necessary to account for such situations can be found in Chapters 3 and 7 of Keeling and Roheni (2008). Furthermore, there are some assumptions in deriving the transmission term βSI that are key for this model; details can be found for example in Chapter 2 of Keeling and Roheni (2008). Briefly, the parameter β can be defined and derived from two additional parameters c as the average number of contacts per unit of time and p as the probability of transmission per contact.

The end outcome, defined as the steady-state solution, for SIR models defined by Eqs. (4)–(6) is that the epidemic will either be cleared and the infected population eventually fully transferred to recovered population. Another possible outcome is that the epidemic will be sustained with coexistence of susceptible, infected and recovered populations (see conditions on this below, but we note that this is not possible in the case of Eqs. (4)–(6) where there is no population growth and no waning of immunity (Fig. 3). A whole theory of mathematical biology is concerned with exploring different constraints necessary for these outcomes and how dependent they are on the parameters β and γ; both Keeling and Roheni (2008) and Murray (1989) Murray (1989) have details of this. Next we revisit this briefly.

The key question when modeling epidemics is, given the values of β and γ and the initial conditions on the populations, will the infection spread or not, how will it behave over time and when will it start to decline and diminish. These questions can be formulated mathematically as the set of Eqs. (4)–(6) plus some initial conditions:

$$S(0) = S_0 > 0, I(0) = I_0 > 0, R(0) = 0$$

Considering Eq. (4), since $\frac{dS}{dt} < 0, S \leq S_0$. Hence we can see that if $S_0 < \frac{\gamma}{\beta}$ then

$$\frac{dI}{dt} = I(\beta S - \gamma) \leq 0, \forall t \geq 0$$

This suggests that $I_0 > I(t)$ as $t \to \infty$ and hence infection will die out in the longtime. But if $S_0 > \frac{\gamma}{\beta}$, then similarly we conclude that $I_0 < I(t)$ as $t \to \infty$, so an epidemic occurs and starts growing. In this sense an epidemic occurs if the infected population increases above its' initial value, i.e., if $I_0 < I(t)$ for some

$t > 0$. In modeling, this gives rise to a threshold phenomenon, i.e., depending on the threshold value of the ratio $\frac{\gamma}{\beta}$ we get two behavioral phenomena: epidemic or not. The inverse ratio $\frac{1}{\frac{\gamma}{\beta}} = \frac{\beta}{\gamma}$ represents the infection contact rate. We note that as epidemiological modeling has been developing over the years, so has the notation. For example, different modelers refer to the contact rate differently and sometime writing $\beta = cp$ or $\beta = -c\,log(1-p)$, where as described above, p represents the probability of risk of transmission upon contact and is related to viral load of the virus, while c represents the number of contacts per unit of time; hence in this case only c is referred to as contact rate.

Whichever notation we use, the interplay between the parameters β and γ leads to the notion of the basic reproduction number which is defined as $R_0 = \frac{\beta}{\gamma}$ that simply represents the number of secondary infections that emerge from one primary infection in a wholly susceptible population. By the definition, if more than one secondary infection is produced from the primary infection this simply means $R_0 > 1$. The basic reproduction number and the condition $R_0 > 1$ has been widely used and discussed in the media as a metric for describing the state of the COVID-19 epidemic and we will revisit this again in our COVID-19 specific case studies in Section 4.

While the basic reproduction number is an important metric to consider in a growing epidemic, this simple example shows that its value is most relevant at the onset of the epidemic, as it depends on the initial pool of susceptible people. As the epidemic develops and the pool of susceptible people changes, the basic reproduction number R_0becomes an effective reproduction number, often written as Reff, Re, Rt or just R. A detailed discussion of the differences between these highlighting what the effective reproduction number can tell us and what it can't in terms of a growing epidemic can be found in Vegvari et al. (n.d.).

It is also important to note that non-pharmaceutical interventions such imposing social distancing rules to reduce the number of contacts (reducing c) or wearing masks to reduce the risk of transmission during a contact (to reduce p) would reduce the transmission rate β. This can in turn reduce the effective reproduction number—we illustrate this in our case studies in Sections 4.1 and 4.2.

2.2.2 *Stochastic models: Adding stochasticity to the SIR framework, branching processes and individual- or agent-based models (IBMs/ABMs)*

Compartmental models with fixed parameters, such as the SIR framework in Section 2.2.2, can answer simple questions around initial epidemic growth and give parameter constraints on expected long-term solutions. But such models have a crucial limitation: since everyone in each compartment is assumed to be the same, these models ignore important aspects of social

interactions and heterogeneous behavior patterns. Various generalisations of the structure of the simple SIR model have been made over the years to address this limitation, which have included modifications of the model structure to include sub cohorts of the population (e.g., different age or risk-groups) or the addition of a mixing matrix to the infection rate beta to account for the interaction between such sub cohorts (see for example Kiss et al., 2006; Lloyd and May, 2001; Rohani et al., 2010). But an alternative means of addressing the assumption of heterogeneity within compartments is to use probability distribution functions for the flow rates between compartments, rather than assuming a fixed value. This allows for uncertainty and variability in the model parameters to be included in compartmental models. For example, the duration of the infectiousness period (described by the parameter gamma in Section 2.2.2) need not be treated as a fixed number but rather as a varying parameter, which takes different values according to a probability distribution. For large populations, this variability averages out, but for small populations, a stochastic treatment is required. Stochastic Differential Equations (SDEs) are one way to capture this variability where there is a natural limiting behavior of the mean field that yields a corresponding ODE system for large populations. Another, more flexible way to capture this variability is by explicitly modeling individuals rather than the population as a whole. This can be done using Agent- or Individual-Based models (ABMs/IBMs).

In ABMs, epidemics are modeled by creating a set of autonomous agents or individuals that follow certain rules and/or decisions and interact with each other in certain defined ways. ABMs allow individual contacts within a network to be modeled, and infectious disease spread within realistic synthetic population to be simulated (Eubank et al., 2004). In addition, agent-based models can more easily allow for uncertainty analysis by allowing the incorporation of stochastic effects, as well as the potential to have well-defined variation between individuals. While compartmental models encourage modelers to simplify the problem at hand, ABMs encourage modelers to think more deeply about the full system. This can pair well with model-driven data collection: rather than building a model simply to use the available data, this approach involves building the model first and then identifying what data gaps remain to be filled. When applied appropriately, this approach can have enormous benefits, since "one can only understand what one is able to build" (Dudai and Evers, 2014). While still a less common approach than compartmental modeling, there are still many examples of where it has been used successfully (Heesterbeek et al., 2015).

There are several disadvantages to using ABMs as well. First, vast stochasticity within ABMs can be a disadvantage: the results from ABMs almost always converge to those of compartmental models if averaged over a large enough numbers of individuals and/or over a large number of trajectories, and in cases where this is the desired outcome, a compartmental model may

be preferable because of its lower computational cost. Second, compared to compartmental models, ABMs are typically much more data-hungry, as they often have many more parameters to inform the behavior of individual agents. It is often hard enough to find sufficient data to parameterise compartmental models appropriately, much less ABMs. In such cases, the better approach is often to opt for a simple model based on available data rather than create a complex model based on many assumptions.

2.3 Challenges of modeling infectious diseases

Developing and using models to inform decision making has challenges. Some of these were highlighted in a series of challenge papers published as a special issue of Epidemics in 2015 led by a consortium of scientists (Lloyd-Smith et al., 2015). A follow-on collection of challenge papers focused on the challenges related to modeling the COVID-19 pandemic and planning for future pandemics is under preparation in late 2020 by another consortium of scientists within the Isaac Newton Institute.

The challenges of modeling infectious diseases can be grouped into four broad categories.

First, it is necessary to find good and reliable data to parametrise the model. Formulating mathematical models requires defining model variables and model parameters. The variables represent the key dynamic quantities that change in time (for example, the number of infected people or the number of disease-related deaths), while the parameters govern the ways in which the state of the model evolves over time. Defining the parameters of a model is one of the main tasks of the modeler, and it often involves a trade-off between simplicity and detail. For example, there are just two parameters (β and γ) in the SIR model presented in Section 2.2.2, while the more detailed ABM we describe in Section 4.2 includes over 500 parameters. Often, the values of the parameters will be informed by data, so a crucial first step in determining the structure and parameterisation of a model is to evaluate the extent of data available. But there are stark limitations on data that can inform the model parameters, especially when modeling new diseases or outbreaks such as COVID-19. This is one of the main challenges of good modeling. In the absence of reliable data sources, the modeler has various options: (a) use a less data-rich model (for example, a simpler SIR model instead of an ABM); (b) propagate the uncertainty stemming from the lack of data through the model to produce distributions of possible predictive outcomes, either by conducting a sensitivity analysis in which the model is simulated numerous times for different parameter values, or by conducting a more formalized statistical analysis.

The second challenge is around having a robust and efficient numerical algorithms to solve the model and produce reliable predictive outcomes. It is the responsibility of the modeler to assure that technically correct and

robust numerical techniques are used for calibration and prediction, and that uncertainty within the model is clearly communicated.

Third, there are challenges associated with developing simulation code that can withstand the tests of time and be reproducible by other users. The modeler needs to have good code etiquette with the code available as open source so it can be understood, reproduced, expanded, adapted and used by other users.

The fourth and final challenge is that models should be able to answer questions from policy decision makers in a timely and informative fashion. This has rarely been more pertinent than in the worldwide pandemic of COVID-19, when mathematical modeling was brought to the forefront of policy making and communication. In the next section, we provide an overview of these models.

3 Models for COVID-19

Since the beginning of the COVID-19 pandemic, mathematical modeling was widely used to help make decisions around the control of COVID-19 spread. COVID-19 presented a unique modeling challenge to the community due to (a) the extreme urgency of generating accurate predictions; (b) the quickly-evolving data; and (c) the large uncertainties, especially early on in the epidemic, around even basic aspects of transmission such as the reproduction number, latent period, and proportion of people who are asymptomatic.

3.1 Compartmental modeling and the SEIR framework: Overview and examples of COVID-19 models

The vast majority of the models that have proliferated in response to the COVID-19 pandemic have been compartmental models, due to their relatively simple requirements for development and the long-standing body of work using them, making them most accessible to epidemiologists. For example, Walker et al. (2020) adopted the basic SIR framework from Section 2.2.2 and used an age-structured stochastic "Susceptible, Exposed, Infectious, Recovered" (SEIR) model to determine the global impact of COVID-19 and the effect of various social distancing interventions to control transmission and reduce health system burden. Read et al. (2020) developed an SEIR model to estimate the basic reproduction number in Wuhan. Keeling et al. (2020) used one to look at the efficacy of contact tracing as a containment measure; and Dehning et al. (2020) used an SIR model to quantify the impact of intervention measures in Germany. In models such as those by Giordano et al. (2020) and Zhao and Chen (2020), compartments are further divided to provide more nuance in simulating progression through different disease states, and have been deployed to study the effects of various population-wide interventions such as social distancing and testing on COVID-19 transmission.

3.2 Agent-based models: Overview and examples of COVID-19 models

ABMs shot to public prominence on the basis of the dire predictions of the Ferguson et al. model, later named CovidSim (Ferguson et al., 2020). At the time, there had been only 6000 COVID-19 deaths globally, but this model predicted half a million deaths in the UK and over 2 million deaths in the US if strong interventions were not implemented. Like many COVID models, CovidSim was based on an earlier influenza model (Ferguson et al., 2006).

Another influential early COVID-19 ABM was that of Koo et al. (2020), who adapted an existing H1N1 model by Chao et al. (2010), in order to explore the impact of interventions on COVID transmission in Singapore. Other ABMs were developed to simulate the spread of COVID-19 transmission and the impact of social distancing measures in Australia (Chang et al., 2020) and the United States (Chao et al., 2020). Due to their flexibility, ABMs can be used to evaluate micro-level policies much more accurately than compartmental models, such as to evaluate the impact of social distancing and contact tracing (Aleta et al., 2020; Kretzschmar et al., 2020; Kucharski et al., 2020) and super-spreading (Lau et al., 2020). Since these models can account for the number of household and non-household contacts (Chao et al., 2020; Kretzschmar et al., 2020; Kucharski et al., 2020); the age and clustering of contacts within households (Aleta et al., 2020; Chao et al., 2020; Kucharski et al., 2020); and the microstructure in schools and workplace settings informed by census and time-use data (Aleta et al., 2020) they can be used to investigate detailed interventions with maximal realism.

3.3 Branching process models: Overview and examples of COVID-19 models

A third type of model, in some senses halfway in between a compartmental model and an ABM, is a branching process model. Although not widely used in the epidemiology community compared to the other two model types, branching process models saw considerable use during the COVID-19 epidemic since in some ways they combine with simplicity of a compartmental model with the detail and stochasticity of an ABM. Branching process models have also been used to investigate the impact of non-pharmaceutical intervention strategies (Hellewell et al., 2020; Peak et al., 2017). Compared to ABMs, which represent both infected and susceptible individuals and their interactions, branching models consider only infected individuals, and use probabilistic algorithms for determining how many new infections each infected individual causes. Although this allows for the incorporation of properties specific to the infected individual, it does not allow for a full treatment of the interactions between infected and susceptible individuals as in an ABM.

4 Applications of modeling of COVID-19: Three case studies

4.1 Case study 1: Application of SEIR-TTI model to the UK COVID-19 epidemic

4.1.1 Overview of SEIR-TTI

For this case study, we illustrate a new method for including the effects of Testing, contact-Tracing and Isolation (TTI) strategies in classic Susceptible-Exposed-Infected-Removed (SEIR) models. The SEIR-TTI model is a direct extension of the SEIR modeling framework that incorporates a probabilistic argument to show how contact tracing at the individual level can be reflected in aggregate on the population level.

4.1.2 SEIR-TTI methodology

Details of the mathematical framework behind the SEIR-TTI model can be found in Sturniolo et al. (2020). Briefly, the SEIR set-up extends the classic SIR model from Section 2.2.1 to include a cohort of individuals exposed (E) to the virus that have been infected with the virus but not yet infectious. A key parameter than describes the latent period when the pathogen reproduces within the host, but the viral load is too low to be categorised as susceptible (S) or infected (I). Hence an intermittent cohort E, and respective mathematical equation is needed to link these. Assuming the average length of the latency period is $1/\alpha$ the system of equations becomes:

$$\frac{dS}{dt} = -\beta cSI/N \tag{7}$$

$$\frac{dE}{dt} = \beta cSI/N - \alpha E \tag{8}$$

$$\frac{dI}{dt} = \alpha E - \gamma I \tag{9}$$

$$\frac{dR}{dt} = \gamma I \tag{10}$$

An important thing to note is that although SIR and SEIR models behave very similarly at steady state, i.e., have the same longtime solutions with the E cohort emerging as an intermediate cohort, the SEIR models have a slower growth rate. This is a consequence of the delayed process of development of infectiousness following the virus/pathogen entering the host system.

The SEIR model, we developed in Sturniolo et al. (2020) has similar structure to Eqs. (7)–(10). The novelty of our work is how we then layer this model with a model for tracing the population that is either exposed or infectious—the E and I cohorts. Existing models that have attempted to do this have incorporated this by asserting that a proportion of exposed individuals become quarantined or via reducing the transmission, sometime with a possible delay.

But neither of these concepts account for finding and tracing both exposed and infectious individuals that have different probabilities of risk of onward transmission. Our SEIR-TTI framework allows us to do that; further details can be found in Sturniolo et al. (2020). Importantly, we validated our SEIR-TTI ODEs based model against two mechanistic ABMs and showed good agreement at far less computational cost.

There are two key concepts in our SEIR-TTI framework:

a) the introduction of overlapping compartments within the SEIR modeling framework;
b) defining the transition rates for people who are traced.

The first of these is achieved by defining overlapping compartments to represent model states that are not mutually exclusive like the S, E, I and R compartments are. Hence, we allow for an individual within our SEIR model to belong in more than one category, e.g., be infected and contact-traced, or exposed and tested. Specifically, as illustrated in Fig. 4A we represent unconfined and isolated individuals simply by doubling the number of states, labeling S_U, E_U, I_U, R_U respectively the undiagnosed S, E, I and R compartments; and similarly, S_D, E_D, I_D, R_D the ones who have been diagnosed or otherwise distanced from the rest of the population, by for example home isolation or hospitalization.

The second point is more complex and explicitly explained in Sturniolo et al. (2020), so we will not go into details here. The key aspect is that we derive transition rates among overlapping compartments, by considering that individuals will be traced proportionally to how quickly the infectious individuals who originally infected them are, themselves, identified. People can be identified via testing, at a rate $\frac{1}{\theta}$, or via tracing at a rate η and success χ. We define a global tracing rate that depends on the probability of an individual of being traced.

In the next section, we give an illustration of how the SEIT-TTI model can be applied to predict the effective reproduction number as a combination of different test-trace strategies and how this changes the epidemic curve.

4.1.3 Application of SEIR-TTI to estimate R

We applied the SEIR-TTI model to simulate the spread of COVID-19 in the UK. We began by creating a total population of 67 million individuals, with 100,000 individuals initially infected (i.e., $I(0) = 100{,}000$). To model the dynamics of transmission, we define the probability of transmission per contact per day to be $P = 0.033$, the number of contacts per day to be $c = 13$, the rate of exposed people becoming infectious to be $\alpha = 0.2$ days^{-1}; and the length of the infectiousness period (recovery time) to be 7 days. These values result in a basic reproduction number of $R_0 = 3$ at the onset.

Assuming the success of tracing $\chi = 0.5$, we then simulate different combinations of tracing and testing levels and project the value of R after 30 days

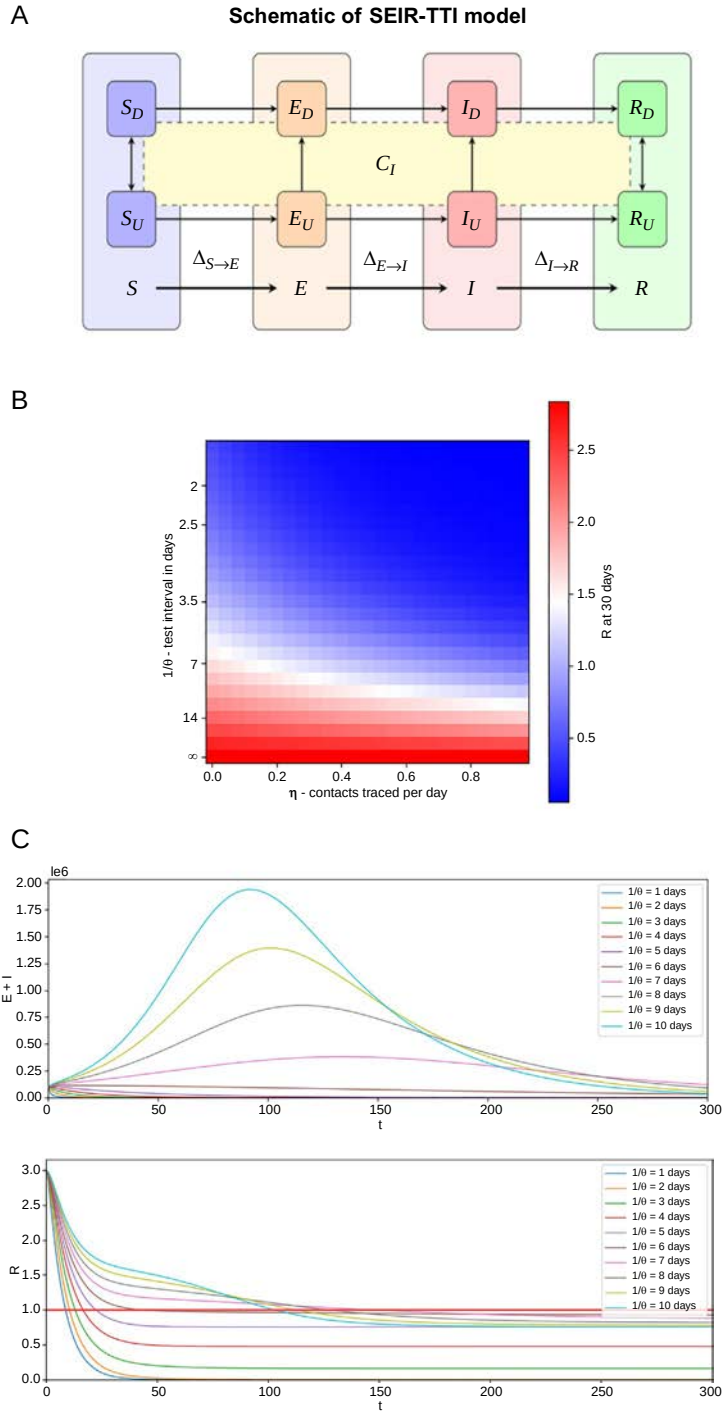


FIG. 4 See figure legend on opposite page.

(Fig. 4B). Our results suggest that sufficiently effective tracing and sufficiently frequent tracing are necessary to control the virus and keep $R<1$ and within the blue region in Fig. 4B. But we also found a non-linear relationship between testing and tracing, implying it would be possible to control the virus with less effective tracing and more frequest testing, or vice versa (Fig. 4C). The epidemic trajectory looks different for different combinations of test-trace levels; in Fig. 4C we illustrate these for different testing levels assuming a tracing rate of $\eta=0.5$.

4.2 Case study 2: Application of Covasim to the UK COVID-19 epidemic

4.2.1 Overview of Covasim

For this case study, we illustrate the application of an open-source agent-based model called Covasim (COVID-19 Agent-based Simulator) developed by the Institute for Disease Modeling. The methodology of the model is contained in Kerr et al. (2020a) with further development and implementation details available at http://docs.covasim.org. Since the onset of the pandemic, Covasim has been applied across a number of studies (Cohen et al., 2020; Kerr et al., 2020b; Panovska-Griffiths et al., 2020a,b; Stuart et al., 2020). Here, we illustrate an application of Covasim to answer questions around the COVID-19 epidemic in the UK.

4.2.2 Covasim methodology

Covasim is an agent-based model with individuals modeled at different stages of their infectiousness, as susceptible to the virus, exposed to it, infected, recovered, or dead. Infectious individuals are additionally categorized as asymptomatic, presymptomatic (before the viral shedding has begun) or with mild, severe or critical symptoms, as illustrated in Fig. 5A.

Covasim is coded in Python with default parameters that are regularly updated based on ongoing literature reviews. It is equipped with demographic data on population age structures and household sizes for different countries,

FIG. 4—CONT'D (A) Schematic of the SEIR-TTI model. Application of the SEIR-TTI model. (B) R phase plane for test-trace levels. Phase plane plot of the effective reproduction number R after 30 days of running the model for different tracing (x-axis) and frequency of testing (y-axis) levels. Larger R, implying higher numbers of new infections is shown in red, while lower values in white, and the blue region representing area where the resurgence of COVID-19 is controlled ($R<1$) with combinations of adequate test-trace strategy. (C and D) Model predictions for different testing levels. SEIR-TTI model predictions of the Exposed and Infected populations (C) and R (D) over time when tracing level is 50% and frequency of testing increases. *Panel A schematic of the SEIR-TTI model reproduced from Sturniolo S, Waites W, Colbourn T, Manheim D, Panovska-Griffiths J. Testing, tracing and isolation in compartmental models." medRxiv. (2020) preprint doi: https:/doi.org/10.1101/2020.05.14.20101808, with permissions.*

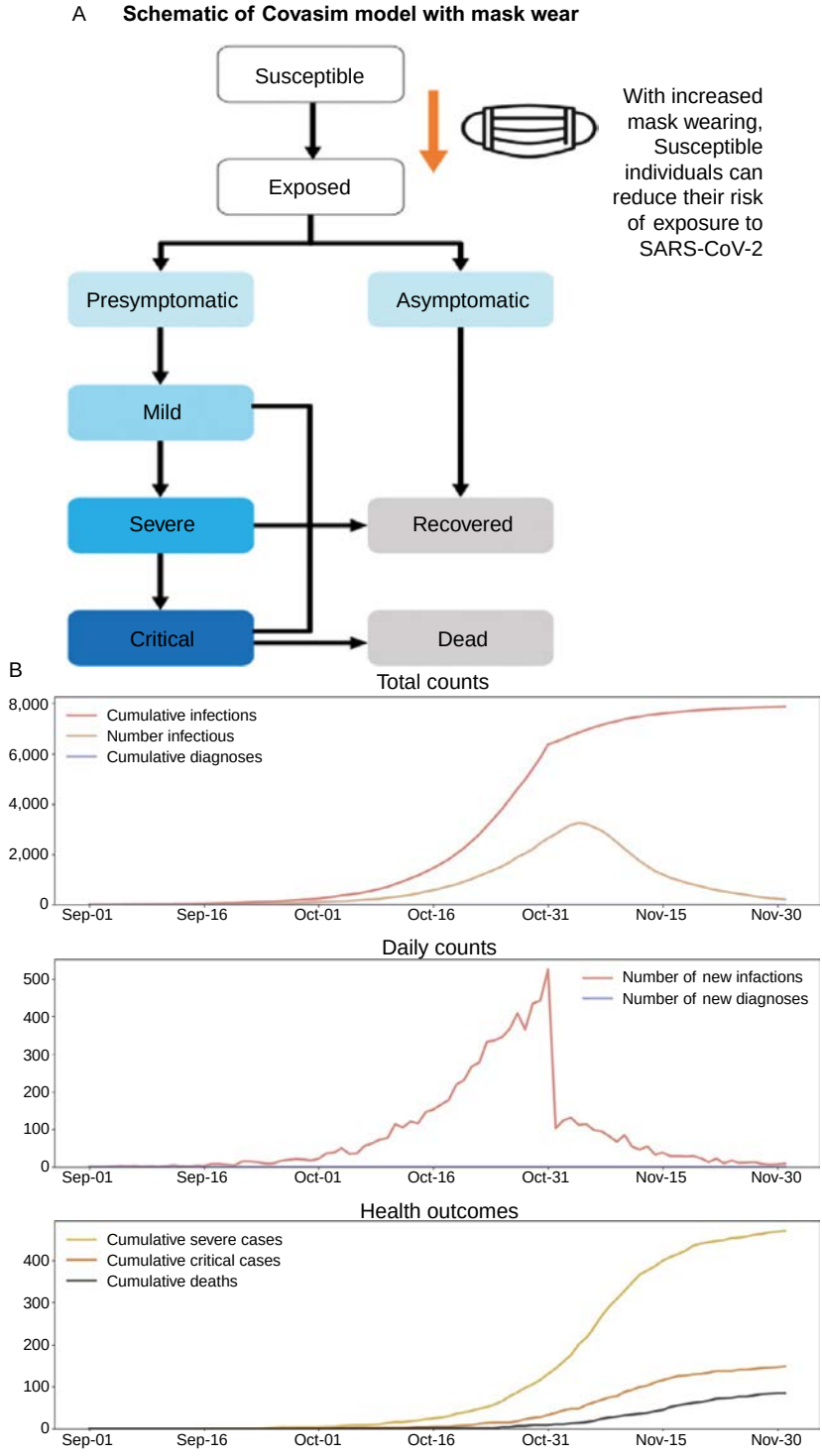


FIG. 5 Illustration of the Covasim model. (A) Schematic of Covasim model with mask usage showing the disease-related states that individuals can be in. (B) Typical model outcomes from Covasim. Model outcomes from a simple example of running Covasim with the defaults values

(Continued)

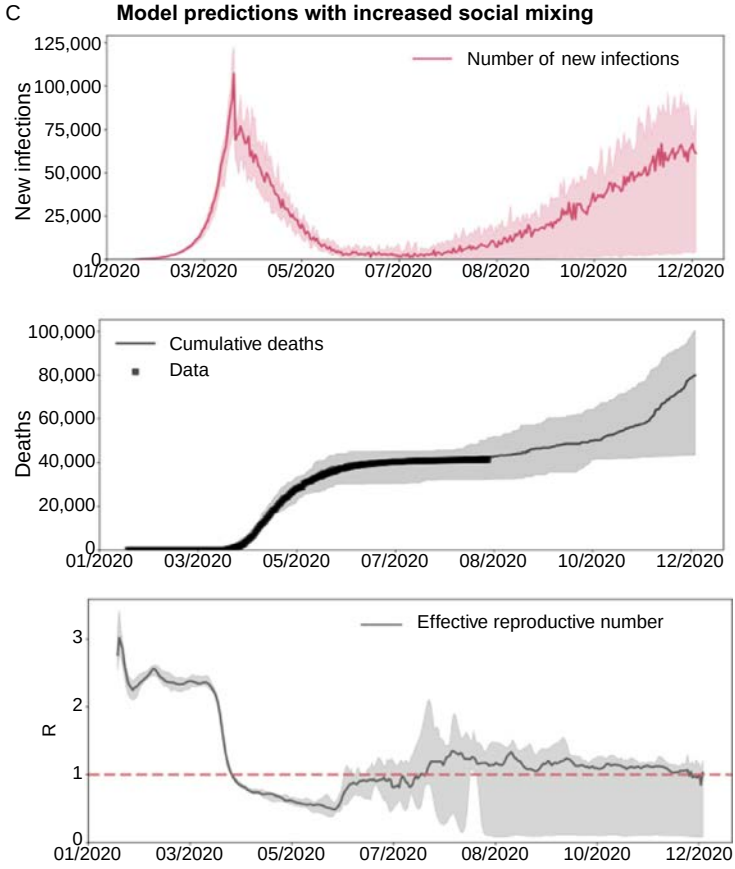


FIG. 5—CONT'D and projecting total and daily counts as well as health outcomes. (C and D) Model predictions with increased social mixing and with reduced social mixing. Outcomes from an application of Covasim to the UK epidemic, calibrated to the reported COVID-19 infections and deaths until August 28, 2020 with model parameters as per (Panovska-Griffits et al., 2020). Medians across 12 simulations are indicated by solid lines and 10% and 90% quantiles by shading.

with individuals interacting across four contact network layers for schools, workplaces, households and community settings. Different epidemics can be modeled by adjusting context-specific parameters, including rates of testing, tracing, isolation compliance, and other non-pharmeceutical interventions. Transmission occurs during contacts between infectious and susceptible individuals, according to a parameter β, which is comparable to the infection rate β in the simple SIR from Section 2.2.2, but stratified across contact network layers and across different risk groups.

Covasim has a flexible framework that can be adapted across settings and for specific conditions. To run a simulation, we can use the default parameters

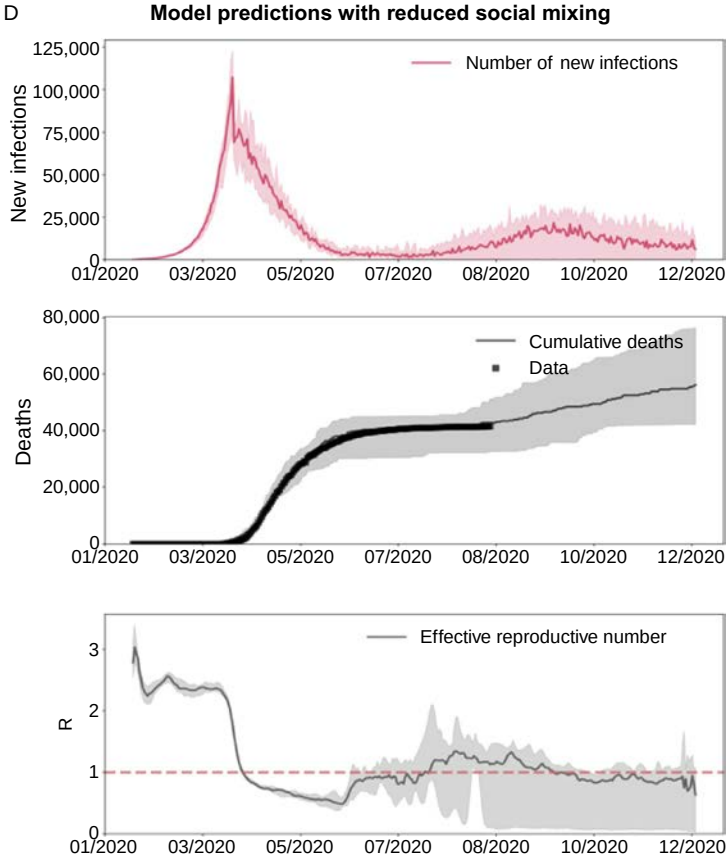


FIG. 5—CONT'D

to generate a population of agents who interact over the four contact networks layers and then change model parameters to calibrate the model to fit the specific epidemic. A simple modification would be to specify the time for which we are interested to run the simulation, or change a model parameter.

In Fig. 5B we show the outcomes of a simple simulation in Covasim with the simulation ran between September 1, 2020 and December 1, 2020 with a change in the infection rate β from 0.016 to 0.2 from November 1, 2020. The simulation shows an increase in infection with increased β, i.e., showcases the impact of a single parameter on the predicted outcomes from November 1, 2020 when β was changed (Fig. 5B).

This example can be built on and expanded within Covasim to simulate and evaluate a large number of interventions. These include non-pharmaceutical

interventions, such as reduced social mixing, hygiene measures and use of face coverings (all of which can be simulated by respective changes in the contact rate c or the infection rate β), different testing interventions, such as testing people with COVID-19 symptoms or asymptomatic testing (for which specific parameters are defined within Covasim), as well as contact tracing and isolation strategies (again changing specific parameters for these within Covasim). It is also possible to simulate different vaccination strategies with Covasim.

The example illustrated above simulates the most basic intervention in Covasim—reducing transmissibility starting on a given day. Transmissibility can mean a reduction in transmissibility per contact (such as through wearing face coverings or maintaining social distance), or a reduction in the number of contacts at households, school, work, or community settings. For example, school closures can be modeled either by setting both of these to 0, while partial closures can be modeled by scaled reductions in either transmissibility per contact or the number of contacts. Our published work (Panovska-Griffiths et al., 2020a) has modeled the impact of reopening school and society on the UK COVID-19 epidemic in a similar, but more complex way to this simple example.

Importantly, Test-Trace-Isolate (TTI) interventions can also be modeled in Covasim. Testing can be modeled either by specifying the probabilities of receiving a test on each day for people with different risk factors and levels of symptoms; or by specifying the number of tests performed on each day directly. Tracing assigns a probability that a contact of a person testing positive can be traced, and allocates a time that it takes to identify and notify contacts. Isolation of persons testing positive and their contacts is the key aspect by which TTI interventions can reduce transmission. In Covasim, people diagnosed with COVID-19 are isolated with an assigned adherence level and a period duration of isolation. In the next section, we illustrate a specific application of Covasim on the UK epidemic during 2020.

4.2.3 Application of Covasim to the UK epidemic

To apply Covasim to the UK epidemic (Panovska-Griffiths et al., 2020a,b), we used the default parameters and generated a population of 100,000 agents across the household, schools, workplaces and community networks. We then calibrated the model outcomes by performing an automated search for the optimal values of the number of infected people on 21 January 2020 (when the UK first COVID-19 case was confirmed), the per-contact transmission probability and the daily testing probabilities for individuals with and without COVID-19 symptoms during May, June, July, August and September (until September 26, 2020). The optimal values determined were the ones that minimised the sum of squared differences between the model's estimates of

confirmed cases and deaths and data on these same two indicators between January 21, 2020 and August 28, 2020. The data we compared against was collated from the UK government's COVID-19 dashboard (https://coronavirus.data.gov.uk). These particular parameters were selected as the most important to estimate because of the considerable uncertainties around them. Details of the exact methodology can be found in Kerr et al. (2020a) and Panovska-Griffiths et al. (2020a,b) and the code used to run all simulations contained in here is available from https://github.com/Jasminapg/Covid-19-Analysis.

We included policies around mask usage and TTI interventions that were part of the policy recommendations in the UK at the time of writing. Specifically, under the policy on masks in September of 2020, face coverings were mandatory in parts of community, such as public transport or in shops, and were recommended in secondary schools from September 1, 2020, but were not mandatory in workplaces. To simulate the impact of the masks policy, we defined effective coverage (a measure for effectiveness of masks) as the product of efficacy of masks (efficacy) and adherence to wearing them (coverage). A systematic review of the efficacy of face coverings suggested that taking in consideration different types of masks, their average efficacy is within the range of 11%–60% (Panovska-Griffits et al., 2020b). To simulate different face masks policy we then reduced the transmission probability of relevant contact network layers by the amount of effective coverage. For example, if masks are worn in 50% of the community settings with efficacy of 60%, then the effective coverage is 30%. To explore the impact of this policy in the model, we then reduced the transmission probability in the community contact network layer by 0.30. To account for masks additionally worn by 50% of those in school (i.e., only secondary school students) we also reduced the transmission risk in this layer by 30% (assuming again efficacy of 60%). We assumed that 60% of workforce were returning to work in September 2020 with the rest of the workforce working from home.

In addition to masks policy, since May 28, 2020 the key non-pharmaceutical intervention in the UK has been the Test-Trace-Isolate (TTI) strategy. For the testing part, the daily testing probabilities for symptomatic and asymptomatic people were fitted between May and September 26, 2020. For the level of contact tracing we collated the publicly available weekly data from NHS Test and Trace reports (NHS, 2020) between May 28, 2020 (when the program started) and September 26, 2020. To generate an average number, we multiplied the percentage of people testing positive that were interviewed, the percentage of those reporting contacts and the percentage of contacts that were traced to generate an overall percentage for contacts of those tested positive that were traced. We derived average monthly levels of contract tracing to be 43% for June, 47% for July, 45% for August and 50% for September (until September 26, 2020). For the isolation part of the TTI strategy, we assumed that 90% of people who are required to isolate do so.

Under these assumptions, the model predicted new infections, cumulative deaths and the effective reproduction number R, as shown in Fig. 5C. To compare these, we additionally simulated a scenario where only 40% of the workforce goes back to work from September 1, 2020 (instead of 60%) and there is additional 20% reduction in transmission in community (i.e., β for this layer was set to 0.4 instead of 0.6 since September 1, 2020). The results are shown in Fig. 5D.

This scenario analysis suggested that allowing fewer people to go back to work and reducing the transmission risk in the community, while keeping schools open, would result in a smaller resurgence of the COVID-19 in the latter part of 2020 than if social mixing was higher (comparing Fig. 5C and D). Fig. 5C and D show the median projections as solid lines, and the range across 12 simulations in shaded area. Although there is an obvious difference in the pattern of the projected epidemic—increasing with less stringent assumption in Fig. 5C and able to be controlled under more stringent assumptions in Fig. 5D—the shaded area highlights the wide range of possibilities across the simulations. Therefore it is important to note a level of uncertainty within the model projections.

4.3 Case study 3: Application of rule-based modeling

4.3.1 Overview

For this case study, we show how Rule-Based modeling (RBM) as a technique from computational molecular biology (Danos and Laneve, 2004) can be applied to the COVID-19 pandemic. Details of the methodology illustrated here can be found in Waites et al. (2020). This technique is based on chemical equations, and hence is similar to compartmental models of infectious disease but generalises reactions in two important ways: it allows arbitrary subsets of the population to be specified in rules, and it allows bonds to be formed between individuals.

4.3.2 Rule-based modeling methodology

One of the characteristics of compartmental models is that they experience rapid combinatorial explosion when adding features. For example, when considering the effect of face masks in the classic SIR model, this can be done via changes in the parameter β, to assume that the effect of masks is to reduce the transmission risk of susceptible people become infected; or to double the number of compartments, e.g., susceptible individuals with and without masks, infectious ones the same, and so on. Introducing vaccines, where an individual can be vaccinated or not, means that another doubling of compartments is required. Adding two features has thus caused the number of compartments to expand from the original 3–12. This explosion in the number of compartments clearly implies an explosion in the number

of transitions: one $I \rightarrow R$ transition is now three, the quadratic $S+I \rightarrow I+I$ transition now requires no less than 16 to completely specify. Each of these rates needs data so using large-scale compartmental models in scarce data scenarios can lead to a large number of assumptions and hence large uncertainty of the predictive modeling.

RBM gives an alternative to this by defining an agent. This is not the kind of agent found in agent-based modeling, but agent by analogy with reagent. It has three internal states: disease progression state, wearing a mask or not, and a binding site for a vaccine:

```
%agent: Person(covid{s i r} mask{y n} vax{y n})
```

Now, if we want to refer to *any* person, we simply write, Person(). If we want to refer to infectious people, we write, Person(covid{i}). If we want to refer to those people who are susceptible and wearing a mask, we can write, Person(covid{s}, mask{y}). We can write the recovery, or removal rule as,

```
Person(covid{i}) -> Person(covid{r}) @ gamma
```

We note that not only is this representation simple, but it is the model, verbatim, meaning that precisely what is written above is provided to the simulator.

The infection rules are somewhat more complicated. Infection involves an infectious person and an unvaccinated susceptible person and infection happens at different rates depending on whether masks are worn or not. So, we have four rules:

```
  Person(covid{i}, mask{n}), Person(covid{s}, mask{n}, vax[.]) ->
  Person(covid{i}, mask{n}), Person(covid{i}, mask{n), vax[.]) @
beta_nn
  Person(covid{i}, mask{y}), Person(covid{s}, mask{n}, vax[.]) ->
  Person(covid{i}, mask{y}), Person(covid{i}, mask{n), vax[.]) @
beta_yn
  Person(covid{i}, mask{n}), Person(covid{s}, mask{y}, vax[.]) ->
  Person(covid{i}, mask{n}). Person(covid{i}, mask{y), vax[.]) @
beta_ny
  Person(covid{i}, mask{y}), Person(covid{s}, mask{y}, vax[.]) ->
  Person(covid{i}, mask{y}), Person(covid{i}, mask{y), vax[.]) @
beta_yy
```

These rules are all nearly identical: an infectious person and a susceptible one come in, two infectious ones come out, for different combinations of masks or no.

The notation `vax[.]` means that there is no vaccine bound to the person's vaccine receptor. What does it mean for a vaccine to become bound? To make use of this, we require another kind of agent. This is the

second feature of RBM that does not have an equivalent in compartmental models. This agent is not a compartment, it does not correspond to a subset of the population but individuals become vaccinated by forming a bond with an available vaccine:

```
Person(vax[.]), Vaccine(p[.]) ->
Person(vax[1]), Vaccine(p[1]) @ vaxrate
```

Within the rule-based modeling framework we also need to specify initial conditions, can model wearing masks and introduce social dynamics, can model different testing and vaccinating strategies. Details of how we do this and simple examples can be found in Waites et al. (2020).

The simulation can in principle be done in a deterministic way and it is possible to produce a system of ordinary differential equations (ODEs) approximated by the mean trajectory of the model when simulated using a stochastic technique such as Gillespie's algorithm. Different software can be used to run the simulations, e.g., KaDE can produce Matlab or Octave code for the differential equations, and KaSim simulates the model stochastically. The expressiveness of the rule-based modeling language makes it very easy to write a rule-based model that results in an unreasonably large (possibly infinite) system of ODEs and for this reason stochastic simulation is normally used.

Next we apply this technique to a question related to testing in resource constraint settings.

4.3.3 Application of RBM to LMICs

We consider a scenario of modeling testing strategies against COVID-19 transmission to answer the question: how do we optimally allocate tests? We focus narrowly on surveillance testing at a relatively high rate and a relatively low rate, aiming to explore if non-pharmaceutical severe social distancing (lockdown)-like intervention is triggered on the rate of positive tests, how much surveillance is required, and how accurate the tests should be. The answer that we get is that this style of outbreak management is relatively insensitive to the accuracy of the tests, and even a low level of surveillance testing will suffice for these purposes.

The setup is as follows. We augment a transmission model with testing. Tests are discrete entities, much like the vaccines in the toy example above. Individuals in the population have two internal states related to testing: a state indicating the correct result for a perfect test, and a state indicating the actual result, for the modeled tests have a certain sensitivity and specificity. The relevant parts of the agent definitions are,

```
%agent: Person(test{p n} result{x p n})
%agent: Test(used{y n})
```

The rules relating to testing are straightforward. If an individual is not in a possession of a result, and their test site is not bound, an unused test may bind to it:

```
Test(used[.]{n}), Person(test[.], result{x}) ->
Test(used[1]{y}), Person(test[1], result{x}) @ testRate
```

The individual in possession of a test receives a result after some time,

```
  Person(test[_]{p}, result{x}) -> Person(test[_]{p}, result{p}) @
r*resultRate
  Person(test[_]{p}, result{x}) -> Person(test[_]{p}, result{n}) @
(1-r)*resultRate
  Person(test[_]{n}, result{x}) -> Person(test[_]{n}, result{n}) @
s*resultRate
  Person(test[_]{n}, result{x}) -> Person(test[_]{n}, result{p}) @
(1-s)*resultRate
```

And the result that they receive is correct or incorrect according to the characteristic sensitivity (recall) and specificity of the test. Finally, an individual becomes eligible for retesting at some rate,

```
Test(used[1]), Person(test[1]) -> Test(used[.]), Person(test[.]) @
retest
```

From this, we compute the rate of positive tests occurring in a window of time given by the retest rate. This is simply the ratio of the number of individuals with positive results bound to a test to the total number of individuals bound to a test with any result:

```
%obs: TPR
|Person(test[_], result{p})| /
( |Person(test[_], result{p})| + |Person(test[_], result{n})| )
```

We use this quantity in a perturbation: when it increases above 2%, a non-pharmaceutical intervention occurs that reduces the contact rate by half. When it falls again below 1%, this intervention is removed.

Our results shown in Fig. 6A suggest that, with a high level of surveillance testing—perhaps unreasonably high where each individual in the population could expect to be tested on average once a year—we see a clearly defined cycle of applying and releasing the non-pharmaceutical intervention. In contrast, with testing an order of magnitude lower as shown in Fig. 6B, where only 10% of the population can expect to be tested in a year, our results suggest this may not be necessary.

As with the previous model, it is useful to show both the mean (solids lines in Figs. 6A and B) and the range of the simulations, hence highlighting the uncertainty of the predictions.

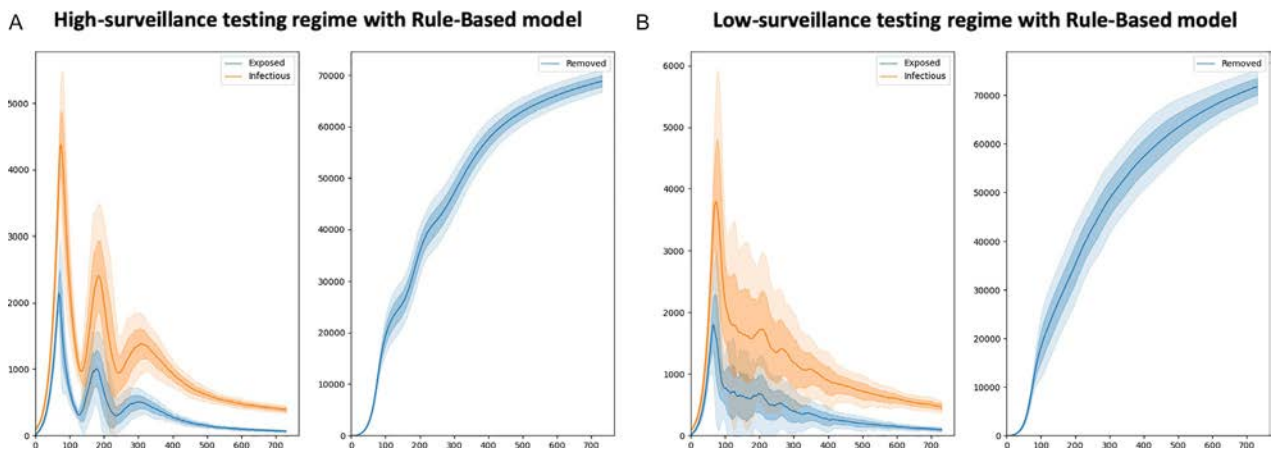


FIG. 6 Application of the Rule-Based Model. (A) High surveillance testing regime. Illustration of the model outcomes under a high-surveillance testing regime showing the clear period solutions of the model. (B) Low-surveillance testing regime. Illustration of the model outcomes under a low-surveillance testing regime showing the higher uncertainty in the model projections.

5 Conclusions

In this chapter, we aimed to highlight the need for and importance of mathematical modeling of infectious diseases, illustrating this with three case studies of how we modeled the COVID-19 pandemic during 2020. We discussed a number of issues related to what modeling is, what it can and cannot do, addressing the issues of model validity, robustness, calibration and different frameworks. By illustrating three conceptually different models applied for the same purpose—to better understand COVID-19 transmission during 2020—we showcased the power of cross-disciplinary and cross-methodologies research in the midst of a pandemic.

As discussed in greater depth in Section 1.2, mathematical modeling can be very useful in evaluating possible interventions and predicting future epidemic curves, but models need to be built well and designed to answer specific questions. There are no models that fit all the questions or can give all the answers; rather, each individual model, assuming it is fit for purpose, can contribute to a collective goal of decision support.

An important development in modeling in infectious diseases over the last 30 years has been the growth in inter- and multi-disciplinary collaborations of mathematicians, computer scientists, and data analysts with biological and medical researchers and policy decision makers. For future applicability and ability of modeling to inform decision making, this cross-disciplinary work needs to continue.

Addressing heterogeneity in both the model framework and within the data is crucial for real-time decision modeling. But behavioral heterogeneities that play an important role in describing infectious disease dynamics and designing and validating infectious diseases models remain challenging. Different modeling frameworks, as illustrated in Section 4, are amenable to modeling infectious diseases, and in this case COVID-19 spread and different interventions. Each has strengths and limitations, as outlined in Section 2.3 and it is important to remain aware of these when modeling real-life epidemics. It is especially important to identify the limits of predictability and potential uncertainty of the model predictions when discussing with policy decision makers.

In summary, modeling remains a very useful technique to support policy decision making and this chapter has included perspectives on modeling of infectious diseases, highlighting different frameworks for modeling COVID-19 and illustrating some of the models that our groups have developed and applied during the COVID-19 pandemic.

Acknowledgments

We thank all contributors to the development and application of the three models we discussed here. For the SEIR-TTI model we thank Simone Sturniolo, David Manheim and Tim Colbourn. For the Covasim model we thank a large team of collaborators: at the Institute for Disease Modeling, Daniel J. Klein, Dina Mistry, Brittany Hagedorn, Katherine

Rosenfeld, Prashanth Selvaraj, Rafael Núñez, Gregory Hart, Carrie Bennette, Marita Zimmermann, Assaf Oron, Dennis Chao, Michael Famulare, and Lauren George; at GitHub, Michał Jastrzębski, Will Fitzgerald, Cory Gwin, Julian Nadeau, Hamel Husain, Rasmus Wriedt Larsen, Aditya Sharad, and Oege de Moor; at Microsoft, William Chen, Scott Ayers, and Rolf Harms; and at the Burnet Institute, Romesh Abeysuriya, Nick Scott, Anna Palmer, Dominic Delport, and Sherrie Kelly. For the RBM model, we thank Matteo Cavalieri and Vincent Danos.

References

Aleta, A., Martin-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N.E., Halloran, M.E., Longini Jr., I.M., Merler, S., Pentland, A., Vespignani, A., Moro, E., Moreno, Y, 2020. Modelling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. Nat. Hum. Behav. 4 (9), 964–971. https://doi.org/10.1038/s41562-020-0931-9.

Anderson, R.M., May, R.M., Anderson, B., 1992. Infectious Diseases of Humans: Dynamics and Control. Wiley Online Library. vol. 28.

Andrianakis, I., Vernon, I.R., McCreesh, N., McKinley, T.J., Oakley, J.E., Nsubuga, R.N., et al., 2015. Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda. PLoS Comput. Biol. 11 (1), e1003968.

Bernoulli, D., 1766. Essai d'une nouvelle analyse de la mortalite causee par la petite verole. Mem. Math. Phys. Acad. Roy. Sci., Paris 1 (Reprinted in: L.P. Bouckaert, B.L. van der Waerden (Eds.), Die Werke von Daniel Bernoulli, Bd. 2 Analysis und Wahrscheinlichkeitsrechnung, Birkh€auser, Basel, 1982, p. 235. English translation entitled 'An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it' in: L. Bradley, Smallpox Inoculation: An Eighteenth Century Mathematical Controversy, Adult Education Department, Nottingham, 1971, p. 21. Reprinted in: S. Haberman, T.A. Sibbett (Eds.) History of Actuarial Science, vol. VIII, Multiple Decrement and Multiple State Models, William Pickering, London, 1995, p. 1.).

Chang, S.L., Harding, N., Zachreson, C., Cliff, O.M., Prokopenko, M., 2020. Modelling Transmission and Control of the COVID-19 Pandemic in Australia. ArXiv 2003. 10218 [Cs, q-Bio], May http://arxiv.org/abs/2003.10218.

Chao, D.L., Halloran, M.E., Obenchain, V.J., Longini Jr., I.M., 2010. FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model. PLoS Comput. Biol. 6 (1), e1000656. https://doi.org/10.1371/journal.pcbi.1000656.

Chao, D.L., Oron, A.P., Srikrishna, D., Famulare, M., 2020. Modeling layered non-pharmaceutical interventions against SARS-CoV-2 in the United States with Corvid. Epidemiology. medRxiv.

Cohen, J.A., Mistry, D., Kerr, C.C., Klein, D.J., 2020. Schools are not islands: Balancing COVID-19 risk and educational benefits using structural and temporal countermeasures. medRxiv. preprint. https://doi.org/10.1101/2020.09.08.20190942.

Danos, V., Laneve, C., 2004. Formal molecular biology. Theoretical Computer Science 325 (1), 69–110. https://doi.org/10.1016/j.tcs.2004.03.065.

Dehning, J., Zierenberg, J., Spitzner, F.P., Wibral, M., Neto, J.P., Wilczek, M., Priesemann, V., 2020. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science 369 (6500), eabb9789. https://doi.org/10.1126/science.abb9789.

Dietz, K., Heesterbeek, J.A.P., 2002. Danie; Bernoulli's epidemiological model revisited. Math. Biosci. 180, 1–21.

Dowe, D.L., Gardner, S., Oppy, G., 2007. Bayes not Bust! Why Simplicity is no problem for Bayesians. Br. J. Philos. Sci 58 (4), 709–754.

Dudai, Y., Evers, K., 2014. To simulate or not to simulate: what are the questions? Neuron 84 (2), 254–261. https://doi.org/10.1016/j.neuron.2014.09.031.

Eubank, S., Guclu, H., Anil Kumar, V.S., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184. https://doi.org/10.1038/nature02541.

Ferguson, N.M., Cummings, D.A.T., Fraser, C., Cajka, J.C., Cooley, P.C., Burke, D.S., 2006. Strategies for mitigating an influenza pandemic. Nature 442 (7101), 448–452.

Ferguson, N.M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., et al., 2020. Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College COVID-19 Response Team, London, p. 16.

Forster, M., 2002. Predictive accuracy as an achievable goal of science. Philos. Sci. 69, S124–S134.

Forster, M., Sober, E., 1994. How to tell when simpler, more unified, or less ad-hoc theories will provide more accurate predictions. Br. J. Philos. Sci. 45, 1–35. MR1277464.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., Colaneri, M., 2020. Modelling the COVID-19 Epidemic and Implementation of Population-Wide Interventions in Italy. Nat. Med., 1–6. https://doi.org/10.1038/s41591-020-0883-7.

Heesterbeek, H., Anderson, R.M., Andreasen, V., et al., 2015. Modeling infectious disease dynamics in the complex landscape of global health. Science 347 (6227), aaa4339. https://doi.org/10.1126/science.aaa4339.

Hellewell, J., Abbott, S., Gimma, A., Bosse, N.I., Jarvis, C.I., Russell, T.W., Munday, J.D., Kucharski, A.J., Edmunds, W.J., 2020. Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Funk, S., Eggo, R.M., Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob. Health 8 (4), e488–e496. https://doi.org/10.1016/S2214-109X(20)30074-7. (Erratum in: Lancet Glob. Health., 2020.).

Hitchcock, C., Sober, E., 2004. Prediction versus accommodation and the risk of overfitting. Br. J. Philos. Sci. 55, 1–34.

Keeling, M.J., Roheni, P., 2008. Modelling Infectious Diseases in Humans and Animals. Princeton University Press. ISBN: 978-0-691-116174.

Keeling, M.J., Hollingsworth, T.D., Read, J.M., 2020. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J. Epidemiol. Community Health 74 (10), 861–866. https://doi.org/10.1136/jech-2020-214051.

Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. J. R.Stat. Soc. Ser. B Stat. Methodol. 63 (3), 425–464.

Kermack, W., McKendrick, A., 1927. A contribution to the mathematical theory of epidemics. Philos. Trans. R. Soc. Lond. A 115, 13–23.

Kermack, K.O., McKendrick, A.G., 1932. Contributions to the mathematical theory of epidemics - ii. The problem of endemicity. Philos. Trans. R. Soc. Lond. A 138, 55–83.

Kermack, K.O., McKendrick, A.G., 1933. Contributions to the mathematical theory of epidemics - iii. Further studies of the problem of endemicity. Philos. Trans. R. Soc. Lond. A.

Kermack, W.O., McKendrick, A.G., 1937. Contributions to the mathematical theory of epidemics: IV. Analysis of experimental epidemics of the virus disease mouse ectromelia. J. Hyg. 37, 172–187.

Kermack, W.O., McKendrick, A.G., 1939. Contributions to the mathematical theory of epidemics: V. Analysis of experimental epidemics of mouse-typhoid; a bacterial disease conferring incomplete immunity. J. Hyg. 39, 271–288.

Kerr, C.C., Stuart, R.M., Mistry, D., et al., 2020a. Covasim: an agent-based model of COVID-19 dynamics and interventions. medRxiv. https://doi.org/10.1101/2020.05.10.20097469.

Kerr, C.C., Mistry, D., Stuart, R.M., et al., 2020b. Controling COVID-19 via test-trace-quarantine. medRxiv. preprint doi https://doi.org/10.1101/2020.07.15.20154765.

Kiss, I.Z., Green, D.M., Kao, R.R., 2006. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. J. R. Soc. Interface 3, 669–677. https://doi.org/10.1098/rsif.2006.0129.

Koo, J.R., Cook, A.R., Park, M., Sun, Y., Sun, H., Lim, J.T., Tam, C., Dickens, B.L., 2020. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. Lancet Infect. Dis. 20 (6), 678–688. https://doi.org/10.1016/S1473-3099(20)30162-6. (Erratum in: Lancet Infect. Dis. 2020 May;20(5):e79.).

Kretzschmar, M.E., Rozhnova, G., Bootsma, M.C.J., van Boven, M., van de Wijgert, J.H.H.M., Bonten, M.J.M., 2020. Impact of Delays on Effectiveness of Contact Tracing Strategies for COVID-19: A Modelling Study. Lancet Public Health 5 (8), e452–e459. https://doi.org/10.1016/S2468-2667(20)30157-2.

Kucharski, A.J., Klepac, P., Conlan, A.J.K., Kissler, S.M., Tang, M.L., Fry, H., Gog, J.R., et al., 2020. Effectiveness of Isolation, Testing, Contact Tracing, and Physical Distancing on Reducing Transmission of SARS-CoV-2 in Different Settings: A Mathematical Modelling Study. Lancet Infect. Dis. 20 (10), 1151–1160. https://doi.org/10.1016/S1473-3099(20)30457-6.

Lau, M.S.Y., Grenfell, B., Thomas, M., Bryan, M., Nelson, K., Lopman, B., 2020. Characterizing Superspreading Events and Age-Specific Infectiousness of SARS-CoV-2 Transmission in Georgia, USA. Proc. Natl. Acad. Sci. 117 (36), 22430–22435. https://doi.org/10.1073/pnas.2011802117.

Lloyd, A.L., May, R.M., 2001. How viruses spread among computers and people. Science 292, 1316–1317. https://doi.org/10.1126/science.1061076.

Lloyd-Smith, J., Mollison, D., Metcalf, J., Klepac, P., Heesterbeek, H., 2015. Challenges in Modelling Infectious Disease Dynamics. Epidemics 10, 1–108. special issue.

Macdonald, G., 1950. The analysis of infection rates in diseases in which superinfection occurs. Trop. Dis. Bull. 47, 907–915.

Macdonald, G., 1952. The analysis of the sporozoite rate. Trop. Dis. Bull. 49, 569–586.

Macdonald, G., 1955. The measurement of malaria transmission. Proc. R. Soc. Med. 48, 295–301.

Macdonald, G., 1956. Epidemiological basis of malaria control. Bull. World Health Organ. 15, 613–626.

Murray, J.D., 1989. Mathematical Biology. Springer-Verlag, Berlin Heidelberg. Second, corrected edition. ISBN 0-387-57204-X.

NHS, 2020. NHS test and trace reports. https://www.gov.uk/government/collections/nhs-test-and-trace-statistics-england-weekly-reports.

Panovska-Griffiths, J., Kerr, C.C., Stuart, R.M., et al., 2020. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study [published online ahead of print, 2020 Aug 3]. Lancet Child Adolesc Health. https://doi.org/10.1016/S2352-4642(20)30250-9. S2352-4642(20)30250-9.

Panovska-Griffits, J., Kerr, C.C., Waited, W., et al., 2020. Modelling the potential impact of mask use in schools and society on COVID-19 control in the UK. medRxiv. preprint doi https://doi.org/10.1101/2020.09.28.20202937.

Peak, C.M., Childs, L.M., Grad, Y.H., Buckee, C.O., 2017. Comparing Nonpharmaceutical Interventions for Containing Emerging Epidemics. Proc. Natl. Acad. Sci. U. S. A. 114 (15), 4023–4028.

Read, J.M., Bridgen, J.R.E., Cummings, D.A.T., Ho, A., Jewell, C.P., 2020. Novel Coronavirus 2019-NCoV: Early Estimation of Epidemiological Parameters and Epidemic Predictions. medRxiv. Infectious Diseases (except HIV/AIDS).

Rice, K., Wynne, B., Martin, V., Ackland, G.J., 2020. Effect of school closures on mortality from coronavirus disease 2019: old and new predictions. BMJ 371. m3588.

Rohani, P., Zhong, X., King, A.A., 2010. Contact network structure explains the changing epidemiology of pertussis. Science 330, 982–985. https://doi.org/10.1126/science.1194134.

Ross, R., 1911. Some quantitative studies in epidemiology. Nature 87, 466–467.

Ross, R., 1916. An application of the theory of probabilities to the study of *a priori* pathometry. Part I. Philos. Trans. R. Soc. Lond. A 92, 204–230.

Ross, R., Hudson, H.P., 1917a. An application of the theory of probabilities to the study of *a priori* pathometry. Part II. Philos. Trans. R. Soc. Lond. A 93, 212–225.

Ross, R., Hudson, H.P., 1917b. An application of the theory of probabilities to the study of *a priori* pathometry. Part III. Philos. Trans. R. Soc. Lond. A 93, 225–240.

Shmueli, G., 2010. To Explain or to Predict. Statistical Science 25 (3), 289–310. https://doi.org/10.1214/10-ST330.

Stuart, R.M., Abeysuriya, R.G., Kerr, C.C., et al., 2020. Robust test and trace strategies can prevent COVID-19 resurgences: a case study from New South Wales, Australia. medRxiv. preprint. https://doi.org/10.1101/2020.10.09.20209429.

Sturniolo S, Waites W, Colbourn T, Manheim D, Panovska-Griffiths J. Testing, tracing and isolation in compartmental models." medRxiv. (2020) preprint. https://doi.org/10.1101/2020.05.14.20101808.

Taylor, D.C.A., Pawar, V., Kruzikas, D., et al., 2010. Methods of Model Calibration. Pharmacoeconomics 28, 995–1000. https://doi.org/10.2165/11538660-000000000-00000.

Varma, V.S., 1977. Exact solutions for a special pre-predator or competing species system. Bull. Math. Biol. 39, 619–622.

Vegvari C. et al. Commentary on the use of the reproduction number R during the COVID-19 pandemic. Statistical Methods in Medical Research. (In submission).

Waites, W., Cavaliere, M., Manheim, D., Panovska-Griffiths, J., Danos, V., 2020. Scaling up epidemiological models with rule-based modelling. arXiv. 12077v3.

Walker, P.G.T., Whittaker, C., Watson, O.J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B.A., Cucunubá, Z., Olivera Mesa, D., Green, W., Thompson, H., Nayagam, S., Ainslie, K.E.C., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Brazeau, N.F., Cattarino, L., Cuomo-Dannenburg, G., Dighe, A., Donnelly, C.A., Dorigatti, I., van Elsland, S.L., FitzJohn, R., Fu, H., Gaythorpe, K.A.M., Geidelberg, L., Grassly, N., Haw, D., Hayes, S., Hinsley, W., Imai, N., Jorgensen, D., Knock, E., Laydon, D., Mishra, S., Nedjati-Gilani, G., Okell, L.C., Unwin, H.J., Verity, R., Vollmer, M., Walters, C.E., Wang, H., Wang, Y., Xi, X., Lalloo, D.G., Ferguson, N.M., Ghani, A.C., 2020. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science 369 (6502), 413–422. https://doi.org/10.1126/science.abc0035.

Wilson, A.J., 1980. On Varma's prey-predator problem Bull. Math. Biol. 42, 599–600.

Zhao, S., Chen, H., 2020. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. Quant. Biol., 1–9. https://doi.org/10.1007/s40484-020-0199-0.