medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Testing, tracing and isolation in compartmental models

Simone Sturniolo[1,◕] William Waites[2,◕] Tim Colbourn[3] David Manheim[4] Jasmina Panovska-Griffiths[3,5,6*]

1 Science and Technology Facilities Council, Swindon, UK
2 School of Informatics, University of Edinburgh, Edinburgh, Scotland, UK
3 UCL Institute for Global Health, London, UK
4 University of Haifa Health and Risk Communication Research Center, Haifa, Israel
5 Department of Applied Health Research, UCL, London, UK
6 The Queen's College, Oxford University, Oxford, UK

◕These authors contributed equally to this work.
* j.panovska-griffiths@ucl.ac.uk

Abstract

Existing compartmental mathematical modelling methods for epidemics, such as SEIR models, cannot accurately represent effects of contact tracing. This makes them inappropriate for evaluating testing and contact tracing strategies to contain an outbreak. An alternative used in practice is the application of agent- or individual-based models (ABM). However ABMs are complex, less well-understood and much more computationally expensive. This paper presents a new method for accurately including the effects of Testing, contact-Tracing and Isolation (TTI) strategies in standard compartmental models. We derive our method using a careful probabilistic argument to show how contact tracing at the individual level is reflected in aggregate on the population level. We show that the resultant SEIR-TTI model accurately approximates the behaviour of a mechanistic agent-based model at far less computational cost. The computationally efficiency is such that it be easily and cheaply used for exploratory modelling to quantify the required levels of testing and tracing, alone and with other interventions, to assist adaptive planning for managing disease outbreaks.

Introduction

Since the beginning of 2020, the World has been in the midst of a COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2. To slow down the spread, many countries, including the UK have imposed social distancing mitigation strategies. However, such measures cannot feasibly be imposed over a long period as this may lead to economic collapse. As a consequence countries need to consider how to ease lockdown measures while controlling SARS-CoV-2 spread.

The World Health Organisation has recently updated their guidance on this, recommending a six point strategy that requires firstly assuring that the pandemic spread has been suppressed, and is followed by detecting, testing, isolating and contact-tracing of infected individuals [1].

Mathematical modelling has figured prominently in decision making around control and containment of Covid-19 spread, including the imposition of physical distancing measures [2]. It provides a logical framework for understanding the propagation of an

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
PREPRINT medRxiv 2020-101808v1 – Not peer-reviewed

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

infectious disease through a population and allows different interventions to be explored, including testing and contact tracing of infected individuals as possible strategies to ease social distancing restrictions. Such models are also necessarily simplifications and understanding of their assumptions and what they do and do not represent is required to correctly interpret them.

Mathematical models have a long history of being used to describe the spread of infectious diseases from plague outbreaks more than a century ago [3] to the more recent SARS [4] and Ebola [5], [6] epidemics, and from making decisions around different vaccination strategies for influenza [7] to modelling HIV [8], and from modelling pandemic influenza [9] to currently facilitating real-time policy decision making around the COVID-19 epidemic [2, 10–14]. There are several common approaches, each with advantages and disadvantages [4, 15]. Compartmental models [4, 16, 17] partition the population into different compartments such as susceptible, exposed to the virus but not infectious, infectious and removed and track the movements of individuals between these groups. Though dynamics of real disease outbreaks are fundamentally stochastic [18–20], this level detail is mainly relevant for early stages or small outbreaks [21]. Commonly within compartmental models a mean-field approximation given by ordinary differential equations (ODE) is used [4, 22, 23]. The latter approach is particularly attractive because it is computationally efficient and can yield informative results. ODE systems can be generalised to explicitly incorporate dependence on system state at some times in the past, yielding delay-differential equations (DDE) [24–26], the analogue for continuous state of Markov processes with finite memory. Such formulations require meticulous care to solve accurately [27, 28] and much of what is known about their behaviour consists of asymptotic results [29–32]. Branching processes are used [23, 33, 34] where more flexibility is desired in representing the timing of transitions among compartments and, for continuous time, are amenable to stochastic differential equation (SDE) treatment. For some choices of distribution, the SDE formulation is Markovian and can be analysed as a continuous-time Markov chain (CTMC) [19, 35]. Finally, individual- or agent-based models (IBM/ABM) explicitly represent each individual in the population and allow for fine-grained modelling of the characteristics of each one such as different contact patterns or susceptibilities to the disease [36–40]. They have been [41], and are being [10–12] widely used for planning and epidemic control. While ABMs allow for maximal flexibility and realism, this comes at a high computational cost and it can be difficult to extract analytical results that relate the fine-grained behaviour to population-level effects. It is generally feasible to conduct agent-based simulations for populations of tens of thousands, but there are salient features of epidemics such as the timing and size of peaks of infectious individuals that depend on population sizes two orders of magnitude larger. An important subset of ABMs are network or graphical models [42–47] where the structure of the population, the possible interactions among its members, are explicitly represented. In addition to the computational cost and analytical difficulties with ABMs, sufficient data to support their fine-grained realism is rarely available. For many purposes, including the one that we are concerned with here, an accurate qualitative understanding of the effect of interventions like testing and contact tracing, cheap, coarse, high-level models are more useful than expensive fine-grained models that rely on vast often not readily available data.

While classic compartmental models can easily be used to simulate some interventions analogous to parameter changes, they cannot readily include contact tracing of infected individuals unless vast assumptions are made. This is because modelling contact-tracing is intrinsically reliant on individual behaviour within a network structure. Previous work on Ebola [6], SARS [48] and COVID-19 used simple approaches to represent contact tracing in a compartmental model: asserting that a constant fraction of exposed individuals becomes isolated due to contact tracing [10, 14, 49, 50] or reducing transmission by a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

constant amount, perhaps after a delay [51]. We believe that this kind of approach is insufficient for the purpose of understanding how the rate and timing of testing and contact tracing affect success in containing outbreaks. The purpose of contact tracing is to attempt to isolate infectious, or soon to be infectious individuals. Therefore, contact tracing should result in the isolation of both infectious and exposed individuals and this is a key assumption that previous work has missed. Contact tracing will also inevitably result in the isolation of susceptible and recovered individuals with the former contributing to a reduced rate of disease propagation. To properly understand this process it is imperative to model the effects of contact tracing with mathematical rigour.

In this paper we develop an extension to the classic Susceptible-Exposed-Infectious-Removed[1] (SEIR) model [16, 52, 53] simulated with ODEs to include testing, contact-tracing, and isolation (TTI) strategies. We call this model SEIR-TTI. This model captures the salient features of the manifestation at the population level of the dynamics of testing and tracing at the individual level. Due to its relative simplicity, SEIR-TTI is applicable across a spectrum of diseases. With appropriate parametrisation, it can be used anywhere a standard SEIR model can be used with the same caveats and limitations.

Though we are clearly motivated by the current COVID-19 pandemic and wish to understand how interventions like TTI can be used to contain it, we do not claim that we are modelling it in particular. Our contribution is a mathematical tool and software implementation that can be used for understanding TTI, not a model of COVID-19.

The method that we present is general and can also be applied to other compartmental models, with the standard caveat that with more compartments comes more work to determine the appropriate rates. We validate our SEIR-TTI ODE model against a mechanistic agent-based model where testing, tracing and isolation of individuals is explicitly represented and show that we can achieve good agreement at far less computational cost. We also provide a flexible software package at `https://github.com/ptti/ptti` with a convenient declarative language for specifying parameters and interventions and implementations of the SEIR-TTI ODE model, mechanistic agent-based model, a second non-mechanistic rule-based model in the κ-language formalism [54, 55], and several related models such as classic SEIR.

Results

We design a compartmentalised model describing the populations of susceptible (S), exposed (E – infected but not infectious), infectious (I) and removed (R) population cohorts.

These models are widely used to describe the spread of various infectious diseases [52]. Within the model framework, disease progression is captured by movement of individuals sequentially between compartments accounting for progression from susceptible individuals (S) being exposed to the virus and becoming infected but not infectious (E), to becoming infectious (I) until they recover (R). A schematic illustrating this model is shown in Fig 1.

The novelty of our model is that we have within each compartment included subgroups of people diagnosed and undiagnosed with the virus, attributable to reported and unreported diagnosis. Individuals in our model are defined to be diagnosed either through testing or putatively through tracing. Diagnosed individuals are then isolated.

[1] Slightly different nomenclature is used by different authors. Exposed means infected but not yet infectious and is sometimes called Latent. Infectious is sometimes called Infective and represents individuals capable of transmitting the disease. Removed is often called Recovered, though we opt for the former as it indicates that those individuals are no longer causing infection but we make no statement about whether they are removed through recovery or death.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

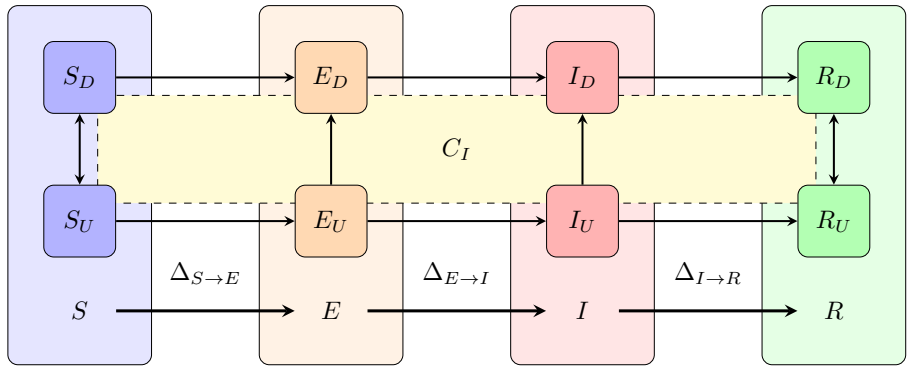

Fig 1. Schematic of an SEIR model with diagnosis described by testing and contact-tracing. SEIR is a compartmentalised model describing susceptible (S), exposed (E – infected but not infectious), infectious (I) and removed (R) population cohorts. Individuals move between these compartments in sequence as they become exposed, infected and infectious during disease progression until recovery. The novelty here is that each compartment comprises diagnosed and undiagnosed individuals with diagnosis leading to isolation. We assume that diagnosis happens through testing or putatively through tracing. Individuals transition between compartments X and Y at rates $\Delta_{X\to Y}$ which we derive in the text.

The effect of testing and isolation alone

Before introducing contact tracing, we examine the standard SEIR model with testing. These results, and those in the following section, use the system of differential equations as described in detail in the Methods. We choose a relatively large initial number of infectious individuals merely for illustrative purposes as it renders the dynamics clearer – the more aggressive testing regimes would result in immediate containment of a small outbreak which would be difficult to see whereas a large outbreak nevertheless takes some time to contain. The parameters have the usual meaning, with values fixed for the purposes of this section: $N = 6.7 \times 10^7$ individuals is the total population, $I(0) = 10^5$ is the initial number of infected individuals, $\hat{\beta} = 0.033$ infections/contact is the probability of transmission; $c = 13$ contacts/day is the contact rate, $\alpha = 0.2$ days^{-1} is the incubation rate, the rate of leaving the exposed state and becoming infectious; and $\gamma = 7^{-1}$ days^{-1} is the rate of recovery, or leaving the infectious state. These values result in a basic reproduction number of $R_0 = 3$. In the simplest case, testing is conducted at random at some rate θ of tests per individual per day and only infectious individuals are tested and immediately isolated.

Representative trajectories from this system for various values of θ are shown in Fig 2. The upper panel shows the time-series for total infections, exposed and infectious, and the lower panel shows the effective reproductive number, $R(t)$. We can observe that while testing the entire population every 20 days ($\theta = 0.05$) results in a lower maximum total number of infections, we require very frequent testing, every 3-4 days ($\theta = 0.3, .25$) in order to control an outbreak and cross the $R(t) = 1$ threshold (red horizontal line). It is straightforward to work out the condition under which testing crosses this threshold by analysing the fixed points in the underlying system of differential equations since the required condition is that there is no change in the number of infectious people as they each infect one other on average and then are removed. Some arithmetic yields $\theta_{\text{crit}} = \hat{\beta}c - \gamma$, the red line in Fig 3.

The above shows that, whilst testing and isolating alone can be sufficient to control an outbreak, it would take a herculean effort on its own. Without any form of distancing

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

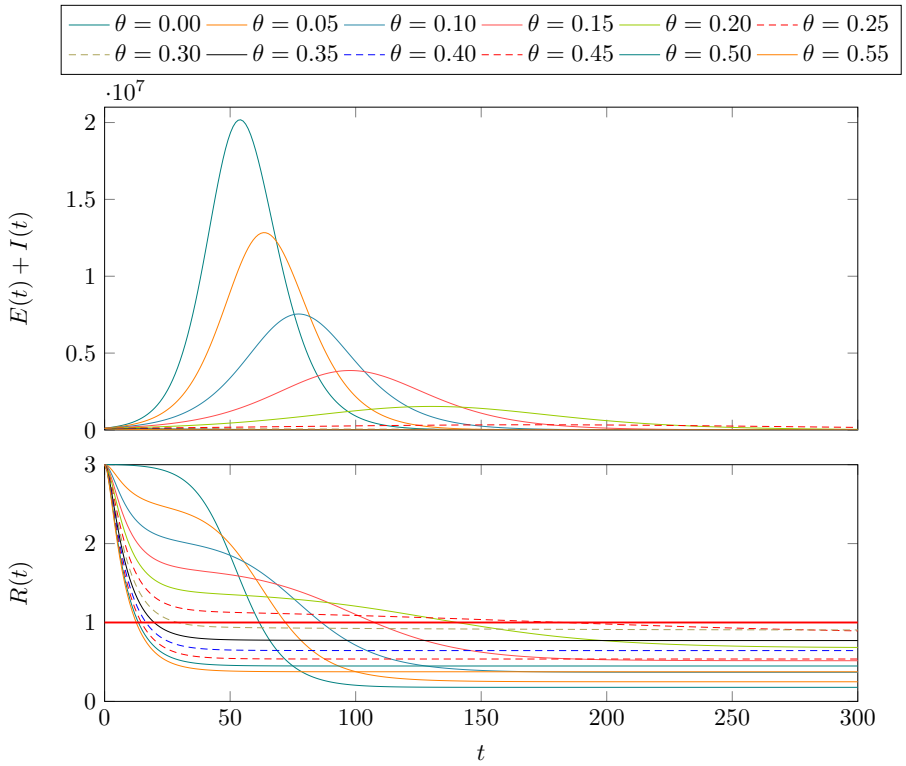

Fig 2. The effect of testing and isolation alone in a hypothetical population. The dynamics represented here are for a scenario with normal contact, $c = 13$, and an initial number of infected individuals, $I(0) = 100,000$. Individuals who test positive are isolated for the duration of their illness. The top plot shows the total infections (exposed and infectious individuals) over time for various testing rates ranging from none, $\theta = 0$, to testing all infectious individuals every two days, $\theta = 0.55$. The bottom plot shows the reproduction number over time for these same scenarios. Observe that even fairly frequent testing, e.g every five days, $\theta = 0.2$, this is only sufficient to reduce peak infections by one order of magnitude from about 20 million to about two million. In the infrequent testing regimes, $\theta \in [0.05, 0.25]$, we can also observe that the curve described by $R(t)$ is not a sigmoid but instead first falls to a value above $R(t) = 1$ before stabilising and then falling again. This is because though testing and isolating does have an effect at those rates, it is not sufficiently frequent to identify all of those who are infectious.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

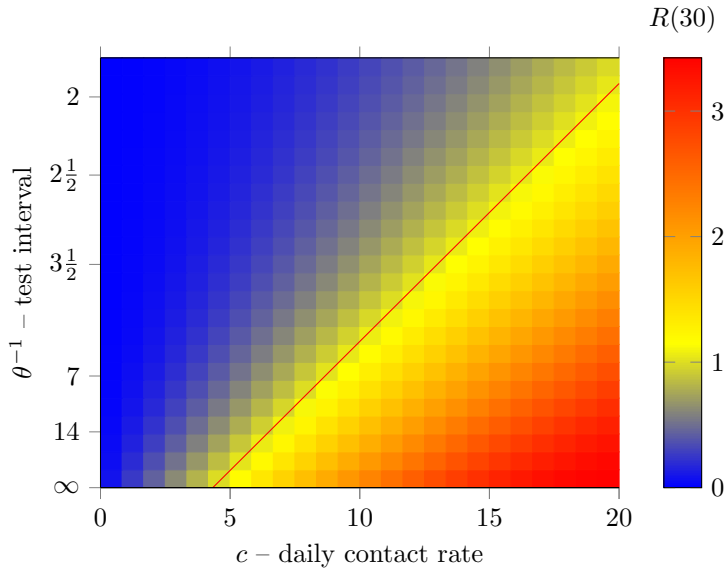

Fig 3. Reproduction number after 30 days for various values of the contact rate, c and the testing rate, θ. The red line is is given by the equation $\theta_{\text{crit}} = \hat{\beta}c - \gamma$. As above, $\hat{\beta} = 0.033$ and $\gamma = 0.1429$.

($c \approx 13$) it is necessary to conduct tests about every 3.5 days. If a sizeable number of infected individuals are asymptomatic, there is no alternative but to test the entire population at this rate. Distancing helps here. If contact rate is cut by half, the required rate is closer to once per fortnight. There is, however, a strategy to avoid regularly sampling the entire population in order to direct tests to those most likely to be infected: contact tracing, which we consider next.

The effect of contact tracing

The central mathematical result is the expression for the rate at which individuals are isolated due to contact tracing,

$$\Delta^{(C_T)}_{X_U \to X_D} = \eta\, \theta \chi \Pr(C_I | X_U) X_U \tag{1}$$

The notation is explained in detail in the methods section, but the intuition is that, for any compartment X, divided into unconfined, X_U, and isolated, X_D, sub-compartments, the rate of moving between them is proportional to the probability of having had contact with an infectious individual conditional on being in X_U.

The effects of contact tracing is shown in Fig 4. The scenario is the same as with testing alone, except that the testing rate is fixed at $\theta = 14^{-1}\text{days}^{-1}$ and the tracing rate is fixed at $\chi = 2^{-1}\text{days}^{-1}$. The interpretation is that, on average, an infectious individual expects to be tested in 7 days and contacts can expect to be traced in 2 days. The choice of these values for illustrative purposes is purposeful. Recall from the previous section that γ, the recovery rate is fixed at 7^{-1}days^{-1}. One would expect that testing and isolating individuals, on average, after they have recovered and it is too late would be insufficient to contain an outbreak. Indeed it is not suffcient, but it does reduce the maximum number of infected individuals somewhat. However, since tracing happens as a consequence of testing, it amplifies its effectiveness. This can be seen in the figure where even a modest tracing success rate of 30-40% results in a substantial reduction of more than half the peak infections.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

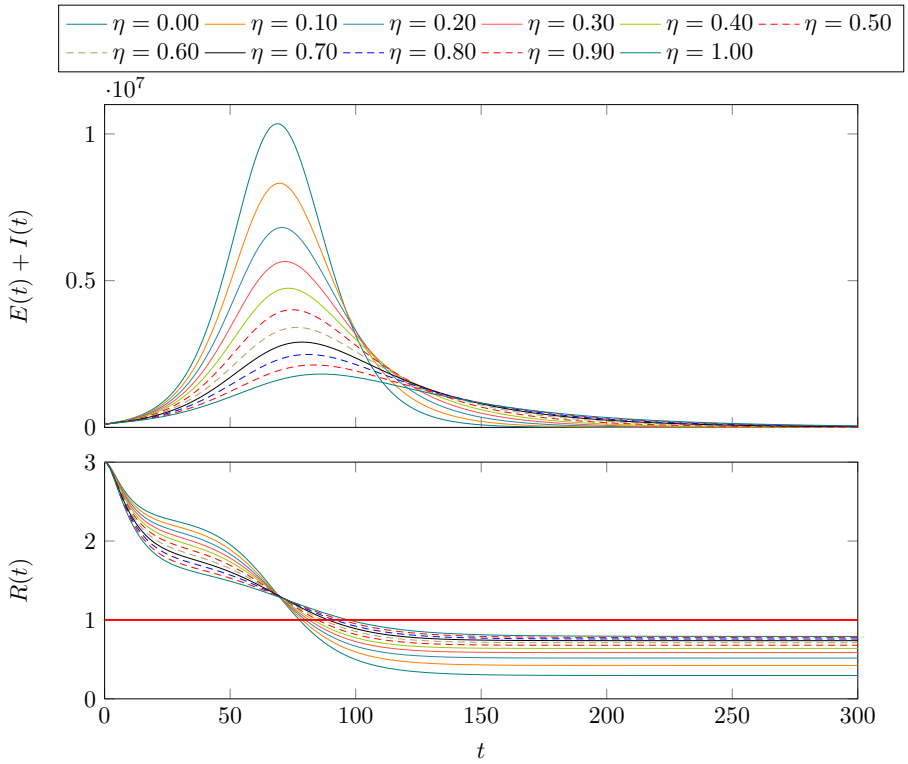

Fig 4. The effect of testing, tracing in a hypothetical population. The dynamics presented here are the same as those of Fig 2 with a testing rate $\theta = 7^{-1}\text{days}^{-1}$, meaning testing of infectious individuals on average once per week. The rate of contact tracing is set at $\chi = 2^{-1}\text{days}^{-1}$, meaning that it takes on average two days to trace a contact. A variety of values of tracing success rate, η are explored. Under these conditions, even a modest success rate of 30-40% enhances testing and results in a maximum of infectious individuals that is about half the magnitude with testing alone. The lower panel is, as above, the corresponding time-series for the reproduction number.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

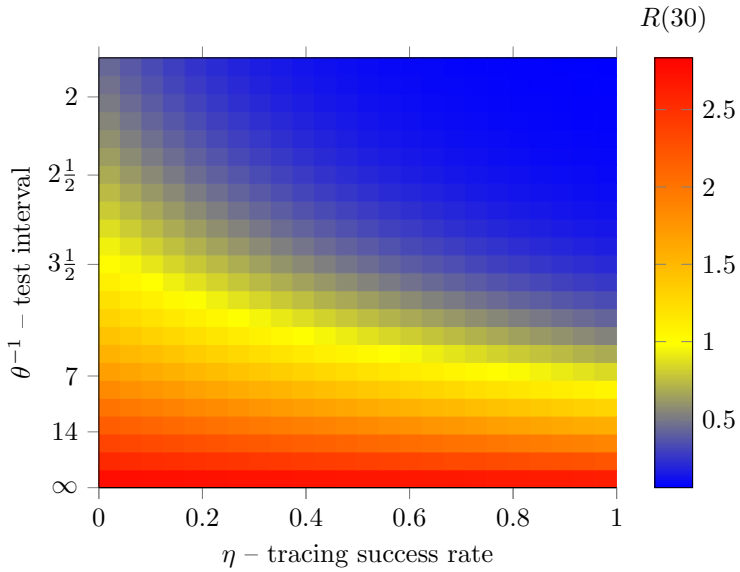

Fig 5. Reproduction number after 30 days for various values of the testing rate, θ and the contact tracing success η.

The relationship between testing rate and tracing rate can be seen from Fig 5. When θ is very small, meaning very little testing, then contact tracing has little effect. This is unsurprising because testing causes tracing. When there is very frequent testing, on the other hand, there is little benefit to contact tracing. When testing happens more frequently on average than an individual can infect another, it is sufficient to control the outbreak on its own. However for intermediate values, contact tracing amplifies the effectiveness of testing. The above result can be seen from this plot as well: when testing of infectious individuals is expected in a week, a modest 40% success rate at tracing contacts in two days is enough to reduce the reproduction number from 2 to less than 1.5, a substantial benefit.

Ordinary differential equations and agent-based models

The central result of this paper is not specific observations about how testing and contact tracing affect the propagation of epidemics, though those are valuable, but a technique to compute these effects efficiently. This technique allows consideration of larger populations than would be possible with agent- or individual-based models allowing for the exploration of many different scenarios. Figs 3 and 5, for example, each contain $25 \times 25 = 525$ data points resulting from a separate simulation. Performing these 1050 total simulations takes under a minute on a regular laptop. This would have not been possible with agent- or individual-based models, with population sizes in the hundreds of thousands or millions.

It could be argued that it is sufficient to capture these dynamics in an agent-based model for modest populations and simply rescale the output for large populations. That approach is not sound for two reasons that are easily seen. First, small outbreaks. Imagine a hypothetical country of 70 million people with 100,000 infections. Proportionally, that is 14.3 infections in a population of 10,000. There is a non-negligible probability that an outbreak of size 14 will die out on its own. This will be accounted for by the ABM but is not a realistic possibility for an outbreak of 100 thousand. Scaling therefore suggests fundamentally different results. Second, without intervention, the number of infectious individuals will reach a maximum as the available pool of susceptible individuals becomes

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

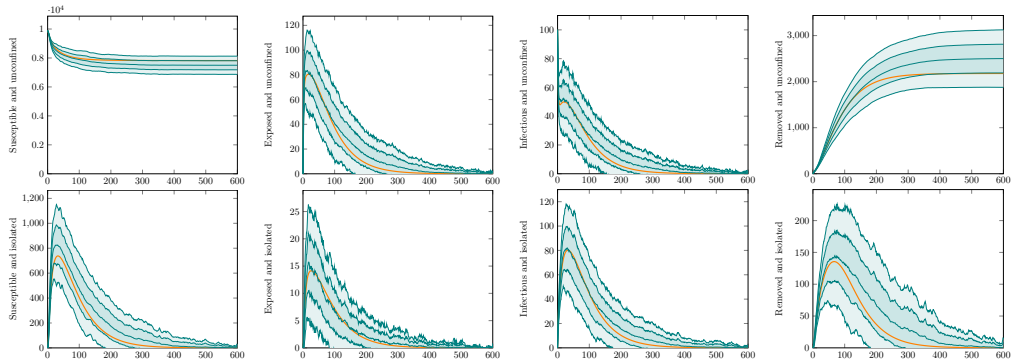

Fig 6. Comparison of differential equation and agent-based model. Here the population size is 10000 individuals with 100 initially infected. The testing rate is $\theta = 7^{-1}\text{days}^{-1}$, and tracing rate and success probability are $\chi = 0.5$, $\eta = 0.5$. The plots show the time-series for each compartment. The top row are the compartments representing unconfined individuals, those in S_U, E_U, I_U and R_U. The bottom row are those representing isolated – diagnosed or distanced – individuals, S_D, E_D, I_D, R_D. The heavy orange curves are the output of the ODE-based simulation. The teal curves are the average output of the agent-based simulation, with envelopes for one and two standard deviations.

depleted. This takes longer in a large population simply because the pool is larger. If timing of the peak of an outbreak is a quantity of interest, a scaled ABM will give the wrong result.

However, doing this requires some approximations and it is important to understand where and how well these approximations hold. To do this, we compare with an agent-based model as described in the methods, and show that our method agrees well for a large range of physically interesting and realistic parameter values. A comparison of the two systems for reasonable parameter values is shown in Fig 6. The figure shows good agreement between the mean trajectory of the ABM and the ODE approximation. The agreement is particularly precise for the exposed and infectious compartment of both varieties. We can observe a slight over-estimate of the number of unconfined susceptible individual and corresponding under-estimate of the unconfined removed ones. These over- and under-estimates are nevertheless acceptably close with a relative error in the magnitude of the susceptible population of under 10%.

There exist extreme scenarios where the ODE performs poorly at reproducing the mean trajectory of the ABM system. An example is shown in Fig 7. One such scenario is when the testing rate is very low. The figure shows when $\theta = 50^{-1}\text{days}^{-1}$. This circumstance violates the assumption underlying Eq 21 that the number of susceptible contacts available for tracing should be much smaller than the total susceptible population. Intuitively, this can be understood as the ODE approximation holding well when testing and tracing are conducted sufficiently rapidly to perform their required purpose. When they do not, the approximation is poor. Even in this extreme scenario, however, where the curve produced by the ODE system is several standard deviations distant from the average trajectory of the ABM, its shape is still similar and realistic.

Methods

We consider the problem of determining the effect of testing and contact tracing in a population, P, consisting of a set of indistinguishable individuals among whom a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

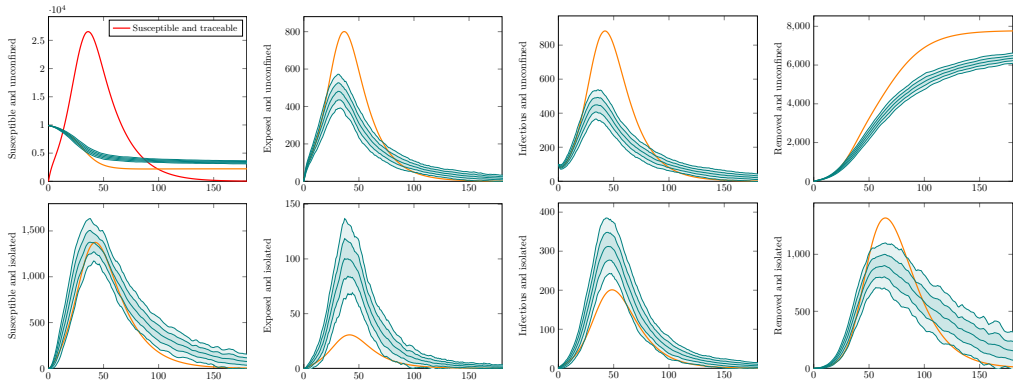

Fig 7. Pathological parameter values. This plot shows the effect of very low levels of testing, $\theta = 50^{-1}\text{days}^{-1}$. In this circumstance, the number of traceable susceptible individuals takes on unphysically high values, shown by the red line in the top left panel. This results in an overestimation of the maximum number of unconfined exposed and infectious individuals and a corresponding underestimation of the effect of contact tracing in preventing infection in this scenario.

disease propagates. To answer this we adapt the standard Susceptible-Exposed-Infectious-Removed (SEIR) compartmental model [16, 52] to incorporate contact tracing as well as testing and isolation of cohorts of people. Our adaptation extends the classic SEIR to not only include progression through disease stages from exposure, via infection to recovery, but to also keeping track of the changing make up of the population as the disease progresses. To achieve this we require our model to have two additional features:

1. to keep track of whether people have been isolated from the rest (either due to testing positive, or having been traced as a contact of someone who tested positive)

2. to keep track of whether people have been in contact with an infectious individual recently enough to be potential targets for tracing.

Ordinary compartment models like SEIR are designed to separate individuals into distinct, non-overlapping groups. This is not a problem for the first feature, as people who are isolated and people who are not constitute entirely distinct sets. We therefore can represent unconfined and isolated individuals simply by doubling the number of states, labeling S_U, E_U, I_U and R_U the Undiagnosed people who are respectively Susceptible, Exposed, Infectious, or Removed, and similarly, S_D, E_D, I_D and R_D the ones who have been Diagnosed or otherwise Distanced from the rest of the population, by means of home isolation, quarantine, hospitalisation and such.

However, dealing with contact tracing is harder, as it can not be achieved with separate compartments. Here we take two approaches. First, we describe an agent-based model that simulates contact tracing with an approximation of how it could take place in real life. This agent-based model serves as our reference. Then we describe fully our compartment model, and, relying on a system of second order Ordinary Differential Equations (ODEs), we introduce the concept of overlapping compartments. Overlapping compartments represent model states that are not mutually exclusive, so that it is possible for an individual to belong in more than one of them e.g. be infected and contact-traced, or exposed and tested. We define equations for this model in order to represent the processes that happen in the agent-based model, providing the comparisons seen above in the Results section.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

An agent based model of contact tracing

Among the possible measures to suppress an epidemic, contact tracing is defined as "an extreme form of targeted control, where the potential next-generation cases are the primary focus" [56]. In other words, contact tracing is the process by which we aim to identify and isolate individuals who have been in contact with an infectious patient in the past and are thus more likely to have been exposed to the disease, in order to remove them from the pool of possible infectious patients before they develop symptoms.

We start by defining our modified SEIR model in agent-based form. The model features N agents each characterised by a state symbolising progression throughout the disease (S, E, I, or R) as well as a single bit characterising whether they are Undiagnosed or Diagnosed/Distanced (U or D). As mentioned above, we label S_U, S_D, E_U, etc. respectively the numbers of individuals in each combination of those states, and S, E, I, R the totals (U and D combined). In addition, we store a contact matrix keeping track of which individuals have been in contact with which infectious members of the population, and an array of all those individuals for whom one past infectious contact has been identified, and thus they can be traced as potentially exposed individuals. We call C_T the total number of such traceable individuals. This contact matrix encapsulates a history of interactions in a way that is realistic but is not possible to represent directly in ODE form. It is specifically the functioning of this individual contact matrix that we claim to reproduce at the population level with our ODE formulation below.

We simulate the model using Gillespie's algorithm [57], which provides a way to sample exact trajectories produced by such stochastic processes. The possible state transitions that can take place are:

1. contact between a random individual and one belonging to I_U, with rate cI_U. The contact is stored in the contact matrix. If the individual happens to belong in S_U, with likelihood $\hat{\beta} \leq 1$, the contact results in exposure, and the S_U individual becomes E_U;
2. progression of the disease for an E individual into I, with rate αE;
3. recovery from the disease, or removal due to hospitalisation or death, for an I individual into R, with rate γI;
4. diagnosis by regular testing of an I_U individual, with rate θI. The individual is moved to I_D; all its past contacts, retrieved from the contact matrix, are marked as traceable with likelihood $\eta \leq 1$. If the individual moved to I_D was marked as traceable, it is unmarked (as they're already in isolation and there is no need to trace them any more);
5. release from isolation of an S_D individual, making them S_U, with rate κS_D;
6. release from isolation of an R_D individual, making them R_U, with rate κR_D;
7. contact tracing of a traceable individual with rate χC_T. The individual is moved from X_U to X_D, where X is whatever state of progression they are in, and they're removed from the list of traceable individuals.

The transitions described above can be intuitively seen as corresponding to the ones that would happen in an idealised real-life version of epidemic spread with testing and contact tracing. The biggest deviation from reality is the perfect mixing of the population implied by the first process. The testing and tracing processes are parametrised by θ, the rate of diagnosis of infectious individuals, η, the likelihood or efficiency with which the tracing process identifies contacts, and χ, the rate at which they are found and isolated. We will describe the meaning and importance of these numbers as we explain how they fit into an ODE model description of the same processes.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The standard SEIR model

We begin by introducing the ODE form of the standard SEIR model [16, 52]. Because of the large number of model compartments and exchange terms between them that will be featured in the full model, we introduce a systematic notation to refer to rates that link them. We refer to $\Delta_{X\to Y}$ as the rate at which members of the population move from compartment X to compartment Y. For example, $\Delta_{S\to E}$ is the rate at which Susceptible members of the population are Exposed to the virus. In addition, for convenience when discussing movements that can happen due to multiple phenomena, we might add a superscript, such as $\Delta^{Z}_{X\to Y}$, to indicate only the part of that rate that can be ascribed to a given process Z.

With this notation, the differential equations that describe the standard SEIR model have the following form,

$$\frac{dS_U}{dt} = -\Delta_{S_U\to E_U} \tag{2}$$

$$\frac{dE_U}{dt} = \Delta_{S_U\to E_U} - \Delta_{E_U\to I_U} \tag{3}$$

$$\frac{dI_U}{dt} = \Delta_{E_U\to I_U} - \Delta_{I_U\to R_U} \tag{4}$$

$$\frac{dR_U}{dt} = \Delta_{I_U\to R_U} \tag{5}$$

Note that all terms involve compartments identified with U subscripts as these equations all apply to the undiagnosed part of our model. They will then be expanded upon to include the effects of isolation and testing in the next section.

The terms in the above differential equations are defined in the usual way as,

$$\Delta_{S_U\to E_U} = \beta S_U \frac{I_U}{N} \tag{6}$$

$$\Delta_{E_U\to I_U} = \alpha E_U \tag{7}$$

$$\Delta_{I_U\to R_U} = \gamma I_U \tag{8}$$

where $\beta = \hat{\beta}c$ is the infection rate, α is the disease progression rate and γ is the disease recovery rate.

While this formulation treats the populations as continuous analytical functions, in general these equations describe the mean trajectory of what is fundamentally a stochastic system. This stochastic system can be simulated with Gillespie's algorithm and, up to this point, is equivalent in the continuous limit to an agent-based model featuring the same compartments and transition rates.

The SEIR model with diagnosis and isolation

Now we add diagnosis to our description. Four more compartments, S_D, E_D, I_D and R_D, are created to keep track of population cohorts who have been identified as potentially infected, and thus isolated from the rest of the population as a measure to limit the spread of the disease. Disease progression is not affected by this process; therefore,

$$\Delta_{E_D\to I_D} = \alpha E_D \tag{9}$$

$$\Delta_{I_D\to R_D} = \gamma I_D \tag{10}$$

Including isolation will change the infection rate, as unlike population I_U, the isolated population I_D does not contribute to further infection. Hence we do not include an

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

infection term here. This is an idealisation. In reality isolation will not be perfect, and we can imagine a reduced 'cross-infection' rate in which some people belonging to S_U are infected by people in I_D. This could happen with medical professionals treating infectious patients or care workers who maintain a quarantine facility. We could even consider infection of people in S_D due to those in I_D, such as a patient in home isolation infecting their family. However, for present purposes, we will work in an ideal situation where isolation is perfect.

Finally, we need to incorporate mechanisms to move individuals between the U and D branches of the model. For this purpose we define a testing rate, θ, which represents the fraction of people belonging in I_U who, each day, are diagnosed with the disease. We note that this parameter does not refer to any specific testing procedure; it just represents the total of people who are recognised as having the disease. It can represent, for example, actual testing for a specific pathogen as well as clinical diagnosis. We only focus on the category of I_U as these are the patients who are most likely to realise they are sick and seek medical help. This generic testing process is described by the equation,

$$\Delta^T_{I_U \to I_D} = \theta I_U \tag{11}$$

In addition, people will be released from isolation after a finite time without symptoms. For this reason, we don't include a mechanism for people in I_D to return to the U branch of the model, as they're likely to be symptomatic or test positive for the pathogen. Instead, we consider that people who have been isolated despite being not infected, or who are still isolated after having recovered, will return to normal conditions at a rate κ,

$$\Delta_{S_D \to S_U} = \kappa S_D \tag{12}$$

$$\Delta_{R_D \to R_U} = \kappa R_D \tag{13}$$

With this model adaptation, a single infected individual can now take two paths:

1. $S_U \to E_U \to I_U \to R_U$, in which they are exposed to the disease, become infectious, and finally recover, without being isolated or diagnosed, as in the normal SEIR model, or,

2. $S_U \to E_U \to I_U \to I_D \to R_D \to R_U$, in which, after becoming infectious, they are identified, isolated, removed from the pool of those who can infect other susceptible people, and after recovering, released from isolation.

Having these two paths allows attainment of some degree of control of the epidemic; however, it must be noted that while we have introduced them, the states S_D and E_D are here left unused. This is because at this stage we associate testing with symptomaticity; there is yet no mechanism other than by diagnosis to identify someone who could be infected. This is especially problematic in terms of the impossibility of isolating exposed people. These are individuals with a latent infection who will soon become infectious. Isolating them pre-emptively would contribute a great deal towards suppressing the epidemic. For this reason, we move on to include contact tracing as a means of preventive isolation.

The SEIR model with testing, tracing and isolation

We've seen previously that it is intuitive how contact tracing can be represented in an agent-based model, in which individuals are simulated and each has an history of contacts with other members of the population. It is not as obvious how to treat contact tracing in a compartment model, where there is no memory of the histories of contacts of specific individuals, but only average quantities. We outline here a probabilistic method for doing this.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Let us define $\Pr(X)$ the probability of an individual of belonging to compartment X of the population. For example, $\Pr(S_U) = S_U/N$ is the probability of an individual to be Susceptible and Undiagnosed. In addition, let us define $\Pr(C_I)$ the probability of an individual of having had contact with an infectious individual in the past where that infectious individual is still infectious. The latter detail is important because here we consider only "next-generation" tracing; in other words, we only try to trace the direct contacts of those infectious individuals who were found to test positive. This is a conservative assumption. It could be possible to make contact tracing more effective by also tracing one generation further (the contacts of the contacts), but because the process requires exponentially more resources with each generation with decreasing likelihood of correctly identifying exposed or infectious individuals, we simply opt to neglect that possibility. Therefore, in this model the only people who can be traced are those whose most recent infectious contact is *still* infectious; once they recover, they cannot be identified as infectious any more, and thus it will be impossible to trace their contacts as well. Finally, we define $\Pr(C_T)$ the probability of an individual of being traced. All these probabilities are functions of time, and quantities that evolve with the model itself.

First, we rewrite the probability of being traced is

$$\Pr(C_T) = \Pr(C_T|C_I)\Pr(C_I) + \Pr(C_T|\neg C_I)\Pr(\neg C_I) \tag{14}$$

where $\Pr(C_T|C_I)$ is the conditional probability of being traced given that one has had an infectious contact in the past, and $\Pr(C_T|\neg C_I)$ the probability of being traced given that one has not. Clearly, $\Pr(\neg C_I) = 1 - \Pr(C_I)$. If we ignore the possibility of false positives, then $\Pr(C_T|\neg C_I) = 0$, namely, a person can only be traced if they did have an infectious contact in the past. If we then set an 'efficiency' parameter η representing the fraction of contacts that we are indeed able to identify, the probability of being traced at a given time is simply

$$\Pr(C_T) = \eta \Pr(C_I) \tag{15}$$

To derive transition rates among compartments, we consider that individuals will be traced proportionally to how quickly the infectious individuals who originally infected them are, themselves, identified. We add a factor χ to account for the speed of the tracing process itself, and we find a global tracing rate,

$$\Delta^{(C_T)} = \chi\theta \Pr(C_T) = \chi\theta\eta \Pr(C_I) \tag{16}$$

It then follows that, for individuals in a given compartment X, the rate at which they're isolated by contact tracing is

$$\Delta^{(C_T)}_{X_U \to X_D} = \eta\,\theta\chi \Pr(X_U|C_I)\Pr(C_I)N = \eta\,\theta\chi \Pr(C_I|X_U)X_U \tag{17}$$

where in the last step we made use of Bayes' theorem [58]. This is our Eq 1, the central mathematical result of this paper.

The difficulty is then computing the exact probabilities. These are functions that, in general, vary in time and require a certain degree of information about the past. We need to define useful assumptions and approximations in order to work with these probabilities in a model that inherently lacks any memory about the individual histories of the elements of its population.

One simple assumption for Exposed and Infectious individuals is

$$\Pr(C_I|E_U) = \Pr(C_I|I_U) = 1 \tag{18}$$

meaning that we assume that if an individual has been Exposed or Infected, they must also have had an infectious contact in the recent past. This is in fact the reason why

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

contact tracing is an effective use of resources: it skews heavily towards identifying those who have in fact been exposed to the disease. We remark that this assumption does not hold in general in circumstances where it is possible for an individual to become infected indirectly, such as by contact with contaminated surfaces. For present purposes we assume that the likelihood of such events is small compared with the likelihood of being infected through contact with another individual.

Another limit of this assumption is that we have defined $\Pr(C_I)$ as the probability of having had an infectious contact *who is still infectious*. For $\alpha \ll \gamma$, or for some infectious individuals who may take a long time to recover, their original infector might have already recovered in the time it takes for them to be tested. However, here we study a model in which $\alpha > \gamma$, and it is reasonable to assume that those infectious individuals who are tested are identified relatively early on in their infection, especially if $\theta > \gamma$. Therefore, we deem the assumption in Equation 18 acceptable at least insofar as these two conditions hold and indirect infection is unlikely.

Estimating $\Pr(C_I|S_U)$ and $\Pr(C_I|R_U)$ is more complicated. One possible approximation is to work as if I_U were constant on the time-scales of interest; in that case we would have

$$\Pr(C_I|S_U) = (1-\beta)c\gamma'^{-1}I_U \tag{19}$$

$$\Pr(C_I|R_U) = c\gamma'^{-1}I_U \tag{20}$$

where γ' is the overall rate at which individuals are removed from the I_U state. Putting together recovery, regular testing, and contact tracing, we find $\gamma' = \gamma + \theta(1+\eta\chi)$. The main difference between the two equations is determined by the fact that someone in S_U might still be infected, and thus only has a probability $1-\beta$ of remaining susceptible after a contact with an infectious member of the population, whereas for recovered individuals this is not an issue any more. Equations 19 and 20 can be used to compute rates of contact tracing by combining them with 1. However, here we try to go beyond the crude approximation of constant I_U, as it may often reflect reality very poorly.

We consider for example the total number of members of S_U who also have had recent infectious contacts, $N(C_I|S_U) = \Pr(C_I|S_U)S_U$. We can describe these in first approximation as

$$N(C_I|S_U)(t) = \int_{-\infty}^{0} (1-\beta)cI_U(t-\tau)S_U(t-\tau)F_I(t,\tau)F_{S_U}(t,\tau)d\tau \tag{21}$$

where the $F_X(t,\tau)$ are the 'survival functions' for the state X. In other words, these are the functions that determine how likely it is that an individual that was in X at time τ still is in the same state at time t. We also used F_I, meaning the survival function of the total number of infectious individuals, $I = I_U + I_D$, because here we focus on overall infectiousness, not the fact that one might have been isolated before recovery. Note, however, that only I_U individuals participate in contacts. The reason that this is an approximation is that we're not excluding the $N(C_I|S_U)$ from the pool of S_U that can be contacted, and thus there is a risk of double counting. That risk will remain negligible as long as $N(C_I|S_U)/S_U$ is small; therefore, this model will perform better in a regime in which there are few infectious individuals, and thus, few contacts. This is in fact the regime in which contact tracing is most likely to be feasible in practice, to control small outbreaks rather than in presence of an uncontrolled epidemic. Regardless, we show in the Results section that even when this approximation does not hold, while it results in oscillatory behaviour early on, it still generally adequately describes the overall trends and long term equilibrium. Equation 21 is equivalent to the integral form of an equation for a compartment model [59]. It can be written in differential form as,

$$\left.\frac{dN(C_I|S_U)}{dt}\right|_t = (1-\beta)cI_U(t-\tau)S_U(t-\tau) - (h_I(t) + h_{S_U}(t))N(C_I|S_U) \tag{22}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

where the $h_X = \frac{1}{F_X}\frac{dF_X}{dt}$ are the 'hazard functions' for the state X. In particular, $h_I = \gamma$.

Given the similarities between these equations and the ones describing the compartment models, it is natural to think of creating a specific compartment for $N(C_I|S_U)$. This is in fact what we do. There is, however, an important difference from regular compartments, because this compartment does not include individuals that exclusively belong to it; rather, it overlaps with S_U. It is more of a device used for book-keeping purposes, to compute the integral in Equation 21 within the confines of the model, than a compartment in the usual sense. We similarly define $N(C_I|E_U)$, $N(C_I|I_U)$ and $N(C_I|R_U)$, which leads, using Equation 1, to the following contact tracing rates,

$$\Delta^{(C_T)}_{S_U \to S_D} = \eta\,\theta\chi N(C_I|S_U) \tag{23}$$

$$\Delta^{(C_T)}_{E_U \to E_D} = \eta\,\theta\chi E_U \tag{24}$$

$$\Delta^{(C_T)}_{I_U \to I_D} = \eta\,\theta\chi I_U \tag{25}$$

$$\Delta^{(C_T)}_{R_U \to R_D} = \eta\,\theta\chi N(C_I|R_U) \tag{26}$$

In addition, we establish the following transition rates between these N compartments,

$$\Delta_{\to N(C_I|S_U)} = (1-\beta)cI_U S_U \tag{27}$$

$$\Delta_{N(C_I|S_U)\to} = (\gamma + \eta\,\theta\chi)N(C_I|S_U) \tag{28}$$

$$\Delta_{N(C_I|S_U)\to N(C_I|E_U)} = \beta c I_U N(C_I|S_U) \tag{29}$$

$$\Delta_{\to N(C_I|E_U)} = cI_U E_U + \beta c I_U S_U \tag{30}$$

$$\Delta_{N(C_I|E_U)\to} = (\gamma + \eta\,\theta\chi)N(C_I|E_U) \tag{31}$$

$$\Delta_{N(C_I|E_U)\to N(C_I|I_U)} = \alpha N(C_I|E_U) \tag{32}$$

$$\Delta_{\to N(C_I|I_U)} = cI_U^2 \tag{33}$$

$$\Delta_{N(C_I|I_U)\to} = (\gamma + \theta + \eta\,\theta\chi)N(C_I|I_U) \tag{34}$$

$$\Delta_{N(C_I|I_U)\to N(C_I|R_U)} = \gamma N(C_I|I_U) \tag{35}$$

$$\Delta_{\to N(C_I|R_U)} = cI_U R_U \tag{36}$$

$$\Delta_{N(C_I|R_U)\to} = (\gamma + \eta\,\theta\chi)N(C_I|R_U) \tag{37}$$

There is a lot going on in Equations 27-37; most importantly, these new compartments do not conserve the total size of the population. Their membership grows as contacts happen and shrinks as time passes. All the key processes can be summed up as follows:

- elements are 'created' for each state proportionally to the rate of contact with individuals belonging to I_U, adjusted with $1-\beta$ in the case of S_U to account for the likelihood that the contact is infective. These terms are 'sources' and can be recognised by having an arrow with nothing on its left in the subscripts;
- elements 'decay' at a rate that amounts to γ (the hazard function for I, which always appears as it refers to the original infector) plus a rate representing the hazard function for the transition $X_U \to X_D$. These terms are 'sinks' and can be recognised by having an arrow with nothing on its right in the subscripts;
- elements move between compartments following the usual transitions that control the dynamics of the SEIR model (infection, progression of the disease, recovery). These terms are analogous to the corresponding ones connecting X_U states, and contribute the remainder of the hazard function for each X_U to eq. 22 and equivalents.

It must also be noted that, in practice, considering Equation 18, it must be $N(C_I|E_U) = E_U$ and $N(C_I|I_U) = I_U$, which removes the need for two of the four compartments above and simplifies the equations to

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Parameter	Description
N	Population size
c	Average contacts per day
$\hat{\beta}$	Transmission rate per contact
α^{-1}	Incubation period (time from exposed to infectious)
γ^{-1}	Recovery period (time from infection to recovery)
θ	Testing rate of infectious individuals
η	Efficiency or success rate of contact tracing
χ	Contact tracing rate

Table 1. Parameters used in the SEIR-TTI model.

$$\Delta_{\to N(C_I|S_U)} = (1-\beta)cI_U S_U \tag{38}$$
$$\Delta_{N(C_I|S_U)\to} = (\gamma + \eta\,\theta\chi + \beta c I_U)N(C_I|S_U) \tag{39}$$
$$\Delta_{\to N(C_I|R_U)} = cI_U R_U + \gamma I_U \tag{40}$$
$$\Delta_{N(C_I|R_U)\to} = (\gamma + \eta\,\theta\chi)N(C_I|R_U) \tag{41}$$

. A few words are necessary on the hazard function for the $X_U \to X_D$ transitions. This is approximated as $\eta\,\theta\chi$ in states S_U and R_U even though that is not precisely correct; the correct hazard function would be $\eta\,\theta\chi N(C_I|X_U)/X_U$, but that introduces a risk of instability for small values of X_U. We justify this choice by the following reasoning. In a weak testing regime ($\eta\,\theta\chi \ll \gamma$), $N(C_I|X_U)/X_U$ might be high due to a great number of infected individuals, but in principle should never be greater than 1 (modulo the point above about double counting). Therefore, the hazard function is dominated by γ. Conversely, in a strong testing regime, the number of infected individuals, and thus $N(C_I|X_U)/X_U$, will be very small, and this assumption will at most end up underestimating the effect of contact tracing (by causing a faster decay in $N(C_I|X_U)$ than otherwise would happen). The examples shown in the Results section illustrate how this affects the simulations – in general, leading to good predictions for the behaviour of the E_U and I_U compartments.

Equations 6-8, 9-10, 11, 23-26 and 27-37, together, define entirely our model. The parameters that appear in these equations are summarised for reference in Table 1.

Software implementation

We implement the above ordinary differential equations and agent-based model in our PTTI Python package (`https://github.com/ptti/ptti`) using the Compyrtment [60] package that facilitates the formulation of initial value problems. It is written for Python 3 and makes use of the scientific computation libraries NumPy and SciPy [61, 62] as well as the optimisation library Numba [63].

The PTTI package provides a declarative language for specifying simulations of models implemented as Python objects. It supports setting of model parameters, simulation hyper-parameters as well as interventions that modify parameters at particular times to conduct piece-wise simulations reflecting changing conditions in a convenient and user-friendly way. We hope that this software formulation will be useful for easy and rapid exploration of the effects of different intervention scenarios for disease outbreak control.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Discussion

Our work outlines a method for extending the classic SEIR model to include Testing, contact-Tracing and Isolation (TTI) strategies. We show that our novel SEIR-TTI model can accurately approximate the behaviour of agent-based models at far less computational cost. Our adaptation is applicable across compartmental models (e.g. SIR, SIS etc) and across infectious diseases. We suggest that the SEIR-TTI model can be applied to the COVID-19 pandemic to understand the impact of possible TTI strategy to control this outbreak.

The importance of modeling to support decision making is widely acknowledged, but models are far more useful when they can accurately represent the classes of interventions that are being considered [15]. The approach described in this paper enables accurate and efficient modelling of contact tracing and testing across a wide range of relevant parameter values. The ability to accurately model TTI strategies across parameter values is vital for controlling disease outbreaks including the current COVID-19 pandemic. Effective testing, contact tracing and isolation strategies have been the key measures that have prevented the epidemic spreading in South Korea [64], New Zealand and Germany [65].

Our work is novel as it is to date, and to the best of our knowledge, the first deterministic model to explicitly incorporate contact tracing. This has been until now only done with agent-based models. An important aspect of our approach is that our ODE formulation *explains* the behaviour of the agent-based model. Namely, agent-based models are formulated in terms of local interactions among individuals and exhibit emergent behaviour at the population level. For interesting agent-based models, it is usually difficult to obtain any explicit connection between the local interactions and the population-level dynamics except through simulation and inspection of the results. We argue that our work here shows such an explicit connection: we have been able to capture the dynamics that arise at a population level from testing and contact tracing. We show that this is correct by demonstrating good agreement with the population-level dynamics that emerge from the agent-based formulation where only local interactions are specified.

The SEIR-TTI model here considers disease propagation in the classical well-mixed setting. This is appropriate especially in circumstances where data are sparse and gives qualitatively similar results to those from fine-grained models that might otherwise provide more quantitatively accurate results if only more detailed data were available. In particular, well-mixed models do not include any notion of the network of contacts across which a contagion spreads in the real world. In reality, individuals in a large population are not equally likely to have contact with one another and it has long been known [42–44, 46, 47, 66–68] that heterogeneity in underlying population structure can have a strong effect [36, 69–71] on disease propagation. Future work will include developing a better understanding of the relationship between network structure and effectiveness of tracing, and mathematical characterisation of the classes of solution available for these models.

Another extension is investigating the extent to which individual decisions about compliance with measures to reduce disease propagation (voluntary distancing, wearing of masks, etc.) affect the success of containment. A game-theoretical approach such as that considered by Zhao et al. [72] may produce useful insights into this question. Insights gained from these extensions can inform policy design for relaxing onerous restrictions on the population.

An important next step in this work is the real-time policy driven application of SEIR-TTI. As our next piece of work we are planning to explore how SEIR-TTI model can be combined with economic analysis to guide decisions around optimal design of a TTI strategy that can suppress the COVID-19 epidemic in the UK.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Conclusion

This paper shows how to extend compartmental models to incorporate testing, contact tracing and isolation. The resulting SEIR-TTI model is a key development in the widely used SEIR models, and an important step if these are to be useful in policy decision making during outbreaks. The long and successful history of testing, contact tracing and isolation in slowing and stopping the spread of infectious diseases is well known [56], with clear immediate importance for COVID-19 control [73].

The design of policies that include a variety of infectious disease control tools, and understanding and applying them in ways that are effective for society at large, is critical. Tools and models that allow policymakers to better understand the policies and the dynamics of a disease are therefore critical. If making policy decisions without evidence is flying blindly, making decisions without understanding the consequences of the various control measures is flying without flight controls. Models like SEIR-TTI can inform policymakers of the role that testing and tracing can play in preventing the spread of disease. Combined with economic and policy analysis, this can enable far better decision making both in the immediate future, and in the longer term. The next step in our work is indeed this: the application of the SEIR-TTI model combined with economic models to investigate the effect of different TTI strategies to conquer the COVID-19 epidemic in the UK.

Acknowledgments

The authors would like to thank Greg Colbourn, Vincent Danos, Gabriel Goh and Rafaele Vardavas for insightful comments on early drafts of this manuscript. This work used the Cirrus UK National Tier-2 HPC Service at EPCC (http://www.cirrus.ac.uk) funded by the University of Edinburgh and EPSRC (EP/P020267/1).

Funding

WW was supported by the Chief Scientist Office Scotland (COV/EDI/20/12). JPG was supported by the National Institute for Health Research (NIHR) Applied Health Research and Care North Thames at Bart's Health NHS Trust (NIHR ARC North Thames). The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Authors contributions

SS, WW and JPG came up with the idea of the study. SS, WW and JPG developed the SEIR-TTI model with input from TC and DM. SS and WW coded the model. WW, SS and JPG drafted the paper with inputs from TC and DM. The final version of the paper was approved by all authors.

References

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 13 April 2020. WHO Director-General Speeches. 2020;.
2. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

and healthcare demand. London: Imperial College London; 2020. 9. Available from: http://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/.

3. Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character. 1927;115(772):700–721. doi:10.1098/rspa.1927.0118.
4. Hollingsworth TD. Controlling infectious disease outbreaks: Lessons from mathematical modelling. Journal of Public Health Policy. 2009;30(3):328–341. doi:10.1057/jphp.2009.13.
5. Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. Journal of theoretical biology. 2015;384:33–49.
6. Whitty CJM. The contribution of biological, mathematical, clinical, engineering and social sciences to combatting the West African Ebola epidemic. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372(1721):20160293. doi:10.1098/rstb.2016.0293.
7. Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ. Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study. PLOS Medicine. 2013;10(10):e1001527. doi:10.1371/journal.pmed.1001527.
8. Stuart RM, Grobicki L, Haghparast-Bidgoli H, Panovska-Griffiths J, Skordis J, Keiser O, et al. How should HIV resources be allocated? Lessons learnt from applying Optima HIV in 23 countries. Journal of the International AIDS Society. 2018;21(4):e25097. doi:10.1002/jia2.25097.
9. Grieco L, Panovska-Griffiths J, Leeuwen Ev, Grove P, Utley M. Exploring the role of mass immunisation in influenza pandemic preparedness: a modelling study for the UK context. bioRxiv. 2019; p. 775270. doi:10.1101/775270.
10. Keeling MJ, Hollingsworth TD, Read JM. The Efficacy of Contact Tracing for the Containment of the 2019 Novel Coronavirus (COVID-19). medRxiv. 2020; p. 2020.02.14.20023036. doi:10.1101/2020.02.14.20023036.
11. Kucharski AJ, Klepac P, Conlan A, Kissler SM, Tang M, Fry H, et al. Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of SARS-CoV-2 in different settings. medRxiv. 2020; p. 2020.04.23.20077024. doi:10.1101/2020.04.23.20077024.
12. Kretzschmar ME, Rozhnova G, Boven MEv. Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. medRxiv. 2020; p. 2020.03.10.20033738. doi:10.1101/2020.03.10.20033738.
13. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Group CCw, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. medRxiv. 2020; p. 2020.03.24.20043018. doi:10.1101/2020.03.24.20043018.
14. Goscé L, Phillips A, Spinola P, Gupta RK, Abubakar I. Modelling SARS-COV2 Spread in London: Approaches to Lift the Lockdown. Preprints. 2020;.
15. Manheim D, Chamberlin M, Osoba OA, Vardavas R, Moore M. Improving Decision Support for Infectious Disease Prevention and Control: Aligning Models and Other Tools with Policymakers' Needs. Santa Monica, California: Rand Corporation; 2016.
16. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Great Clarendon Street, Oxford, OX2 6DP: Oxford university press; 1992.
17. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, Angelis DD, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. Science. 2015;347(6227). doi:10.1126/science.aaa4339.
18. Frauenthal JC. Stochastic Epidemic Models. In: Frauenthal JC, editor. Mathematical Modeling in Epidemiology. Universitext. Berlin, Heidelberg: Springer; 1980. p. 25–40. Available from: https://doi.org/10.1007/978-3-642-67795-3_3.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

19. Allen LJS. An Introduction to Stochastic Epidemic Models. In: Brauer F, van den Driessche P, Wu J, editors. Mathematical Epidemiology. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer; 2008. p. 81–130. Available from: https://doi.org/10.1007/978-3-540-78911-6_3.

20. Britton T. Stochastic epidemic models: A survey. Mathematical Biosciences. 2010;225(1):24–35. doi:10.1016/j.mbs.2010.01.006.

21. De Serres G, Gay NJ, Farrington CP. Epidemiology of Transmissible Diseases after Elimination. American Journal of Epidemiology. 2000;151(11):1039–1048. doi:10.1093/oxfordjournals.aje.a010145.

22. Frauenthal JC. Deterministic Epidemic Models. In: Frauenthal JC, editor. Mathematical Modeling in Epidemiology. Universitext. Berlin, Heidelberg: Springer; 1980. p. 1–11. Available from: https://doi.org/10.1007/978-3-642-67795-3_1.

23. Müller J, Kuttler C. Epidemiology. In: Müller J, Kuttler C, editors. Methods and Models in Mathematical Biology: Deterministic and Stochastic Approaches. Lecture Notes on Mathematical Modelling in the Life Sciences. Berlin, Heidelberg: Springer; 2015. p. 415–481. Available from: https://doi.org/10.1007/978-3-642-27251-6_4.

24. van den Driessche P. Some epidemiological models with delays. University of Victoria; 1994. Available from: https://dspace.library.uvic.ca/handle/1828/1767.

25. Castillo-Chavez C, Blower S, van den Driessche P, Kirschner D, Yakubu AA. Mathematical approaches for emerging and reemerging infectious diseases: an introduction. vol. 1. Springer Science & Business Media; 2002.

26. Arino J, van den Driessche P. Time delays in epidemic models. In: Arino O, Hbid ML, Dads EA, editors. Delay Differential Equations and Applications. NATO Science Series. Dordrecht: Springer Netherlands; 2006. p. 539–578.

27. Busenberg S, Cooke KL. The effect of integral conditions in certain equations modelling epidemics and population growth. Journal of Mathematical Biology. 1980;10(1):13–32. doi:10.1007/BF00276393.

28. Shakeri F, Dehghan M. Solution of delay differential equations via a homotopy perturbation method. Mathematical and Computer Modelling. 2008;48(3):486–498. doi:10.1016/j.mcm.2007.09.016.

29. Beretta E, Takeuchi Y. Global stability of an SIR epidemic model with time delays. Journal of Mathematical Biology. 1995;33(3):250–260. doi:10.1007/BF00169563.

30. Beretta E, Hara T, Ma W, Takeuchi Y. Global asymptotic stability of an SIR epidemic model with distributed time delay. Nonlinear Analysis: Theory, Methods & Applications. 2001;47(6):4107–4115. doi:10.1016/S0362-546X(01)00528-4.

31. Wang W. Global behavior of an SEIRS epidemic model with time delays. Applied Mathematics Letters. 2002;15(4):423–428. doi:10.1016/S0893-9659(01)00153-7.

32. Zhang T, Teng Z. Global behavior and permanence of SIRS epidemic model with time delay. Nonlinear Analysis: Real World Applications. 2008;9(4):1409–1424. doi:10.1016/j.nonrwa.2007.03.010.

33. Becker N. Estimation for Discrete Time Branching Processes with Application to Epidemics. Biometrics. 1977;33(3):515–522. doi:10.2307/2529366.

34. Frauenthal JC. Branching Process Model. In: Frauenthal JC, editor. Mathematical Modeling in Epidemiology. Universitext. Berlin, Heidelberg: Springer; 1980. p. 54–65. Available from: https://doi.org/10.1007/978-3-642-67795-3_5.

35. Allen LJS. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infectious Disease Modelling. 2017;2(2):128–142. doi:10.1016/j.idm.2017.03.001.

36. Keeling MJ, Grenfell BT. Individual-based Perspectives on R0. Journal of Theoretical Biology. 2000;203(1):51–61. doi:10.1006/jtbi.1999.1064.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

37. Patlolla P, Gunupudi V, Mikler AR, Jacob RT. Agent-Based Simulation Tools in Computational Epidemiology. In: Böhme T, Larios Rosillo VM, Unger H, Unger H, editors. Innovative Internet Community Systems. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2006. p. 212–223.

38. Marshall BDL, Galea S. Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology. American Journal of Epidemiology. 2015;181(2):92–99. doi:10.1093/aje/kwu274.

39. Hunter E, Mac Namee B, Kelleher JD. A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology. Journal of Artificial Societies and Social Simulation. 2017;20(3). doi:10.18564/jasss.3414.

40. Tracy M, Cerda M, Keyes KM. Agent-Based Modeling in Public Health: Current Applications and Future Directions. Annu Rev Public Health. 2018;39:77094. doi:10.1146/annurev-publhealth-040617-014317.

41. Burke DS, Epstein JM, Cummings DAT, Parker JI, Cline KC, Singa RM, et al. Individual-based Computational Modeling of Smallpox Epidemic Control Strategies. Academic Emergency Medicine. 2006;13(11):1142–1149. doi:10.1197/j.aem.2006.07.017.

42. Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. Physical Review E. 2001;63(6):066117. doi:10.1103/PhysRevE.63.066117.

43. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. Journal of The Royal Society Interface. 2007;4(16):879–891. doi:10.1098/rsif.2007.1100.

44. Meyers L. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. Bulletin of the American Mathematical Society. 2007;44(1):63–86. doi:10.1090/S0273-0979-06-01148-7.

45. Easley D, Kleinber J. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press; 2010. Available from: http://www.cs.cornell.edu/home/kleinber/networks-book/.

46. Wang L, Li X. Spatial epidemiology of networked metapopulation: an overview. Chinese Science Bulletin. 2014;59(28):3511–3522. doi:10.1007/s11434-014-0499-8.

47. Kiss IZ, Miller J, Simon PL. Mathematics of Epidemics on Networks: From Exact to Approximate Models. Interdisciplinary Applied Mathematics. Springer International Publishing; 2017. Available from: https://www.springer.com/gp/book/9783319508047.

48. Gumel AB, Ruan S, Day T, Watmough J, Brauer F, van den Driessche P, et al. Modelling strategies for controlling SARS outbreaks. Proceedings of the Royal Society of London Series B: Biological Sciences. 2004;271(1554):2223–2232. doi:10.1098/rspb.2004.2800.

49. Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. Institute for Biocomputation and Physics of Complex Systems Preprint. 2020;.

50. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine. 2020; p. 1–6. doi:10.1038/s41591-020-0883-7.

51. Kissler SM, Tedijanto C, Lipsitch M, Grad Y. Social distancing strategies for curbing the COVID-19 epidemic. medRxiv. 2020; p. 2020.03.22.20041079. doi:10.1101/2020.03.22.20041079.

52. Aron JL, Schwartz IB. Seasonality and Period-doubling Bifurcations in an Epidemic Model. Journal of Theoretical Biology. 1984;110:665–679.

53. Hethcote HW. The Mathematics of Infectious Diseases. SIAM Review. 2000;42(4):599–653. doi:10.1137/S0036144500371907.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.14.20101808; this version posted May 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

55. Boutillier P, Feret J, Krivine J, Fontana W. The Kappa Language and Tools; 2020. Available from: https://kappalanguage.org/.
56. Eames KTD, Keeling MJ. Contact tracing and disease control. Proceedings of the Royal Society of London Series B: Biological Sciences. 2003;270(1533):2565–2571. doi:10.1098/rspb.2003.2554.
57. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977;81(25):2340–2361. doi:10.1021/j100540a008.
58. Bayes T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philosophical Transactions of the Royal Society of London. 1763;53:370–418. doi:10.1098/rstl.1763.0053.
59. Greenhalgh S, Day T. Time-varying and state-dependent recovery rates in epidemiological models. Infectious Disease Modelling. 2017;2(4):419–430. doi:10.1016/j.idm.2017.09.002.
60. Sturniolo S. Compyrtment; 2020. Available from: https://github.com/stur86/compyrtment.
61. Oliphant TE. Python for Scientific Computing. Computing in Science Engineering. 2007;9(3):10–20. doi:10.1109/MCSE.2007.58.
62. Oliphant TE. Python for Scientific Computing. Computing in Science Engineering. 2007;9(3):10–20. doi:10.1109/MCSE.2007.58.
63. Lam SK, Pitrou A, Seibert S. Numba: a LLVM-based Python JIT compiler. In: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM '15. Austin, Texas: Association for Computing Machinery; 2015. p. 1–6. Available from: https://doi.org/10.1145/2833157.2833162.
64. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. International Journal of Infectious Diseases. 2020;.
65. Cohen J, Kupferschmidt K. Countries test tactics in 'war' against COVID-19; 2020.
66. Bharucha-Reid AT. Comparison of Populations Whose Growth Can Be Described by a Branching Stochastic Process: With Special Reference to a Problem in Epidemiology. Sankhyā: The Indian Journal of Statistics (1933-1960). 1958;19(1/2):1–14.
67. Nold A. Heterogeneity in disease-transmission modeling. Mathematical Biosciences. 1980;52(3):227–240. doi:10.1016/0025-5564(80)90069-3.
68. Wang Y, Chakrabarti D, Wang C, Faloutsos C. Epidemic spreading in real networks: an eigenvalue viewpoint. In: 22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.; 2003. p. 25–34.
69. Reich O, Shalev G, Kalvari T. Modeling COVID-19 on a network: super-spreaders, testing and containment; 2020. Available from: https://www.medrxiv.org/content/10.1101/2020.04.30.20081828v1.
70. Britton T, Ball F, Trapman P. The disease-induced herd immunity level for Covid-19 is substantially lower than the classical herd immunity level. arXiv:200503085 [physics, q-bio]. 2020;.
71. Gomes MGM, Aguas R, Corder RM, King JG, Langwig KE, Souto-Maior C, et al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. medRxiv. 2020; p. 2020.04.27.20081893. doi:10.1101/2020.04.27.20081893.
72. Zhao S, Bauch CT, He D. Strategic decision making about travel during disease outbreaks: a game theoretical approach. Journal of The Royal Society Interface. 2018;15(146):20180515. doi:10.1098/rsif.2018.0515.
73. Peto J, Alwan NA, Godfrey KM, Burgess RA, Hunter DJ, Riboli E, et al. Universal weekly testing as the UK COVID-19 lockdown exit strategy. Lancet (London, England). 2020;.