Learning transmission dynamics modelling of COVID-19 using **comomodels**

Solveig A. van der Vegt[a,b,1], Liangti Dai[a,c,1], Ioana Bouros[a,d,1], Hui Jia Farm[d,1], Richard Creswell[d,1], Oscar Dimdore-Miles[g,1], Idil Cazimoglu[a,1], Sumali Bajaj[a], Lyle Hopkins[a,d], David Seiferth[a], Fergus Cooper[a], Chon Lok Lei[e,f], David Gavaghan[a,d], Ben Lambert[a]

[a]*Doctoral Training Centre, University of Oxford, Oxford, U.K.*
[b]*Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford, U.K.*
[c]*MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, U.K.*
[d]*Department of Computer Science, University of Oxford, Oxford, U.K.*
[e]*Institute of Translational Medicine, Faculty of Health Sciences, University of Macau, Macau, China*
[f]*Department of Biomedical Sciences, Faculty of Health Sciences, University of Macau, Macau, China*
[g]*Atmospheric, Oceanic and Planetary Physics Department, University of Oxford, Oxford, U.K.*

Abstract

The COVID-19 epidemic continues to rage in many parts of the world. In the UK alone, an array of mathematical models have played a prominent role in guiding policymaking. Whilst considerable pedagogical material exists for understanding the basics of transmission dynamics modelling, there is a substantial gap between the relatively simple models used for exposition of the theory and those used in practice to model the transmission dynamics of COVID-19. Understanding these models requires considerable prerequisite knowledge and presents challenges to those new to the field of epidemiological modelling. In this paper, we introduce an open-source R package, **comomodels**, which can be used to understand the complexities of modelling the transmission dynamics of COVID-19 through a series of differential equation models. Alongside the base package, we describe a host of learning resources,

---

[1]These authors contributed equally.

including detailed tutorials and an interactive web-based interface allowing dynamic investigation of the model properties. We then use **comomodels** to illustrate three key lessons in the transmission of COVID-19 within R Markdown vignettes.

Highlights

- Compartmental models of transmission dynamics have been important determinants of public health policy for COVID-19.

- Many important characteristics of the spread of COVID-19 can be deduced from relatively simple models.

- The **comomodels** package and its associated GUI allow users to learn the characteristics of complex compartmental models in incremental fashion.

- Estimating model parameters from data is generally difficult and hence sensitivity analyses are key.

1. Introduction

Mathematical modelling of COVID-19 epidemiology has been a crucial determinant of policymaking during the pandemic. For example, in the UK, the release of "Report 9" by Imperial College London (Ferguson *et al.*, 2020) is widely believed to have been influential in the UK government's decision to institute a first lockdown on 23$^{\text{rd}}$ March 2020 (IFG, 2021). *Compartmental models* of transmission dynamics (see, for example, Anderson & May (1992); Brauer (2008)), in particular, have featured prominently. In these models, individual people are allocated to compartments which, amongst other things, indicate their disease state. Popular variants of compartmental models used during the pandemic include Susceptible-Exposed-Infectious-Recovered (SEIR) and Susceptible-Exposed-Infectious-Recovered-Dead (SEIRD) models, where individuals naive to infection are labelled *susceptible* (S); individuals who are infected but not yet infectious to others are labelled *exposed* (E); exposed individuals then eventually become *infectious* (I) and subsequently recover (R) or die (in the SEIRD model only with those individuals falling into the "Dead" (D) compartment).

The advice given to the UK government through the Chief Scientific Advisor has been guided by the Scientific Pandemic Influenza Group on Modelling (SPI-M), and a diversity of models have been used to inform this (Anderson *et al.*, 2020). These models include: the "Warwick model", an ordinary differential equation (ODE) model with SEIR compartments with the infected compartments broken down into symptomatic and asymptomatic cases (Keeling *et al.*, 2021). The model incorporates compartments representing admissions to hospital and intensive care units (ICUs), and is age- and spatially structured by region of the UK; the "Cambridge / Public Health England model" is an ODE model with an SEIR base structure extended to include two exposed and two infectious compartments (Birrell *et al.*, 2021). In addition, the model is structured by age, incorporating age-dependent contact structures which vary according to UK region which can change over time in line with published mobility data; the model underlying Imperial College's Report 9 (Ferguson *et al.*, 2020), which is a large stochastic, individual-based model with SEIR compartments, which accounts for mixing within-households, at specific places (e.g. work or school) or randomly in the community. The model is spatial and age-structured. In it, the infectious individuals are stratified into asymptomatic, mild and "influenza-like illness" groups. From this last, more severe symptomatic group, individuals may be admitted to hospital, with a subset of hospitalised individuals requiring care in ICUs; the "Manchester model", which is a deterministic non-age-structured SEIR model (Anderson *et al.*, 2020). This model allows a piece-wise transmission rate, which can change in response to non-pharmaceutical interventions such as a lockdown. Apart from these mechanistic models, a host of more statistical approaches, primarily based around renewal models (Flaxman *et al.*, 2020; Nouvellet *et al.*, 2021) have also been used within SPI-M. This latter class of models typically require fewer assumptions to be made about the transmission dynamics and, because of this, are straightforward to fit to data. A cost of their relative simplicity is that they are less able to explore a diversity of future scenarios compared to mechanistic models. In this paper and in our software, we currently focus on mechanistic models of transmission, and we plan to introduce other model classes in the future.

Additionally, there have been a host of other models that have been influential in informing policymaking worldwide. Many of these models have incorporated detailed information bespoke to particular vaccines. For instance, Hogan *et al.* (2020) extend a deterministic, age-structured SEIR compartment model (described in Walker *et al.* (2020)) to include vaccinated com-

partments. The modelling approach assumes that those who are currently infected do not receive the vaccine and that vaccines offer partial protection which can wane over time at rates specific to particular COVID-19 vaccines. An international consortium of modellers named the "COVID-19 Modelling (CoMo) Consortium" (Aguas *et al.*, 2020) use a deterministic SEIRD-type model structured by age which also includes, amongst other components, compartments corresponding to individuals in quarantine, vaccinated individuals, and hospitalised individuals. This model inspired the development of our package, hence we chose to name our package **comomodels**.

The larger models used in transmission dynamics modelling of COVID-19 comprise a complex mix of many different small modelling motifs. For an individual new to this type of modelling, such as a graduate student or a policymaker, understanding how these various components function and interact is key to building intuition about how these models work. It is also important in ensuring that the models themselves or their outputs are not misused. Whilst there are ample pedagogical materials on basic SEIR models, to our knowledge, there are no widely available tools used to teach understanding of the more complex types of models often used to decide public health policy for diseases like COVID-19. Others have recognised the need for such materials: relatively early on during the pandemic, The Royal Society published a rapid review of the science of estimating the reproduction number for COVID-19, which included toy models used to explain facets of transmission dynamics modelling of COVID-19 (Anderson *et al.*, 2020).

Historically, mathematical modelling expertise has been concentrated in high income countries (HICs), with several infectious disease modelling groups forming consortia. These consortia inform policy for HICs but also consult for policymakers in low and middle income countries (LMICs). This asymmetric group of HIC modellers and LMIC policymakers that ultimately determines policy for LMICs is inequitable and may lead to models far removed from reality (Aguas *et al.*, 2020). Part of this asymmetry likely stems from a lack of capacity in LMICs, and we argue that part of this is due to a lack of widely available pedagogical tools that teach how modelling is applied in practice.

This stimulated us to create an open-source pedagogical R package called **comomodels**, which aims to produce accessible materials pertinent to applied COVID-19 modellers. **comomodels** was developed using robust software development principles including unit testing and continuous integration testing. The main workhorses of the package are a series of deterministic compartmental models that represent the transmission dynamics of infectious

4

diseases which are spread human-to-human, such as COVID-19. The starting point of the package is the basic SEIRD model. Upon this foundation, we build a series of models of increasing complexity (see Figure 1), each with a tutorial vignette where that model is applied to explain elements of COVID-19 transmission dynamics. By working incrementally through these tutorials, an individual can build a foundation in modelling required for understanding much of the applied work on mechanistic modelling of COVID-19. Alongside the package, we have created an interactive web-based interface (The Como-DTC Development Team, 2021), which visualises the structure of the model and plots key outputs.

In this article, in §2, we briefly describe the models currently available in the package. In §3, we describe package design choices and functionality and outline the approach taken for development. Then in dynamic vignettes described in §4, written in R Markdown, we use our package to illustrate three key lessons in COVID-19 modelling.

2. Models

comomodels contains a series of deterministic ODE-based transmission dynamics models of infectious diseases which are passed directly between humans (see Figure 1). The set of models is tailored to highlight key elements of models of COVID-19 infection dynamics in the literature. In this section, we provide equations and a brief motivation for each model. For much more detailed information, we recommend consulting the accompanying website[2] which contains detailed tutorials for each model. This pedagogical material is a keystone of the package and aims to help a user to build intuition behind transmission dynamics modelling of COVID-19. In doing so, these tutorials also explain how to effectively use the package functionality. These tutorials can be run by cloning the Github repository. Alternatively, the rendered markdown files can be viewed on the package website.

In what follows, each of the model systems described is closed with a set of initial conditions, and we omit them in what follows for brevity. Across all models, we assume that the states are normalised to 1: this means, for example, in the SEIRD model, that $S(t) + E(t) + I(t) + R(t) + D(t) = 1$ for all times $t \geq 0$.

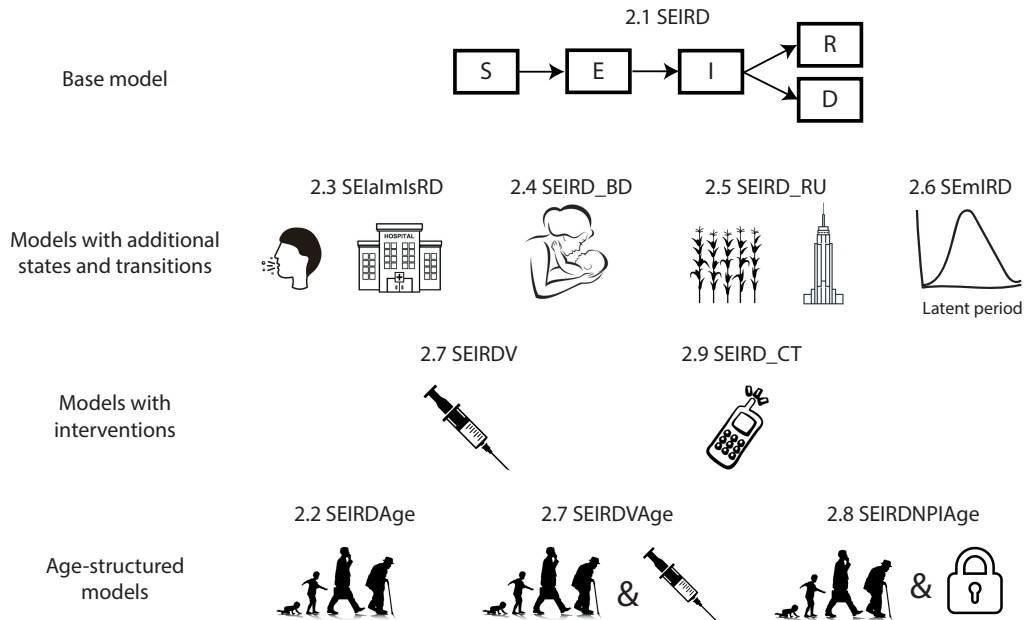[2]See https://como-dtc-collaboration.github.io/como-models/.

Figure 1: **The set of models within the comomodels package.** text above pictures corresponds to model names, and the number corresponds to particular sections in §2. All models are extensions of the SEIRD model (top; see §2.1). The models on the second row have additional state variables and transitions between states compared to the SEIRD model; the third row shows non-age-structured models which model the effect of interventions on COVID-19 transmission; the bottom row shows the age-structured models.

2.1. SEIRD model

All the models in the package are extensions of the SEIRD model. This model considers the dynamics of individuals within five compartments: Susceptibles (S), Exposed (E), Infectious (I), Recovered (R) and Dead (D) indi-

viduals. The model is described by the following system of coupled ODEs:

$$\frac{dS}{dt} = -\beta SI,$$
$$\frac{dE}{dt} = \beta SI - \kappa E,$$
$$\frac{dI}{dt} = \kappa E - (\gamma + \mu)I, \tag{1}$$
$$\frac{dR}{dt} = \gamma I,$$
$$\frac{dD}{dt} = \mu I.$$

Here, $\beta > 0$ quantifies the rate at which a susceptible individual becomes infected (entering the exposed compartment) after coming into close contact with an infectious individual; $\kappa > 0$ represents the rate at which exposed individuals become infectious; $\gamma > 0$ indicates the rate that infectious individuals recover; and $\mu > 0$ indicates the rate at which infectious individuals die.

The SEIRD model tutorial describes how to parameterise the model to simulate COVID-19 transmission dynamics. It then explains how R_0, the basic reproduction number (Diekmann *et al.*, 1990), is calculated for this model. Finally, it performs a sensitivity analysis, illustrating how changes to β affect transmission.

2.2. SEIRDAge: an age-structured model

A natural extension to the SEIRD model in §2.1 is to add age-structure to the population. This is equivalent to splitting each of the current compartments into separate ones for each age group considered. This is done because sociological – e.g. contact rates and patterns – and biological – e.g. mortality and pathogen clearance rates – characteristics may vary, on average, according to the age of individuals.

In these, so called, *age-structured* models, individual age groups have different characteristics, which are reflected in varying parameter values, such as mortality and recovery rates. Additionally, heterogeneous mixing between different age groups is introduced to allow certain combinations of age groups to meet more often than others. For example, a newborn child will typically spend most of a given day with their parents; a teenager may mix frequently and mostly with other teenagers; an elderly person may mostly mix with

7

younger members of their family; and so on. In age-structured models, so-called *contact matrices* keep track of the expected number of contacts between people of different ages. It is also possible to estimate the number of contacts which occur at various locations where people spend time: for instance, the number of contacts occurring at home, school and work. Moreover, contact matrices can vary by geography to handle differences in social structures typical of diverse societies. As part of **comomodels**, we make the contact matrices estimated in Prem *et al.* (2017) for 152 countries available. In Figure 2, we show a set of contact matrices for India, where each panel represents the average number of daily contacts between individuals of different age groups for a particular location type (home, school, work and "other"). In this case, the contact matrix used in simulations would be the sum of these location-specific (l) matrices, $C_{i,j}^l$:

$$C_{i,j} = \sum_l C_{i,j}^l \tag{2}$$

where $l \in \{\text{home}, \text{school}, \text{work}, \text{other}\}$. The elements of the contact matrix $C_{i,j}^l \geq 0$ indicate the expected number of contacts an individual of age class i experiences with age class j per day in location l. The matrix $C_{i,j}$ is then the total number of daily contacts per individual in age class i with those in age class j, on average. If all the elements of the $C_{i,j}$ matrix are 1s and the death rates are set to be age-independent (i.e. $\mu_i = \mu$), the infection dynamics recapitulate those from the base SEIRD. Running counter to intuition, contact matrices are typically not symmetric: that is, $C_{i,j} \neq C_{j,i}$ for reasons that we explain in the "S2" supplementary vignette.

The system of ODEs that describe our age-structured SEIRD model is

given by:

$$\frac{dS_i}{dt} = -\beta S_i \sum_j C_{i,j} I_j,$$

$$\frac{dE_i}{dt} = \beta S_i \sum_j C_{i,j} I_j - \kappa E_i,$$

$$\frac{dI_i}{dt} = \kappa E_i - (\gamma_i + \mu_i) I_i, \tag{3}$$

$$\frac{dR_i}{dt} = \gamma_i I_i,$$

$$\frac{dD_i}{dt} = \mu_i I_i.$$

where $i \in \mathcal{I}$ indicates the categorical age group to which an individual is assigned. For example, an individual may belong to the 0-5 year old group. A user within age group i may be in any of the infection states: if they are susceptible, they are counted towards S_i; and so on. Overall, this means that the system consists of $5|\mathcal{I}|$ coupled ODEs.

The parameters β and κ have the same meaning as in the SEIRD model (see §2.1). We allow an age group-specific death rate, μ_i, since, in COVID-19, death rates are known to vary substantially according to age (Verity *et al.*, 2020). We also allow age group-specific recovery rates, γ_i, since this facilitates flexible parameterisation for each group.

Note that this model should be used only to handle infection dynamics over a relatively short period of time (i.e. up a few years), since individuals do not age throughout the course of simulations.

The tutorial for this model explains how eqs. (3) allow the model to be parameterised using the mortality risk estimates for COVID-19 in Verity *et al.* (2020). It also visualises and explains the use of the contact matrices available in the package, and shows how to calculate R_0 for this age-structured model. Finally, simulations are used to show how the transmission dynamics of a COVID-19-like disease are affected by the age composition of the initial infected population.

2.3. SEIaImIsRD: Varying infectiousness compartments model

COVID-19 is known to evoke a range of symptoms of differing levels of severity across individuals, and individuals who are asymptomatic are thought to play a significant role in driving transmission of COVID-19 (Gandhi
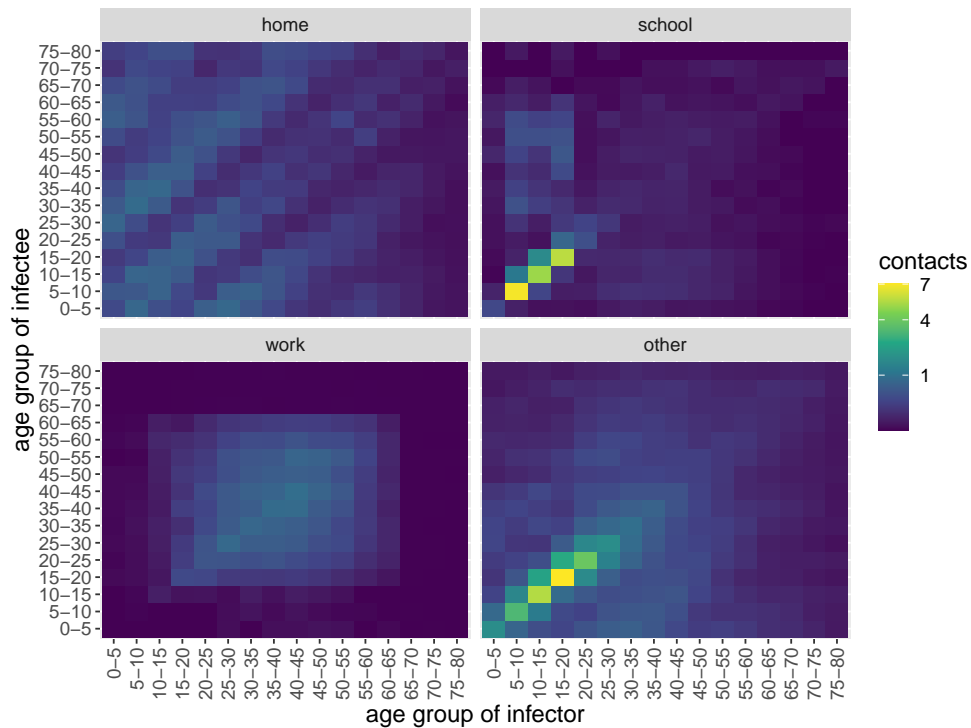
9

Figure 2: **India contact matrices broken down by those which occur at home, school, work and "other" locations.** Each plot displays the mean number of daily contacts per individual by age group (via a colour scale). These data are from Prem *et al.* (2017).

et al., 2020; Johansson *et al.*, 2021; Byambasuren *et al.*, 2020). This group is particularly problematic for transmission, since they are unlikely to change their behaviour whilst infectious, unlike those experiencing more severe infections who may isolate or potentially enter care. Indeed, another reason for stratifying by symptom severity is to be able to forecast the demand for clinical care, such as intensive care unit beds.

As such, we include the SEIaImIsRD model, in the package which considers separately individuals with different degrees of infection severity. This

model is described by the following set of ODEs:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S(\beta_a I_a + \beta_m I_m + \beta_a I_s),$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = S(\beta_a I_a + \beta_m I_m + \beta_s I_s) - \kappa E,$$

$$\frac{dI_a}{dt} = \eta_a \kappa E - (\gamma_a + \mu_a)I_a,$$

$$\frac{dI_m}{dt} = \eta_m \kappa E - (\gamma_m + \mu_m)I_m, \tag{4}$$

$$\frac{dI_s}{dt} = \eta_s \kappa E - (\gamma_s + \mu_s)I_s,$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma_a I_a + \gamma_m I_m + \gamma_s I_s,$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \mu_a I_a + \mu_m I_m + \mu_s I_s.$$

where I_a, I_m and I_s indicate infectious individuals that are asymptomatic, or have mild or severe symptoms. The proportions of individuals entering each of these compartments is governed by the η parameters, and it is assumed that exposed individuals must enter one of these three infectious states, so that $\eta_a + \eta_m + \eta_s = 1$. The recovery and death rate parameters vary according to the symptom category, and it is assumed that $\mu_s > \mu_m \geq \mu_a \approx 0$.

The tutorial for this model provides a detailed explanation of how to calculate R_0 for it. We then assess the sensitivity of total deaths in the epidemic to R_0 and η_s.

2.4. SEIRD_BD: Births and natural deaths and waning natural immunity

The current pandemic has raged for about two years meaning that births and natural deaths (i.e. those not caused by COVID-19) are unlikely to have influenced transmission dynamics. Over longer periods of time, births and natural deaths can lead to replenishment and depletion of susceptibles, altering the level of herd immunity in the population. Considering COVID-19 transmission, it appears that infection confers a protection against reinfection for several months up to around a year post-infection (Leidi *et al.*, 2021; Abu-Raddad *et al.*, 2021; Gallais *et al.*, 2021), but this protection is thought to eventually wane. Waning immunity, like births, can also supplement the susceptible population and lead to conditions where COVID-19 can reinvade

11

a population. We include a model incorporating these three processes:

$$
\begin{aligned}
\frac{dS}{dt} &= \lambda + \delta R - \beta SI - \nu S, \\
\frac{dE}{dt} &= \beta SI - (\kappa + \nu)E, \\
\frac{dI}{dt} &= \kappa E - (\gamma + \mu + \nu)I, \\
\frac{dR}{dt} &= \gamma I - (\nu + \delta)R, \\
\frac{dD}{dt} &= \mu I + \nu(S + E + I + R).
\end{aligned}
\tag{5}
$$

Here, $\lambda > 0$ represents the birth rate, which is assumed constant; newborns are assumed not to have maternal immunity, so enter into the susceptible compartment. $\nu > 0$ represents the rate of natural death which is assumed the same across all compartments. $\delta > 0$ is the rate at which recovered individuals become susceptible again.

The SEIRD_BD model tutorial shows how waves of infection can repeatedly sweep through the population. It also explains how this model, unlike the SEIRD model, allows an equilibrium where the disease is endemic, and it explores how many daily cases may be expected for COVID-19 at equilibrium.

2.5. SEIRD_RU: an urban-rural metapopulation model

Although the pandemic has reached just about every corner of the world, the global population does not behave as a fully well-mixed group of individuals. It is thus important to consider how well different population centres are connected and how this affects the spread of infection. This is often done with a metapopulation model, in which multiple connected communities are considered (Calvetti *et al.*, 2020; Kramer *et al.*, 2020). As a first step towards this, we include a model which consists of two interacting communities, a rural (Y) and an urban (U) community:

$$\frac{\mathrm{d}S_U}{\mathrm{d}t} = -\beta S_U \left((I_U + I_Y)(\phi_U N_U + \phi_Y N_Y)C + \frac{I_U}{\phi_U} N_U (1-C) \right),$$

$$\frac{\mathrm{d}S_Y}{\mathrm{d}t} = -\beta S_Y \left((I_U + I_Y)(\phi_U N_U + \phi_Y N_Y)C + \frac{I_Y}{\phi_Y} N_Y (1-C) \right),$$

$$\frac{\mathrm{d}E_U}{\mathrm{d}t} = \beta S_U \left((I_U + I_Y)(\phi_U N_U + \phi_Y N_Y)C + \frac{I_U}{\phi_U} N_U (1-C) \right) - \kappa E_U,$$

$$\frac{\mathrm{d}E_Y}{\mathrm{d}t} = \beta S_Y \left((I_U + I_Y)(\phi_U N_U + \phi_Y N_Y)C + \frac{I_Y}{\phi_Y} N_Y (1-C) \right) - \kappa E_Y,$$

$$\frac{\mathrm{d}I_U}{\mathrm{d}t} = \kappa E_U - (\gamma + \mu)I_U,$$

$$\frac{\mathrm{d}I_Y}{\mathrm{d}t} = \kappa E_Y - (\gamma + \mu)I_Y,$$

$$\frac{\mathrm{d}R_U}{\mathrm{d}t} = \gamma I_U,$$

$$\frac{\mathrm{d}R_Y}{\mathrm{d}t} = \gamma I_Y,$$

$$\frac{\mathrm{d}D_U}{\mathrm{d}t} = \mu I_U,$$

$$\frac{\mathrm{d}D_Y}{\mathrm{d}t} = \mu I_Y,$$

$$(6)$$

where $0 \le \phi_U \le 1$ is the fraction of the population that lives in an urban environment; $\phi_Y = 1 - \phi_U$ is the fraction of the population that lives in a rural environment; N_U is the average number of contacts that an urban individual has in a day; N_Y is the average number of contacts that a rural individual has in a day. C is the "connectedness" parameter, where $C = 1$ means the whole population is well-mixed, and $C = 0$ means the two communities are isolated from one another.

In this model, individuals do not permanently move to the other community but can interact with individuals in the other community. This should thus be seen as short-term travel, e.g. commuting to work, quick shopping trips, and so on.

Compared to a single-community model, splitting the population into two communities primarily affects how susceptible individuals become infected, i.e. the functional forms in equations $\frac{\mathrm{d}S_U}{\mathrm{d}t}$ and $\frac{\mathrm{d}S_Y}{\mathrm{d}t}$. These equations are based on the model employed by Tun *et al.* (2021), where the infectious

group that the susceptible group interacts with, which is represented as I in a one-community model, now needs to take into consideration the degree to which the communities are separated. If $C = 1$ and the communities are completely well mixed, then we can multiple the average number of contacts that a person in that community has in a day $(\phi_U N_U + \phi_Y N_Y)$ by the fraction of those contacts that are infectious $(I_U + I_Y)$ to give us the number of infectious contacts for a susceptible individual. If the communities are completely separate, i.e. $C = 0$, then a susceptible individual only contacts people in their own community so we multiply the number of contacts an individual has, N_U or N_Y, by the fraction of their community that is infectious, I_U/ϕ_U or I_Y/ϕ_Y. For any level of connectedness C, we weight these two scenarios accordingly to obtain the number of infectious contacts for a susceptible individual.

The tutorial for this model shows how it can be parameterised using contact and demographic data, and illustrates the sensitivity of the outputs to the connectedness parameter, C. A key point of this vignette is to illustrate how spatial structure generally leads to slower spread of a pathogen through a population compared to spatially homogeneous models. As such, interventions such as travel bans and travel restrictions may reduce connectivity between subgroups of a population and thus can be used to effectively slow disease spread.

2.6. SEmIRD model: a model with multiple exposed compartments

In the SEIRD model, there is a single exposed compartment, which implies that the latent period follows an exponential distribution. An implication of the exponential distribution is that the rate at which infected individuals become infectious is greatest at the point when they first become infected. This contrasts with laboratory studies, which indicate that there is typically a non-negligible delay between when an individual becomes infected with SARS-CoV-2 and when they become infectious to others (Xin *et al.*, 2021). One way to introduce delays in such ODE models is to break up a compartment into multiple subcompartments. In the SEmIRD model, we include a user-specified number of intermediate exposed compartments,

E_1, E_2,...,E_n resulting in a model structure of the form:

$$\frac{dS}{dt} = -\beta SI,$$

$$\frac{\mathrm{d}E_i}{\mathrm{d}t} = \begin{cases} \beta SI - \kappa E_i, & \text{if } i = 1 \\ \kappa E_{i-1} - \kappa E_i, & \text{otherwise} \end{cases}$$

$$\frac{dI}{dt} = \kappa E_n - (\gamma + \mu)I, \qquad (7)$$

$$\frac{dR}{dt} = \gamma I,$$

$$\frac{dD}{dt} = \mu I.$$

The tutorial for this model shows that the model eqs. (7) results in a latent period distribution which is closer to what is observed experimentally.

2.7. SEIRDV and SEIRDVAge: models including vaccinations

A key intervention that has been used extensively by governments to curb COVID-19 transmission is mass vaccination campaigns, and, here, we present two different models for understanding their impact: the first is the "base" SEIRDV model; the second is an age-structured version of this model (SEIRDVAge). In these models, we add two compartments: one for vaccinated individuals who have not previously been infected (V); and another for those who have (V_R). The SEIRDV model takes the form:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI - \nu \mathrm{Inter}(t)S + \delta_V V + \delta_R R + \delta_{V_R} V_R,$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta SI - \kappa E,$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \kappa E - (\gamma + \mu)I,$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \delta_R R - \nu \mathrm{Inter}(t)R, \qquad (8)$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \mu I,$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \nu \mathrm{Inter}(t)S - \delta_V V,$$

$$\frac{\mathrm{d}V_R}{\mathrm{d}t} = \nu \mathrm{Inter}(t)R - \delta_{V_R} V_R.$$

15

Here, we assume that vaccination provides perfect protection against infection while the person remains in the vaccinated compartment, but vaccinated individuals eventually lose immunity and return to susceptible. We also assume all (alive) non-vaccination compartments can be vaccinated with the exception of infectious individuals, whom we assume are symptomatic and do not participate at that time in vaccination campaigns.

The maximum rate at which individuals can be vaccinated is given by $\nu > 0$; the model allows dynamic changes in the vaccination campaign, and this is controlled by a "vaccination protocol" function $\text{Inter}(t)$ (whose outputs are between 0 and 1) which modulates the intensity of the campaign. In this model, the rate of loss of individuals' immunity to reinfection is controlled by parameters δ_V, δ_R and δ_{V_R} for those previously vaccinated, those who have previously been infected and those who were vaccinated after they were infected respectively.

Faced with vaccinating an entire country, governments have chosen to prioritise vaccinations for those whom are most vulnerable. One dominant characteristic of these campaigns has been to prioritise vaccinations for the older age groups who are at higher risk of COVID-19 complications (Verity *et al.* (2020)). To model age-specific vaccination campaigns, we extend eqs. (8) to incorporate age-structuring as in eqs. (3):

$$
\begin{aligned}
\frac{\mathrm{d}S_i}{\mathrm{d}t} &= -\beta S_i \sum_j C_{ij} I_j - \nu \text{Inter}_i(t) S_i + \delta_V V_i + \delta_R R_i + \delta_{V_R} V_{Ri}, \\
\frac{\mathrm{d}E_i}{\mathrm{d}t} &= \beta S_i \sum_j C_{ij} I_j - \kappa E_i, \\
\frac{\mathrm{d}I_i}{\mathrm{d}t} &= \kappa E_i - (\gamma + \mu) I_i, \\
\frac{\mathrm{d}R_i}{\mathrm{d}t} &= \gamma I_i - \delta_R R_i - \nu \text{Inter}_i(t) R_i, \\
\frac{\mathrm{d}V_i}{\mathrm{d}t} &= \nu \text{Inter}_i(t) S_i - \delta_V V_i, \\
\frac{\mathrm{d}V_{Ri}}{\mathrm{d}t} &= \nu \text{Inter}_i(t) R_i - \delta_{V_R} V_{Ri},
\end{aligned}
\tag{9}
$$

where $\text{Inter}_i(t)$ indicates an age-specific vaccination protocol.

Tutorials for the SEIRDV and SEIRDVAge models illustrate the importance of mass vaccination campaigns for controlling the current pandemic.

2.8. SEIRDNPIAge: a model including non-pharmaceutical interventions

Apart from pharmaceutical interventions such as vaccinations, a host of non-pharmaceutical interventions (NPIs) have been used by governments around the world (Hale *et al.*, 2021). These include physical distancing, shielding vulnerable groups (for example, the elderly), school closures, and lockdowns. Different types of NPIs may introduce changes in social contact matrices and/or the transmission rate β (Davies *et al.*, 2020), which leads to the following ODEs representation:

$$
\begin{aligned}
\frac{\mathrm{d}S_i}{\mathrm{d}t} &= -\beta(t)S_i \sum_j C_{ij}(t)I_j, \\
\frac{\mathrm{d}E_i}{\mathrm{d}t} &= \beta(t)S_i \sum_j C_{ij}(t)I_j - \kappa E_i, \\
\frac{\mathrm{d}I_i}{\mathrm{d}t} &= \kappa E_i - (\gamma_i + \mu_i)I_i, \\
\frac{\mathrm{d}R_i}{\mathrm{d}t} &= \gamma_i I_i, \\
\frac{\mathrm{d}D_i}{\mathrm{d}t} &= \mu_i I_i.
\end{aligned}
\tag{10}
$$

In this model, there are two ways to allow a dynamic change in transmission rates: the first is through contact matrices, with elements $C_{ij}(t)$, which can vary through time – these could, for example, allow for those elderly age groups to isolate themselves for a period of time; the second is through dynamic transmission rate parameter, $\beta(t)$, which could, for example, be used to represent the impact of lockdowns when social contacts across all ages are approximated as being reduced by a single common factor.

The tutorial for this model shows how this model can be used to investigate the effect of a lockdown similar to the one instituted in the UK on 23rd March 2020. In doing so, this illustrates how this NPI can "flatten" the epidemic curve.

2.9. SEIRD_CT

Contact tracing, testing and isolation are key elements of many governments' responses to the COVID-19 pandemic. In this model, we consider how these types of NPIs can be used to control the spread of COVID-19. The model has the form:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -(\beta(P + I) + \beta_a A)S,$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \beta(1 - \chi)(P + I)S + \beta_a AS - \omega E,$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = (1 - \eta_a)\omega E - \psi P,$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \eta_a \omega E - \gamma A,$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = (1 - \phi)\psi P - (\gamma + \mu)I,$$

$$\frac{\mathrm{d}E^T}{\mathrm{d}t} = \beta \chi(P + I)S - \omega E^T, \tag{11}$$

$$\frac{\mathrm{d}P^T}{\mathrm{d}t} = (1 - \eta_a)\omega E^T - \psi P^T,$$

$$\frac{\mathrm{d}A^T}{\mathrm{d}t} = \eta_a \omega E^T - \gamma A^T,$$

$$\frac{\mathrm{d}I^T}{\mathrm{d}t} = \psi P^T + \phi \psi P - (\gamma + \mu)I^T,$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma(I + I^T + A + A^T),$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = \mu(I + I^T).$$

The effectiveness of contact tracing and quarantining hinges crucially on the timing of when individuals are isolated relative to their profile of infectiousness. It also depends on the difficulty of identifying infected and/or infectious individuals in the first place. To adequately account for these difficulties, we introduce a number of additions over and above the SEIRD model: since COVID-19 is thought to be transmissible *before* symptoms appear (He *et al.*, 2020), we include a presymptomatic infectious compartment, P; we also introduce an asymptomatic infectious compartment, A (making up a corresponding fraction $0 \leq \eta_a \leq 1$ of overall cases). If infected individuals are successfully isolated, they are assumed not to transmit infections from the moment of isolation onwards. They then follow a series of transitions of the same form as unisolated individuals: here, we index the isolated states by a superscript T. This model considers two types of NPIs: self isolation upon appearance of symptoms with a proportion $0 \leq \phi \leq 1$ of symptomatic

individuals doing so; and isolation following being traced as having come into contact with an infected individual: due to the difficulty in identifying asymptomatic individuals, only contacts of individuals who go on to be symptomatic cases are deemed traceable. Since the proportion of susceptible individuals forced to isolate is likely small relative to the overall susceptible pool, we do not include a compartment for susceptible individuals who are quarantined. Here, $0 \leq \chi \leq 1$ determines the proportion of contacts of symptomatic cases that are traced and isolated. This model structure is visualised in Figure 3.

The tutorial for this model considers the relative effectiveness of these two types of NPIs and investigates their usefulness for controlling COVID-19 epidemics induced by ancestral SARS-CoV-2 and the Delta and Omicron variants.

3. Methods

3.1. Package structure

The **comomodels** package uses an object-oriented approach to design. In particular, it uses R's S4 classes to encapsulate each of the models described in §2. These classes allow a user to instantiate an object, corresponding to a particular compartmental model. Operations can then be performed in a consistent manner between different models. For example, a user may set, change or ask for the transmission parameters underpinning the model; a user may set, change or ask for the initial conditions of the system; they may run the model; and they may ask a model to return its basic reproduction number.

Understanding a model's structure can be made easier if the system's states and flows between them are visualised using a compartmental diagram. In **comomodels**, we provide this functionality to a user: calling the `ode_structure_diagram` method plots the compartmental structure (see Figure 3).

Apart from deciding on a model structure, the most important consideration in modelling the spread of any infectious disease is choice of parameter values, since the (typically) non-linear ODE system permits a variety of dynamics, and only a subset of these is epidemiologically realistic. Since this package is specific to COVID-19 modelling, we provide users with some predetermined parameter sets characterising the spread of ancestral SARS-CoV-2 and two of its variants: Delta and Omicron. Whilst there
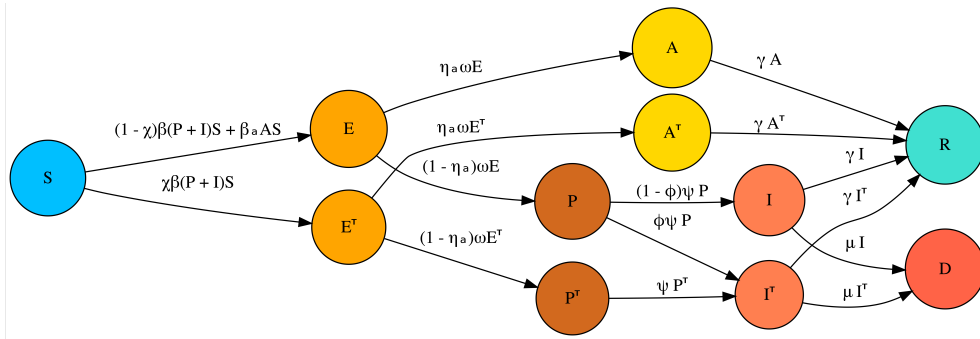
Figure 3: **Compartmental model structure of the SEIRD_CT model.** See 2.9 for a description of this model.

is ample uncertainty in these parameter values (particularly for Omicron), these parameter sets represent reasonable starting points for exploring how the different variants spread. These parameter sets are accessed through the `covid_transmission_parameters` function, which can, in addition, return age group-specific parameters if the model being parameterised is age-structured. The associated function documentation provides information on the studies and assumptions made in deriving these parameter value choices.

To run the age-structured models and the SEIRD_RU model, further inputs are required: an age-structured model requires contact matrices (see §2.2 and Figure 2) and to set the initial conditions requires fractions of the population in each age group; the urban-rural model requires fractions of individuals living in rural versus urban environments. As part of the package, we provide datasets for each of these elements across a diverse range of countries.

On running each model, an R list is returned with two named dataframe objects: "states", which contains the dynamic evolution of the state variables of the ODE (for example, for the SEIRD model, there are time series for susceptibles, exposed, infectious, recovered and dead individuals); and "changes", which calculate the number of individuals becoming infected or dying on a daily basis. This latter dataframe is particularly useful for comparison with observed epidemiological data. Each of these dataframes is in the *long* format to facilitate quick and powerful graphing through the ggplot2 package (Wickham, 2016). The dataframes each have column headings: "time" (in days), "compartment", "age_range" and "value". If the model is not age-structured, an age range of 0-150 is assigned to each data point. Fig-

20

ure 4 plots the "states" from running the age-structured SEIRD model for a parameter set typical for transmission of ancestral SARS-CoV-2 (see the tutorial for this model for more information).
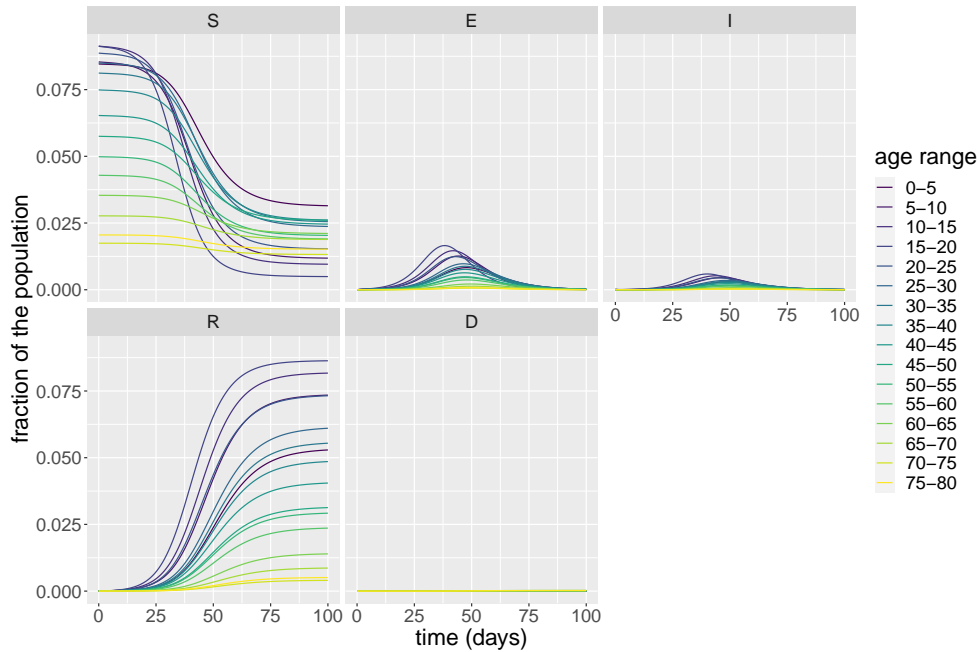

Figure 4: **Example visualisation: modelled states for an age-structured model.** Each panel corresponds to a different state (see §2.2); each line corresponds to an age group where the colours are given by the legend.

Along with the raw package functionality, the package comes with a series of detailed tutorials[3] – one for each model described in §2.

3.2. The ComoModel GUI

To complement the **comomodels** package, we have also produced an R shiny application (The Como-DTC Development Team, 2021) as an openly accessible web-based interface to the model. This tool allows users to select from the suite of models within the package, producing an interactive visualisation of the model structure. For each of the submodels, we introduce the model structure through a compartment structure diagram and its

---

[3]Note that these tutorials are different to those described in §4.

full ODE system. To use the application interactively, users first access a target model by selecting a header tab. Then by either clicking on the corresponding parameter on the diagram or directly navigating through the side bar, the user can tune the model's parameters (see Fig 5). Additionally, for age-structured models, users can upload social contact matrices from a host of countries (Prem *et al.*, 2017) and population demography files (United Nations, 2019). Each time a parameter is modified, a simulation is run automatically, and the outputs are interactively visualised using plotly (Plotly Technologies Inc. (2015)). The model states and the daily incidence and deaths are visualised in separate panels. These simulation results can then be downloaded as comma-separated values (csv) files.
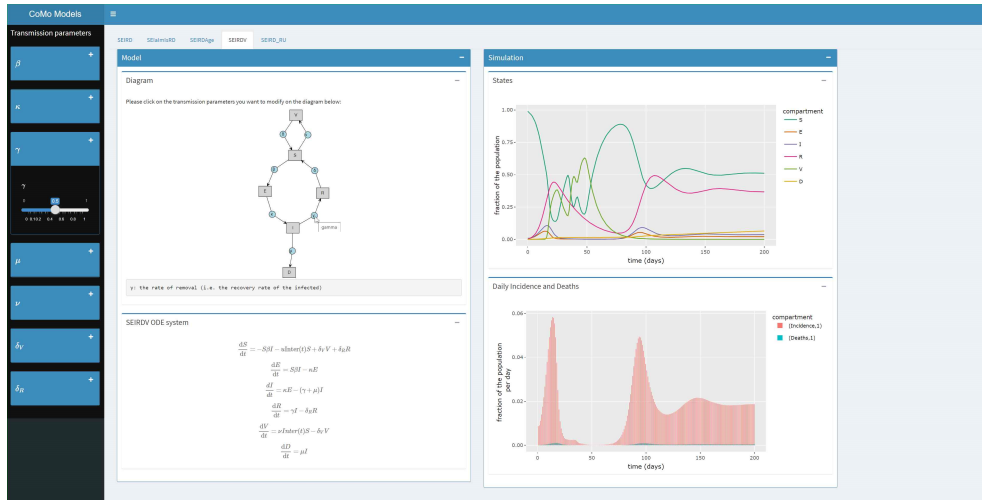

Figure 5: **Using comomodels-explore to investigate the SEIRDV submodel.** Here, the user selects the parameter γ (see §2.7) by clicking on the ODE structure diagram, which allows them to change this input.

3.3. Software engineering approach

The model underpinning the influential Imperial College Report 9 (Ferguson *et al.*, 2020) received substantial criticism for its code and software engineering practices (Baker, 2020). The bulk of these criticisms were unfair and ill-spirited given the time pressure the Imperial team were under to deliver a message of key importance for public health policymaking. As a silver lining, however, there has subsequently been a call for an increased focus on the *quality* of research software used to direct policy in public health

22

(Horner & Symons, 2020; Habli *et al.*, 2020). We aimed to create software according to robust software engineering practices, which included collaborative development over Github using pull requests with reviews, class and function documentation, pedagogical vignettes (as discussed above), style checking of code, unit testing of individual functions and continuous integration testing. Before we began to design the software, the team had a mix of skill levels in software engineering, with some having not done any formal development previously, including the use of Github; others had little or no experience in transmission dynamics modelling. As such, the process of creating **comomodels** produced a useful byproduct – it taught a host of (mostly) early career scientists how to do software engineering and introduced them to epidemiological modelling in the midst of a pandemic.

4. Results

We used the **comomodels** package to produce three key lessons in understanding the transmission of COVID-19, which are contained within dynamic R markdown vignettes included as supplementary materials. In this section, we provide a brief description of these.

4.1. On the numerical solution of compartmental models

All of the models considered in **comomodels** are systems of ODEs which do not have general analytical solutions. Thus, for any set of parameter values, the equations are solved approximately using numerical methods. In the vignette labelled "S1", we first describe a simple approach to solving differential equations using a fixed step size and demonstrate the importance of choosing an appropriate step size for the SEIRD model. In particular, we show that simulations of daily cases and deaths obtained using a daily time step differ significantly from those obtained with smaller steps. Next, we provide a brief overview of adaptive step size solvers and show how these are used in **comomodels** to obtain accurate solutions with reasonable runtimes. Finally, we illustrate how allowing discontinuous changes in intervention variables, such as vaccination coverage, can cause problems for numerical solution of compartmental models, and we describe approaches to avoid some of these issues.

Overall, these analyses demonstrate the importance of considering inbuilt ODE solvers not as oracular – rather as a set of tools which must be yielded

with care and consideration. Of particular import when doing applied modelling is to experiment with the numerical solver step size to investigate if substantial changes to the outputs result when step size is decreased. If so, this signifies that the ODE solution has likely been poorly approximated.

4.2. The importance of uncertainty in age-specific contact patterns for quantifying COVID-19 risk

Contact matrices are an important element of age-structured transmission dynamics models, where they are used to dictate intra- and inter-generational mixing (eq. (3)). In the "S2" vignette, we perform a type of sensitivity analysis on this element of the model. To do so, we use the **socialmixr** package (with POLYMOD contact data (Mossong *et al.*, 2008)) to generate bootstrapped samples of the contact matrix in the United Kingdom. By performing multiple simulations of the SEIRDAge model (once for each contact matrix sample), we propagate the uncertainty in the contact matrix and estimate the corresponding uncertainty in epidemic trajectories and deaths. The results show substantial uncertainty in both of these quantities, motivating the importance of incorporating uncertainty in the contact matrix either when making projections or when performing inference on age-structured models.

4.3. The importance of sensitivity analysis and model inference

Epidemiological models allow us to infer key model parameters of the disease from data, such as the infection rate. Using these parameter estimates, the models can be used to predict the trajectory of an ongoing epidemic. The reliability of such predictions, however, rests on (a) whether the model itself is a faithful approximation of reality and (b) whether the data are sufficient to identify parameters.

In the "S3" vignette, we show how using limited data, typical of that encountered early on during an epidemic, can lead to considerable uncertainty in key model parameter estimates. In particular, we use early COVID-19 case incidence and deaths data for London (UK Health Security Agency (UK HSA), 2021) collected early on during the first wave in 2020 to illustrate that many diverse parameter sets are compatible with the observed data. A consequence of this is that, at that stage in the epidemic, there were many plausible future epidemic trajectories.

To further illustrate this issue with model identifiability, we conduct a profile likelihood analysis (Murphy & Van der Vaart, 2000; Kreutz *et al.*,

2013) using simulated data. This analysis highlights the difficulties involved in estimating the parameters of the SEIRD model directly from observed cases and deaths, particularly early on during the epidemic. This means that, in reality, a number of key parameters are typically estimated using other sources of data – for instance, studies of patients admitted to hospital with COVID-19.

5. Discussion and conclusion

The mathematical models used to inform public health policy for COVID-19 are often large and intricate, due to the inherent complexities of the manifold ways that this epidemic has affected society. Understanding these models can be challenging, particularly for those new to infectious disease modelling. Our software, **comomodels**, helps people to understand the common modelling motifs that appear in these models: bridging the gap between introductory SEIRD models and more intricate models used for policymaking.

As models become more complex, there is a greater chance of coding errors. Accordingly, we designed the **comomodels** package using robust software development practices, including unit testing and continuous integration testing.

The package thus far focuses on deterministic compartmental models of COVID-19 transmission dynamics. A number of other modelling frameworks have been used, including stochastic compartmental models (e.g. Ferguson *et al.* (2020)) and stochastic renewal models (e.g. Flaxman *et al.* (2020); Nouvellet *et al.* (2021)). Compartmental models have also been extended spatially, to allow spatial variation in transmission and to explore how inter- and intra-country mobility patterns influence transmission (Ferguson *et al.*, 2020; Keeling *et al.*, 2021; Danon *et al.*, 2021). In the future, we intend to expand the package to explore further these modelling frameworks.

6. Acknowledgements

References

Abu-Raddad, Laith J, Chemaitelly, Hiam, Coyle, Peter, Malek, Joel A, Ahmed, Ayeda A, Mohamoud, Yasmin A, Younuskunju, Shameem, Ayoub, Houssein H, Al Kanaani, Zaina, Al Kuwari, Einas, *et al.* 2021. SARS-CoV-2 antibody-positivity protects against reinfection for at least seven months with 95% efficacy. *EClinicalMedicine*, **35**, 100861.

Aguas, Ricardo, White, Lisa, Hupert, Nathaniel, Shretta, Rima, Pan-Ngum, Wirichada, Celhay, Olivier, Moldokmatova, Ainura, Arifi, Fatima, Mirza-zadeh, Ali, Sharifi, Hamid, Adib, Keyrellous, Sahak, Mohammad Nadir, Franco, Caroline, Coutinho, Renato, & CoMo Consortium. 2020. Modelling the COVID-19 pandemic in context: an international participatory approach. *BMJ Global Health*, **5**(12).

Anderson, R, Donnelly, C, Hollingsworth, D, Keeling, M, Vegvari, C, Baggaley, R, & Maddren, R. 2020. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. *The Royal Society*, **2020**.

Anderson, Roy M., & May, Robert M. 1992. *Infectious diseases of humans: dynamics and control.* Oxford University Press.

Baker, Simon. 2020. *Pandemic response shines spotlight on coding in science.* https://www.timeshighereducation.com/news/pandemic-response-shines-spotlight-coding-science. Accessed: 2021-12-06.

Birrell, Paul, Blake, Joshua, van Leeuwen, Edwin, Gent, Nick, & De Angelis, Daniela. 2021. Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Philosophical Transactions of the Royal Society B*, **376**(1829), 20200279.

Brauer, Fred. 2008. Compartmental models in epidemiology. *Pages 19–79 of: Mathematical epidemiology.* Springer.

Byambasuren, Oyungerel, Cardona, Magnolia, Bell, Katy, Clark, Justin, McLaws, Mary-Louise, & Glasziou, Paul. 2020. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission:

systematic review and meta-analysis. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada*, **5**(4), 223–234.

Calvetti, Daniela, Hoover, Alexander P., Rose, Johnie, & Somersalo, Erkki. 2020. Metapopulation Network Models for Understanding, Predicting, and Managing the Coronavirus Disease COVID-19. *Frontiers in Physics*, **8**, 261.

Danon, Leon, Brooks-Pollock, Ellen, Bailey, Mick, & Keeling, Matt. 2021. A spatial model of COVID-19 transmission in England and Wales: early spread, peak timing and the impact of seasonality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**(1829), 20200272.

Davies, Nicholas G., Kucharski, Adam J, Eggo, Rosalind M., Gimma, Amy, Edmunds, W. John, Jombart, Thibaut, O'Reilly, Kathleen, Endo, Akira, Hellewell, Joel, Nightingale, Emily S., Quilty, Billy J., Jarvis, Christopher I., Russell, Timothy W., Klepac, Petra, Bosse, Nikos I., Funk, Sebastian, Abbott, Sam, Medley, Graham F., Gibbs, Hamish, Pearson, Carl A.B., Flasche, Stefan, Jit, Mark, Clifford, Samuel, Prem, Kiesha, Diamond, Charlie, Emery, Jon, Deol, Arminder K., Procter, Simon R., van Zandvoort, Kevin, Sun, Yueqian Fiona, Munday, James D., Rosello, Alicia, Auzenbergs, Megan, Knight, Gwen, Houben, Rein M.G.J, & Liu, Yang. 2020. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, **5**(7), e375–e385.

Diekmann, Odo, Heesterbeek, JAP, & Metz, Johan AJ. 1990. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, **28**(4), 365–382.

Ferguson, Neil M., Laydon, Daniel, Nedjati-Gilani, Gemma, Imai, Natsuko, Ainslie, Kylie, Baguelin, Marc, Bhatia, Sangeeta, Boonyasiri, Adhiratha, Cucunubá, Zulma, Cuomo-Dannenburg, Gina, Dighe, Amy, Dorigatti, Ilaria, Fu, Han, Gaythorpe, Katy, Green, Will, Hamlet, Arran, Hinsley, Wes, Okell, Lucy C., van Elsland, Sabine, Thompson, Hayley, Verity, Robert, Volz, Erik, Wang, Haowei, Wang, Yuanrong, Walker, Patrick G.T., Walters, Caroline, Winskill, Peter, Whittaker, Charles, Donnelly, Christl A., Riley, Steven, & Ghani, Azra C. 2020. *Report 9: Impact*

of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.

Flaxman, Seth, Mishra, Swapnil, Gandy, Axel, Unwin, H Juliette T, Mellan, Thomas A, Coupland, Helen, Whittaker, Charles, Zhu, Harrison, Berah, Tresnia, Eaton, Jeffrey W, *et al.* 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, **584**(7820), 257–261.

Gallais, Floriane, Gantner, Pierre, Bruel, Timothée, Velay, Aurélie, Planas, Delphine, Wendling, Marie-Josée, Bayer, Sophie, Solis, Morgane, Laugel, Elodie, Reix, Nathalie, *et al.* 2021. Evolution of antibody responses up to 13 months after SARS-CoV-2 infection and risk of reinfection. *EBioMedicine*, **71**, 103561.

Gandhi, Monica, Yokoe, Deborah S., & Havlir, Diane V. 2020. Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19. *New England Journal of Medicine*, **382**(22), 2158–2160.

Habli, Ibrahim, Alexander, Rob, Hawkins, Richard, Sujan, Mark, McDermid, John, Picardi, Chiara, & Lawton, Tom. 2020. Enhancing COVID-19 decision making by creating an assurance case for epidemiological models. *BMJ Health & Care Informatics*, **27**(3).

Hale, Thomas, Angrist, Noam, Goldszmidt, Rafael, Kira, Beatriz, Petherick, Anna, Phillips, Toby, Webster, Samuel, Cameron-Blake, Emily, Hallas, Laura, Majumdar, Saptarshi, *et al.* 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, **5**(4), 529–538.

He, Xi, Lau, Eric HY, Wu, Peng, Deng, Xilong, Wang, Jian, Hao, Xinxin, Lau, Yiu Chung, Wong, Jessica Y, Guan, Yujuan, Tan, Xinghua, *et al.* 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*, **26**(5), 672–675.

Hogan, A, Winskill, P, Watson, O, Walker, P, Whittaker, C, Baguelin, M, Haw, D, Lochen, A, Gaythorpe, K, Ainslie, K, *et al.* 2020. Report 33: modelling the allocation and impact of a COVID-19 vaccine.

Horner, Jack K., & Symons, John F. 2020. Software engineering standards for epidemiological models. *History and Philosophy of the Life Sciences*, **42**(4), 1–24.

IFG. 2021. *Timeline of UK coronavirus lockdowns, March 2020 to March 2021.* Accessed: January 3rd 2022.

Johansson, Michael A., Quandelacy, Talia M., Kada, Sarah, Prasad, Pragati Venkata, Steele, Molly, Brooks, John T., Slayton, Rachel B., Biggerstaff, Matthew, & Butler, Jay C. 2021. SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Network Open*, **4**(1), e2035057–e2035057.

Keeling, Matt J, Hill, Edward M, Gorsich, Erin E, Penman, Bridget, Guyver-Fletcher, Glen, Holmes, Alex, Leng, Trystan, McKimm, Hector, Tamborrino, Massimiliano, Dyson, Louise, *et al.* 2021. Predictions of COVID-19 dynamics in the UK: short-term forecasting and analysis of potential exit strategies. *PLOS Computational Biology*, **17**(1), e1008619.

Kramer, Sarah C., Pei, Sen, & Shaman, Jeffrey. 2020. Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel. *PLOS Computational Biology*, **16**(10), 1–21.

Kreutz, Clemens, Raue, Andreas, Kaschek, Daniel, & Timmer, Jens. 2013. Profile likelihood in systems biology. *The FEBS Journal*, **280**(11), 2564–2571.

Leidi, Antonio, Koegler, Flora, Dumont, Roxane, Dubos, Richard, Zaballa, María-Eugenia, Piumatti, Giovanni, Coen, Matteo, Berner, Amandine, Farhoumand, Pauline Darbellay, Vetter, Pauline, *et al.* 2021. Risk of reinfection after seroconversion to SARS-CoV-2: A population-based propensity-score matched cohort study. *MedRxiv*.

Mossong, Joël, Hens, Niel, Jit, Mark, Beutels, Philippe, Auranen, Kari, Mikolajczyk, Rafael, Massari, Marco, Salmaso, Stefania, Tomba, Gianpaolo Scalia, Wallinga, Jacco, Heijne, Janneke, Sadkowska-Todys, Malgorzata, Rosinska, Magdalena, & Edmunds, W. John. 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, **5**(3), e74.

Murphy, Susan A, & Van der Vaart, Aad W. 2000. On profile likelihood. *Journal of the American Statistical Association*, **95**(450), 449–465.

Nouvellet, Pierre, Bhatia, Sangeeta, Cori, Anne, Ainslie, Kylie EC, Baguelin, Marc, Bhatt, Samir, Boonyasiri, Adhiratha, Brazeau, Nicholas F, Cattarino, Lorenzo, Cooper, Laura V, *et al.* 2021. Reduction in mobility and COVID-19 transmission. *Nature Communications*, **12**(1), 1–9.

Plotly Technologies Inc. 2015. *Collaborative data science.*

Prem, Kiesha, Cook, Alex R., & Jit, Mark. 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Computational Biology*, **13**(9), e1005697.

The Como-DTC Development Team. 2021. *Comomodels GUI.* `https://comodtc.shinyapps.io/comomodels-explore`.

Tun, Sai Thein Than, Parker, Daniel M., Aguas, Ricardo, & White, Lisa J. 2021. The assembly effect: the connectedness between populations is a double-edged sword for public health interventions. *Malaria Journal*, **20**, 189.

UK Health Security Agency (UK HSA). 2021. *Cases in London: Cases by specimen date.* `https://coronavirus.data.gov.uk/details/cases?areaType=region&areaName=London`. [Online; accessed 08-October-2021].

United Nations. 2019. *World Population Prospects 2019, Online Edition. Rev. 1.*

Verity, Robert, Okell, Lucy C, Dorigatti, Ilaria, Winskill, Peter, Whittaker, Charles, Imai, Natsuko, Cuomo-Dannenburg, Gina, Thompson, Hayley, Walker, Patrick G.T., Fu, Han, Dighe, Amy, Griffin, Jamie T., Baguelin, Marc, Bhatia, Sangeeta, Boonyasiri, Adhiratha, Cori, Anne, Cucunubá, Zulma, FitzJohn, Rich, Gaythorpe, Katy, Green, Will, Hamlet, Arran, Hinsley, Wes, Laydon, Daniel, Nedjati-Gilani, Gemma, Riley, Steven, van Elsland, Sabine, Volz, Erik, Wang, Haowei, Wang, Yuanrong, Xi, Xiaoyue, Donnelly, Christl A., Ghani, Azra C., & Ferguson, Neil M. 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, **20**(6), 669–677.

Walker, Patrick GT, Whittaker, Charles, Watson, Oliver J, Baguelin, Marc, Winskill, Peter, Hamlet, Arran, Djafaara, Bimandra A, Cucunubá, Zulma, Mesa, Daniela Olivera, Green, Will, *et al.* 2020. The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*, **369**(6502), 413–422.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Xin, Hualei, Li, Yu, Wu, Peng, Li, Zhili, Lau, Eric HY, Qin, Ying, Wang, Liping, Cowling, Benjamin J, Tsang, Tim, & Li, Zhongjie. 2021. Estimating the latent period of coronavirus disease 2019 (COVID-19). *Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America.*