A population-level SEIR model for COVID-19 scenarios

IEMAG Epidemiology Modelling subgroup

11 May 2020

Abstract

This technical note summarizes the current version of the population-level susceptible-exposed-infected-removed (SEIR) model that is used by the Irish Epidemiological Modelling Advisory Group (IEMAG) reporting to the National Public Health Emergency Team (NPHET). The model is continuously updated, so this note supersedes previous versions, and will be superseded by future updates.

1 Description of the model

Population-level SEIR models [1–3] assume fully-mixed, homogeneous populations. Despite this simplification, they provide useful information for scenario-based planning, with the potential for further extension of the structure (e.g., to disaggregated age cohorts) in more advanced models.

At each moment of time, each individual in the population is considered to be in one of a discrete number of compartments. The structure of the compartments, and the timescales for individuals to move in and out of compartments, are based on the current understanding of the epidemiology of COVID-19, as evidenced by the extensive literature review and evidence synthesis conducted by [4–6].

1.1 Compartments and parameters

The structure of the model compartments is shown in Figure 1. Informed by the literature review, each parameter is chosen from a uniform distribution with upper and lower limits as listed below. (All timescales are expressed in days. Day 0 of the Irish epidemic is February 28, 2020.)

- L: average latent period; range assumed is 3.9 to 5.9.
- C: average incubation period; range assumed is $\max(L, 4.8)$ to 6.8 (lower limit ensures $C - L$ is nonnegative).
- D: average infectious period; range assumed is $\max(C - L, 5.0)$ to 9.0 (lower limit ensures $D - (C - L)$ is nonnegative).
- h: multiplicative factor for reduction of effective transmission from the Asymptomatic Infected compartment, relative to Symptomatic Infected; range assumed is 0.01 to 0.5.
- i: multiplicative factor for reduction of effective transmission from the Immediate Isolation compartment, relative to Symptomatic Infected; range assumed is 0 to 0.1.
- j: multiplicative factor for reduction of effective transmission from the Post-test isolation compartment, relative to Symptomatic Infected; range assumed is 0 to 0.1.

- f: fraction of infected who are Asymptomatic; range assumed is 0.18 to 0.82.
- τ: fraction of symptomatic cases that are tested; range assumed is 0.5 to 1.0.
- q: fraction of symptomatic cases that self-quarantine from appearance of symptoms until recovery; range assumed is 0 to $1-\tau$ (upper limit ensures $1-q-\tau$ is nonnegative).
- T: average wait for test results; range assumed is 1.0 to $\max(5.0, D-C+L)$ (upper limit ensures $D-C+L-T$ is nonnegative).
- N: population, assumed to be 4.9×10^6.

The transmission rate β is assumed to be time-dependent, in order to model the impact of interventions. The relationship between β and the basic reproductive number R_0 for this model is derived in the Appendix.

1.2 Equations

Each compartment in Figure 1 has a corresponding time-dependent variable, which gives the number of individuals that are in that compartment at time t. The dynamics evolve according to the following differential equations:

$$\frac{dS}{dt} = -\beta S\left(I_p + hI_a + iI_i + I_{t1} + jI_{t2} + I_n\right)/N \tag{1}$$

$$\frac{dE}{dt} = \beta S\left(I_p + hI_a + iI_i + I_{t1} + jI_{t2} + I_n\right)/N - \frac{1}{L}E \tag{2}$$

$$\frac{dI_p}{dt} = \frac{(1-f)}{L}E - \frac{1}{C-L}I_p \tag{3}$$

$$\frac{dI_a}{dt} = \frac{f}{L}E - \frac{1}{D}I_a \tag{4}$$

$$\frac{dI_i}{dt} = \frac{q}{C-L}I_p - \frac{1}{D-C+L}I_i \tag{5}$$

$$\frac{dI_{t1}}{dt} = \frac{\tau}{C-L}I_p - \frac{1}{T}I_{t1} \tag{6}$$

$$\frac{dI_{t2}}{dt} = \frac{1}{T}I_{t1} - \frac{1}{D-C+L-T}I_{t2} \tag{7}$$

$$\frac{dI_n}{dt} = \frac{(1-q-\tau)}{C-L}I_p - \frac{1}{D-C+L}I_n \tag{8}$$

$$\frac{dR}{dt} = \frac{1}{D}I_a + \frac{1}{D-C+L}I_i + \frac{1}{D-C+L-T}I_{t2} + \frac{1}{D-C+L}I_n, \tag{9}$$

where $S(t)$ is the number of susceptible individuals, $E(t)$ is the number who are exposed, $I_p(t)$ is the number who are pre-symptomatic infected, $I_a(t)$ is the number who are asymptomatic infected, $I_i(t)$ is the number who are symptomatic and self-isolating (without testing), $I_{t1}(t)$ is the number who are symptomatic and waiting for testing, $I_{t2}(t)$ is the number who are in post-test self-isolation, $I_n(t)$ is the number who are symptomatic and not isolating and $R(t)$ is the number who are removed (i.e., recovered from the virus or dead). In addition, we define $C_c(t)$ to be the cumulative number of new cases reported by time t, given by integrating the flux out of the I_{t1} (waiting-for-test) compartment:

$$\frac{dC_c}{dt} = \frac{1}{T}I_{t1} \tag{10}$$

and we also report the number of new daily cases on day n, defined by $c_c(n) = C_c(n) - C_c(n-1)$.

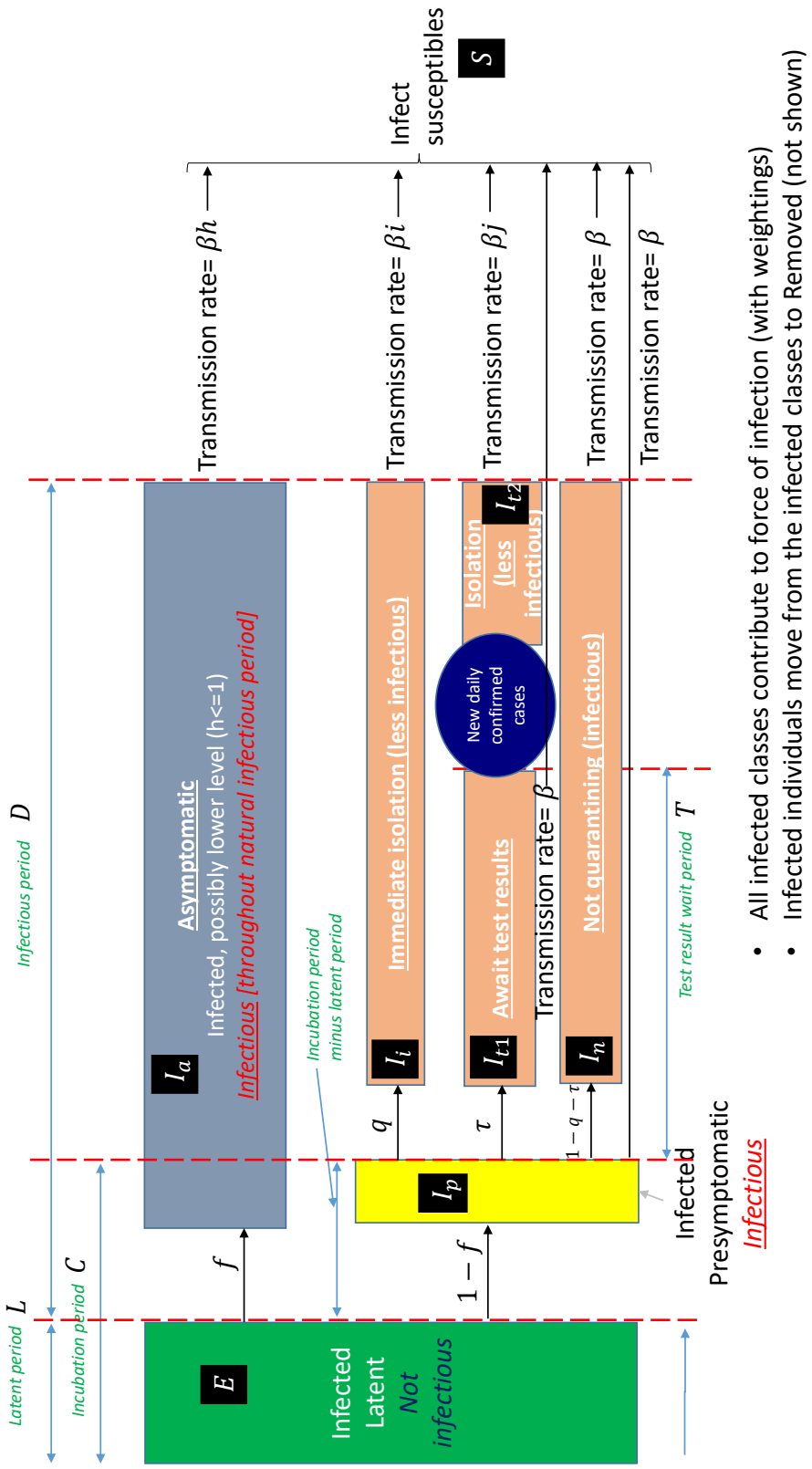

Figure 1: Model compartment structure.

2 Calibration of the model

2.1 Data

The confirmed positive case data was extracted from the Computerised Infectious Disease Reporting (CIDR) database hosted by the Heath Protection Surveillance Centre (HPSC). For each observation, the event date and lab specimen collected date were stored, however for 2.5% of cases the lab specimen collected date were missing. For each sample with a missing lab specimen collected date, the event date was considered and the missing value was imputed using the average lab specimen collected date for all samples with that event date. The lab specimen collected date of the cases was used for modelling purposes.

2.2 Poisson model with rate following a Gompertz function

The daily case counts $c_c(t)$ at day t, as described in Section 2.1, is modelled by a Poisson random variable

$$c_c(t) \sim \mathsf{Po}(g(t)). \tag{11}$$

To model the rate parameter of the Poisson distribution we first use a Gompertz growth function

$$G(t) = \theta_1 e^{-\theta_2 \theta_3^t}, \tag{12}$$

where $\theta_1, \theta_2, \theta_3 > 0$, to model the mean rate of accumulated cases each day. The derivative of $G(t)$ then models the daily rate of cases:

$$\frac{d}{dt}G(t) = \theta_1 \theta_2 \log(1/\theta_3)(\theta_3^t) e^{-\theta_2 \theta_3^t}. \tag{13}$$

We employ a Bayesian model to infer the parameters $\theta_1, \theta_2, \theta_3$. To do so, we specify prior distributions for each parameter as

$$\begin{aligned} \theta_1 &\sim \text{half-}N(10000, 100000), \\ \theta_2 &\sim \text{half-}N(0, 100), \\ \theta_2 &\sim \text{half-}N(0, 10). \end{aligned} \tag{14}$$

The Bayesian model then yields the posterior distribution of $\theta_1, \theta_2, \theta_3$ given the daily data $y_0, \ldots, y_t$. The Rstan package [7] was used to generate samples from the posterior distribution

2.3 Calibration algorithm

We calibrate the SEIR model with data from February 29th to May 6th, using the Gompertz fit described in Section 2.2. For each choice of parameter values, i.e., each realization $G(t)$ of the data-fitting function, the SEIR equations can be inverted to determine the time-varying transmission rate $\beta(t)$ that would yield precisely $C_c(t) = G(t)$. The steps in this inversion algorithm are as follows:

1. Assuming $C_c(t) = G(t)$, solve the algebraic equation (10) for $I_{t1}(t)$.
2. Using $I_{t1}(t)$ from step 1, solve the differential equation (7) for $I_{t2}(t)$.
3. Using $I_{t1}(t)$ from step 1, solve the algebraic equation (6) for $I_p(t)$.
4. Using $I_p(t)$ from step 3, solve the differential equation (5) for $I_i(t)$.

5. Using $I_p(t)$ from step 3, solve the differential equation (8) for $I_n(t)$.

6. Using $I_p(t)$ from step 3, solve the algebraic equation (3) for $E(t)$.

7. Using $E(t)$ from step 6, solve the differential equation (4) for $I_a(t)$.

8. Combine equations (1) and (2) to give

$$\frac{dS}{dt} = -\frac{dE}{dt} - \frac{1}{L}E \tag{15}$$

and, using $E(t)$ from step 6, solve this differential equation for $S(t)$.

9. Using the results of steps 2, 3, 4, 5, 7 and 8, solve equation (1) for the inferred time-dependent transmission rate $\beta(t)$:

$$\beta(t) = -\frac{N}{S}\frac{dS}{dt}\left(I_p + hI_a + iI_i + I_{t1} + jI_{t2} + I_n\right)^{-1}. \tag{16}$$

Using the linear relationship between transmission rate and reproductive number for this model, given by equation (18) in the Appendix, each realization of the inferred $\beta(t)$ can be converted to an equivalent time-varying model-inferred reproductive number R_t. Figure 2 shows the results of this calculation, using 10000 independent draws for the model parameters and for the Gompertz fitting parameters. This method of estimating a reproductive number is specific to the model, and the estimates should be understood as applying within the context of the SEIR model only. Similar to more standard measures for reproductive number, when the model-inferred R_t exceeds one the number of infections will rise exponentially, while the number of infections decays when R_t is less than one. Note the very high values of this model-inferred R_t in the early stage of the epidemic: this is because the SEIR model is not accurate in the early stages of an epidemic, and the model-inferred R_t values should not be taken as indicative of the true value of the reproductive number in the early stages. However, as the figures in Section 3 demonstrate, the SEIR approach yields good fits to the data once the number of cases exceeds approximately 100, and so it can be used for comparison of possible future scenarios.

3 Scenario analysis

For each set of model parameters, the calibration algorithm detailed in Section 2 yields a time-varying transmission rate $\beta(t)$ that is consistent with the Gompertz-fitted data up to the current date. Forward predictions of the model may then be examined under assumptions about how the transmission rate (or, equivalently, the reproductive number) may behave in the future.

To numerically solve the differential equations we require initial conditions that are consistent with the data. Differential equation models for populations are known to be inaccurate when the number of cases is low, so we aim to match the numerical solution for the cumulative number of cases $C_c(t)$ to the Gompertz fitting function $G(t)$ at some time $t_1 > 0$ (we choose $t_1 = 10$ days). We use the following two-step algorithm:

1. Set $I_p(0) = n_s$, $S(0) = N - n_s$ and all other variables to be initially zero, with the number of "seed" cases set to $n_s = 1$. This models a situation where there is exactly one individual who is pre-symptomatic infected on day 0. Solving the differential equations with this initial condition gives a solution where the cumulative number of new cases is denoted $\widetilde{C_c}(t)$.

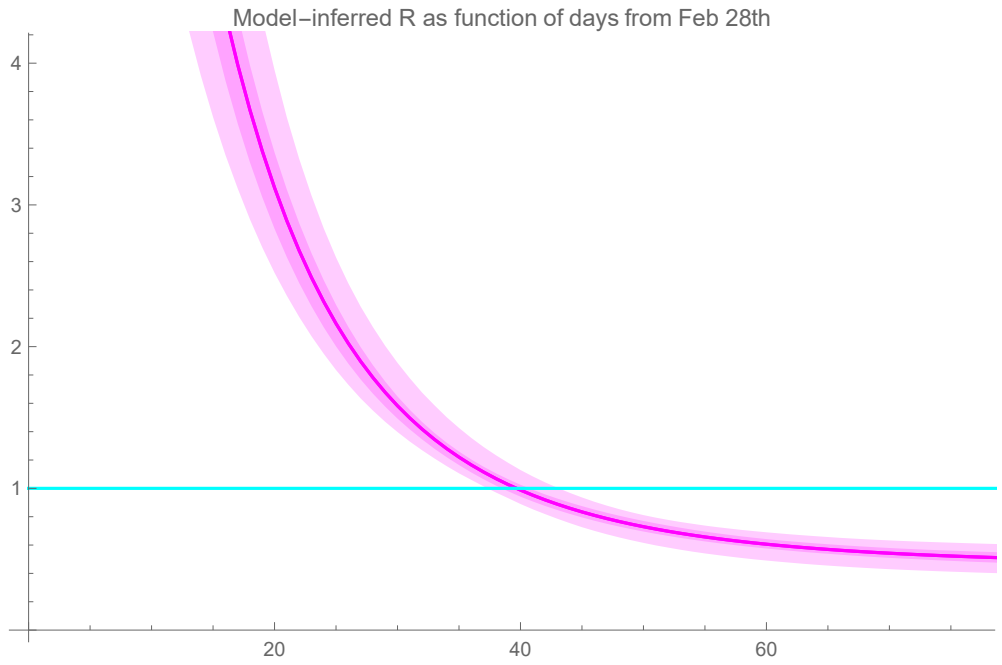

Figure 2: Model-inferred R_t as described in Section 2. The curve indicates the mean (over 10000 parameter-set selections); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) intervals. In this and subsequent figures time is measured in days from February 28th, 2020.

2. We update n_s, the expected number of seed cases, to the value $n_s = G(t_1)/\widetilde{C_c}(t_1)$. Since the system (1)-(10) is very close to being linear (this is because $S(t)/N$ does not vary appreciably from its initial value of 1), the scaling of the number of seeds causes an equal scaling in the value $C_c(t)$ of the cumulative number of cases.

Using this updated value for the initial number of seeds n_s, we solve the differential equations under several assumed scenarios by controlling the time-varying transmission rate $\beta(t)$.

3.1 Scenario 1: $R = 1.6$ after May 18th

Figure 3 shows results for a scenario where the effective reproductive number R_t is given by the inversion of the model (as in Fig. 2) up to May 11th, with R_t then assumed to be 0.6 until May 18th. On May 18th, the value of R_t is assumed to increase to 1.6, and to remain at this level for four weeks, with restrictions being reimposed on June 15th to cause R_t to be held at 0.6 thereafter.

3.2 Scenario 2: $R = 1.2$ after May 18th

Scenario 2 has R_t changing value at the same dates as Scenario 1, but assumes that the value of R_t from May 18th to June 15th is 1.2, which gives a slower growth in number of cases than Scenario 1.

3.3 Scenario 2: $R = 0.9$ after May 18th

In this scenario, the value of R_t from May 18th is assumed to be 0.9. As this number is below 1, the number of cases continues to decay over time and no reimposition of restrictions is required.

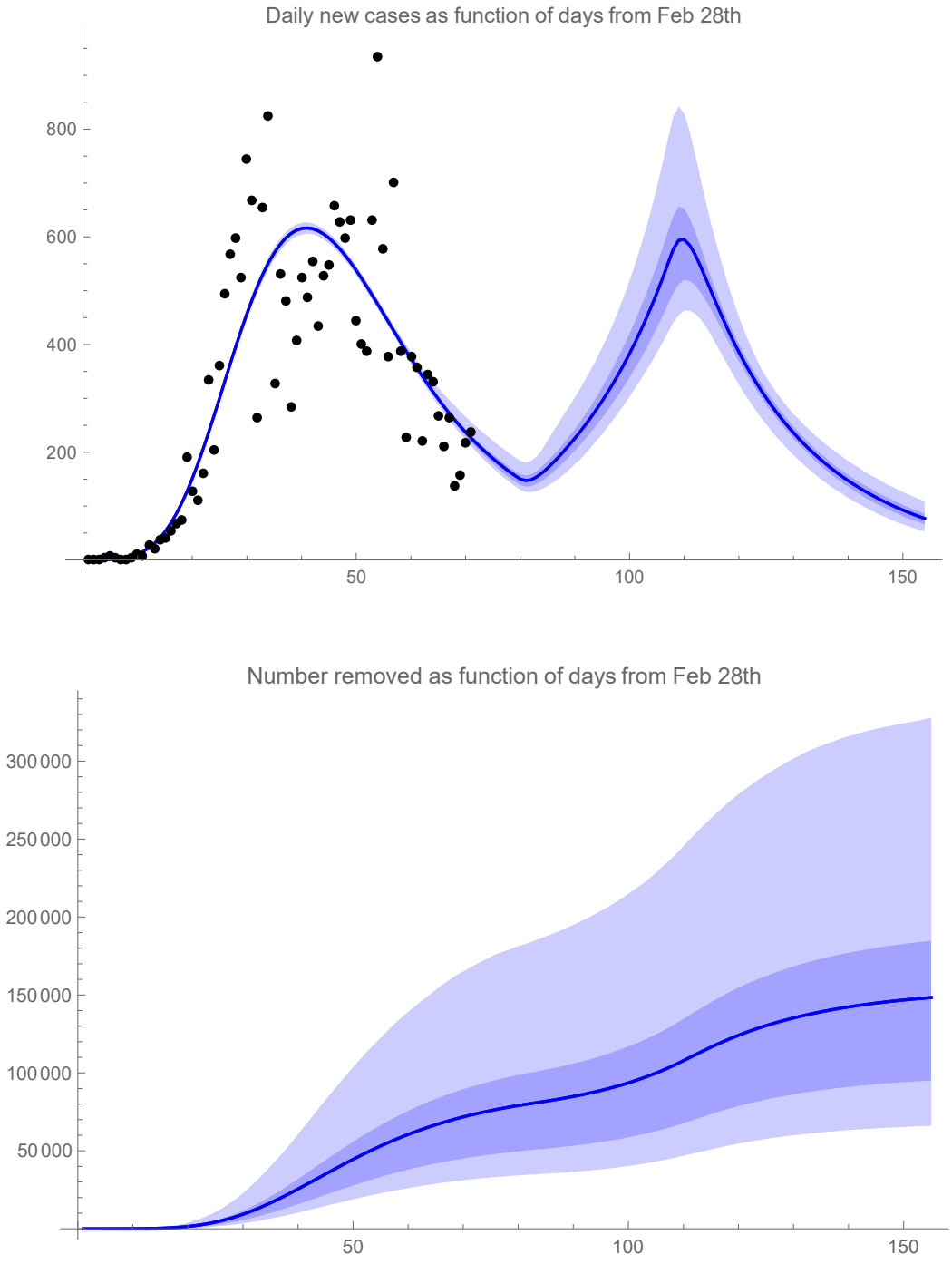

Figure 3: Scenario 1 assumes $R = 1.6$ from May 18th, for four weeks, with restrictions reimposed (assumed to give $R = 0.6$) from June 15th. Top panel shows confirmed new cases per day; bottom panel is the number of removed individuals (this includes both deaths and those who recovered from the virus). The curve indicates the mean (over 10000 parameter-set selections); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) intervals.

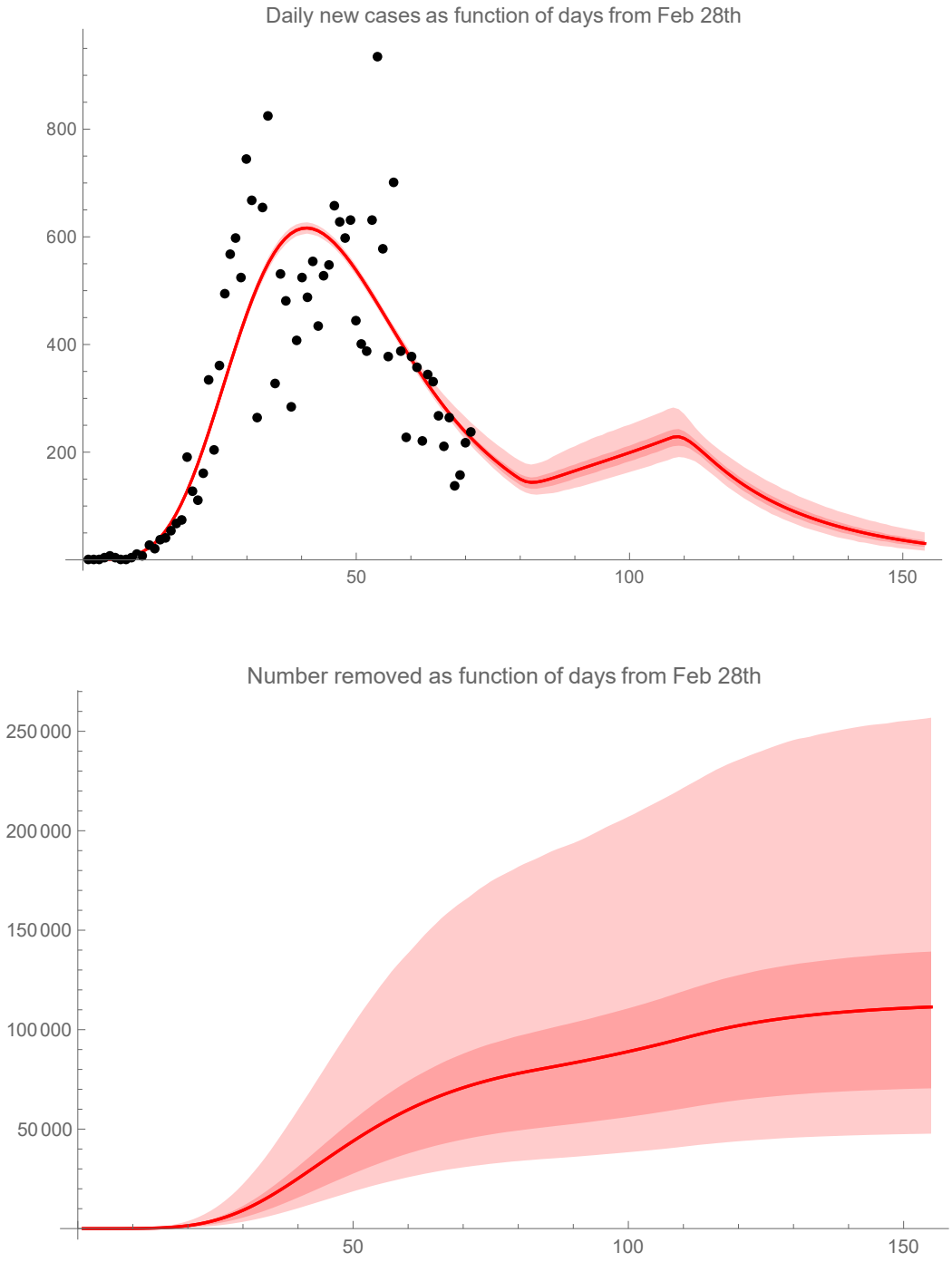

Figure 4: Scenario 2 assumes $R = 1.2$ from May 18th, for four weeks, with restrictions reimposed (assumed to give $R = 0.6$) from June 15th. Top panel shows confirmed new cases per day; bottom panel is the number of removed individuals (this includes both deaths and those who recovered from the virus). The curve indicates the mean (over 10000 parameter-set selections); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) intervals.

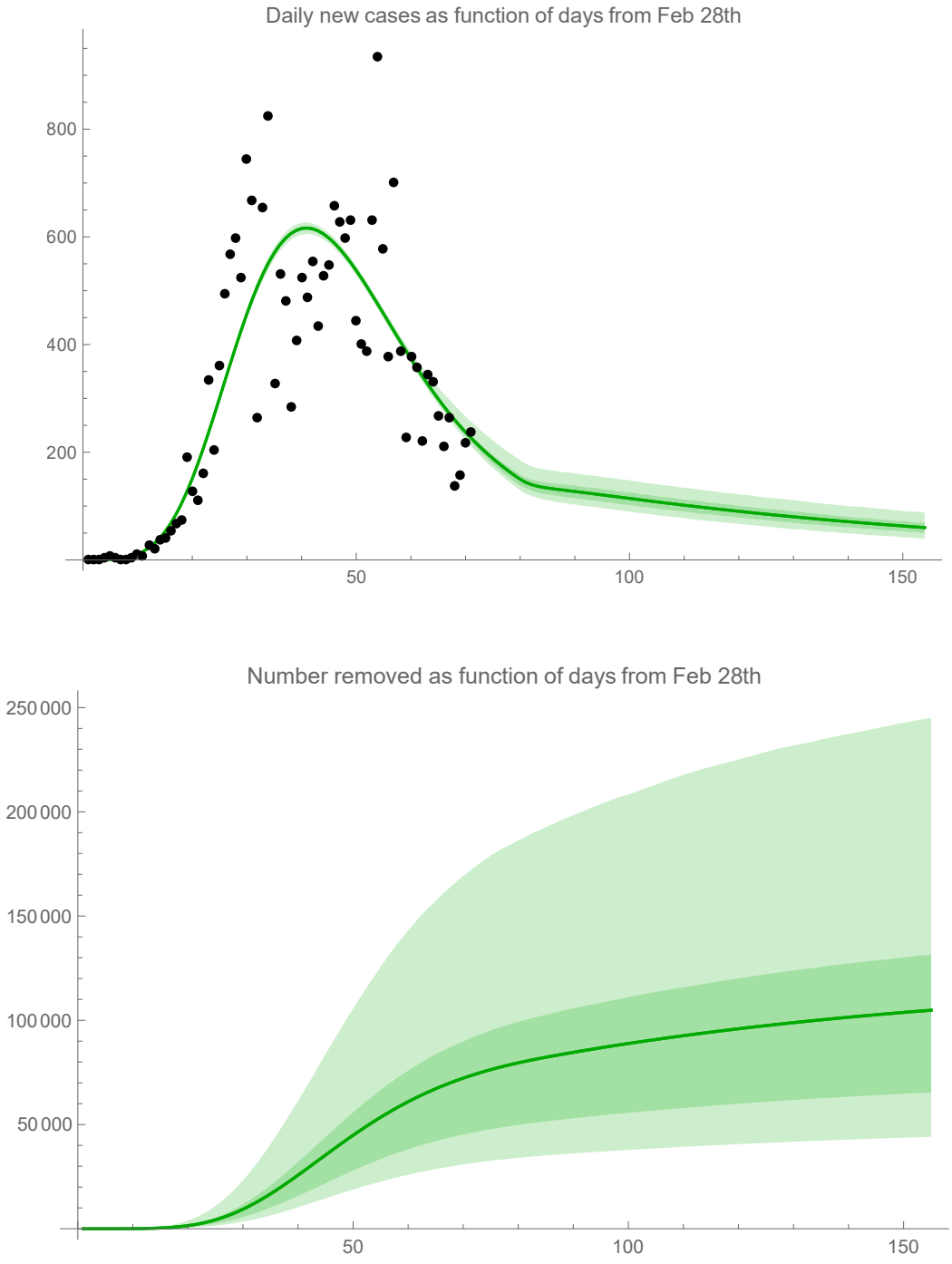

Figure 5: Scenario 3 assumes $R = 0.9$ from May 18th with no reimposition of restrictions. Top panel shows confirmed new cases per day; bottom panel is the number of removed individuals (this includes both deaths and those who recovered from the virus). The curve indicates the mean (over 10000 parameter-set selections); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) intervals.

The uncertainty shown with the quantiles in Figures 3 through 5 is due to the ranges of parameter values as specified in Section 1.1. However, it is important to note that in each case it is assumed that the reproductive number is exactly as specified in the scenario description. Comparing the results for the three scenarios, it is clear that the effect of uncertainty in future R values dominates the uncertainty due to the range of possible values for the other parameters of the model.

Acknowledgements

This work includes input from researchers funded by Science Foundation Ireland (grant numbers SFI/16/IA/4470, SFI/18/CRT/6049, SFI/12/RC/2289_P2, SFI/16/RC/3918, SFI/12/RC/3835, SFI/12/RC/2275_P2, SFI/15/IA/3090) and by the Health Research Board (grant number HRB RL2013/04).

Appendix

To calculate the basic reproductive ratio for the model we follow the approach explained in Section 2.2 of [8]. The value of R_0 is given by the spectral radius of the next generation matrix, which can be written as FV^{-1}. Defining the vector of relevant dynamical variables as $\mathbf{x} = \{E, I_p, I_a, I_i, I_{t1}, I_{t2}, I_n\}$, the matrix F is zero except for its first row, which is $(0, \beta, h\beta, i\beta, \beta, j\beta, \beta)$. The corresponding matrix V is

$$\begin{pmatrix} \frac{1}{L} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{(1-f)}{L} & \frac{1}{C-L} & 0 & 0 & 0 & 0 & 0 \\ -\frac{f}{L} & 0 & \frac{1}{D} & 0 & 0 & 0 & 0 \\ 0 & -\frac{q}{C-L} & 0 & \frac{1}{D-C+L} & 0 & 0 & 0 \\ 0 & -\frac{\tau}{C-L} & 0 & 0 & \frac{1}{T} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{T} & \frac{1}{D-C+L-T} & 0 \\ 0 & -\frac{(1-q-\tau)}{C-L} & 0 & 0 & 0 & 0 & \frac{1}{D-C+L} \end{pmatrix}. \tag{17}$$

Calculating the eigenvalues of FV^{-1} then yields the relationship

$$\begin{aligned} \frac{R_0}{\beta} &= (f-1)((i-1)q(C-L) + (j-1)\tau(C-L+T)) + \\ &\quad + D(f(h - iq - j\tau + q + \tau - 1) + (i-1)q + (j-1)\tau + 1). \end{aligned} \tag{18}$$

It can prove useful to also consider this formula is terms of contributions from each of the infected compartments. With a contribution to R_0 defined as the fraction of infectives moving through the compartment multiplied by the average time spend in the compartment multiplied by the transmission rate for that compartment, the respective contribution from each compartment is as follows:

- Contribution from I_a is $fDh\beta$.
- Contribution from I_p is $(1-f)(C-L)\beta$.

- Contribution from I_i is $(1-f)q(D-C+L)i\beta$.
- Contribution from I_{t1} is $(1-f)\tau T\beta$.
- Contribution from I_{t2} is $(1-f)\tau(D-C+L-T)j\beta$.
- Contribution from I_n is $(1-f)(1-q-\tau)(D-C+L)\beta$.

Summing these contributions gives equation (18).

References

[1] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[2] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. Oxford Univeristy Press, 2010.

[3] Mason A Porter and James P Gleeson. Dynamical systems on networks. *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, 4, 2016.

[4] John M Griffin, Aine Collins, Kevin Hunt, David McEvoy, Miriam Casey, Andrew Byrne, Conor G McAloon, Ann Barber, Elizabeth Lane, David McEvoy, and Simon J More. A rapid review of available evidence on the serial interval and generation time of COVID-19. *medRxiv doi:10.1101/2020.05.08.20095075*, 2020.

[5] Conor G McAloon, Aine Collins, Kevin Hunt, Ann Barber, Andrew Byrne, Francis Butler, Miriam Casey, John M Griffin, Elizabeth Lane, David McEvoy, Patrick Wall, Martin J Green, Luke O'Grady, and Simon J More. The incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *medRxiv doi:10.1101/2020.04.24.20073957*, 2020.

[6] Andrew W Byrne, David McEvoy, Aine Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth Lane, Conor McAloon, Kirsty O'Brien, Patrick Wall, Kieran Walsh, and Simon J More. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic covid-19 cases. *medRxiv doi:10.1101/2020.04.25.20079889*, 2020.

[7] Stan Development Team. RStan: the R interface to Stan, 2020. R package version 2.21.1.

[8] Jane M Heffernan, Robert J Smith, and Lindi M Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.