SEIR model of COVID-19 transmission dynamics

A SEIR Model is a well-known compartmental model of viral transmission framed in ordinary differential equations, and representing a homogeneously mixing population divided among susceptible, exposed, infectious and recovered or died (SEIR), as follows:

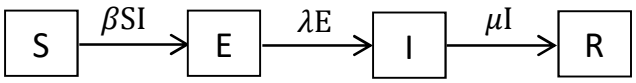

Fig S1. The SEIR model. The compartments denote those in the population that are Susceptible, Exposed, Infected and Recovered.

The variables S, E, I and R satisfy the ordinary differential equations:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \quad (1)$$

$$\frac{dE}{dt} = \beta S(t)I(t) - \lambda E(t) \quad (2)$$

$$\frac{dI}{dt} = \lambda E(t) - \mu I(t) \quad (3)$$

$$\frac{dR}{dt} = \mu I(t) \quad (4)$$

A convenient recent reference is Ma (1), though we have adjusted the notation here, reserving certain symbols for use in other standard ways.

As this discussion focuses on the number that die due to the epidemic we adjust the SEIR model by splitting R, those that recover, to distinguish between Z, those that die, from those that recover well. This arrangement is depicted in Fig S2, and is a simplified version of that used by Dagpunar (2), which discusses the outcomes of hospitalization.

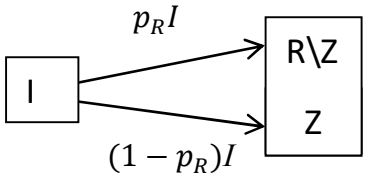

Fig S2. Adjustment of the SEIR model where R is divided into two compartments, R \ Z, those that recover and Z, those that die; where p_R is the proportion that recover.

We allow this adjusted SEIR model to depend on certain parameters with the understanding that once these parameters are known the behaviour of the SEIR model is completely specified. There are nine parameters in the model:

Symbol	Definition	Value for Switzerland in

		Fig 1
D = μ^{-1}	mean duration of infectiousness	7
T = β^{-1}	mean interval between infectious contacts	2, 3 or 4 R_0 = D/T
M = λ^{-1}	mean latent period	0.102
t_0	number of days from start of epidemic before observations began	5.351
e_0	initial number of exposed	6.00E-05
s_0	initial population size	5.09E+04 or 2.64E+04 to scale the epidemic
σ	standard deviation of observational error	68.44
p_R	probability of recovering	0.963
τ	mean time between the end of infectiousness and recovery or death	7.636

The last parameter τ is used to modify the last differential equation (4) to:

$$\frac{dR}{dt} = \mu I(t-\tau) \qquad (5)$$

We denote by $\boldsymbol{\theta} = (b1, b2, \dots, b9)$, the vector of parameters. With $\boldsymbol{\theta}$ given, the four differential equations (1), (2), (3) and (5) can be solved by numerical integration to give the trajectories

$$S(t,\boldsymbol{\theta}), E(t,\boldsymbol{\theta}), I(t,\boldsymbol{\theta}), R(t,\boldsymbol{\theta}), Z(t,\boldsymbol{\theta}) \quad \text{for } t = 1,2,\dots,N \qquad (6)$$

where *t* is the day and *N* is the number of days of interest. However, the parameters are assumed unknown. We therefore used the standard method of Maximum Likelihood (ML) as given for example in Cheng (3) to estimate the parameter values.

Here we outline the approach (called fitting the model) used to estimate the parameters from a sample of observed daily deaths, this being used to prepare Fig 1 in the main text. Let the sample of observed number of daily deaths be denoted by

$$\mathbf{Z} = \{z_t \;\; t = 1,2,\dots N\} \qquad (7)$$

where z_t is the number of deaths on day *t* and N is the number of days observed. If the observations were made without error and if, with the right parameter values are correct for $\boldsymbol{\theta}$, then the death trajectory $\{Z(t,\boldsymbol{\theta}) \;\; t = 1,2,\dots,N\}$ would match the observed deaths **Z** in (7). So the model would be clearly successful in explaining deaths.

To include statistical uncertainty in the model we assume instead

$$z_t = z(t,\boldsymbol{\theta}) + e(t) \quad t = 1,2,\dots,N \qquad (8)$$

where $e(t)$ is random error. For simplicity the $e(t)$ are assumed to be normally and independently distributed (NID) with standard deviation σ, i.e.

$$e(t) \sim \text{NID}(0, \sigma^2), \text{ so that } \quad z_t - z(t,\boldsymbol{\theta}) \sim \text{NID}(0, \sigma^2) \qquad (9)$$

The logarithm of the distribution of the sample is then

$$\mathrm{L}(\mathbf{Z}|\boldsymbol{\theta}) = -\ (N/2)\ln(2\pi) - N\ln\sigma - \ [1/(2\sigma^2)]\ \sum_{i=1}^{N}[z_t - \ z(t,\boldsymbol{\theta})]^2 \qquad (10)$$

where $\mathbf{Z}$ is the random argument, and the parameters $\boldsymbol{\theta}$ are fixed. In ML estimation, this is turned on its head so that **Z** is simply the known sample of observations now regarded as fixed and we write L as $\mathrm{L}(\mathbf{Z}|\boldsymbol{\theta}) = \mathrm{L}(\boldsymbol{\theta}|\mathbf{Z})$) calling it the (log)likelihood to indicate that it is now treated as a function of $\boldsymbol{\theta}$. The ML estimator $\widehat{\boldsymbol{\theta}}$ is simply the value of $\boldsymbol{\theta}$ at which L($\boldsymbol{\theta}$| **Z**) is maximized. i.e.

$$\widehat{\boldsymbol{\theta}} = \mathrm{argmax}_{\boldsymbol{\theta}}\{(\mathrm{L}(\boldsymbol{\theta}|\ \mathbf{Z})\}. \qquad (11)$$

Nelder-Mead numerical search for the maximum was used. This goes through different $\boldsymbol{\theta}_i$ *i*=1, 2, 3,… comparing the different $\mathrm{L}(\boldsymbol{\theta}_i, |\mathbf{Z})$ to find $\widehat{\boldsymbol{\theta}}$, the best $\boldsymbol{\theta}$.

To simplify description of the estimation process, only fitting to deaths data, **Z** as in (7) has been described, but the method extends straightforwardly to include other data samples. For example

$$\mathbf{Y} = \{y_t \quad t = 1,2, \dots, N\} \qquad (12)$$

where y_t is the number of active cases on day *t*. Fitting simultaneously to both **Y** and **Z** can be carried out by adding to the right-hand side of (10) a corresponding set of terms for **Y**. Numerical solution of the differential equations requires initial values for *S*, *E*, *I*, *R*. These are essentially scale invariant with $(S + E + I + R)$ constant and independent of *t.* So the numerical integration can conveniently be done using S(0, $\boldsymbol{\theta}$) = 1, E(0, $\boldsymbol{\theta}$) some small quantity subsequently adjustable as its initial value e_0 is a parameter; with *I* and *R* also initially zero. The population size, also a parameter, is only needed to provide scaled values *S*, *E*, *I*, *R* at each step for comparison with the data **Y** and **Z**.

References

1. J. Ma, Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling* **5**, 129-141 (2020).
2. J. S. Dagpunar, Sensitivity of UK Covid-19 deaths to the timing of suppression measures and their relaxation. *medRxiv*, 2020.2005.2009.20096859 (2020).
3. R. C. H. Cheng, *Non-Standard Parametric Statistical Inference*. (Oxford University Press, Oxford, 2017).