



OPEN

## The approximately universal shapes of epidemic curves in the Susceptible-Exposed-Infectious-Recovered (SEIR) model

Kevin Heng<sup>1,2✉</sup> & Christian L. Althaus<sup>3</sup>

Compartmental transmission models have become an invaluable tool to study the dynamics of infectious diseases. The Susceptible-Infectious-Recovered (SIR) model is known to have an exact semi-analytical solution. In the current study, the approach of Harko et al. (Appl. Math. Comput. 236:184–194, 2014) is generalised to obtain an approximate semi-analytical solution of the Susceptible-Exposed-Infectious-Recovered (SEIR) model. The SEIR model curves have nearly the same shapes as the SIR ones, but with a stretch factor applied to them across time that is related to the ratio of the incubation to infectious periods. This finding implies an approximate characteristic timescale, scaled by this stretch factor, that is universal to all SEIR models, which only depends on the basic reproduction number and initial fraction of the population that is infectious.

Compartmental models provide a key tool in infectious disease epidemiology for studying the transmission dynamics of various pathogens<sup>1–3</sup>. The Susceptible-Infectious-Recovered (SIR) model is known to have an exact semi-analytical solution<sup>4–6</sup>. No such solution exists for the Susceptible-Exposed-Infectious-Recovered (SEIR) model, although some of its properties have been examined using an approximate analytical approach<sup>7</sup>. In the current study, the approach of<sup>5</sup> is generalised to demonstrate that, while no exact semi-analytical solution is possible, an approximate one does exist.

It will be demonstrated that this approximate solution of the SEIR model implies the curves of *all* SEIR models are simply stretched or compressed relative to one another by the factor,

$$\alpha = \frac{\sigma}{\sigma + \gamma}, \quad (1)$$

where the incubation period is  $1/\sigma$ , the infectious period is  $1/\gamma$  and the generation time is  $1/\sigma + 1/\gamma$ . The SIR model is a special case with  $\alpha = 1$ . This property implies the time taken for the infectious curve to peak is approximately universal for the SEIR model when scaled by  $\alpha$ .

In “The SIR model” section, the SIR model is concisely reviewed and extended. In “The SEIR model” section, approximate solutions of the SEIR model and their implications are elucidated. A concise summary is provided in “Summary” section.

### The SIR model

In the SIR model, the fraction of the population that is susceptible ( $S$ ) becomes infected at a rate  $\beta = \mathcal{R}_0\gamma$ , where  $\mathcal{R}_0$  is the basic reproduction number. There is no incubation period. The fraction of the population that is infected is immediately infectious ( $I$ ) for a period of  $1/\gamma$ , after which a fraction of the population recovers ( $R$ ). The SIR model is described by the following set of coupled ordinary differential equations<sup>4,5</sup>,

<sup>1</sup>Center for Space and Habitability, University of Bern, Gesellschaftsstrasse 6, 3012 Bern, Switzerland. <sup>2</sup>Department of Physics, Astronomy & Astrophysics Group, University of Warwick, Coventry CV4 7AL, United Kingdom. <sup>3</sup>Institute of Social and Preventive Medicine, University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland. ✉email: kevin.heng@csh.unibe.ch

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I,\end{aligned}\quad (2)$$

where  $t$  represents the time. Since this set of equations does not consider births or deaths, we have  $S + I + R = 1$ .

**Review of Harko et al.<sup>5</sup>** As a starting point, the derivation of<sup>5</sup> is made more compact and cast in the mathematical notation of the current study. By taking the derivative of the first equation of (2) with respect to time, one obtains equation (12) of<sup>5</sup>,

$$I' = -\frac{1}{\beta} \left[ \frac{S''}{S} - \left( \frac{S'}{S} \right)^2 \right], \quad (3)$$

where for convenience one uses shorthand notation for the derivatives with respect to time,

$$I' \equiv \frac{dI}{dt}, \quad S' \equiv \frac{dS}{dt}, \quad S'' \equiv \frac{d^2S}{dt^2}. \quad (4)$$

By combining Eq. (3) with the second equation in (2), one obtains equation (13) of<sup>5</sup>,

$$\frac{S''}{S} - \left( \frac{S'}{S} \right)^2 + \frac{\gamma S'}{S} - \beta S' = 0. \quad (5)$$

By using the change of variables,

$$S' = \phi^{-1}, \quad S'' = -\phi^{-3} \frac{d\phi}{dS}, \quad (6)$$

one obtains from Eq. (5) an expression that is equivalent, but not identical, to equation (24) of<sup>5</sup>,

$$\frac{d\phi}{dS} + \frac{\phi}{S} + (\beta S - \gamma)\phi^2 = 0, \quad (7)$$

because one has chosen to work directly with  $S$  (and not  $S/S_0$ ) as the independent variable. The preceding expression is recognised as a Bernoulli differential equation, which may be solved to obtain an expression that is equivalent, but not identical, to equation (25) of<sup>5</sup>,

$$\phi^{-1} = S \left[ \beta(S - S_0 - I_0) - \gamma \ln \left( \frac{S}{S_0} \right) \right], \quad (8)$$

where the initial value of  $S$  is denoted as  $S_0$ . The constant of integration is set by demanding that  $S + I + R = 1$ . Recalling the definition of  $\phi$ , an expression that is equivalent to equation (26) of<sup>5</sup> follows,

$$t - t_0 = \int_{S_0}^S \frac{1}{s \left[ \beta(s - S_0 - I_0) - \gamma \ln \left( \frac{s}{S_0} \right) \right]} ds, \quad (9)$$

where  $t_0$  is the initial time. The preceding integral has no exact analytical (closed-form) solution and needs to be evaluated numerically, which is why it is strictly speaking an exact semi-analytical solution of the SIR model.

The first and third equation of (2) may be combined to obtain

$$R = \frac{\gamma}{\beta} \ln \left( \frac{S_0}{S} \right), \quad (10)$$

where the initial fraction of the population that has recovered is chosen to be  $R_0 = 0$ , which in turn implies that the initial fraction of the population that is infectious is  $I_0 = 1 - S_0$ .

**Extension of Harko et al.<sup>5</sup>** By setting  $I' = 0$  in Eq. (2), one realizes that the infectious curve  $I$  peaks at  $S = \gamma/\beta = 1/\mathcal{R}_0$ . Thus, Eq. (9) may be used to express the time taken for  $I$  to peak,

$$\gamma \Delta t \approx \int_{S_0}^{1/\mathcal{R}_0} \frac{1}{S \left[ \mathcal{R}_0(S - S_0) - \ln \left( \frac{S}{S_0} \right) \right]} dS, \quad (11)$$

where one assumes  $I_0 \ll 1$ . The quantity  $\gamma \Delta t$  is the time interval expressed in terms of the infectious period and depends only on two parameters:  $\mathcal{R}_0$  and  $I_0$ . Variations in  $I_0$  shift the  $S$ ,  $I$  and  $R$  curves back and forth in time without changing their shapes. We emphasize a subtle choice of notation:  $R_0$  is the initial fraction of the

population that has recovered (and is always set to zero in the current study), whereas  $\mathcal{R}_0$  is the basic reproduction number.

When the infectious curve  $I$  first starts to rise from its initial value, the logarithm term in the integral of Eq. (9) may be approximated as  $\ln(S/S_0) \approx S/S_0 - 1$ , which allows the integral to be evaluated analytically. It follows that

$$S \approx \Lambda \left[ \gamma \left( \mathcal{R}_0 - \frac{1}{S_0} \right) + \frac{\gamma \mathcal{R}_0 I_0}{S_0} e^{\Lambda(t-t_0)} \right]^{-1},$$

$$I \approx 1 - \frac{1}{\mathcal{R}_0} - \left( 1 - \frac{1}{S_0 \mathcal{R}_0} \right) S,$$
(12)

where we have defined the epidemic growth rate as

$$\Lambda \equiv \gamma(\mathcal{R}_0 - 1),$$
(13)

from which one obtains the known relationship between the basic reproduction number and the growth rate<sup>1,8</sup>,

$$\mathcal{R}_0 = 1 + \Lambda D,$$
(14)

where  $D \equiv 1/\gamma$  is the infectious period.

### The SEIR model

**Seeking an approximate semi-analytical solution.** The SEIR model builds on the SIR model by considering an additional compartment for the fraction of the population that is exposed ( $E$ ): infected but not yet infectious. The incubation period is  $1/\sigma$ . The SEIR model is described by the following set of coupled ordinary differential equations<sup>1</sup>,

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS, \\ \frac{dE}{dt} &= \beta IS - \sigma E, \\ \frac{dI}{dt} &= \sigma E - \gamma I, \\ \frac{dR}{dt} &= \gamma I. \end{aligned}$$
(15)

Since this set of equations does not consider births or deaths, we have  $S + E + I + R = 1$ .

The first and fourth equations may be combined to obtain

$$R = \frac{\gamma}{\beta} \ln \left( \frac{S_0}{S} \right),$$
(16)

which is identical to the SIR model. Again, the choice of  $R_0 = 0$  is made with no loss of generality.

By combining all four equations, one obtains

$$\frac{d^3 R}{dt^3} + (\sigma + \gamma) \frac{d^2 R}{dt^2} + \sigma \gamma \left( \frac{dR}{dt} + \frac{dS}{dt} \right) = 0.$$
(17)

The approximation is taken that the rate of change of the acceleration of  $R$  is vanishingly small,

$$R''' \equiv \frac{d^3 R}{dt^3} = 0.$$
(18)

This yields

$$\frac{d^2 R}{dt^2} + \alpha \gamma \left( \frac{dR}{dt} + \frac{dS}{dt} \right) = 0,$$
(19)

where one defines  $\alpha \equiv \sigma/(\sigma + \gamma)$ . When  $\alpha = 1$ , one recovers equation (19) of<sup>5</sup> for the SIR model.

One generalises equation (13) of<sup>5</sup>,

$$\frac{S''}{S} - \left( \frac{S'}{S} \right)^2 + \frac{\alpha \gamma S'}{S} - \alpha \beta S' = 0,$$
(20)

from which the familiar Bernoulli equation follows,

$$\frac{d\phi}{dS} + \frac{\phi}{S} + \alpha(\beta S - \gamma)\phi^2 = 0.$$
(21)

Retaining the  $R'''$  term in Eq. (17) would lead to a second-order, non-linear ordinary differential equation of  $\phi(S)$  with no known analytical solution.

Solving for  $\phi$  as in “Review of Harko et al.<sup>5</sup>” section yields

$$\phi^{-1} = S \left[ \frac{1}{S_0 \phi_0} + \alpha \beta (S - S_0) - \alpha \gamma \ln \left( \frac{S}{S_0} \right) \right], \quad (22)$$

where  $\phi_0$  is the initial value of  $\phi$ . The preceding expression leads to an expression for  $I$ , in terms of  $S$ , with a yet unspecified constant of integration ( $\phi_0$ ),

$$I = -\frac{1}{\beta S_0 \phi_0} - \alpha (S - S_0) + \frac{\alpha \gamma}{\beta} \ln \left( \frac{S}{S_0} \right). \quad (23)$$

Let the initial fraction of the population that is exposed be  $E_0$ . Demanding that  $S_0 + E_0 + I_0 + R_0 = 1$  yields

$$I_0 = -\frac{1}{\beta S_0 \phi_0} = 1 - S_0 - E_0. \quad (24)$$

Expressions for  $E$  and  $I$ , in terms of  $S$ , are obtained

$$\begin{aligned} E &= 1 - I_0 - \alpha S_0 + (\alpha - 1) \left[ S - \frac{\gamma}{\beta} \ln \left( \frac{S}{S_0} \right) \right], \\ I &= I_0 - \alpha (S - S_0) + \frac{\alpha \gamma}{\beta} \ln \left( \frac{S}{S_0} \right). \end{aligned} \quad (25)$$

Finally,  $S$  can be expressed in terms of  $t$  via the following integral,

$$t - t_0 = \int_{S_0}^S \frac{1}{s \left\{ \beta [-I_0 + \alpha (s - S_0)] - \alpha \gamma \ln \left( \frac{s}{S_0} \right) \right\}} ds. \quad (26)$$

Since  $I_0 \ll 1$ , the time taken for  $I$  to peak is

$$\alpha \gamma \Delta t \approx \int_{S_0}^{1/\mathcal{R}_0} \frac{1}{S \left[ \mathcal{R}_0 (S - S_0) - \ln \left( \frac{S}{S_0} \right) \right]} dS. \quad (27)$$

The preceding expression is identical to Eq. (11) of the SIR model, except for the extra factor of  $\alpha$ . It should be noted that the upper limit of the integral ( $1/\mathcal{R}_0$ ) assumes the approximation  $I' = E' = 0$ . However, Eq. (27) is not used to compute the peak times in Fig. 2. Its only purpose is to demonstrate that one may factor out  $\alpha \gamma$  from the integral. Stating the upper limit of the integral in Eq. (27) more accurately does not alter the main conclusion of the current study.

The relationship between the growth rate and the basic reproduction number can again be derived. Using the same series expansion of the logarithm term in the integral of Eq. (26), one obtains

$$\begin{aligned} S &\approx \Lambda \left[ \alpha \gamma \left( \mathcal{R}_0 - \frac{1}{S_0} \right) + \frac{\gamma \mathcal{R}_0 I_0}{S_0} e^{\Lambda(t-t_0)} \right]^{-1}, \\ I &\approx I_0 + \alpha \left( S_0 - \frac{1}{\mathcal{R}_0} \right) - \left( 1 - \frac{1}{S_0 \mathcal{R}_0} \right) \alpha S, \end{aligned} \quad (28)$$

albeit with a different definition of the growth rate,

$$\Lambda \equiv \gamma \mathcal{R}_0 (I_0 + \alpha S_0) - \alpha \gamma. \quad (29)$$

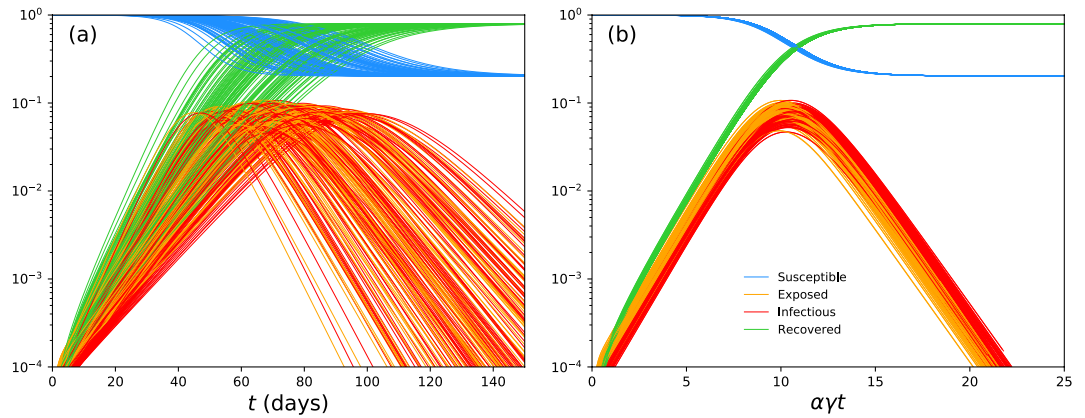
It follows that

$$\mathcal{R}_0 = \frac{\alpha + \Lambda D}{I_0 + \alpha S_0} = \frac{1 + \Lambda (D' + D)}{S_0 + I_0 \left( 1 + \frac{D'}{D} \right)}, \quad (30)$$

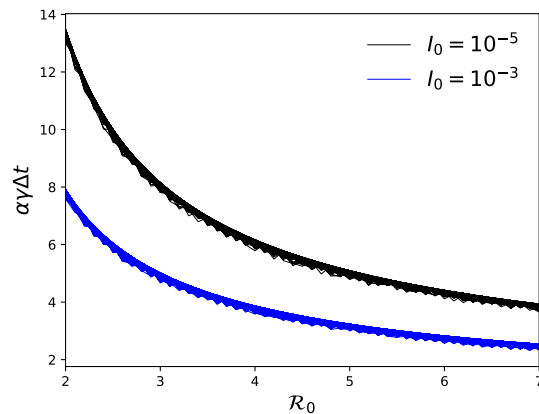
where  $D' \equiv 1/\sigma$  is the incubation period. When  $\alpha = 1$ , the expression for the SIR model in Eq. (14) is recovered. If  $S_0 \approx 1$  and  $I_0 \ll 1$ , then one obtains  $\mathcal{R}_0 \approx 1 + \Lambda (D' + D)$ .

The exact relationship between the growth rate and  $\mathcal{R}_0$  has been derived in various ways<sup>8</sup> (and references therein) and is given by  $\mathcal{R}_0 = (1 + \Lambda D')(1 + \Lambda D)$ . This equation accounts for the characteristic generation time distribution of SEIR models, which is a convolution of the exponentially distributed incubation and infectious periods with mean durations of  $D'$  and  $D$ , respectively. The approximate solution of Eq. (30) lacks the term  $\Lambda^2 D' D$ . Hence, it corresponds to the case of an exponentially distributed generation time with mean duration  $D' + D$ , which is the same as the solution for the SIR model assuming an infectious period of  $D' + D$ .

**Implications.** Equation (27) has non-trivial implications. It suggests that the susceptible, exposed, infectious and recovered curves of SEIR models, with different values of  $D'$  and  $D$ , follow approximately universal shapes that are stretched by a factor of  $1/\alpha = 1 + D'/D$  relative to one another. To demonstrate this property, the full set of coupled equations in (15) is solved numerically using the `solve_ivp` routine of the Python programming



**Figure 1.** Solution curves of 100 SEIR models as a (a) function of time and (b) time scaled by  $\alpha\gamma$ . For illustration, the basic reproduction number has been set to  $\mathcal{R}_0 = 2$  and the initial fraction of the population that is infectious has been set to  $I_0 = 10^{-4}$ . Each set of curves is generated using 100 random realisations of the incubation and infectious periods, each drawn from an interval between 2 and 5 days for illustration.



**Figure 2.** Time until the infectious curve ( $I$ ) peaks as a function of the basic reproduction number  $\mathcal{R}_0$ . In the SEIR model, the time to the epidemic peak ( $\Delta t$ ) scales approximately with  $\alpha$  and  $\gamma$ . For illustration, two values of the initial fraction of population that is infectious ( $I_0$ ) are considered. Each set of curves is generated using 10,000 random draws of the incubation and infectious periods from an interval between 2 and 5 days.

language suite<sup>9</sup>. For illustration, one assumes  $\mathcal{R}_0 = 2$  and  $I_0 = 10^{-4}$ . Figure 1 shows the solution curves of 100 SEIR models, where the values of the incubation ( $D' \equiv 1/\sigma$ ) and infectious ( $D \equiv 1/\gamma$ ) periods are randomly drawn from an interval between 2 and 5 days. When time is scaled by the factor  $\alpha\gamma$ , the 100 susceptible, exposed, infectious and recovered curves lie approximately on top of one another.

The second implication is that the time taken for the infectious curve to peak is approximately universal for all SEIR models when scaled by  $\alpha$  and expressed in terms of the infectious period. In other words,  $\alpha\gamma\Delta t$  should only depend on  $\mathcal{R}_0$  and  $I_0$ . To demonstrate this property, the full set of equations in (15) is again solved numerically for 10,000 random draws of  $1/\sigma$  and  $1/\gamma$  and for  $\mathcal{R}_0 = 2$  to 7. For each SEIR model, the time taken for the infectious curve to peak ( $\Delta t$ ) is calculated numerically. All 10,000 values of  $\Delta t$  are multiplied by  $\alpha\gamma$ ; two sets of curves with different  $I_0$  values are shown in Fig. 2 for illustration. For all 10,000 SEIR models, the  $\alpha\gamma\Delta t$  values lie approximately on the same curve across  $\mathcal{R}_0$  for a given value of  $I_0$ , demonstrating that  $\alpha\gamma\Delta t$  is a dimensionless (with no physical units), approximately universal timescale of the SEIR model.

### Summary

In the current study, approximate semi-analytical solutions of the SEIR model are found by generalising a previous approach for deriving an exact solution of the SIR model. This finding implies that the entire family of susceptible, exposed, infectious and recovered curves of the SEIR model follow approximately universal shapes that are stretched or compressed, relative to one another, by a factor consisting of the incubation and infectious periods. The time taken for the infectious curve to peak is the characteristic timescale of the system and depends only on the basic reproduction number and the initial fraction of the population that is infectious when scaled by the infectious period and this stretch factor.

Received: 2 September 2020; Accepted: 30 October 2020

Published online: 09 November 2020

## References

1. Vynnycky, E. & White, R. G. *An Introduction to Infectious Disease Modelling* (Oxford University Press, Oxford, 2010).
2. Karrer, B. & Newman, M. E. J. Message passing approach for general epidemic models. *Phys. Rev. E* **82**, 016101–1–9 (2010).
3. Wilkinson, R. R. & Sharkey, K. J. Impact of the infectious period on epidemics. *Phys. Rev. E* **97**, 052403–1–12 (2018).
4. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A Math. Phys.* **115**, 700–721 (1927).
5. Harko, T., Lobo, F. S. N. & Mak, M. K. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* **236**, 184–194 (2014).
6. Miller, J. C. Mathematical models of SIR disease spread with combined non-sexual and sexual transmission route. *Infect. Dis. Model.* **2**, 35–55 (2017).
7. Pioveella, N. Analytical solution of SEIR model describing the free spread of the COVID-19 pandemic. *Chaos Solitons Fractals* **140**, 110243 (2020).
8. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604 (2007).
9. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

## Author contributions

K.H. formulated the problem, derived the equations, performed the numerical calculations and wrote the manuscript. C.L.A. made the link between the reproduction number and growth rate, checked the equations and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to K.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020