Journal of Theoretical Biology 313 (2012) 12–19

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Richards model revisited: Validation by and application to infection dynamics

Xiang-Sheng Wang [a,b], Jianhong Wu [a,b,*], Yong Yang [a,b]

[a] Mprime Centre for Disease Modelling, York Institute for Health Research, York University, Toronto, Canada M3J 1P3
[b] Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, Canada M3J 1P3

HIGHLIGHTS

- Revisit Richards model from SIR model.
- Four parameters in either Richards or SIR model are redundant in data fitting.
- Propose a constraint to tackle the over fitting problem.
- Conduct numerical simulations for real time prediction of several diseases.
- Provide stable forecast of final outbreak size and basic reproduction number.

ARTICLE INFO

Article history:
Received 16 November 2011
Received in revised form
22 July 2012
Accepted 24 July 2012
Available online 7 August 2012

Keywords:
Richards model
SIR model
Asymptotic approximation
Basic reproduction number
Over fitting

ABSTRACT

Ever since Richards proposed his flexible growth function more than half a century ago, it has been a mystery that this empirical function has made many incredible coincidences with real ecological or epidemic data even though one of its parameters (i.e., the exponential term) does not seem to have clear biological meaning. It is therefore a natural challenge to mathematical biologists to provide an explanation of the interesting coincidences and a biological interpretation of the parameter. Here we start from a simple epidemic SIR model to revisit Richards model via an intrinsic relation between both models. Especially, we prove that the exponential term in the Richards model has a one-to-one nonlinear correspondence to the basic reproduction number of the SIR model. This one-to-one relation provides us an explicit formula in calculating the basic reproduction number. Another biological significance of our study is the observation that the peak time is approximately just a serial interval after the turning point. Moreover, we provide an explicit relation between final outbreak size, basic reproduction number and the peak epidemic size which means that we can predict the final outbreak size shortly after the peak time. Finally, we introduce a constraint in Richards model to address over fitting problem observed in the existing studies and then apply our method with constraint to conduct some validation analysis using the data of recent outbreaks of prototype infectious diseases such as Canada 2009 H1N1 outbreak, GTA 2003 SARS outbreak, Singapore 2005 dengue outbreak, and Taiwan 2003 SARS outbreak. Our new formula gives much more stable and precise estimate of model parameters and key epidemic characteristics such as the final outbreak size, the basic reproduction number, and the turning point, compared with earlier simulations without constraints.

1. Introduction

In a well known study of malaria prevention, Ross (1911) introduced the concept of basic reproduction number which turns out to be a crucial index measuring the disease outbreak potential. This index, normally denoted by R_0, is defined as the number of secondary infections generated by an introduction of a primary infection into the total population previously unexposed to the disease. The basic reproduction number has a clear biological significance as it describes the speed of the disease spreading through the susceptible population. It is thus important to estimate the basic reproduction number as early as possible during the outbreak of a disease. One of the methods for estimating the reproductive number is to use Richards empirical growth function to estimate the growth rate of cumulative infected incidence and then express R_0 as an exponential function involving both the infectious growth rate and another parameter

* Corresponding author. Department of Mathematics and Statistics, York University, Toronto, Canada M3J 1P3. Tel.: +1 416 736 2100x33116.
E-mail address: wujh@mathstat.yorku.ca (J. Wu).

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.07.024

defined as the generation time from primary infected case to a secondary case.

In spite of its many successful applications in real-time data fitting and prediction of infection dynamics, there seems no clear biological explanation of the Richards model. Here, we provide an intrinsic connection between the Richards model and a simple SIR model with standard incidence function. Especially, we prove that the basic reproduction number can be explicitly determined from parameters in the Richards model, and all parameters in the Richards model can be linked to the parameters of the SIR model which have explicit biological interpretation.

During a disease outbreak, we are interested in the peak time with maximal infected incidence and the turning point when the growth rate reaches its maximum. It is reasonable to expect a time lag from turning point to peak time. In this paper, we will show that this time delay can be approximated by the serial interval. The biological significance of this result is that as long as we have detected the turning point of an epidemic outbreak, we know the peak time will be just approximately one serial interval away. Moreover, we will show that the ratio of peak epidemic size to the final outbreak size is an explicit function of the basic reproduction number. This relation enables us to predict the final size using the peak epidemic value and the basic reproduction number.

The Richards model originates from the standard logistic model proposed by Verhulst (1838) who incorporated self-regulation in the study of population growth. Let $N(t)$ represent the population size at time t and assume it evolves as follows:

$$N'(t) = rN(t)\left[1-\frac{N(t)}{K}\right], \tag{1}$$

where r is the intrinsic growth rate and K is the carrying capacity. There is extensive literature discussing and extending the logistic model and we refer the reader to the book by May (2001) and references therein. This model was later generalized by Richards (1959) with one additional freedom (i.e., the exponent of deviation a) in density dependence as

$$C'(t) = \tilde{r}C(t)\left[1-\left(\frac{C(t)}{K}\right)^{a}\right]. \tag{2}$$

The Richards model, also named as theta-logistic model in some literatures (Gilpin and Ayala, 1973; Ross, 2009) was initially introduced for ecological population growth, but recently adapted in epidemiology for real-time prediction of outbreak of diseases such as SARS (cf. Hsieh, 2009; Hsieh and Cheng, 2006; Hsieh et al., 2004), dengue fever (cf. Hsieh and Chen, 2009; Hsieh and Ma, 2009) and pandemic influenza H1N1 (cf. Hsieh, 2010; Hsieh et al., 2010).

A simple calculation shows that the solution to (2) can be explicitly given by

$$C(t) = K[1+ae^{-a\tilde{r}(t-t_c)}]^{-1/a}, \tag{3}$$

where t_c is the turning point defined as the time when the second derivative of $C(t)$ vanishes, or equivalently, when $C(t)$ takes the value $K(1+a)^{-1/a}$. Set

$$r := a\tilde{r}. \tag{4}$$

Then formula (3) becomes

$$C(t) = K[1+ae^{-r(t-t_c)}]^{-1/a}. \tag{5}$$

We have to mention that in the existing literature, (5), instead of (3), was used to simulate disease outbreaks. From the mathematical point of view, these two formulas are exactly the same with a scaling of parameters; see (4). However, from the biological point of view, we should be careful in applying formula (5) for practical use because the growth rate here should be r/a, not r. There is a common minor mistake in Hsieh (2009, 2010), Hsieh and Chen (2009), Hsieh and Cheng (2006), Hsieh et al. (2010, 2004), and Hsieh and Ma (2009) where r was regarded as the growth rate.

The Richards model is an empirical model, and one of its parameters (i.e., the exponential term a) seems to have no clear biological meaning. Here we intend to revisit the Richards formula based on a simple SIR model, and to provide epidemiological interpretations to the parameters (for instance, the exponential term a) in the Richards model. Especially, we will show that the exponential term a has a clear one-to-one relationship with the basic reproduction number R_0.

Moreover, it has been observed that fitting Richards model to some data is not ideal and the parameters could "play-off against each other" so that their values become extreme and biologically implausible (Clark et al., 2010). We believe this is a problem of over fitting and propose a constraint to reduce the number of parameters by one. Numerical simulations demonstrate that fitting with constraints provides more stable and precise estimation of some parameter values.

The remaining part of this paper is organized as follows. In the next section, we will provide some mathematical analysis on how to revisit and formally derive the Richards model from the SIR model. Then, we discuss the over fitting problem and address it by introducing a constraint. We will also conduct some real data fittings using the reported data of a few diseases to validate our approach. The conclusion and discussion are given in the last section.

2. Revisiting the Richards model from SIR model

Compartmental models are built and analyzed for epidemic outbreaks. The population is separated into compartments and assumed to be transferred from one compartment to another with certain rates. In general, the terminology SIR is used to describe the phenomenon that the individual is moving from the susceptible class to the infective class I and finally flows into the recovery class R. In a deterministic formulation, the behavior of the disease dynamics is determined totally by its current status and by the rules which describe the rate of change of population in each compartment (Brauer, 2008).

Let $S(t)$, $I(t)$ and $R(t)$ be the numbers of individuals in the susceptible, infected and removed (i.e., recovered or quarantined) class at time t, respectively. Based on the aforementioned transmission dynamics, we consider the following SIR model:

$$S' = -\beta S\frac{I}{S+I}, \tag{6}$$

$$I' = \beta S\frac{I}{S+I}-\gamma I, \tag{7}$$

$$R' = \gamma I, \tag{8}$$

where β denotes the transmission rate and γ is the recovery rate, or more precisely, the rate of individuals removed from the infected class by recovery or being quarantined. Here we only consider the simple case when there is no birth or death. Also, we assume the population in the infected class will restrict their social activity after being recovered/quarantined. Therefore, the denominator in the transmission term is the sum of susceptible and infected individuals $S+I$, not the total population $S+I+R$. Throughout this paper, we will still name γ to be the recovery rate.

Note that in what follows, the initial value $S(0)$ is not the total population of the community. Instead, we regard $S(0)$ as the number of individuals in the susceptible class who will eventually

be infected, namely, the value of $S(t)$ vanishes as t goes to infinity. The initial value $I(0)$ is positive but relatively small compared with $S(0)$, and the initial value $R(0)$ is set to be zero. Using the next generation method (Diekmann et al., 1990, 2010; van den Driessche and Watmough, 2002; Heffernan et al., 2005), it is easy to calculate the basic reproductive ratio of this model as

$$R_0 = \beta/\gamma. \tag{9}$$

From (6) and (7), it follows that

$$\frac{d(S+I)}{dS} = \frac{\gamma(S+I)}{\beta S}.$$

Solving the above equation, we obtain

$$S(t)+I(t) = cS(t)^{\gamma/\beta}, \tag{10}$$

where

$$c := [S(0)+I(0)]S(0)^{-\gamma/\beta}. \tag{11}$$

A combination of (6) and (10) gives

$$S' = -\beta S[1-(S/L)^{\alpha}], \tag{12}$$

where

$$\alpha := 1-\gamma/\beta \tag{13}$$

and

$$L := c^{1/(1-\gamma/\beta)}. \tag{14}$$

Note that Eq. (12) has exactly the same form as Eq. (2). It can be solved explicitly by

$$S(t) = L[1+\alpha e^{b(t-t_j)}]^{-1/\alpha}, \tag{15}$$

where

$$b := \alpha\beta \tag{16}$$

and t_j is the turning point defined as the finite time where $S''(t)=0$, or equivalently,

$$S(t_j) = L(1+\alpha)^{-1/\alpha}.$$

As mentioned before, we set the initial value $S(0)$ to be the population in the susceptible class that will eventually be exposed to virus. This is the reason why we observe from (15) that

$$\lim_{t\to\infty} S(t) = 0.$$

We are interested in the number of the accumulated cases defined as

$$\begin{aligned} J(t) &:= I(t)+R(t) \\ &= N-S(t) \\ &= N-L[1+\alpha e^{b(t-t_j)}]^{-1/\alpha}, \end{aligned} \tag{17}$$

where N is the constant total population. It is easily seen from (11), (14) and the equation $N = S(0)+I(0)$ that

$$N \sim L \tag{18}$$

as $I(0)/S(0)\to 0$. Consequently,

$$J(t) \sim L-L[1+\alpha e^{b(t-t_j)}]^{-1/\alpha}. \tag{19}$$

As in the Richards model, here the parameter L is (asymptotically equal to) the final outbreak size and t_j is the turning point. Note from (9), (13) and (16) that $\alpha = 1-1/R_0$ is related to the basic reproductive ratio, and $b = \beta-\gamma$ is the infectious rate.

Now we are ready to formulate some explicit relations between the parameters in (5) and (19). Firstly, we hope that the final size and the turning point should be the same in both formulas, namely, we set

$$K = L \tag{20}$$

and

$$t_c = t_j. \tag{21}$$

Furthermore, we require the accumulated cases at the turning point be the same, leading to

$$C(t_c) = J(t_j).$$

From (5) and (19), we obtain

$$(1+a)^{-1/a}+(1+\alpha)^{-1/\alpha} = 1. \tag{22}$$

It can be shown that viewing as a function of α, $a = a(\alpha)$ is a decreasing function which maps [0,1] into $[1,\infty)$; see Appendix. With the aid of Lambert W-function (DLMF, 2011, Section 4.13), we have the following explicit formula (cf. Appendix):

$$a = \frac{W_m W_p^{-1}[\ln(1-(1+\alpha)^{-1/\alpha})]}{\ln(1-(1+\alpha)^{-1/\alpha})}-1, \tag{23}$$

where $W_p(x)$ and $W_m(x)$ are two branches of solutions to the equation $We^W = x$. Noting that $\alpha = 1-1/R_0$, the above equation provides an explicit one-to-one connection between the basic reproduction number and the exponential term a. Finally, we require the derivatives of $C(t)$ and $J(t)$ at the turning point $t = t_c = t_j$ are the same, which gives rise to

$$\frac{r}{(1+a)^{1/a+1}} = \frac{b}{(1+\alpha)^{1/\alpha+1}}. \tag{24}$$

In view of (4), (13) and (16), Eqs. (22) and (24) provide us an explicit and one-to-one connection between the parameters $(\tilde{r},a)$ in the Richards model (2) and (β,γ) in the SIR model (6). Define

$$f(t) := (1+ae^{-rt})^{-1/a} \tag{25}$$

and

$$g(t) := 1-(1+\alpha e^{bt})^{-1/\alpha}. \tag{26}$$

It then follows from (5) and (19) that

$$C(t) = Kf(t-t_c), \quad J(t) \sim Lg(t-t_j). \tag{27}$$

Given any $\alpha\in[0,1]$ (i.e., $R_0 = 1/(1-\alpha)\geq 1$) and $b>0$, we set

$$v(\alpha,b,t) := f(t)-g(t)$$

or equivalently,

$$v(\alpha,b,t) = (1+ae^{-rt})^{-1/a}+(1+\alpha e^{bt})^{-1/\alpha}-1, \tag{28}$$

where a and r are functions of α and b determined by the relations (22) and (24). It is readily seen from (24) and (28) that

$$v(\alpha,b,t) = v(\alpha,1,bt).$$

The functions $v(\alpha,1,t)$ for different values of α are plotted in Fig. 1. We observe from numerical calculations that $v(\alpha,1,t)<0.022$ for

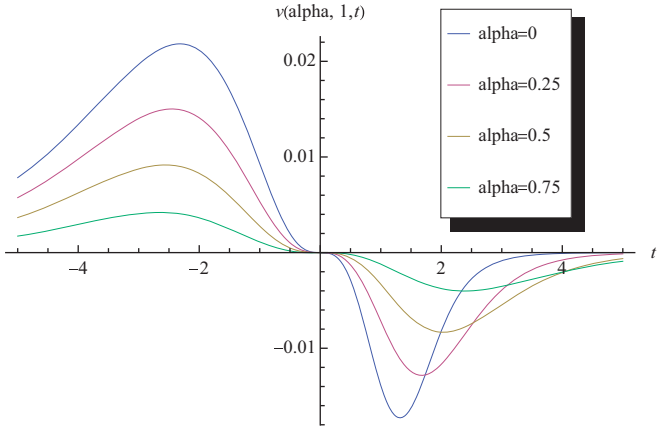

Fig. 1. The functions $v(\alpha,1,t)$ for $\alpha = 0$, 0.25, 0.5 and 0.75, respectively.

all $\alpha \in [0,1]$ and $t \in \mathbb{R}$; see Table 1. Thus, we have

$$\sup_{\alpha \in [0,1], b>0, t \in \mathbb{R}} \nu(\alpha,b,t) < 0.022.$$

Therefore, we conclude from (27) and (28) that the functions $C(t)$ in (5) and $J(t)$ in (19) are close to each other for all $t \in \mathbb{R}$ provided the four connection formulas (20), (21), (22) and (24) are satisfied. We would like to mention that the connection formulas (22) and (24) can also be obtained by matching the local linearizations of the functions $f(t)$ in (25) and $g(t)$ in (26) near zero. (When t is far away from zero, the difference between these two functions is exponentially small.)

Before ending this section, we provide two explicit formulas for peak time and peak epidemic value in terms of parameters in the SIR model. The peak time t_i is defined as the time when $I'(t_i)=0$, which coupling with Eq. (7) gives

$$\frac{\beta S(t_i)}{S(t_i)+I(t_i)} = \gamma.$$

In view of (10), (13) and (14), we then obtain

$$S(t_i) = L(\beta/\gamma)^{-1/\alpha}. \tag{29}$$

Coupling the above two equations yields

$$I(t_i) = LR_0^{-R_0/(R_0-1)}(R_0-1), \tag{30}$$

where we have used the definition of basic reproduction number (9). Thus we have found an intrinsic relation among the peak value, the final size and the basic reproduction number. Consequently, we are able to predict the final outbreak size from the above expression with the knowledge of peak epidemic value and basic reproduction number. Moreover, substituting (29) and (15) gives

$$1+\alpha e^{b(t_i-t_j)} = \beta/\gamma.$$

Table 1
Numerical calculations of a, r and $\|\nu\| = \sup_{t\in\mathbb{R}}|\nu(\alpha,1,t)|$ for $b=1$ and α varies from 0 to 1.

α	a	r	$\|\nu\|$
0	3.04886	2.35634	0.02183
0.1	2.66933	2.09303	0.01899
0.2	2.35535	1.87871	0.01630
0.3	2.09135	1.70123	0.01377
0.4	1.86631	1.55209	0.01139
0.5	1.67222	1.42519	0.00916
0.6	1.50313	1.31603	0.00708
0.7	1.35450	1.22124	0.00512
0.8	1.22284	1.13824	0.00330
0.9	1.10540	1.06502	0.00159
1	1	1	0

Taking (9) and (13) into account, we have

$$R_0 = e^{b(t_i-t_j)}. \tag{31}$$

When $b=\alpha\beta=\beta-\gamma$ is small (i.e., $R_0 \sim 1$), we can approximate the exponential function on the right by a linear function $1+b(t_i-t_j)$. On the other hand, we have $R_0 = 1+b/\gamma$ from (9) and (16). It follows that

$$t_i - t_j \sim 1/\gamma, \tag{32}$$

which means that the time interval between the turning point and the peak time is approximately the serial interval $T=1/\gamma$.

3. Over fitting problem

In fitting the real data by our formula (19), we found that the best fit sometimes lead to estimation of α outside of the interval (0,1), which is biologically unreasonable. A similar phenomenon occurred when the Richards formula is used. This is due to the redundancy of parameters which causes the over fitting. To address this over fitting issue, we introduce one constraint to reduce the number of parameters in (19) from four to three. In practise, one may be able to obtain an estimate of the recovery rate at the early stage of a disease outbreak. We choose to fix this parameter and use our formula (19) for real time prediction of disease outbreaks. Later on, we carry out some sensitivity analysis to see how our results depend on the choice of this parameter.

We should mention that our simulations do produce different results from those reported in the existing literature because we have added one constraint. However, the numerical simulations (cf. Table 2) based on both Richards and our formulas with the same constraints should not vary too much because these two formulas have intrinsic connections with each other, as shown in the previous section.

4. Validation and application

In this section, we validate our formula (19) by applying it to four real data: Canada 2009 H1N1 two-stage epidemic outbreak data, GTA 2003 SARS two-stage epidemic outbreak data, Singapore 2005 dengue data, and Taiwan 2003 SARS data. We wish to show that our formula does provide a reliable tool for simulation of a disease outbreak, and for detecting the turning point and multiple waves/phases.

Case 1: *Canada* 2009 *H1N1 outbreak*. The Canada 2009 H1N1 outbreak occurred from April 12 to June 19, 2010 and it involved two stages: 4/12-5/4 and 5/4-6/19. In order to fit the data by using our formula, we have to determine the recovery rate which is the reciprocal of serial interval; see Lipsitch et al. (2003) for the

Table 2
Results for the 2009 pandemic H1N1 cases in Canada by using different approaches with fixed recovery rate $\gamma=0.3$. From $a=1.69349$ and $(1+a)^a+(1+\alpha)^\alpha=1$ we can compute $\alpha=0.488328$. Similarly, from $\alpha=0.477159$ and $(1+a)^a+(1+\alpha)^\alpha=1$ we can compute $a=1.71416$.

Parameters	Richards best fit	Computation from SIR best fit and our four relationships	Absolute difference	Relative difference (%)
K	316.551	319.435	2.884	0.9
a	1.69349	1.71416	0.02067	1.2
r	0.412008	0.397669	0.014339	3.5
t_c	16.4575	16.5552	0.0977	0.6
	SIR best fit	Computation from Richards best fit and our four relationships	Absolute difference	Relative difference (%)
L	319.435	316.551	2.884	0.9
α	0.477159	0.488328	0.011169	2.3
b	0.273789	0.286314	0.012525	4.6
t_j	16.5552	16.4575	0.0977	0.6

Table 3
Results for the 2009 H1N1 cases in Canada using our formula with fixed recovery rate $\gamma = 0.3$.

Time period	Basic reproduction number	Turning point	Final size
4/12-5/4	1.9127 [1.8370, 1.9883]	16.56 [16.00, 17.11]	319.4 [301.0, 337.9]
5/4-6/19	1.3577 [1.3467 1.3687]	31.81 [31.32, 32.29]	5093 [4996, 5189]

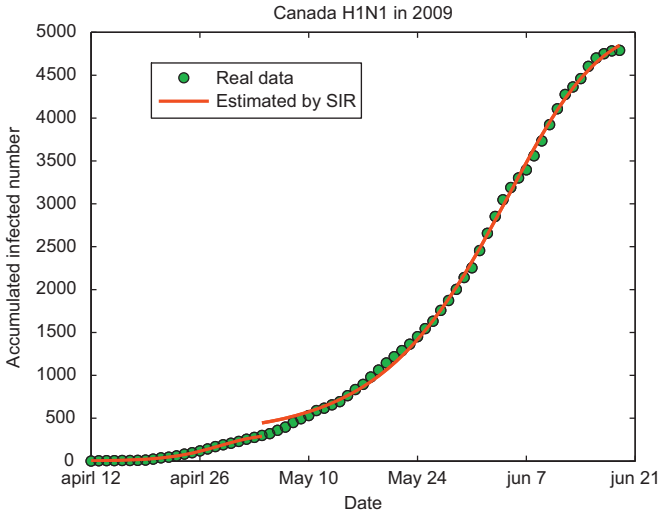

Fig. 2. Numerical simulation for the Canada 2009 H1N1 outbreak with two waves: the first stage is from April 14 to May 4, 2009 and the second stage is from May 4 to June 16, 2009.

definition of serial interval. Cowling et al. (2009) estimated the 95% confidence interval for serial interval of seasonal influenza (in days) is [2.9, 4.3]. Hence, we fix the recovery rate in our formula to be 0.3. Sensitivity analysis shows that the results do not vary much if we choose the recovery rate between 0.23 and 0.34; see Appendix B.

We list in Table 3 the estimates obtained by using our formula: the basic reproduction ratio R_0 is 1.91 for the first stage and 1.36 for the second; the turning point of first wave occurred on April 29 (t_j=16.56) and the turning point of second wave is June 5 (t_j=31.81); the final sizes in each stage are 319.4 and 5093, respectively. The fitted curve is shown in Fig. 2. Our estimates on turning points coincide with those obtained by Hsieh et al. (2010) but their estimates on final sizes are slightly larger than ours. If we introduce the same constraint $\gamma = 0.3$ to the Richards model, then the difference between two models is insignificant; see Table 2. We would also like to remark that the basic reproduction number in Hsieh et al. (2010) was computed by using the formula $R_0 = \exp(rT)$ where T is the serial interval. The more accurate formula should be $R_0 = \exp(rT/a)$ because r/a, instead of r, is the growth rate; see discussion in the Introduction.

Case 2: *GTA* 2003 *SARS outbreak*. Health Canada pinpointed that the first phase for the Great Toronto Area (GTA) SARS started from February 23 and ended on April 21, 2003. Moreover, it was reported that the actual infected number of the first stage is 141. Here we will apply our formula to fit the GTA 2003 SARS data. First of all, we fix the recovery rate to be $\gamma = 1/T$ where T=8.4 is the serial interval of SARS (Lipsitch et al., 2003). In Table 4, we list numerical estimates of model parameters by using the data in different time intervals. It is remarkable that our estimates for the basic reproduction number, turning point and final size are very stable. By using a similar method as in Hsieh and Cheng (2006), we detect that the first stage ended near April 18, which is very close to the Health Canada's assessment (April 21). Based on the data of first wave (2/23-4/18), we estimate the

Table 4
Parameter estimates (with recovery rate fixed as $\gamma = 1/8.4$) during the period 2/23-4/30 of the GTA 2003 SARS outbreak.

Time period	Basic reproduction number	Turning point	Final size
2/23-4/04	2.42 [2.31 2.53]	30.34 [29.50 31.19]	123.4 [116.8 130.0]
2/23-4/14	2.30 [2.24 2.36]	31.75 [31.33 32.18]	135.6 [133.0 138.1]
2/23-4/18	2.27 [2.21 2.33]	32.02 [31.65 32.40]	137.5 [135.5 139.6]
2/23-4/20	2.26 [2.20 2.31]	32.16 [31.79 32.51]	138.4 [136.6 140.3]
2/23-4/22	2.25 [2.19 2.31]	32.25 [31.91 32.60]	139.1 [137.4 140.8]
2/23-4/24	2.24 [2.18 2.29]	32.33 [31.99, 32.66]	140.0 [138.0 141.2]
2/23-4/26	2.23 [2.18 2.28]	32.40 [32.08 32.73]	140.2 [138.7 141.6]
2/23-4/28	2.22 [2.17 2.27]	32.46 [32.15 32.78]	140.5 [139.1 141.9]
2/23-4/30	2.21 [2.16 2.27]	32.54 [32.22 32.85]	141.0 [139.7 142.4]

Table 5
Parameter estimates (with recovery rate fixed as $\gamma = 1/8.4$) during the period 4/18-6/11 of the GTA 2003 SARS outbreak.

Time period	Basic reproduction number	Turning point	Final size
4/18-5/31	2.29 [2.21 2.37]	37.82 [36.67 38.95]	276.0 [264.9 287.0]
4/18-6/02	2.37 [2.29 2.46]	36.53 [35.76 37.30]	263.3 [257.2 269.5]
4/18-6/04	2.44 [2.35 2.52]	35.88 [35.30 36.46]	257.6 [253.6 261.7]
4/18-6/06	2.48 [2.39 2.56]	35.48 [35.01 35.95]	254.4 [251.5 257.4]
4/18-6/08	2.51 [2.42 2.59]	35.27 [34.87 35.68]	252.8 [250.5 255.2]
4/18-6/11	2.53 [2.45 2.61]	35.12 [34.78 35.46]	251.7 [249.9 253.5]

basic reproduction number as 2.27 with 95% confidence interval (CI) [2.21, 2.33]. The turning point for the first outbreak is March 27 (t_j=32.02), and the final size is 137.5 [135.5, 139.6]. This is similar to the result obtained by Hsieh and Cheng (2006).

For the second stage (4/18-6/11), we estimate that (cf. Table 5) the basic reproduction number is 2.53 [2.45, 2.61], the turning point is May 23 (t_j=35.12), and the final size is 251.7 [249.9, 253.5]. This result again coincides with that obtained in Hsieh and Cheng (2006) where they estimated that the turning point is May 24 and the final size is 249. It is worth to mention that if we use the formula $R_0 = \exp(rT)$ to calculate the basic reproduction number, it will give unreasonable value 21.8. However, if we use the corrected formula $R_0 = \exp(rT/a)$, the result becomes 3.3; noting that r=0.367, a=2.576 in the last line of Table 3 in Hsieh and Cheng (2006) and T=8.4 (cf. Lipsitch et al., 2003; Hsieh, 2009). This justifies that the growth rate in the Richards model should be r/a, instead of r.

In Table 5, we also use the data with different end points to demonstrate that our estimates of model parameters are stable. The epidemic curve is illustrated in Fig. 3.

Case 3: *Singapore* 2005 *dengue outbreak*. We apply our formula to Singapore 2005 dengue outbreak data from e-week 17 to e-week 52. We choose the serial interval to be T=19/7 weeks; see Aldstad (2007) and Hsieh and Ma (2009). Hence, the recovery rate is fixed at $\gamma = 1/T = 7/19$. The basic reproduction ratio R_0 is estimated to be 1.45. The turning point fitted by using our formula is e-week 35 (t_j=18) and the final size is 10,971. This coincides with the result obtained by using the Richards model (Hsieh and Ma, 2009). Note that the basic reproduction number

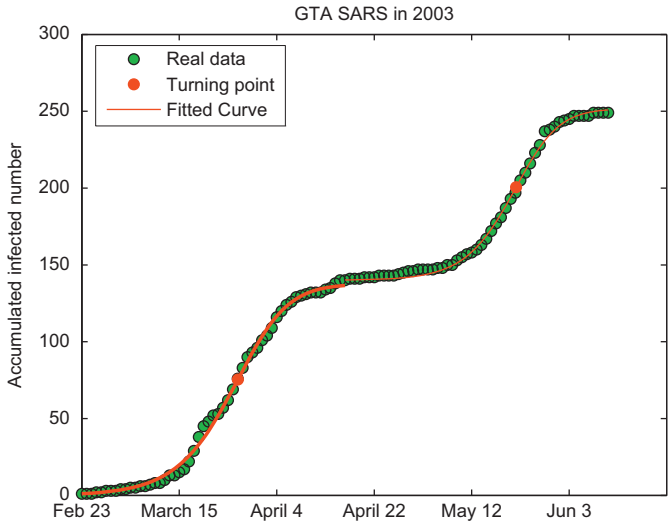

Fig. 3. Numerical simulation for the GTA 2003 SARS outbreak with two stages: 2/23-4/18 and 4/18-6/11.

Table 6
Results for the Singapore 2005 weekly dengue outbreak data (e-week 17–52). The recovery rate is fixed as $\gamma = 7/19$.

Time period	Basic reproduction number	Turning point	Final size
17–44	1.46 [1.43 1.49]	18.74 [18.05 19.44]	10,979 [10,475 11,483]
17–45	1.46 [1.43 1.49]	18.68 [18.11 19.26]	10,928 [10,530 11,327]
17–46	1.46 [1.43 1.49]	18.64 [18.15 19.13]	10,892 [10,569 11,215]
17–47	1.46 [1.44 1.48]	18.63 [18.19 19.05]	10,880 [10,611 11,149]
17–48	1.46 [1.44 1.48]	18.63 [18.24 19.00]	10,884 [10,655 11,114]
17–52	1.45 [1.44 1.47]	18.75 [18.47 19.04]	10,971 [10,824 11,119]

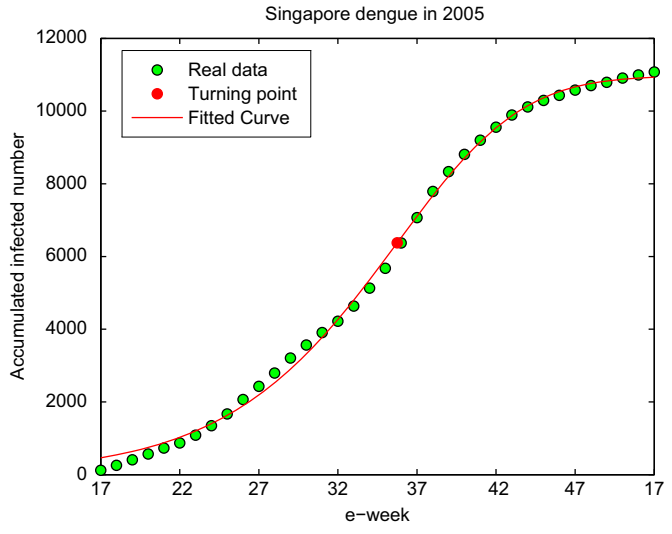

Fig. 4. Epidemic curve for the cumulative dengue cases in Singapore during e-weeks 17–52, 2005.

(1.89) calculated in Hsieh and Ma (2009) was using the formula $R_0 = \exp(rT)$. If we use the corrected formula $R_0 = \exp(rT/a)$, then the number would be 1.56, which is very close to our estimate. We would also like to mention that the parameters in our formula are stable by using different end weeks; see Table 6. The numerical simulations are also illustrated in Fig. 4.

Table 7
Estimated results for the 2003 SARS cases in Taiwan. The recovery rate is fixed as $\gamma = 1/8.4$.

Time period	Basic reproduction number	Turning point	Final size
2/25-6/25	1.88 [1.85 1.91]	69.18 [68.88 69.49]	453.2 [450.3 456.0]

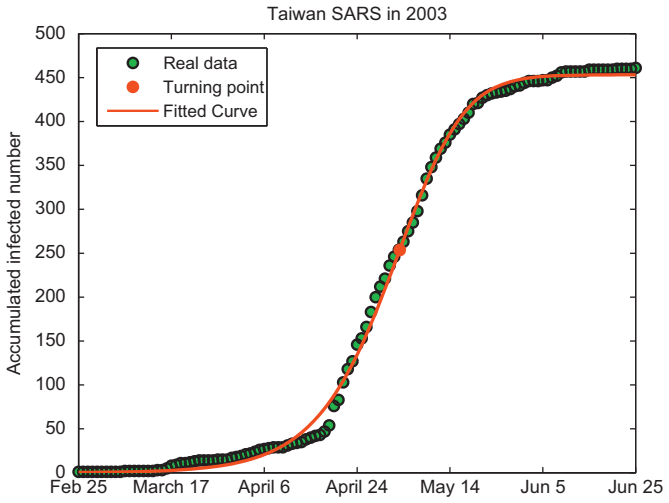

Fig. 5. Epidemic curve for the 2003 Taiwan SARS cases.

Case 4: *Taiwan* 2003 *SARS outbreak*. We conduct a numerical estimation for 2003 SARS outbreak in Taiwan. As before, we set the recovery rate to be $\gamma = 1/T$ where T=8.4 days is the serial interval of SARS; see Lipsitch et al. (2003) and Hsieh (2009). The basic reproduction number is estimated as 1.88 with 95% CI [1.85, 1.91]. The turning point is May 5 (t_j=69.18) and the final size is 453.2 [450.3, 456.0]; see Table 7. The epidemic curve is also illustrated in Fig. 5.

5. Conclusion and discussion

In this paper, we derived from the SIR model an explicit formula for the accumulated infected cases in terms of four parameters: final size, turning point, basic reproduction number and infectious rate. We showed that our formula has an intrinsic connection with the Richards empirical formula. In other words, we are able to provide biological interpretations to all of the parameters in the Richards model, especially the exponential term a.

Furthermore, we observed that four parameters in either the Richards model or our simulation based on the SIR model are too much to do data fitting and we propose one constraint that one parameter such as the recovery rate is fixed. This assumption is reasonable because in practice we may obtain the value of this parameter shortly after the disease outbreak.

Moreover, we conducted some numerical simulations for real time prediction of several diseases (H1N1, SARS, dengue) and provided more stable and precise forecast of final outbreak size and the basic reproduction number, compared with previous simulations with no constraint.

It is an important issue to be addressed whether our formula (and thus the Richards model) provide equally good prediction of turning points and multiple waves when there are more disease stages rather than the setting of SIR.

Acknowledgements

This work was partially supported by CRC, IDRC, GEOIDE, Mitacs, Mprime and NSERC. We are grateful to Professor Ying-Hen Hsieh for sharing with us the epidemic data which were used to conduct our numerical simulations. We would also like to thank two anonymous referees for their insightful comments and helpful suggestions.

Appendix A. The relation between a and α

From (22) we have

$$(1+a)^{-1/a}+(1+\alpha)^{-1/\alpha}=1.$$

Note that

$$(1+a)^{-1/a}=\exp\left[-\frac{1}{a}\ln(1+a)\right]$$

and the function

$$h(x):=\frac{1}{x}\ln(1+x)$$

is a decreasing function on the positive real line because

$$h'(x)=\frac{x-(1+x)\ln(1+x)}{x^2(1+x)}<0$$

for any $x>0$. Hence, the function $(1+a)^{1/a}$ is increasing for $a>0$. Similarly, the function $(1+\alpha)^{-1/\alpha}$ is increasing for $\alpha>0$. Consequently, viewing as a function of α, $a=a(\alpha)$ is decreasing for $\alpha\in(0,1)$; recall from (9) and (13) that $\alpha=1-1/R_0\in(0,1)$ since $R_0>1$. Numerical computation shows that when α increases from 0 to 1, a should be decreasing from 3.04886 (precise to the fifth decimal place) to 1; see the first two columns in Table 1.

Now we are ready to prove (23), namely,

$$a=\frac{W_mW_p^{-1}[\ln(1-(1+\alpha)^{-1/\alpha})]}{\ln(1-(1+\alpha)^{-1/\alpha})}-1.$$

Note from (DLMF, 2011, Section 4.13) that W_m and W_p are two solutions to the equation $We^W=x$ and $W_p(x)$ is an increasing function which maps $[-1/e,\infty)$ onto $[-1,\infty)$, while $W_m(x)$ is a decreasing function which maps $[-1/e,0)$ onto $[-1,-\infty)$. For $\alpha\in(0,1)$, it is easily seen that $(1+\alpha)^{-1/\alpha}\in(1/e,1/2)$ and thus $\ln(1-(1+\alpha)^{-1/\alpha})\in(-1,0)$. Consequently, $W_p^{-1}[\ln(1-(1+\alpha)^{-1/\alpha})]\in(-1/e,0)$ and $W_mW_p^{-1}[\ln(1-(1+\alpha)^{-1/\alpha})]\in(-1,-\infty)$. Especially,

$$W_mW_p^{-1}[\ln(1-(1+\alpha)^{-1/\alpha})]<\ln(1-(1+\alpha)^{-1/\alpha})<0,$$

which implies $a>0$. By the monotonicity of $(1+a)^{1/a}$ for $a>0$, we are left to verify that a given above is indeed a (and thus the unique positive) root to the equation

$$(1+a)^{-1/a}+(1+\alpha)^{-1/\alpha}=1.$$

Firstly, it can be easily shown that

$$W_m^{-1}[(1+a)\ln(1-(1+\alpha)^{-1/\alpha})]=W_p^{-1}[\ln(1-(1+\alpha)^{-1/\alpha})].$$

The left-hand side is

$$(1+a)\ln(1-(1+\alpha)^{-1/\alpha})\cdot(1-(1+\alpha)^{-1/\alpha})^{1+a}$$

and the right-hand side is

$$(1-(1+\alpha)^{-1/\alpha})\ln(1-(1+\alpha)^{-1/\alpha}).$$

Hence, we obtain

$$(1+a)(1-(1+\alpha)^{-1/\alpha})^a=1,$$

which yields

$$1-(1+\alpha)^{-1/\alpha}=(1+a)^{-1/a}.$$

This ends our proof of (23).

Appendix B. Statistical test and sensitivity analysis

We use the first stage (4/12-5/4) of Canada 2009 H1N1 outbreak data to conduct some statistical test. Assuming the daily infected case follows a Poisson distribution with mean being the same as the reported number, we generate 5000 samples and fit each data set by applying formulas derived from the SIR model and the Richards model, respectively. As before, we fix the recovery rate to be $\gamma=0.3$. The SIR model estimates that the final size has mean 320.4 with standard deviation 22.4, and the turning point has mean 16.57 with standard deviation 0.57. The Richards model with the same constraint $\gamma=0.3$ estimates that the mean final size is 323.4 (std 21.4) and the mean turning point is 16.24 (std 0.59). It is also noted that the Richards model without constraint gives mean final size 465.5 (std 91.5) and mean turning point 16.84 (1.35); see Table B1. Therefore, we conclude that the Richards model with four parameters may induce over fitting problem and this problem can be solved by introducing one constraint (fixing recovery rate).

The serial interval of seasonal influenza has 95% CI [2.9, 4.3] (cf. Cowling et al., 2009), which implies that the recovery rate has 95% CI [0.23 0.34]. We increase the recovery rate γ from 0.23 to 0.34 and conduct numerical simulation for each γ. It is shown that the final size ranges from 317.8 to 323.0, the turning point lies in the interval [16.52, 16.58], and the basic reproduction number varies from 1.80 to 2.20. Thus, the parameters are insensitive to the choice of recovery rate.

Table B1
Statistical test with 5000 samples generated by Poisson distribution. Richards (constraint) means the Richards model with constraint $\gamma=0.3$, and the last line is obtained by applying the Richards model without constraint.

Models	Final size (std)	Turning point (std)
SIR	320.4 (22.4)	16.57 (0.57)
Richards (constraint)	323.4 (21.4)	16.24 (0.59)
Richards	465.5 (91.5)	16.84 (1.35)

References

Aldstad, J., 2007. An incremental Knox test for the determination of the serial interval between successive cases of an infectious disease. Stoch. Environ. Res. Risk Assess. 21, 487–500.
Brauer, F., 2008. Compartmental models in epidemiology. In: Mathematical Epidemiology, Lecture Notes in Mathematics. Springer, Berlin, 1945, pp. 19–79.
Clark, F., Brook, B.W., Delean, S., Akcakaya, H.R., Bradshaw, C.J.A., 2010. The theta-logistic is unreliable for modelling most census data. Methods Ecol. Evol. 1, 253–262.
Cowling, B.J., Fang, V.J., Riley, S, Malik Peiris, J.S., Leung, G.M., 2009. Estimation of the serial interval of influenza. Epidemiology 20 (3), 344–347.
Digital Library of Mathematical Functions 2011-08-29. National Institute of Standards and Technology from ⟨http://dlmf.nist.gov/⟩.
Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J., 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28 (4), 365–382.
Diekmann, O., Heesterbeek, J.A.P., Roberts, M.G., 2010. The construction of next-generation matrices for compartmental epidemic models. J. R. Soc. Interface 47, 873–885.
Gilpin, M.E., Ayala, F.J., 1973. Global models of growth and competition. Proc. Natl. Acad. Sci. U.S.A. 70, 3590–3593.
Heffernan, J.M., Smith, R.J., Wahl, L.M., 2005. Perspectives on the basic reproductive ratio. J. R. Soc. Interface 47, 873–885.
Hsieh, Y.-H., 2009. Richards model: a simple procedure for real-time prediction of outbreak severity. In: Ma, Z., Wu, J., Zhou, Y. (Eds.), Modeling and Dynamics of

Infectious Diseases Series in Contemporary Applied Mathematics (CAM), vol. 11. Higher Education Press, Beijing, pp. 216–236.
Hsieh, Y.-H., 2010. Pandemic influenza A (H1N1) during winter influenza season in the southern hemisphere. Influenza Other Respir. Viruses 4, 187–197.
Hsieh, Y.-H., Chen, C.W.S., 2009. Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks. Trop. Med. Int. Health 14, 628–638.
Hsieh, Y.-H., Cheng, Y.-S., 2006. Real-time forecast of multiphase outbreak. Emerging Infect. Dis. 12, 122–127.
Hsieh, Y.-H., Fisman, D.N., Wu, J., 2010. On epidemic modeling in real time: an application to the 2009 novel A (H1N1) influenza outbreak in Canada. BMC Res. Notes 3, 283.
Hsieh, Y.-H., Lee, J.-Y., Chang, H.-L., 2004. SARS epidemiology modeling. Emerging Infect. Dis. 10, 1165–1168.
Hsieh, Y.-H., Ma, S., 2009. Intervention measures, turning point, and reproduction number for dengue, Singapore, 2005. Am. J. Trop. Med. Hyg. 80, 66–71.
Lipsitch, M., Cohen, T., Cooper, B., Robins, J.M., Ma, S., James, L., Gopalakrishna, G., Chew, S.K., Tan, C.C., Samore, M.H., Fisman, D., Murray, M., 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300, 1966–1970.
May, R.M., 2001. Stability and Complexity in Model Ecosystems. Princeton University Press.
Richards, F.J., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10, 290–300.
Ross, R., 1911. The Prevention of Malaria, 2nd ed. John Murray, London.
Ross, J.V., 2009. A note on density dependence in population models. Ecol. Modelling 220, 3472–3474.
van den Driessche, P., Watmough, J., 2002. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 180, 29–48.
Verhulst, P.-F., 1838. Notice sur la loi que la population poursuit dans son accroissement. Corresp. Math. Phys. 10, 113–121.