Identifying the Number of Unreported Cases in SIR Epidemic Models

A. Ducrot[1], P. Magal[2], T. Nguyen[2] and G. F. Webb[3]

1 *Normandie Univ, UNIHAVRE, LMAH, FR-CNRS-3335, ISCN, 76600 Le Havre, France*

2 *Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, Talence, France*

3 *Mathematics Department, Vanderbilt University, Nashville, TN*

May 13, 2019

Abstract

A SIR epidemic model is analyzed with respect to the identification of its parameters and initial values, based upon reported case data from public health sources. The objective of the analysis is to understand the relationship of unreported cases to reported cases. In many epidemic diseases the reported cases are a small fraction of the unreported cases. This fraction can be estimated by the identification of parameters for the model from reported case data. The analysis is applied to the Hong Kong seasonal influenza epidemic in New York City in 1968-1969.

Key words. epidemic model, transmission rate, reported cases, unreported cases, turning point.

2000 Mathematical Subject Classification. 92D30, 92C50, 92B05

1 Introduction

Mathematical models of epidemics have a long history [1, 4, 5, 6, 7, 11, 17, 19, 22, 28, 32]. One of the most important considerations of epidemic models is the identification of parameters needed for applications. The parameter identification problem for SIR model has been investigated by many researchers, including [2, 3, 8, 9, 10, 12, 13, 14, 15, 16, 18, 20, 21, 23, 24, 25, 27, 29, 30, 31, 33]. Our objective here is to continue the investigation in [26] of the parameter identification problem for the standard SIR ordinary differential equations model of an outbreak epidemic:

$$\begin{cases} S'(t) = -\tau S(t)I(t), \\ I'(t) = \tau S(t)I(t) - \nu I(t). \end{cases} \tag{1.1}$$

Here $S(t)$ and $I(t)$ denote the number of susceptible and infected individuals, respectively, at time $t > 0$. The parameter $\tau > 0$ corresponds to the disease

transmission rate and the parameter $\nu > 0$ corresponds to the removal rate of infected individuals. The initial conditions of the model are

$$S(0) = S_0 > 0 \text{ and } I(0) = I_0 > 0. \tag{1.2}$$

For specific applications, the parameters τ, ν, as well as the initial conditions, S_0, I_0, are usually not known. Our objective here is to identity these values from specific time data of reported infective cases.

Typically, the reported cases are only a small fraction of the total number of cases, since only the most severe symptomatic cases are reported. Our approach is based on knowledge of the data of new reported cases (typically weekly) over the time-course of the epidemic. This known data consists of the cumulative reported cases at time t, denoted by $CR(t)$, that correspond to the total number of reported infection cases up to time t. To handle these data we assume that these cumulative reported cases at time t consist in a constant fraction along time of the total number of infected cases up to to time t. In other words we assume that the removal rate ν takes the following form $\nu = \nu_1 + \nu_2$, where ν_1 is the removal rate of reported infected individuals, and ν_2 is the removal rate of infected individuals due to all other causes, such as mortality, recovery, or other reasons. With this assumption and notation, the cumulative reported cases are related to the number of infected by the following formula

$$CR(t) = \nu_1 \int_0^t I(s)ds, \tag{1.3}$$

where $\nu_1 > 0$ is an unknown parameter. We formulate our problem as follows:

Problem 1.1 *How can we identify the parameter set* $\Theta = \{(S_0, I_0, \tau, \nu, \nu_1)\} \subset (0, \infty)^5$ *if we know the cumulative reported cases* $CR(t)$ *for all time* $t > 0$*?*

We will show that knowledge of the cumulative reported cases $CR(t)$ is not sufficient to recover the five-dimensional parameter set Θ. Roughly speaking, this set is defined up to one degree of freedom. More precisely, under suitable hypotheses on the reported case data $CR(t)$, $t \geq 0$, only the following combination of parameters and initial values can be reconstructed:

$$\frac{I_0}{S_0},\ S_0\tau,\ S_0\nu_1 \text{ and } \nu. \tag{1.4}$$

As a consequence, knowledge about the value of S_0, the number of susceptibles in the population before the epidemic outbreak, allows us to obtain precise information about the values, I_0, τ, and ν_1. Then the *basic reproductive number* of the epidemic

$$\mathcal{R}_0 = \frac{S_0\tau}{\nu}, \tag{1.5}$$

can be obtained from (1.4). The interpretation of $\mathcal{R}_0$ is that if $\mathcal{R}_0 < 1$, then the epidemic subsides, and if $\mathcal{R}_0 > 1$, then the epidemic worsens. We will describe a simple methodology to compute the parameter set in (1.4), and apply this method to specific epidemic data.

2 Parameter identifiability

In this section we show that the parameter set Θ is not identifiable from the reported case data $CR(t), t \geq 0$. We refer to to Evans *et al.* [13] for more results on this topic. Here we use another approach.

Consider the parameter set $p := (\tau, \nu, S_0, I_0, \nu_1) \in (0, +\infty)^5$. Define $(S(t,p), I(t,p))$, as the unique solution of (1.1)-(1.2), and $CR(t,p)$ as the output function (1.3), for a given value of the parameter set $p \in (0, +\infty)^5$. We have the following proposition:

Proposition 2.1 *Suppose that $(S(t,p), I(t,p))$ and $(S(t,\overline{p}), I(t,\overline{p}))$ are the two solutions of (1.1)-(1.2) for the parameter set $p = (\tau, \nu, S_0, I_0, \nu_1)$ and the parameter set $\overline{p} = (\bar{\tau}, \bar{\nu}, \bar{S}_0\overline{I}_0, \bar{\nu}_1)$, respectively. Then*

$$CR(t,p) = CR(t,\overline{p}) \tag{2.1}$$

for every $t \geq 0$, if and only if

$$\nu = \bar{\nu}, \frac{\tau}{\nu_1} = \frac{\bar{\tau}}{\bar{\nu}_1}, \tau S_0 = \bar{\tau}\overline{S}_0, \tau I_0 = \bar{\tau}\overline{I}_0. \tag{2.2}$$

Proof. (Proof of $\Rightarrow$) Assume first that (2.1) holds and differentiate both sides with respect to t to obtain

$$CR'(t,p) = CR'(t,\overline{p}), \forall t \geq 0,$$

which is equivalent to

$$\nu_1 I(t,p) = \bar{\nu}_1 I(t,\overline{p}), \forall t \geq 0. \tag{2.3}$$

Differentiate both sides of (2.3) with respect to t, to obtain

$$\nu_1 I'(t,p) = \bar{\nu}_1 I'(t,\overline{p}), \forall t \geq 0.$$

Replacing $I'(t)$ by its formula in (1.1) on both sides of the above equality, we obtain

$$\begin{aligned}
&\nu_1 I'(t,p) = \bar{\nu}_1 I'(t,\overline{p}) \\
\Leftrightarrow\ &\nu_1\Big(\tau S(t,p)I(t,p) - \nu I(t,p)\Big) = \bar{\nu}_1\Big(\bar{\tau} S(t,\overline{p})I(t,\overline{p}) - \bar{\nu} I(t,\overline{p})\Big) \\
\Leftrightarrow\ &\nu_1 I(t,p)\Big(\tau S(t,p) - \nu\Big) = \bar{\nu}_1 I(t,\overline{p})\Big(\bar{\tau} S(t,\overline{p}) - \bar{\nu}\Big),
\end{aligned}$$

and (2.3) implies that

$$\tau S(t,p) - \nu = \bar{\tau} S(t,\overline{p}) - \bar{\nu}, \forall t \geq 0. \tag{2.4}$$

Differentiating both side of (2.4) with respect to t, we obtain

$$\begin{aligned}
&\tau S'(t,p) = \bar{\tau} S'(t,\overline{p}) \\
\Leftrightarrow\ &\tau\Big(-\tau S(t,p)I(t,p)\Big) = \bar{\tau}\Big(-\bar{\tau} S(t,\overline{p})I(t,\overline{p})\Big) \\
\Leftrightarrow\ &\tau^2 S(t,p)I(t,p) = \bar{\tau}^2 S(t,\overline{p})I(t,\overline{p})
\end{aligned}$$

and by using again (2.3), we obtain

$$\frac{\bar{\nu}_1}{\nu_1}\tau^2 S(t,p) = \bar{\tau}^2 S(t,\overline{p}), \forall t \geq 0. \tag{2.5}$$

Then, by using (2.4), and replacing $S(t,\overline{p})$ in (2.5), we obtain

$$\frac{\bar{\nu}_1}{\nu_1}\tau^2 S(t,p) = \bar{\tau}^2 (\frac{\bar{\nu} - \nu + \tau S(t,p)}{\bar{\tau}}).$$

Therefore,

$$(\frac{\bar{\nu}_1}{\nu_1}\tau - \bar{\tau})S(t,p) = \frac{\bar{\tau}}{\tau}(\bar{\nu} - \nu), \forall t \geq 0. \tag{2.6}$$

Differentiating both sides of equation (2.3) with respect to t, we obtain

$$(\frac{\bar{\nu}_1}{\nu_1}\tau - \bar{\tau})S'(t,p) = 0,$$

which implies that

$$(\frac{\bar{\nu}_1}{\nu_1}\tau - \bar{\tau})(-\tau S(t,p)I(t,p)) = 0, \forall t \geq 0. \tag{2.7}$$

Setting $t = 0$ in equations (2.3), (2.4), (2.6) and (2.7) we obtain the following system of equations

$$\begin{cases} I_0\nu_1 = \overline{I}_0\bar{\nu}_1 \\ \tau S_0 - \nu = \bar{\tau}\overline{S}_0 - \bar{\nu} \\ (\frac{\bar{\nu}_1}{\nu_1}\tau - \bar{\tau})S_0 = \frac{\bar{\tau}}{\tau}(\bar{\nu} - \nu) \\ -\tau S_0 I_0 (\frac{\bar{\nu}_1}{\nu_1}\tau - \bar{\tau}) = 0 \end{cases} \tag{2.8}$$

and (2.2) follows.
(Proof of $\Leftarrow$) To prove the converse implication, let $(S(t), I(t))$ be a solution of (1.1)-(1.2). Let $\overline{S}_0 > 0$, $\overline{I}_0 > 0$, and set

$$\overline{S}(t) := \frac{\overline{S}_0}{S_0}S(t) \text{ and } \overline{I}(t) := \frac{\overline{I}_0}{I_0}I(t).$$

Since $(S(t), I(t))$ satisfies (1.1)-(1.2), we obtain by replacing $S(t)$ and $I(t)$ with the above formulas,

$$\begin{cases} \frac{S_0}{\overline{S}_0}\overline{S}'(t) = -\tau\frac{S_0}{\overline{S}_0}\overline{S}(t)\frac{I_0}{\overline{I}_0}\overline{I}(t) \\ \frac{I_0}{\overline{I}_0}\overline{I}'(t) = \tau\frac{S_0}{\overline{S}_0}\overline{S}(t)\frac{I_0}{\overline{I}_0}\overline{I}(t) - \nu\frac{I_0}{\overline{I}_0}\overline{I}(t) \\ \overline{S}(0) = \overline{S}_0 \\ \overline{I}(0) = \overline{I}_0. \end{cases}$$

After simplifying,

$$\begin{cases} \overline{S}'(t) = -\tau \dfrac{I_0}{\overline{I}_0} \overline{S}(t)\overline{I}(t) \\ \overline{I}'(t) = \tau \dfrac{S_0}{\overline{S}_0} \overline{S}(t)\overline{I}(t) - \nu \overline{I}(t) \\ \overline{S}(0) = \overline{S}_0 \\ \overline{I}(0) = \overline{I}_0 \end{cases} \tag{2.9}$$

and by using (2.2) we deduce that $\tau \dfrac{I_0}{\overline{I}_0} = \tau \dfrac{S_0}{\overline{S}_0} = \bar{\tau}$ and $\nu = \bar{\nu}$. Therefore, $(\overline{S}(t), \overline{I}(t))$ satisfies (1.1)-(1.2) with the new parameter set $\overline{p}$. Moreover, the cumulative reported cases function for the parameter set $\bar{p}$ satisfies

$$CR(t, \overline{p}) := \bar{\nu}_1 \int_0^t \overline{I}(s)ds = \bar{\nu}_1 \frac{\overline{I}_0}{I_0} \int_0^t I(s)ds, \forall t \geq 0,$$

and by using (2.2) we deduce that $\dfrac{\tau}{\nu_1} = \dfrac{\bar{\tau}}{\bar{\nu}_1}, \tau I_0 = \bar{\tau}\overline{I}_0$. Therefore,

$$CR(t, \overline{p}) = CR(t, p), \forall t \geq 0.$$

■

Remark 2.2 *The cumulative reported cases function $CR(t)$ in (1.3) is uniquely determined for the parameter set*

$$p = \Big(\frac{\tau}{a}, \nu, aS_0, aI_0, \frac{\nu_1}{a}\Big),$$

whenever $a > \nu_1/\nu$.

3 Computation of the combined parameter set

3.1 System of equations to identify the parameters

In this section, we consider the SIR model (1.1)-(1.2) in the case where the basic reproduction number $\mathcal{R}_0 = \tau S_0/\nu > 1$. The cumulative reported case function $CR(t) := \nu_1 \int_0^t I(s)ds$, $t \geq 0$ is assumed to be known. The goal is to provide a simple method to identify the parameter set (1.4).

Recall that, since $\mathcal{R}_0 > 1$, the solutions of (1.1)-(1.2) have a typical outbreak behavior [26]:

(i) The function $t \mapsto S(t)$ is nonincreasing on $[0, \infty)$ with $S(0) = S_0$ and $S(\infty) > 0$;

(ii) There exists a unique turning point $t_p > 0$ such that $I'(t_p) = 0$, with I nondecreasing on $[0, t_p]$ and nonincreasing on $[t_p, \infty)$. Moreover $I(\infty) = 0$ and $t \mapsto I(t)$ is integrable on $[0, \infty)$.

In addition to the turning point t_p, the above properties allow us to defined several important quantities related to the function CR:

$$CR(t_p),\ CR'(t_p) \text{ and } CR(\infty). \tag{3.1}$$

As it will be seen later, these four quantities will be sufficient to compute the combined parameter set in (1.4). To compute these four combined parameters we will provide four independent equations. Three of them has been derived in [26]. Following the notations introduced in [26], we set

$$c := CR(\infty),\ r := \frac{CR(t_p)}{CR(\infty)}.$$

Next, by setting

$$X := c\frac{\tau}{\nu_1}, \tag{3.2}$$

and then by multiplying both sides by $S_0\nu_1$ we deduce that

$$X \times (S_0\nu_1) = cS_0\tau. \tag{3.3}$$

Moreover, by using respectively, equations (3.3), (3.7) and (3.9) in [26], we derive the three following independent equations

$$e^{-X} + Xe^{-rX} = 1 + \frac{I_0}{S_0}, \tag{3.4}$$

$$(S_0\nu_1) \times \left[1 + \frac{I_0}{S_0} - e^{-rX}\left(1 + rX\right)\right] = CR'(t_p) \tag{3.5}$$

and

$$\nu = (S_0\tau) \times e^{-rX}. \tag{3.6}$$

We recall first, that by Proposition 3.1 in [26], equation (3.4) implies the following compatibility condition with the data: $r \in (0, 1/2)$. Thus, for the model (1.1)-(1.2) more than half of all cases occur after the turning point t_p. As noted in [26], some outbreak epidemics have more than half of all cases occurring before the turning point, and the model (1.1)-(1.2) is not applicable to these examples.

3.2 Derivation of the equation for the turning point

In order to define an equation for the turning point, we first introduce the function

$$F(X) := e^{-X} + Xe^{-rX} - 1.$$

Lemma 3.1 $r \geq \frac{1}{2} \Rightarrow F(X) < 0, \forall X > 0.$

Proof. We have $F(0) = 0$ and $F'(X) = e^{-rX}G(x)$, with

$$G(X) := 1 - rX - e^{-(1-r)X}.$$

Then $G(0) = 0$ and $G'(X) = -r + (1-r)e^{-(1-r)X}$. So if $r \geq 1/2$ we have

$$G'(X) < 0, \forall X > 0.$$

The result follows. ∎

Lemma 3.2 *Assume that $r \in (0, 1/2)$. There exists a unique strictly positive solution $X(r) > 0$ of equation*

$$F(X) = 0 \Leftrightarrow e^{-X} + Xe^{-rX} - 1 = 0. \tag{3.7}$$

Moreover, there exists $X_{\max} \in (0, X(r))$, such that the function $F(X)$ is strictly increasing on $(0, X_{\max})$ and strictly decreasing on $(X_{\max}, X(r))$.
Furthermore,

$$\begin{cases} F(X) > 0, \text{ if } X \in (0, X(r)), \\ F(X) < 0, \text{ if } X \in (X(r), \infty). \end{cases} \tag{3.8}$$

Proof. We have $F(0) = 0$ and $F'(X) = e^{-rX}G(x)$, with

$$G(X) := 1 - rX - e^{-(1-r)X}.$$

Then $G(0) = 0$ and $G'(X) = -r + (1-r)e^{-(1-r)X}$. Moreover, we have

$$G'(X) = 0 \Leftrightarrow X = \frac{1}{1-r} \ln\left(\frac{1}{r} - 1\right) := X^* > 0.$$

Thus, $G'(X) > 0$ for $X \in (0, X^*)$ and $G'(X) < 0$ for $X > X^*$. We also have

$$\lim_{X \to \infty} G(X) = -\infty.$$

Let $X_{\max} > X^*$ be the unique value in $(0, +\infty)$ such that $G(X_{\max}) = 0$. Moreover, $F'(X) > 0$ on $(0, X_{\max})$, $F'(X) < 0$ on $(X_{\max}, \infty)$, and $F'(0) = F'(X_{\max}) = 0$. Hence, $F(X_{\max}) > 0$ is the maximum of F. Next we observe that

$$\lim_{X \to \infty} F(X) = -1$$

Therefore there exists a unique $X(r) \in (X_{\max}, \infty)$ such that $F(X(r)) = 0$. ∎

The above results are summarized in Figure 1.

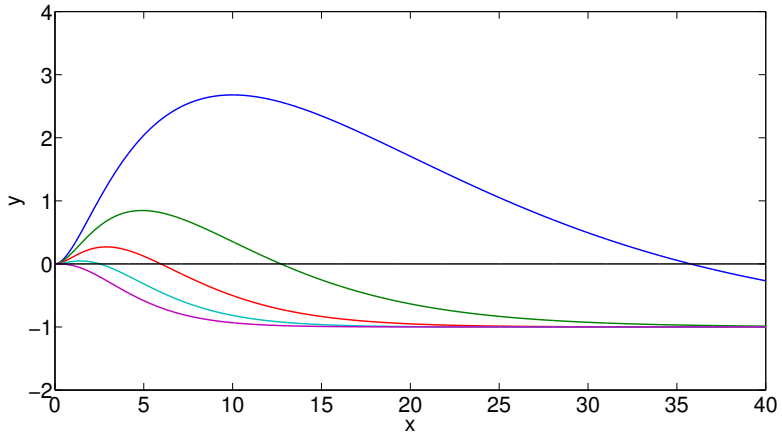

Figure 1: *We plot* $X \to F(X)$ *whenever* $r = 0.1, 0.2, 0.3, 0.4$ *and* 0.5 *respectively in blue, green, red, clear blue and purple.*

Next, we derive an additional independent equation involving the turning point t_p. To that aim, recall that

$$CR'(t) = \nu_1 I(t),\ \forall t > 0 \text{ and } CR(0) = 0.$$

As a consequence, we obtain from (1.1)-(1.2),

$$\frac{d}{dt}\left(S(t) + I(t) + \frac{\nu}{\nu_1} CR(t)\right) = 0,\ \forall t > 0,$$

so that

$$S(t) + I(t) + \frac{\nu}{\nu_1} CR(t) = S_0 + I_0,\ \forall t \geq 0.$$

Hence, we obtain

$$CR'(t) = \nu_1 I(t) = \nu_1 \left[S_0 + I_0 - S(t) - \frac{\nu}{\nu_1} CR(t)\right],\ \forall t \geq 0.$$

However, by using (1.1), we may eliminate $S(t)$, since $S(t) = S_0 e^{-\frac{\tau}{\nu_1} CR(t)}$. As a consequence, $CR(t)$ satisfies the equation

$$CR'(t) = \nu_1 I(t) = S_0 \nu_1 \left[1 + \frac{I_0}{S_0} - e^{-\tau CR(t)} - \frac{\nu}{S_0 \nu_1} CR(t)\right],\ \forall t \geq 0.$$

Now note that since $CR'(t) = \nu_1 I(t) > 0$, the function $t \to CR(t)$ must be increasing on $(0, \infty)$. We thus have

$$1 + \frac{I_0}{S_0} - e^{-\frac{\tau}{\nu_1} y} - \frac{\nu}{S_0 \nu_1} y > 0,\ \forall y \in [0, CR(\infty)).$$

Remark 3.3 *The above inequality gives a condition on the ultimate number of cumulative cases* $c = CR(\infty)$.

Integration of the differential equation for $CR(t)$ above, from $t = 0$ to $t = t_p$, yields

$$\int_0^{t_p} \frac{CR'(t)}{1 + \frac{I_0}{S_0} - e^{-\frac{\tau}{\nu_1} CR(t)} - \frac{\nu}{S_0 \nu_1} CR(t)} dt = (S_0 \nu_1) t_p.$$

Set $s = CR(t)$ and we obtain

$$\int_0^{CR(t_p)} \frac{1}{1 + \frac{I_0}{S_0} - e^{-\frac{\tau}{\nu_1} s} - \frac{\nu}{S_0 \nu_1} s} ds = (S_0 \nu_1) t_p.$$

Now recalling that $CR(t_p) = rc$, the change of variable $s = c\sigma$ yields

$$\int_0^{r} \frac{1}{1 + \frac{I_0}{S_0} - e^{-\frac{c\tau}{\nu_1} \sigma} - \frac{c\nu}{S_0 \nu_1} \sigma} d\sigma = (S_0 \nu_1) \frac{t_p}{c}.$$

By (3.2), that is $X = \frac{c\tau}{\nu_1}$, we deduce that

$$\int_0^{r} \frac{1}{1 + \frac{I_0}{S_0} - e^{-X\sigma} - \frac{\nu}{\tau S_0} X \sigma} d\sigma = (S_0 \nu_1) \frac{t_p}{c}.$$

By (3.6), that is $\nu = S_0 \tau e^{-rX}$, we have

$$\int_0^{r} \frac{d\sigma}{1 + \frac{I_0}{S_0} - e^{-X\sigma} - X e^{-rX} \sigma} = (S_0 \nu_1) \frac{t_p}{c}.$$

By (3.5), that is $S_0 \nu_1 \left[1 + \frac{I_0}{S_0} - e^{-rX} (1 + rX)\right] = CR'(t_p)$, we obtain

$$\left[1 + \frac{I_0}{S_0} - e^{-rX} (1 + rX)\right] \times \int_0^{r} \frac{d\sigma}{1 + \frac{I_0}{S_0} - e^{-X\sigma} - X e^{-rX} \sigma} = CR'(t_p) \frac{t_p}{c}.$$

Finally, by (3.4), that is $e^{-X} + X e^{-rX} = 1 + \frac{I_0}{S_0}$ and Lemma 3.2, we deduce that

$$0 < X < X(r) \tag{3.9}$$

and that X must satisfy the *turning point equation*

$$T(X, r) = \frac{CR'(t_p) t_p}{c}, \tag{3.10}$$

where the mapping T is defined by

$$T(X, r) := \int_0^{r} \frac{H(X, r)}{H(X, \sigma)} d\sigma, \tag{3.11}$$

where the right hand side is an improper integral and the function H is defined by

$$H(X, \sigma) := e^{-X} + X e^{-rX} - e^{-X\sigma} - X e^{-rX} \sigma. \tag{3.12}$$

Remark 3.4 *We observe that* $H(X, 0) = e^{-X} + X e^{-rX} - 1 = F(X)$.

3.3 Analysis of the turning point equation

Lemma 3.5 *If $r \in (0, \frac{1}{2})$, and $X \in (0, X(r))$, then*

$$0 < F(X) = H(X, 0) \leqslant H(X, \sigma) \leqslant H(X, r)$$

for every $\sigma \in [0, r]$.

Proof. Suppose that $r \in (0, \frac{1}{2})$, $X \in (0, X(r))$, and consider

$$\partial_\sigma H(X, \sigma) = X(e^{-\sigma X} - e^{-rX}) \geqslant 0, \forall \sigma \in [0, r],$$

which means that $H(X, \sigma)$ is increasing with respect to σ. ∎

Lemma 3.6 *The function $X \to T(X, r)$ is well-defined on the open interval $(0, X(r))$. Moreover,*

$$\lim_{X \to X(r)^-} T(X, r) = \infty, \tag{3.13}$$

and

$$\lim_{X \to 0^+} T(X, r) = \frac{r-1}{2} \ln(1 - 2r). \tag{3.14}$$

Proof. Proof of (3.13): By Lemma 3.5, $T(X, r)$ is well-defined on the interval $(0, X(r))$. Since $X(r)$ is the unique positive solution of the equation (3.7), it follows that $H(X(r), r) = 1 - e^{-rX(r)} - r(1 - e^{-rX(r)})$. Let $k(x, r) = 1 - e^{-rx} - r(1 - e^{-x})$ on $[0, \infty)$. Then, $\partial_x k(x, r) = r(e^{-rx} - e^{-x}) > 0$ for every $x \in [0, \infty)$, and $r \in (1, \frac{1}{2})$. This means that $k(x, r) > k(0, r) = 0$ for every $x \in (0, \infty)$, and $r \in (1, \frac{1}{2})$. Therefore, $H(X(r), r) = 1 - e^{-rX(r)} - r(1 - e^{-rX(r)}) > 0$ for every $r \in (1, \frac{1}{2})$.

Moreover,

$$\begin{aligned}\lim_{\sigma \to 0^+} \frac{H(X(r), \sigma)}{\sigma} &= \lim_{\sigma \to 0^+} \frac{1 - e^{-\sigma X(r)} - rX(r)e^{-rX(r)}\sigma}{\sigma} \\ &= \lim_{\sigma \to 0^+} \frac{1 - e^{-\sigma X(r)}}{\sigma} - rX(r)e^{-rX(r)} \\ &= X(r) - X(r)e^{-rX(r)} = X(r)(1 - e^{-rX(r)}) > 0.\end{aligned}$$

This means that $\int_0^r \lim_{X \to X(r)^-} \frac{H(X, r)}{H(X, \sigma)} d\sigma = \infty$, and by Fatou's Lemma, we have $\lim_{X \to X(r)^-} T(X, r) = \infty$.

Proof of (3.14): Next, taking the Taylor's expansions of the functions e^{-X}, e^{-rX}, e^{-hX} at $X = 0$, and letting $\sigma \in [0, r] \subset (0, \frac{1}{2})$, we obtain

$$\begin{aligned}e^{-\sigma X} &= 1 - \sigma X + \frac{1}{2}\sigma^2 X^2 + o((\sigma X)^3) = 1 - \sigma X + \frac{1}{2}\sigma^2 X^2 + o(X^3), \\ e^{-rX} &= 1 - rX + \frac{1}{2}r^2 X^2 + o(X^3), \text{ and } e^{-X} = 1 - X + \frac{1}{2}X^2 + o(X^3),\end{aligned}$$

where $o(X^3)$ does not depend on σ. Thus,

$$H(X,\sigma)=\frac{1-2r-\sigma^2+2r\sigma}{2}X^2+o(X^3),\ H(X,r)=\frac{(1-r)^2}{2}X^2+o(X^3).$$

Hence,

$$\frac{H(X,r)}{H(X,\sigma)}=\frac{\frac{(1-r)^2}{2}X^2+o(X^3)}{\frac{1-2r-\sigma^2+2r\sigma}{2}X^2+o(X^3)}\ =\frac{(1-r)^2+o(X)}{1-2r-\sigma^2+2r\sigma+o(X)}.$$

Since $o(X)$ does not depend on σ, when X tends to 0^+ the function $\dfrac{H(X,r)}{H(X,\sigma)}$ is uniformly convergent to $h(\sigma)=\dfrac{(1-r)^2}{1-2r-\sigma^2+2r\sigma}$ on $[0,r]$. Thus,

$$\lim_{X\to 0^+}T(X,r)=\int_0^r\frac{(1-r)^2}{1-2r-\sigma^2+2r\sigma}\mathrm{d}\sigma$$

$$=(1-r)^2\int_0^r\frac{1}{(2r-1-\sigma)(\sigma-1)}\mathrm{d}\sigma\ =\frac{(1-r)}{2}\ln(\frac{1}{1-2r}).$$

■

In Figure 2 we plot the mapping $r\to X(r)$, where $X=X(r)$ is the solution of (3.7), as r varies in $(0,\frac{1}{2})$. In Figure 3 we plot the mapping $x\to T(xX(r),r)$, where T is defined by (3.11), as x varies in $(0,1)$, for different values of r. From Figure 3 we observe that numerically, the mappings $X\to T(X,r)$ are all monotone increasing for each value of r. As a consequence we can conclude (numerically) that equation (3.10) has a unique solution $X\in(0,X(r))$.

Remark 3.7 *From Figure 3 we can also visualize the minimum value for $T(X,r)$, which is given by (3.14). By using (3.10) we deduce that we must have*

$$\frac{CR'(t_p)t_p}{CR(\infty)}>\frac{r-1}{2}\ln(1-2r).$$

where $r=\dfrac{CR(t_p)}{CR(\infty)}$. Therefore, we obtain a new relationship constraining the values $t_p, CR(t_p), CR'(t_p), CR(\infty)$.

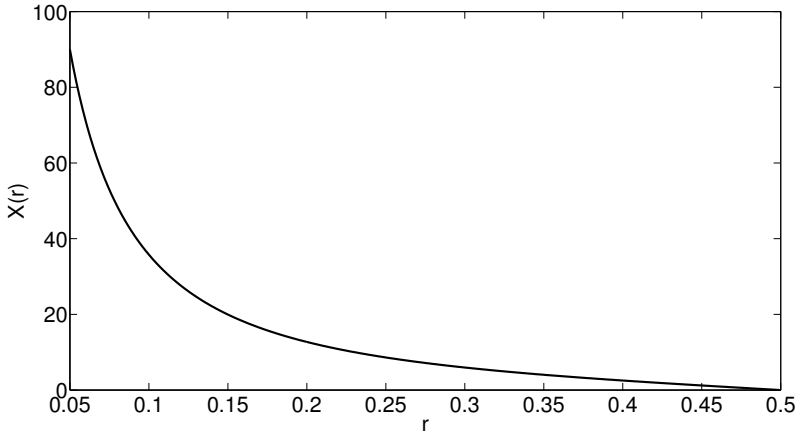

Figure 2: *The mapping* $r \to X(r)$*, where* r *varies in* $(0, \frac{1}{2})$.

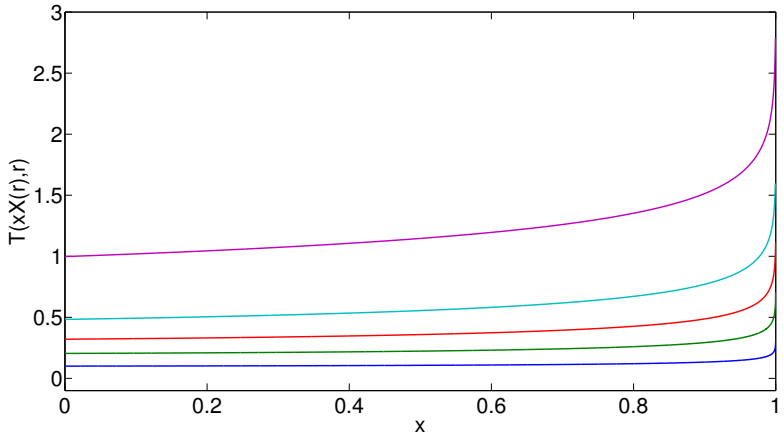

Figure 3: *The mapping* $x \to T(xX(r), r)$*, where* x *varies in* $(0, 1)$*. The blue, green, red, clear blue and purple curves correspond (from the bottom to the top) to* $r = 0.1$*,* $r = 0.2$*,* $r = 0.3$*,* $r = 0.4$ *and* $r = 0.49$*, respectively.*

As a consequence of Proposition 2.1 we have the following theorem.

Theorem 3.8 *Assume that the equation (3.10) has a unique solution* X *in* $(0, X(r))$*. Then the function* $t \to CR(t)$ *is uniquely determined by the turning point* t_p*,* $CR(t_p)$*,* $CR'(t_p)$ *and* $CR(\infty)$*.*

Remark 3.9 *Assume that the equation (3.10) has a unique solution* X *in*

$(0, X(r))$. *Then the basic reproduction number* $\mathcal{R}_0$ *is uniquely determined by the turning point* t_p, $CR(t_p)$, $CR'(t_p)$ *and* $CR(\infty)$.

4 An identification method and an application to an outbreak epidemic

4.1 Description of the method

By Proposition 2.1, the combination of parameters $\frac{I_0}{S_0}$, $\nu_1 S_0$, $S_0\tau$, $\nu_1 I_0$, and ν for the system (1.1)-(1.2) are uniquely determined by the cumulative reported cases function $CR(t), t \geq 0$ for the parameter set $p = (\tau, \nu, S_0, I_0, \nu_1)$. Moreover, the analysis in Section 3 allows us to derive a method to compute this combination by the following four steps: Assume the values of t_p, $CR(t_p)$, $CR'(t_p)$, and $CR(\infty)$ are known, and set $c = CR(\infty)$ and $r = \frac{CR(t_p)}{CR(\infty)}$.

Step 1: Solve the equation (3.7), $e^{-X} + Xe^{-rX} - 1 = 0$, to obtain the unique positive solution $X(r)$.

Step 2: Solve the turning point equation (3.10)

$$\int_0^r \frac{e^{-X} + Xe^{-rX} - e^{-rX} - rXe^{-rX}}{e^{-X} + Xe^{-rX} - e^{-\sigma X} - \sigma Xe^{-rX}} d\sigma = \frac{CR'(t_p)t_p}{c},$$

with the condition (3.9), $0 < X < X(r)$, to obtain the value $X = c\frac{\tau}{\nu_1}$.

Step 3:

i. Compute the value of $a_1 := \frac{\tau}{\nu_1} = \frac{X}{c}$ by the formula (3.2);

ii. Compute the value of $a_2 := \frac{I_0}{S_0} = e^{-X} + Xe^{-rX} - 1$ by the formula (3.4);

iii. Compute the value of $a_3 := \nu_1 S_0 = \frac{CR'(t_p)}{1 + a_2 - e^{-rX}(1 + rX)}$ by the formula (3.5);

iv. Compute the value of $a_4 := \tau S_0 = a_3 a_1$ by the formula (3.3);

v. Compute the value of $a_5 := \nu_1 I_0 = a_2 a_3$;

vi. Compute the value of $a_6 := \nu = a_4 e^{-rX}$ by the formula (3.6).

Remark 4.1 *By the fact that $\nu_1 < \nu$, we obtain the following evaluations:*

$$\begin{aligned} S_0 &> S_0\frac{\nu_1}{\nu} = \frac{a_3}{a_6} \\ I_0 &> I_0\frac{\nu_1}{\nu} = \frac{a_5}{a_6} \\ \tau &< \nu\frac{\tau}{\nu_1} = a_1a_6. \end{aligned} \tag{4.1}$$

Moreover, the basic reproduction number is

$$\mathcal{R}_0 = \frac{\tau S_0}{\nu} = \frac{a_4}{a_6}. \tag{4.2}$$

Remark 4.2 *If S_0 is known, the epidemic final size is written as follows:*

$$\begin{aligned} C(\infty) &= CR(\infty) + CU(\infty) = CR(\infty) + \frac{\nu_2}{\nu_1}CR(\infty) \\ &= \frac{\nu}{\nu_1}CR(\infty) = \nu\frac{S_0}{\nu_1 S_0}CR(\infty) = a_6\frac{c}{a_3}S_0. \end{aligned} \tag{4.3}$$

Denote by $N = S_0 + I_0$, the number of individuals involved in the epidemic,which is typically smaller than the total number individuals in the population, since some people have immunity. Then S_0 satisfies

$$\frac{a_3}{a_6} < S_0 \leq N - I_0 = N - S_0\frac{I_0}{S_0},$$

which implies

$$\frac{a_3}{a_6} < S_0 \leq \frac{N}{1 + \frac{I_0}{S_0}} = \frac{N}{1 + a_2}. \tag{4.4}$$

Moreover, the number of susceptible individuals at the end of epidemic can be computed by the following formula:

$$S(\infty) = S_0 + I_0 - C(\infty) = S_0\Big(1 + a_2 - a_6\frac{c}{a_3}\Big). \tag{4.5}$$

4.2 Application to Hong-Kong influenza in New York City in 1968-1969

In this application, we have the values $CR(\infty) = 1080$, $CR(t_p) = 500$, $t_p = 6.15$, $CR'(t_p) = 190$ (see [26]). The total population of New York City in 1968 is $7,900,000$. Consider the equation

$$\int_0^{\frac{500}{1080}} \frac{e^{-X} + Xe^{-\frac{500}{1080}X} - e^{-\frac{500}{1080}X} + \frac{500}{1080}Xe^{-\frac{500}{1080}X}}{e^{-X} + Xe^{-\frac{500}{1080}X} - e^{-\sigma X} - \sigma Xe^{-\frac{500}{1080}X}}d\sigma = \frac{11685}{10800}. \tag{4.6}$$

for $X \in (0, X(r))$ where $X(r) = 0.89478$ is the positive solution of the equation

$$e^{-X} + Xe^{-\frac{500}{1080}X} - 1 = 0. \tag{4.7}$$

First, solve the equation (4.7), and obtain the value of $X(r) = 0.89478$. This value corresponds to the positive zero of the function in Figure 4 - left side.

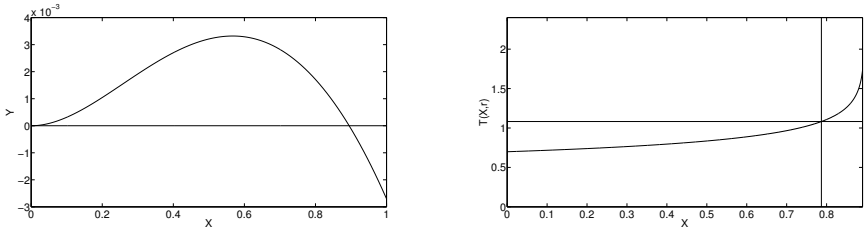

Figure 4: *Left side: The graph of* $Y = e^{-X} + Xe^{-\frac{500}{1080}X} - 1$. *Right side: The intersection of* $Y = T(X, \frac{500}{1080})$ *and* $Y = \frac{t_p CR'(t_p)}{CR(\infty)}$. *We find* $X(r) = 0.89478$, $X = 0.7869$ *and* $\dfrac{I_0}{S_0} = e^{-X} + Xe^{-\frac{500}{1080}X} - 1 = 0.0019$.

Finally, by applying Step 4 of the method described above we obtain the following table of values:

Variable	Description	Estimated value
X	$c\tau/\nu_1$	0.79
a_1	τ/ν_1	7.3×10^{-4}
a_2	I_0/S_0	0.002
a_3	$\nu_1 S_0$	3509.1
a_4	τS_0	2.56
a_5	$\nu_1 I_0$	6.65
a_6	ν	1.78

Table 1: *List of a combination of parameters obtained for Hong-Kong influenza in New York in 1968-1969.*

Remark 4.3 *From the reported case data for the Hong Kong influenza in New York City in 1968-1969, there are at least three infected individuals at the beginning of the epidemic.*

If the value of the initial susceptible individuals S_0 is given, then all the parameters can be obtained. The following table gives these values when $S_0 = 1,976$, $S_0 = 4,000,000$, and $S_0 = 7,885,047$:

Variable	Estimated value 1	Estimated value 2	Estimated value 3
S_0	1976	$4,000,000$	$7,885,047$
I_0	3.7472	$7,586$	$14,953$
τ	1.3×10^{-3}	6.4×10^{-7}	3.2×10^{-7}
ν_1	1.78	0.88×10^{-3}	4.5×10^{-4}
ν_2	3.1×10^{-4}	1.78	1.78
$C(\infty)$	1080	2.19×10^6	4.31×10^6
$S(\infty)$	899	1.82×10^6	3.59×10^6

Table 2: *List of combinations of parameters obtained for Hong-Kong influenza in New York in 1968-1969. In this table we vary the value of* S_0 *between the minimal value* 1976 *up to the maximal value* $7,885,047$ *and we computed the corresponding estimated parameters values.*

In Figures 5,6, and 7 we provide model (1.1)-(1.2) output for the Hong Kong influenza epidemic in New York City in 1968-1968 for the parameters in Table 4.2 and the values under $S_0 = 4,000,000$ in Table 2 (see [26]). In Figure 5 we compare the model output to the reported case data (see [26]). In Figure 6 we illustrate the epidemic final size as a function of the initial number of susceptibles S_0. In Figure 7 we illustrate the epidemic final size as a function of the turning point of the epidemic.

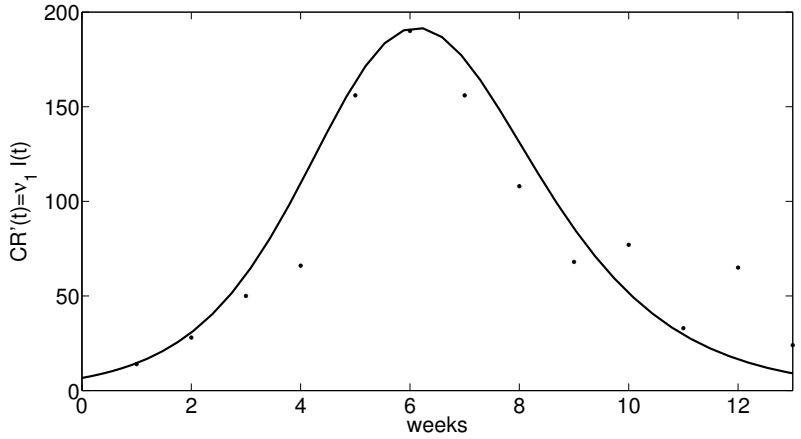

Figure 5: *Hong Kong influenza epidemic in New York City in 1968-1969. The weekly reported mortality case data and cumulative reported case data (blue stars), and the model output graph* $CR'(t)$ *(red). To run this simulation we fix* $S_0 = 4,000,000$.

Moreover, the basic reproduction number is

$$\mathcal{R}_0 = \frac{\tau S_0}{\nu} = \frac{a_4}{a_6} = 1.4. \tag{4.8}$$

From Remark 4.2 whenever S_0 is known the final size of the epidemic $C(\infty)$ is expressed linearly in function of S_0

$$C(\infty) = \frac{a_6 CR(\infty)}{a_3} S_0, \tag{4.9}$$

and we have the following upper and lower bounded for S_0

$$\frac{a_3}{a_6} < S_0 \leq \frac{N}{1+a_2}. \tag{4.10}$$

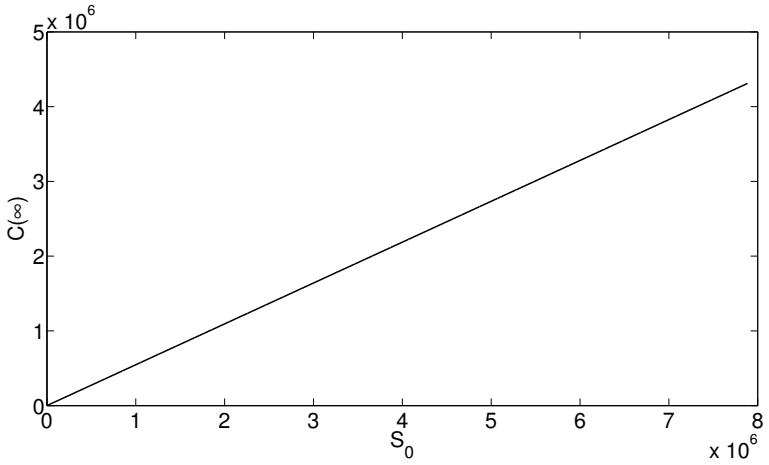

Figure 6: *The relationship between the total case number at the end of epidemic* $C(\infty)$ *and* S_0*. Here* S_0 *varies from* $1,976$ *(which is strictly larger than the minimal value* $\frac{a_3}{a_6} = 1,975$*) up to the maximal value* $S_0 = \frac{N}{1+a_2} = 7,885,047$ *which corresponds to* $I_0 = 14,953$*.*

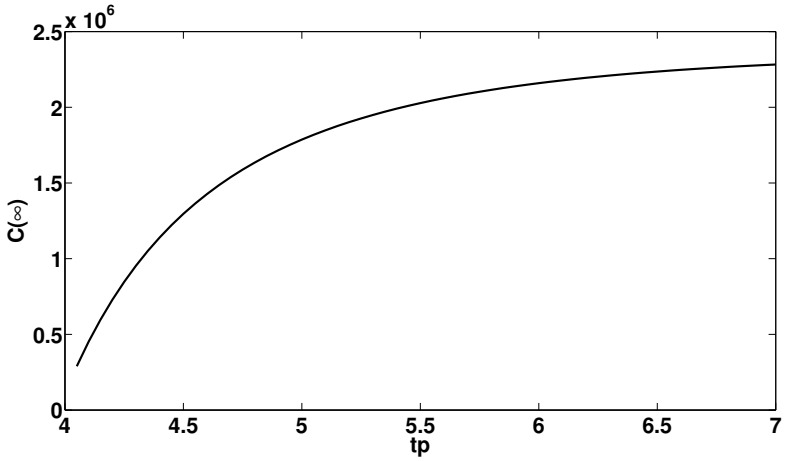

Figure 7: *In this figure we fix* $S_0 = 4,000,000$*, vary the turning point* t_p *and plot the final size* $C(\infty)$ *of the epidemic as a function of the turning point.*

4.3 Discussion

We have investigated the parameter identification problem for the well-known SIR model (1.1)-(1.2). Although this model has a long history (see our references), there is a need for further development of the role of parameters for its applications. A major difficulty in applying this model is the lack of precise data for most epidemic outbreaks, including influenza, cholera, Ebola, and other current epidemics. Epidemic data provided by public health agencies consists of reported cases, which are typically only a fraction of all cases. We have provided a methodology to estimate parameters and initial conditions for this model, which allows realistic applications for its data input and output simulations. One future goal will be to extend this simple model to more complex formulations of an epidemic progression. In particular, the issues of disease age progression, incomplete immunity, vaccination, quarantine, and social interventions are all of vital importance in controlling an outbreak epidemic. A second future goal will be to extend the analysis to real-time predictions of epidemic progression from early to final stages. With improved epidemic tracking and mathematical model forecasting, the toll of future outbreak epidemics on human society can be greatly reduced.

References

[1] R.M. Anderson and R.M. May, *Infective Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991.

[2] V. Andreasen, The final size of an epidemic and its relation to the basic reproduction number, *Bull. Math. Biol.*, **73(10)** (2011), 2305-2321.

[3] J. Arino, F. Brauer, P. Van Den Driessche, J. Watmough and J. Wu, A final size relation for epidemic models, *Math. Biosci. Eng.*, **4(2)** (2007), 159-175.

[4] N.T.J. Bailey, *The Mathematical Theory of Epidemics*, Charles Griffin, London, 1957

[5] F. Brauer, P. van den Driessche and J. Wu (eds.) Mathematical epidemiology, Springer, Berlin, 2008.

[6] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, Springer, New York, 2000.

[7] S. Busenberg and K. Cooke, *Vertically Transmitted Diseases: Models and Dynamics*, Lecture Notes in Biomathematics,**23**, Springer-Verlag, Berlin, 1993.

[8] M. Capistran, M. Moreles, and B. Lara, Parameter estimation of some epidemic models. The case of recurrent epidemics caused by respiratory syncytial virus, *Bull. Math Biol.*, **71** (2009), 1890-1901.

[9] G. Chowell, E. Shim, F. Brauer, P. Diaz-Dueñas, J.M. Hyman, C. Castillo-Chavez, Modelling the transmission dynamics of acute haemorrhagic conjunctivitis: Application to the 2003 outbreak in Mexico, *Stat. Med.*, **25**, (2006), 1840-1857.

[10] G. Chowell, P. Diaz-Dueñas, J.C. Miller, A. Alcazar-Velazco, J.M. Hyman, P.W. Fenimore, and C. Castillo-Chavez, Estimation of the reproduction number of dengue fever from spatial epidemic data, *Math. Biosci.*, **208**, (2007), 571-589.

[11] O. Diekmann, H. Heesterbeek and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton University Press, Princeton, 2013.

[12] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz, 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** (1990), 365-382.

[13] N.D. Evans, L.J. White, M.J. Chapman, K.R. Godfrey, and M. Chappell, The structural identifiability of the susceptible infected recovered model with seasonal forcing, *Math. Biosci.*, **194** (2005), 175-197.

[14] N. Grassly and C. Fraser, Seasonal infectious disease epidemiology, *Proc. Roy. Soc. Lond. B: Biol. Sci.*, **273** (2006), 2541-2550.

[15] K.P. Hadeler, Parameter identification in epidemic models, *Math. Biosci.*, **229** (2011),185-189.

[16] K.P. Hadeler, Parameter estimation in epidemic models: simplified formulas, *Can. Appl. Math. Q.*, **19** (2011), 343-356.

[17] H.W. Hethcote, Qualitative analyses of communicable disease models, *Math. Biosci.*, **28** (1976), 335-356.

[18] H. Hethcote, Modeling heterogeneous mixing in infectious disease dynamics, in V. Isham and G. Medley (eds.), *Models for infectious human diseases: their structure and relation to data*, Cambridge University Press, Cambridge, 1996.

[19] H.W. Hethcote, The mathematics of infectious diseases, *SIAM Review*, **42(4)** (2000), 599-653.

[20] G. Hooker, S.P. Ellner, L. De Vargas Roditi, and D.J.D. Earn, Parameterizing state space models for infectious disease dynamics by generalized profiling: measles in Ontario, *J. Roy. Soc. Interface* **8**, (2011), 961-974.

[21] Y-H. Hsieh, D. Fisma, and J. Wu, On epidemic modeling in real time: An application to the 2009 Novel A (H1N1)influenza outbreak in Canada, *BMC Research Notes*, **3** (2010), 283.

[22] M. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*, Princeton University Press, Princeton, 2007.

[23] A. Lange, Reconstruction of disease transmission rates: Applications to measles, dengue, and influenza, *J. Theor. Biol.* **400** (2016), 138-153.

[24] J. Li and Y. Lou, Characteristics of an epidemic outbreak with a large initial infection size, *J. Biol. Dyn.* **10** (2016), 366-378.

[25] J. Ma and D.J.D. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bull. Math. Biol.*, **68** (2006), 679-702.

[26] P. Magal and G. Webb, The parameter identification problem for SIR epidemic models: Identifying Unreported Cases, *J. Math. Biol.* (2018). `https://doi.org/10.1007/s00285-017-1203-9`

[27] A. Mummert, Studying the recovery procedure for the time-dependent transmission rate(s) in epidemic models, *J. Math. Biol.*, **67** (2013), 483-507.

[28] J.D. Murray, *Mathematical Biology*, Springer, Berlin, 1993.

[29] L. Pellis, N.M. Ferguson, and C. Fraser, Threshold parameters for a model of epidemic spread among households and workplaces, *J. Roy. Soc. Interface* **6**, (2009), 979-987.

[30] M. Pollicott, H. Wang H, and H. Weiss, Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem, *J. Biol. Dyn.* **6** (2012), 509-523.

[31] L.I.W. Roeger, Z. Feng and C. Castillo-Chavez, Modeling TB and HIV co-infections, *Math. Biosci. Eng.*, **6(4)** (2009), 815-837.

[32] H.R. Thieme, *Mathematics in Population Biology*, Princeton University Press, Princeton, 2003.

[33] P. Van den Driessche and J. Watmough, Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.* **180** (2002) 29-48.