

# The Parameter Identification Problem for SIR Epidemic Models: Identifying Unreported Cases

PIERRE MAGAL<sup>(\*)</sup> AND GLENN WEBB<sup>(\*\*)</sup>

<sup>(\*)</sup> *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France*

*CNRS, IMB, UMR 5251, F-33400 Talence, France.*

<sup>(\*\*)</sup> *Mathematics Department, Vanderbilt University, Nashville, TN 37240, USA*

December 25, 2017

*This article is dedicated to the memory of Karl Hadeler.*

## Abstract

A SIR epidemic model is analyzed with respect to identification of its parameters, based upon reported case data from public health sources. The objective of the analysis is to understand the relation of unreported cases to reported cases. In many epidemic diseases the ratio of unreported to reported cases is very high, and of major importance in implementing measures for controlling the epidemic. This ratio can be estimated by the identification of parameters for the model from reported case data. The analysis is applied to three examples: (1) the Hong Kong seasonal influenza epidemic in New York City in 1968-1969, (2) the bubonic plague epidemic in Bombay, India in 1906, and (3) the seasonal influenza epidemic in Puerto Rico in 2016-2017.

**Key words.** epidemic models, transmission rate, reported cases, unreported cases.

## 1 Introduction

A major challenge in the application of epidemic models is the determination of model parameters. The challenge is made difficult by the lack of complete data available in the course of a typical epidemic disease. In the United States, typical epidemic data consist of Morbidity and Mortality Weekly Reports (MMWR) published by the Centers for Disease Control and Prevention. For many diseases, such as seasonal influenza, these data are very incomplete, and record only a small fraction of total cases. In general, reported cases are

not representative of all cases, with respect to symptoms severity, infectiousness level, demographic classification, and social context. Identifying the fraction of unreported cases requires that the unknown becomes known. Mathematical models of an epidemic's evolution provide a way to make known this unknown. The distinction can be incorporated into the model parameterization, and the known case data can be related to the number of unknown cases through the transmission dynamics in the model construction.

We assume that reported case data provide a time-sequential record of the number of cases reported in some specified formulation by public health officials. We assume that these cases are removed from the infected class and that they no longer have capability to infect susceptible individuals (due to mortality, hospitalization, isolation, social instruction, or other reasons). We assume that a typical susceptible individual has the potential to be infected by an infected individual independently of whether or not the infected individual will ultimately be reported.

We note several examples of case data for specific epidemics. In [9] data for the 2010-2011 influenza season in the United States were obtained by landline telephone survey of approximately 90,000 people. Of these 8.9% of adults and 33.9% of children reported influenza like symptoms. In [48] a statistical estimator model was used to estimate the ratio of unreported to reported cases for the H1N1 influenza epidemic in the United States from April to July 2009 as 79 to 1, and the ratio of total cases to confirmed cases as 140 to 1. In [51] it is estimated that approximately 60.8 million cases occurred in the United States during this epidemic, with the population approximately 300 million during this time period.

In Section 2 we present the SIR model and formulate a system of equations connecting its parameters to reported case data. In Section 3 we solve this system of equations and identify the model parameters. In Section 4 we present three examples of epidemics with reported case data, and determine the fraction of their unreported cases. In Section 5 we discuss our results and further work.

## 2 The deterministic SIR epidemic model and the identification of its parameters

The first SIR deterministic epidemic models were developed by Bernoulli [8, 20, 21], Ross [50], Kermack-McKendrick [34, 35, 36], and Macdonald [40]. Investigations of these models and their extensions have been developed by many researchers, including [1, 6, 11, 12, 13, 18, 28, 30, 33, 45, 54]. The parameter identification problem for SIR model has been investigated by many researchers, including [3, 4, 14, 16, 17, 19, 22, 25, 26, 27, 29, 31, 32, 37, 39, 41, 44, 46, 47, 49, 56].

The SIR model we analyze is the following: at time  $t$ , let  $S(t)$  be the number of susceptible,  $I(t)$  the number of infected, and  $R(t)$  the number of removed and

permanently immune. The equations for the SIR model are

$$\frac{d}{dt}S(t) = -\tau I(t)S(t) \quad (2.1)$$

$$\frac{d}{dt}I(t) = \tau I(t)S(t) - (\nu_1 + \nu_2)I(t) \quad (2.2)$$

$$\frac{d}{dt}R(t) = (\nu_1 + \nu_2)I(t) \quad (2.3)$$

with initial conditions  $S(0) = S_0 > 0$ ,  $I(0) = I_0 > 0$ ,  $R(0) = R_0 \geq 0$ . Here  $\tau$  is the transmission rate,  $\nu_1$  is the removal rate of reported infected individuals, and  $\nu_2$  is the removal rate of infected individuals due to all other unreported causes such as mortality, recovery, or other reasons. It is assumed that a reported case, known to medical care-givers or public health authorities, produces no further cases.

Although the analysis of (2.1), (2.2), (2.3) is well-known, we provide details here for the sake of completeness. First,  $S(t)$  and  $I(t)$  are nonnegative when  $S_0 > 0$  and  $I_0 > 0$ , which implies  $S(t)$  is decreasing. The *basic reproduction number*  $\mathbb{R}_0$  is defined as  $\tau S_0 / (\nu_1 + \nu_2)$ . If  $\mathbb{R}_0 < 1$ , then  $I(t)$  is decreasing. If  $\mathbb{R}_0 > 1$ , then  $I(t)$  is initially increasing and then decreasing. Add (2.1) and (2.2) to obtain

$$S(t) + I(t) + (\nu_1 + \nu_2) \int_0^t I(s) ds = S_0 + I_0, \quad (2.4)$$

which implies  $I(t) < \infty$  for  $t \geq 0$ , and  $\int_0^\infty I(t) dt < \infty$ . Also, (2.1) implies

$$\log \left( \frac{S(t)}{S_0} \right) = -\tau \int_0^t I(s) ds \iff S(t) = S_0 \exp \left( -\tau \int_0^t I(s) ds \right). \quad (2.5)$$

Since  $S(t)$  is decreasing and  $\int_0^\infty I(t) dt < \infty$  by (2.4),  $\lim_{t \rightarrow \infty} S(t) = S_\infty > 0$ . Then (2.4) implies  $\lim_{t \rightarrow \infty} I(t) = 0$ . Further, (2.4) and (2.5) imply

$$S_\infty = S_0 + I_0 + \frac{\nu_1 + \nu_2}{\tau} \log \left( \frac{S_\infty}{S_0} \right). \quad (2.6)$$

It is possible to use the data of cumulative reported cases to determine the total number of cases  $S(0) - S_\infty$  over the course of the epidemic, as well as the parameters  $\tau$ ,  $\nu_1$ , and  $\nu_2$ . The cumulative number of reported cases at time  $t$  is

$$CR(t) = \nu_1 \int_0^t I(s) ds = \frac{\nu_1}{\tau} \log \left( \frac{S_0}{S(t)} \right),$$

and the total cumulative number of reported cases at the end of the epidemic is

$$CR_\infty = \lim_{t \rightarrow \infty} CR(t) = \frac{\nu_1}{\tau} \log \left( \frac{S_0}{S_\infty} \right), \quad (2.7)$$

which implies

$$S_\infty = S_0 \exp \left[ -\frac{\tau}{\nu_1} CR_\infty \right]. \quad (2.8)$$

Then (2.6) implies

$$S_0 \exp \left[ -\frac{\tau}{\nu_1} CR_\infty \right] = S_0 + I_0 - \frac{\nu_1 + \nu_2}{\nu_1} CR_\infty. \quad (2.9)$$

The cumulative number of both reported and unreported cases at time  $t$  is  $C(t) = S_0 + I_0 - S(t)$ , and the cumulative number of unreported cases at time  $t$  is  $CU(t) = C(t) - CR(t)$ . Since the total number of cases (the epidemic *final size*) is

$$CR_\infty + CU_\infty = S_0 + I_0 - S_\infty,$$

the total number of unreported cases is

$$CU_\infty = S_0 + I_0 - S_\infty - CR_\infty.$$

The epidemic *attack ratio* (defined as the fraction of the susceptible population that becomes infected) is

$$(S_0 + I_0 - S_\infty)/(S_0 + I_0) = (CR_\infty + CU_\infty)/(S_0 + I_0).$$

By symmetry, the ratio of reported cases to unreported cases is

$$\frac{CR(t)}{CU(t)} = \frac{\nu_1}{\nu_2}, \text{ since } CR(t) = \nu_1 \int_0^t I(s) ds \text{ and } CU(t) = \nu_2 \int_0^t I(s) ds.$$

Assume that the reported cases have a maximum at time  $tp$ , which we call the *turning point*. Since (2.2) implies  $\nu_1 I'(tp) = \nu_1 (\tau S(tp) - (\nu_1 + \nu_2) I(tp)) = 0$ , we obtain  $S(tp) = (\nu_1 + \nu_2)/\tau$ . From (2.4 and (2.5) we obtain

$$CR(tp) = \frac{\nu_1}{\tau} \log \left( \frac{\tau S_0}{\nu_1 + \nu_2} \right) \iff \frac{\nu_1 + \nu_2}{\tau} = S_0 \exp \left[ -\frac{\tau}{\nu_1} CR(tp) \right], \quad (2.10)$$

$$S_0 + I_0 - \left( \nu_1 I(tp) \right) \left( \frac{1}{\nu_1} \right) = \frac{\nu_1 + \nu_2}{\tau} \left( 1 - \log \left( \frac{\nu_1 + \nu_2}{\tau S_0} \right) \right). \quad (2.11)$$

### 3 Analysis of Equations (2.9), (2.10), (2.11)

The Equations (2.9), (2.10), (2.11) provide an algorithm for identifying the parameters  $S_0$ ,  $I_0$  and  $\tau$ ,  $\nu_1$  and  $\nu_2$ . Assume that  $tp$ ,  $CR'(tp)$ ,  $CR(tp)$ , and  $CR_\infty$  are known. The algorithm is based on the following equation:

$$S_0 + I_0 = S_0 \exp \left[ -\frac{\tau}{\nu_1} CR_\infty \right] + \frac{\tau}{\nu_1} S_0 \exp \left[ -\frac{\tau}{\nu_1} CR(tp) \right] CR_\infty. \quad (3.1)$$

Equation (3.1) is obtained by using equation (2.10) to obtain

$$\nu_1 + \nu_2 = \tau S_0 \exp \left( -\frac{\tau}{\nu_1} CR(tp) \right),$$

and then substituting  $\nu_1 + \nu_2$  into equation (2.9). Re-write (3.1) as

$$\exp\left[-\frac{\tau}{\nu_1}CR_\infty\right] + \frac{\tau}{\nu_1} \exp\left[-\frac{\tau}{\nu_1}CR(tp)\right]CR_\infty - 1 = \frac{I_0}{S_0}. \quad (3.2)$$

Equation (3.2) is equivalent to

$$F(X) = \frac{I_0}{S_0}, \text{ where } F(X) = e^{-cX} + cXe^{-rcX} - 1, X \geq 0, \quad (3.3)$$

with  $X = \frac{\tau}{\nu_1}$ , and  $c = CR_\infty$ , and  $r = \frac{CR(tp)}{CR_\infty}$ .

**Proposition 3.1** *Let  $c > 0$  and  $r \in (0, 1)$ . We have the following alternative:*

- i) If  $2r \geq 1$ , then  $F(X) \leq 0, \forall X \geq 0$ .*
- ii) If  $2r < 1$ , then  $Y_{\max} := \sup_{X \geq 0} F(X) > 0$ . Moreover there exists a unique  $X_{\max} > 0$  such that  $Y_{\max} := F(X_{\max})$ , and  $F$  is strictly increasing on  $(0, X_{\max})$  and strictly decreasing on  $(X_{\max}, +\infty)$ . Furthermore, for each  $Y_0 \in (0, Y_{\max})$  there exists  $X^* \in (0, X_{\max})$  and  $X^{**} \in (X_{\max}, +\infty)$  such that  $F(X^*) = F(X^{**}) = Y_0$ .*

*Proof.* Observe that  $F(0) = 0$  and  $\lim_{X \rightarrow +\infty} F(X) = -1$ . Moreover

$$\begin{aligned} F'(X) &= -ce^{-cX} + ce^{-crX} - c^2rXe^{-crX} \\ &= ce^{-crX} (1 - e^{-c[1-r]X} - crX), \end{aligned}$$

and

$$\begin{aligned} F''(X) &= c^2e^{-cX} - 2c^2re^{-crX} - c^3r^2Xe^{-crX} \\ &= c^2e^{-crX} (e^{-c(1-r)X} - 2r + cr^2X). \end{aligned}$$

Therefore,

$$F'(0) = 0 \text{ and } F''(0) = c^2(1 - 2r). \quad (3.4)$$

Define  $G(X) := 1 - e^{-c[1-r]X} - crX$ , and we then have

$$F'(X) = ce^{-crX}G(X), \quad (3.5)$$

and  $F'(X) = 0 \iff G(X) = 0$ . Thus,  $G(0) = 0$ ,  $\lim_{X \rightarrow +\infty} G(X) = -\infty$ , and  $G'(X) = c[1-r]e^{-c[1-r]X} - cr$ . By assumption  $r < 1$ , so the mapping  $X \rightarrow G'(X)$  is strictly decreasing,  $G'(0) = c(1 - 2r)$ , and  $\lim_{X \rightarrow +\infty} G'(X) = -cr < 0$ .

*Proof of i).* Assume that  $1 - 2r \leq 0$ . Then  $G'(0) = c(1 - 2r) \leq 0$ , and since  $G'(x)$  is strictly decreasing,  $G'(X) < 0, \forall X > 0$ . But  $G(X) = 0$ , therefore  $G(X) < 0, \forall X > 0$ . By using (3.5), we deduce that  $F'(X) < 0, \forall X > 0$  and *i)* follows from the fact that  $F(0) = 0$ .

*Proof of ii).* Assume that  $1 - 2r > 0$ . Then  $G'(0) = c(1 - 2r) > 0$ , and the mapping  $G$  has exactly one point  $X_{\max} > 0$  such that  $G(X_{\max}) = 0$ , and

$$G(X) > 0, \forall X \in (0, X_{\max}), \text{ and } G(X) < 0, \forall X > X_{\max}. \quad (3.6)$$

By using again (3.5), we deduce that  $X_{\max} > 0$  is a critical point of  $F(X)$  (i.e.  $F'(X_{\max}) = 0$ ), and *ii*) follows from (3.5) and (3.7). ■

**Remark 3.2** *Proposition 3.1 implies that the cumulative reported case data  $CR(tp)$  and  $CR_{\infty}$  are compatible with the solution of model (2.1), (2.2) only if  $CR(tp)/CR_{\infty} < 1/2$ . The condition  $2CR(tp) < CR_{\infty}$  means that more than half of the reported cases will be reported after the turning point. The multiplier  $\nu_1$  in the definition of  $CR(t) = \nu_1 \int_0^t I(s)ds$  can be replaced by  $\nu_2$  or  $\nu_1 + \nu_2$ , and the turning point is unchanged. Then, the cumulative unreported cases satisfy  $2CU(tp) < CU_{\infty}$  (where  $CU(t) = \nu_2 \int_0^t I(s)ds$ ), and all the cumulative cases satisfy  $2CT(tp) < CT_{\infty}$  (where  $CT(t) = (\nu_1 + \nu_2) \int_0^t I(s)ds$ ). Thus,  $CR(t)$ ,  $CU(t)$ ,  $CT(t)$  all have the same the turning point  $tp$ . Illustrations of Proposition 3.1 are given in Figures 1 and 2.*

The algorithm to determine the initial conditions  $S_0, I_0$  and the parameters  $\tau, \nu_1, \nu_2$  from the reported case data is as follows:

(1) Assume the reported case data  $\nu_1 I(t)$  and the turning point  $tp$  are known, and  $2CR(tp) < CR_{\infty}$ . Define  $F(X) = I_0/S_0$  as in (3.3) (where  $F$  depends only on  $CR(tp)$ ,  $CR_{\infty}$ , and not on  $S_0, I_0, tp$ ).

(2) The value of the ratio  $I_0/S_0$  in (3.3) must be adjusted so that the model simulation of (2.1), (2.2) gives agreement with the reported case data and its turning point  $tp$ . Let trial values of  $S_0$  and  $I_0$  be given. Then  $F(X) = I_0/S_0$  has two positive solutions  $X = X^*$  and  $X^{**}$ , provided  $I_0/S_0 < \sup_{X \geq 0} F(X)$ .

(3) Set either  $X^*$  or  $X^{**}$  to  $\tau/\nu_1$ , and use (2.10), (2.11) to obtain  $\nu_1, \tau$ , and  $\nu_2$ , for each from

$$\nu_1 = \frac{CR'(tp)}{S_0 + I_0 - S_0 \exp\left(-\frac{\tau}{\nu_1} CR(tp)\right) \left(1 + \frac{\tau}{\nu_1} CR(tp)\right)}, \quad (3.7)$$

$$\tau = \left(\frac{\tau}{\nu_1}\right) \nu_1, \quad (3.8)$$

$$\nu_2 = \tau S_0 \exp\left(-\frac{\tau}{\nu_1} CR(tp)\right) - \nu_1. \quad (3.9)$$

(4) One value of  $X^*$  or  $X^{**}$  for  $\tau/\nu_1$  will yield parameters  $\tau, \nu_1, \nu_2$  such that the corresponding numerical simulation of (2.1), (2.2) has a graph with the shape of the reported case data, but possibly not the same turning point  $tp$  of the data. The value of the ratio  $I_0/S_0$  must be adjusted so that the value of  $X^*$  or  $X^{**}$  for  $\tau/\nu_1$  will yield parameters  $\tau, \nu_1, \nu_2$  such that the corresponding numerical simulation has a graph with the shape of the reported case data and the same turning point as the data.

**Remark 3.3** We note that when  $I_0$  is small relative to  $S_0$ , the turning point of the model simulation can be approximated as a decreasing function of the ratio  $I_0/S_0$  according to the following formula:

$$tp(k) \approx tp \frac{1 + \log(kI_0/S_0)}{1 + \log(I_0/S_0)},$$

where  $tp$  is the turning point obtained by the algorithm for  $I_0/S_0$  and  $tp(k)$  is the turning point obtained by the algorithm for  $kI_0/S_0$  with  $k > 0$ .

**Remark 3.4** For  $I_0/S_0 = (cI_0)/(cS_0)$ ,  $c > 0$  given, the algorithm yields parameters  $\tau(c), \nu_1(c), \nu_2(c)$  that depend on the scaling factor  $c$ , but with turning point independent of  $c$ . Further,  $S_0c\nu_1(c)$ ,  $S_0c\tau(c)$ ,  $\nu_1(c) + \nu_2(c)$ , and  $\mathbb{R}_0(c)$  are independent of  $c$ . More precisely, assume that  $S_0 = c\hat{S}_0$  and  $I_0 = c\hat{I}_0$  for some constant  $c > 0$ , where  $\hat{S}_0 > 0$ ,  $\hat{I}_0 > 0$ , and  $CR(t)$  are given. Then, by solving (3.2), we deduce that  $\frac{\tau(c)}{\nu_1(c)}$  is independent of  $c$ . By using (3.7), we deduce that  $\nu_1(c) = \frac{\hat{\nu}_1}{c}$ , where

$$\hat{\nu}_1 := \frac{CR'(tp)}{\hat{S}_0 + \hat{I}_0 - \hat{S}_0 \exp\left(-\frac{\tau(c)}{\nu_1(c)}CR(tp)\right)\left(1 + \frac{\tau(c)}{\nu_1(c)}CR(tp)\right)}$$

is independent of  $c$ . By (3.8)  $\tau(c) = \frac{\tau(c)}{\nu_1(c)}\nu_1(c)$ , and we obtain  $\tau(c) = \frac{\hat{\tau}}{c}$  (where  $\hat{\tau} := \frac{\tau(c)}{\nu_1(c)}\hat{\nu}_1$  is independent of  $c$ ). Thus,  $\tau S_0 = \hat{\tau}\hat{S}_0$  is independent of  $c$ , and by using (3.9) we deduce that  $\nu_2(c) = \eta - \frac{\hat{\nu}_1}{c}$  with  $\eta := \tau S_0 \exp(-\frac{\tau}{\nu_1}CR(tp))$ , is independent of  $c$ . Now by replacing the values for  $\tau(c), \nu_1(c), \nu_2(c)$  with these formulas in (2.1), (2.2), we obtain

$$\begin{cases} S'(t) = -\frac{\hat{\tau}}{c}S(t)I(t) \\ I'(t) = \frac{\hat{\tau}}{c}S(t)I(t) - \eta I(t) \\ S(0) = c\hat{S}_0, I(0) = c\hat{I}_0. \end{cases} \quad (3.10)$$

By setting  $\hat{S}(t) := \frac{S(t)}{c}$  and  $\hat{I}(t) := \frac{I(t)}{c}$  we obtain

$$\begin{cases} \hat{S}'(t) = -\hat{\tau}\hat{S}(t)\hat{I}(t) \\ \hat{I}'(t) = \hat{\tau}\hat{S}(t)\hat{I}(t) - \eta\hat{I}(t) \\ \hat{S}(0) = \hat{S}_0, \hat{I}(0) = \hat{I}_0. \end{cases} \quad (3.11)$$

Moreover, since  $\hat{I}(t) = \frac{I(t)}{c}$ , we obtain  $\hat{I}'(t) = 0 \Leftrightarrow I'(t) = 0$ . Since (3.10) is independent of  $c$ , we deduce that the turning point of (3.10) is the same as the turning point of (3.11), which is independent of  $c$ .

From (2.10)  $(\nu_1(c) + \nu_2(c))/(S_0 c \tau(c)) = \mathbb{R}_0(c) = \exp[-\tau(c)CR(tp)/\nu_1(c)]$ , independently of  $c$ . Thus,  $\mathbb{R}_0(c)$  is independent of  $c$ . Further, (2.11) (divided by  $S_0 c$ ) implies

$$1 + \frac{I_0 c}{S_0 c} - \frac{\nu_1(c)I(tp)}{S_0 c \nu_1(c)} = \left( \frac{\nu_1(c) + \nu_2(c)}{S_0 c \tau(c)} \right) \left( 1 - \log \left( \frac{\nu_1(c) + \nu_2(c)}{S_0 c \tau(c)} \right) \right),$$

which yields  $S_0 c \nu_1(c)$  independent of  $c$ . Then,  $S_0 c \tau(c) = S_0 c \nu_1(c) \tau(c)/\nu_1(c)$  and  $\nu_1(c) + \nu_2(c) = S_0 c \tau(c) \mathbb{R}_0(c)$  are independent of  $c$ .

**Remark 3.5** *The total number of initial cases  $I_0$  may be known, both reported and unreported, but the initial number of susceptible individuals  $S_0$  may not be known. This is the case if a significant fraction of the demographic population has acquired immunity from a prior infection. The value of  $S_0$  can be obtained from the algorithm by varying  $S_0$  and comparing the model output to the reported case data and its turning point  $tp$ . The turning point in the model simulation, with  $I_0$  fixed, increases approximately linearly with  $\log(S_0)$ , so the correct value of  $S_0$  can be matched to the turning point in the model and the data. An illustration is given in Figure 10.*

**Remark 3.6** *The initial number of reported cases may be known from the data, but the initial number of all cases  $I_0$ , including unreported cases, may not be known. If  $S_0$  is known, however, the algorithm can be extended to determine  $I_0$ : vary  $I_0$  in the algorithm and compare the model output to the reported case data and its turning point  $tp$ . The turning point in the model simulation, with  $S_0$  fixed, decreases approximately linearly with  $\log(I_0)$ , so the correct value of  $I_0$  can be matched to the turning point of the data. An illustration is given in Figure 10.*

**Remark 3.7** *In practice neither  $I_0$  nor  $S_0$  may be known. The algorithm can be used in this case by varying the ratio  $I_0/S_0$ . As  $I_0/S_0$  increases, the turning point in the model simulation decreases. The correct value of the ratio  $I_0/S_0$  can be identified by matching the model turning point to the data and its turning point  $tp$ . For this value of  $I_0/S_0$  any scaling  $cI_0/cS_0$  of this ratio will also match the data and its turning point  $tp$ . For  $c$  given,  $\mathbb{R}_0$  is independent of  $c$ , and the ratio  $\nu_2(c)/\nu_1(c)$  is a linearly increasing function of  $c$ . Thus, the scaling factor  $c$  can be varied and the values of  $cS_0, cI_0, \tau(c), \nu_1(c), \nu_2(c)$  can be compared to known information about the epidemic to obtain a realistic value of  $c$ .*

## 4 Examples of the Algorithm for Identifying the Model Parameters

**Example 1.** The Hong Kong influenza epidemic in New York City in 1968-1969.

We apply the parameter identification algorithm to the Hong Kong influenza epidemic in New York City in 1968-1969. The reported case data consist of



weekly reported mortality cases in excess of typical reported mortality (assumed to be due to the influenza epidemic) for the same time period in previous years [53]. The cumulative mortality is obtained by adding these weekly values over the 13 week duration of the epidemic. The time units are weeks. We take  $S(0) = 7,900,000$  (the population of New York City in the 1970 census) and  $I(0) = 15,000$ . The weekly reported data and the cumulative reported excess mortality data are illustrated by the black dots in Figure 3. From these data we estimate  $CR_\infty \approx 1,080$ ,  $tp \approx 6.15$  weeks,  $CR(tp) \approx 500$ , and  $\nu_1 I(tp) = 190$ . Notice that  $2CR(tp) < CR_\infty$ .

The two solutions of  $F(X) = I_0/S_0$ , with  $F(X)$  given in (3.3), are  $X = 0.000276$  and  $X = 0.000728$ . The solution that gives agreement with the data is  $X = 0.000728$ .  $F(X)$  has its maximum at  $F(0.000524) \approx 0.00332$ . If  $I_0 > 26,220$ , then  $I_0/S_0 > 0.00332$ , and the reported case data are not consistent with these initial conditions. For  $X = 0.000728$ , the numerical solutions of equations (2.9),(2.10),(2.11) are  $\tau \approx 3.24 \times 10^{-7}$ ,  $\nu_1 \approx 0.00044$ ,  $\nu_2 \approx 1.78$ , which corresponds to an average infectious period in non-fatal cases of  $\approx 1/1.78$  weeks  $\approx 4$  days, and a ratio of unreported to reported cases of  $\approx 4045$  to 1. The graphs of the reported data and model reported cases output are given in Figure 3. The epidemic peak size is  $I(tp) \approx 198,000$  at  $tp \approx 6.15$  weeks. From model output we obtain  $S_\infty \approx 3,620,000$ . The epidemic final size is  $S_0 + I_0 - S_\infty \approx 4,295,000$ . The epidemic attack ratio is  $(S_0 + I_0 - S_\infty)/(S_0 + I_0) \approx 54\%$ . The basic reproduction number  $\mathbb{R}_0 \approx 1.44$ . The graphs of  $S(t)$ ,  $I(t)$ ,  $R(t)$  are given in Figure 4, and the graph of  $F(X)$  is given in Figure 5(a). The value of  $\nu_1 I(t)$  obtained from the numerical solution of model, with parameters obtained from  $X = 0.000276$  as the solution of  $F(X) = I_0/S_0$ , is graphed in Figure 5(b), and is not in agreement with the reported data.

The case fatality ratio  $CR_\infty/(S_0 + I_0 - S_\infty) = 1,080/4,295,000 \approx 0.00025 = 0.025\%$  is lower than commonly claimed values for seasonal influenza, typically 0.1 to 0.2 % [38]. If the initial susceptible population  $S_0$  is reduced, then the fatality ratio is increased proportionately.

**Example 2.** The plague epidemic in Bombay, India in 1906.

We apply the parameter identification algorithm to the bubonic plague epidemic in Bombay, India in 1906. This epidemic has been modeled many times, beginning with one of the most famous works in mathematical epidemiology, due to Kermack and McKendrick [34]. These reported case data consist of weekly reported mortality cases over a period of 30 weeks beginning in January, 1906. These data of reported mortality cases are available at <https://www.math.psu.edu/treluga/misc.html>, and are graphed Figure 6(a) (black dots). The population of Bombay in 1906 was approximately 1,000,000, but not all residents had equal likelihood of infection. Infection did not spread from person to person, but was strongly associated with shared dwellings, workplaces, or other localities [55]. Additionally, many residents fled the city during the seasonal plague epidemics of that time period, as for example, 850,000 in 1896 (<http://en.wikipedia.org/wiki/Mumbai>). The case-fatality rate was as high

as 90%, and the typical duration of a fatal infection was  $\approx 5$  days with an incubation period of  $\approx 3$  days [5].

For the model (2.1),(2.2) we take  $S(0) = 100,000$  and  $I(0) = 8$ . The cumulative reported mortality data are illustrated by the black dots in Figure 6(b). From these data we estimate  $CR_\infty \approx 8,840$ ,  $tp \approx 13.5$  weeks,  $CR(tp) \approx 4,330$ , and  $\nu_1 I(tp) = 770$ . Notice that  $2CR(tp) < CR_\infty$ . The two solutions of  $F(X) = I_0/S_0$ , with  $F(X)$  given in (3.3), are  $X = 0.0000159$  and  $X = 0.0000204$ . The solution that gives agreement with the data is  $X = 0.0000159$ .

For  $X = 0.0000159$  the numerical solutions of (2.9),(2.10),(2.11) are  $\tau \approx 5.2 \times 10^{-5}$ ,  $\nu_1 \approx 3.3$ ,  $\nu_2 \approx 1.6$ , which correspond to an average infectious period of  $\approx 1/4.9$  weeks  $\approx 1.4$  days, and a ratio of unreported to reported cases of  $\approx 0.48$  to 1. From (2.8) we estimate  $S_\infty \approx 87,000$  with corresponding total cases  $\approx 13,000$ , an attack ratio  $\approx 13\%$ , and a reported removal rate of  $\approx 70\%$  (the total mortality rate would be increased by unreported removal cases). The epidemic peak size is  $I(tp) \approx 235$  at  $tp \approx 13.5$  weeks. The basic reproduction number  $\mathbb{R}_0 \approx 1.07$ . The graph of the model output of the weekly reported cases  $\nu_1 I(t)$  is given in Figure 6(a). The graph of model output of cumulative weekly reported cases is given in Figure 6(b).

**Example 3.** The seasonal influenza epidemic in Puerto Rico in 2016-2017.

We apply the parameter identification algorithm to the seasonal influenza epidemic in Puerto Rico in 2016-2017. The reported case data consist of weekly reported cases from Departamento de Salud, Puerto Rico [57] from week 36 in 2016 to week 23 in 2017 (Figure 7). The cumulative reported cases are obtained by adding these weekly values over the 37 weeks duration of the epidemic. We take  $S(0) = 2,500,000$  and  $I(0) = 57$ . The 2010 census of Puerto Rico was  $\approx 3,500,000$ , so we assume  $\approx 1/3$  of the population has acquired immunity. The weekly reported data and the cumulative reported data are illustrated by the black dots in Figure 8. From this data we estimate  $CR_\infty \approx 45,300$ ,  $tp \approx 14$  weeks,  $CR(tp) \approx 20,400$ , and  $\nu_1 I(tp) \approx 6,200$ . Notice that  $2CR(tp) < CR_\infty$ .

The two solutions of  $F(X) = I_0/S_0$ , with  $F(X)$  given in (3.3), are  $X = 4.78 \times 10^{-7}$  and  $X = 2.67 \times 10^{-5}$ . The solution that gives agreement with the data is  $X = 2.67 \times 10^{-5}$ , and the numerical solutions of equations (2.9), (2.10) are  $\tau \approx 6.36 \times 10^{-7}$ ,  $\nu_1 \approx 0.024$ ,  $\nu_2 \approx 0.90$ , which correspond to an average infectious period in non-fatal cases of  $\approx 1.1$  weeks, and a ratio of unreported to reported cases of  $\approx 37.7$  to 1. The graphs of the reported data and model output for reported cases are given in Figure 8. From the model output we obtain the epidemic final size  $S_0 + I_0 - S_\infty \approx 1,756,000$ . The attack ratio of the susceptible population is  $(S_0 + I_0 - S_\infty)/(S_0 + I_0) \approx 70\%$ . The attack ratio as a fraction of the total population of Puerto Rico is  $(S_0 + I_0 - S_\infty)/3,500,000 \approx 48.8\%$ . The reported case data include year-round background cases, which may over estimate the reported cases and the attack ratio. The basic reproduction number  $\mathbb{R}_0 \approx 1.73$ . The epidemic peak size is  $I(tp) \approx 237,000$  at  $tp \approx 14$  weeks. The graphs of  $S(t)$ ,  $I(t)$ ,  $R(t)$  are given in Figure 9.

## 5 Discussion

For the SIR model (2.1), (2.2), (2.3) we have constructed an algorithm to identify unreported cases using reported case data  $\nu_1 I(t)$ . This SIR model yields a simple rise and fall of the infected cases over the course of the epidemic, with a turning point at time  $tp$ . A necessary condition for the algorithm is that the data satisfy  $CR(tp)/CR_\infty < 1/2$ . The algorithm uses the cumulative reported case data  $CR(tp)$  at the turning point and cumulative reported case data  $CR_\infty$  at the end of the epidemic to obtain model parameters  $\nu_1, \nu_2$ , and  $\tau$ , that fit the reported case data and data turning point  $tp$ , assuming  $S_0, I_0$  are known.  $S_0$ , however, may not be known, since the number of initially susceptible individuals  $S_0$  may be less than the demographic population of the region due to some individuals having acquired immunity from previous infections.  $I_0$  may not be known, since the number of initially infected individuals  $I_0$  may be much higher than the number of initially reported infected individuals. If one of  $S_0$  and  $I_0$  is known, then the algorithm allows identification of the other using the turning point of the data. If neither  $S_0$  nor  $I_0$  is known, then the ratio  $I_0/S_0$  in the algorithm can be adjusted to identify parameters  $\tau, \nu_1, \nu_2$  that agree with the reported case data and its turning point  $tp$ . In this case any scaling  $cI_0/cS_0$  will yield parameters that also fit the reported case data and turning point. The ratio  $S_0/I_0$  is central in our algorithm for identifying the parameters and initial conditions of the model.

Other methods provide alternative approaches for the parameter identification problem for epidemic models, including formal least squares methods and stochastic methods [2, 7, 24, 38]. Stochastic SIR models account for probabilistic individual behavior, but typically have much greater computational requirements for large population sets. Deterministic SIR models provide only mean distribution outputs, but are typically much more computationally efficient. Our deterministic modeling approach emphasizes the role of the initial conditions  $I_0$  and  $S_0$ , which for many epidemic diseases such as seasonal influenza, are largely unknown. Our approach also emphasizes the role of reported cases, which for diseases such as seasonal influenza, are a very small fraction of the total number of cases.

### 5.1 The SEIR epidemic model

A generalization of the SIR model (2.9), (2.10), (2.11) is the SEIR model, which allows for an incubation period of newly infected individuals before they become infectious. The equations of the SEIR model are

$$\frac{d}{dt}S(t) = -\tau I(t)S(t) \quad (5.1)$$

$$\frac{d}{dt}E(t) = \tau I(t)S(t) - \sigma E(t) \quad (5.2)$$

$$\frac{d}{dt}I(t) = \sigma E(t) - (\nu_1 + \nu_2)I(t) \quad (5.3)$$

$$\frac{d}{dt}R(t) = (\nu_1 + \nu_2)I(t) \quad (5.4)$$

with initial conditions  $S(0) = S_0 > 0$ ,  $E(0) = E_0 \geq 0$ ,  $I(0) = I_0 \geq 0$ ,  $R(0) = R_0 \geq 0$ . Here  $E(t)$  is the number of pre-infectious infected individuals at time  $t$ , the parameters  $\tau$ ,  $\nu_1$ , and  $\nu_2$  are as before, and  $\sigma$  is the rate at which incubating infected individuals become infectious. The *basic reproduction number*  $\mathbb{R}_0$  is  $\tau S_0 / (\nu_1 + \nu_2)$ , the same as the SIR model, and  $S_\infty = \lim_{t \rightarrow \infty} S(t)$  satisfies

$$S_\infty = S_0 + E_0 + I_0 + \frac{\nu_1 + \nu_2}{\tau} \log \left( \frac{S_\infty}{S_0} \right) \quad (5.5)$$

( $\mathbb{R}_0$  and  $S_\infty$  are independent of  $\sigma$ ). If  $\mathbb{R}_0 > 1$ ,  $I(t)$  may first decrease before increasing and then decreasing at a turning point  $tp$  (depending on  $S_0, E_0, I_0$ ). The SIR analysis carries over to a similar analysis for the SEIR model, with equations (2.9), (2.10), (2.11) the same, except that  $I_0$  is replaced by  $I_0 + E_0$ , and  $F(X) = (I_0 + E_0)/S_0$  in (3.3). With the values of  $\tau, \nu_1, \nu_2$  obtained from the modified equations (2.9), (2.10), (2.11), and (3.3), the equations (5.1), (5.2), (5.3), are solved numerically with values of  $\sigma$  chosen so that the output matches the reported case data and the turning point  $tp$ . For seasonal influenza in Puerto Rico in 2016-2017,  $\tau = 4.54 \times 10^{-7}$ ,  $\nu_1 = 0.017$ , and  $\nu_2 = 0.90$  as in the SIR model (Example 3) and  $E_0 = 70$ ,  $I_0 = 70$ , and  $\sigma = 21$ . Thus, the incubation period is  $\approx 1/21$  weeks = 0.33 days, which is consistent with incubation times for influenza in volunteer studies measured by viral shedding in [15].

## 5.2 The general incidence SIR epidemic model

The SIR model (2.9), (2.10), (2.11) does not take into account changes in social behavior or public health policies as the epidemic unfolds, which may reduce transmission. A model that incorporates such change is the general incidence SIR model:

$$\frac{d}{dt}S(t) = -\frac{\tau I(t)^p}{1 + \kappa I(t)^q} S(t) \quad (5.6)$$

$$\frac{d}{dt}I(t) = \frac{\tau I(t)^p}{1 + \kappa I(t)^q} S(t) - \nu I(t) \quad (5.7)$$

$$\frac{d}{dt}R(t) = \nu I(t) \quad (5.8)$$

with initial conditions  $S(0) = S_0 > 0$ ,  $I(0) = I_0 > 0$ ,  $R(0) = R_0 \geq 0$ . The parameter  $\kappa$  accounts for reduced transmission as  $I(t)$  increases, and the positive exponents  $p$  and  $q$  satisfy  $1 \leq p \leq q+1$ . The *basic reproduction number* is  $\mathbb{R}_0 = \tau I_0^{p-1} S_0 / (1 + \kappa I_0^q)$ . If  $R_0 < 1$ , then  $S(t)$  decreases to a limiting value  $S_\infty > 0$  and  $I(t)$  decreases to 0. If  $R_0 > 1$ , then  $S(t)$  decreases to a limiting value  $S_\infty > 0$  and  $I(t)$  first increases, then decreases to 0 [43].

The condition  $CR(tp)/CR_\infty < 1/2$  may be violated for the general incidence model. If the epidemic data do not satisfy this condition, then the SIR model (2.9), (2.10), (2.11) may not be valid for epidemic analysis. A numerical example of a general incidence model not satisfying  $CR(tp)/CR_\infty < 1/2$  is given in Figure 11. Examples of epidemic reported case data not satisfying  $CR(tp)/CR_\infty < 1/2$  are the 2013-2014 and 2014-2015 seasonal influenza epidemics in Puerto Rico [58] illustrated in Figure 12.

### 5.3 More general epidemic models

In many cases epidemic data do not show a simple rise and fall of the reported cases. One example is found in models with multi-group infected populations. In [42] a two-group model, with super-spreaders and ordinary spreaders, yielded a two-peak output of the total infected population for the SARS influenza outbreak in Singapore in 2003. Spatial heterogeneity is also important in modeling the spread of epidemics. One example is given in [23] of the 2015-2016 Zika epidemic in Rio de Janeiro, Brazil. Another example is given in [43] of the geographical spread of the 2015-2016 seasonal influenza epidemic in Puerto Rico, in which the reported infected cases show an early high peak and a later low peak (Figure 7). The multiple peaks in this case can be attributed to spatial variation in the course of epidemic. Models incorporating the disease phase of infected individuals with a continuum disease age variable can track the rise of infectiousness, and in particular the pre-symptomatic periods of infectiousness [10]. Models involving time dependent transmission and removal rates, corresponding to public behavioral changes are also important, and considerations of time dependent parameters were treated by K.P. Haderl in [26, 27] and T. Stader *et al.* in [52]. These examples illustrate the need for extension of the work here to identify parameters in epidemic models that incorporate additional features of outbreak epidemic diseases.

## References

- [1] R.M. Anderson and R.M. May, *Infective Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991.
- [2] H. Andersson and T. Britton, Stochastic epidemic models and their statistical analysis, *Springer Lecture Notes in Statistics*, **151** (2000), Springer-Verlag, New York.
- [3] V. Andreasen, The final size of an epidemic and its relation to the basic reproduction number, *Bull. Math. Biol.*, **73**(10) (2011), 2305-2321.
- [4] 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.
- [5] N. Bacaër, The model of Kermack and McKendrick for the plague epidemic in Bombay and the type reproductive number with seasonality, *J. Math. Biol.*, **64** (2012), 403-422.
- [6] N.T.J. Bailey, *The Mathematical Theory of Epidemics*, Charles Griffin, London, 1957
- [7] N. Becker, *Analysis of Infectious Disease Data*, Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1989.
- [8] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir, *Mém. Math. Phys. Acad. Roy. Sci., Paris* (1760), 1-45.
- [9] M. Biggerstaff and L. Balluz, Self-reported influenza-like illness during the 2009 H1N1 influenza pandemic, *United States, Morbid Mortal Weekly Rep., September 2009-March 2010*, **60** (2011), 60:37.
- [10] M. Blaser, Y-H. Hsieh, G.F. Webb, and J. Wu, Pre-symptomatic influenza transmission, surveillance, and school closings: Implications for novel influenza A(H1N1), *Math. Mod. Nat. Phen.*, **3** (2010), 191-205.
- [11] F. Brauer, P. van den Driessche and J. Wu (eds.) *Mathematical epidemiology*, Springer, Berlin, 2008.
- [12] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, Springer, New York, 2000.
- [13] S. Busenberg and K. Cooke, *Vertically Transmitted Diseases: Models and Dynamics*, Lecture Notes in Biomathematics, **23**, Springer-Verlag, Berlin, 1993.
- [14] 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.

- [15] F. Carrat, E. Vergu, N.M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A-J. Valleron, Time lines of infection and disease in human influenza: A review of volunteer challenge studies, *Amer. J. Epid.*, **167**, (2008), 775-785.
- [16] 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.
- [17] 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.
- [18] O. Diekmann, Hans Heesterbeek, and Tom Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton University Press, Princeton, 2013.
- [19] O. Diekmann, J.A.P. Heesterbeek, and J.A.J. Metz, On the definition and the computation of the basic reproduction ratio  $\mathbb{R}_0$  in models for infectious diseases in heterogeneous populations, *J. Math. Biol.* **28** (1990), 365-382.
- [20] K. Dietz and J.A.P. Heesterbeek, Bernoulli was ahead of modern epidemiology, *Nature* **408** (2000), 513-514.
- [21] K. Dietz and J.A.P. Heesterbeek, Daniel Bernoulli's epidemiological model revisited, *Math. Biosci.*, **180** (2002), 1-21.
- [22] 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.
- [23] W.E. Fitzgibbon, J.J. Morgan, and G.F. Webb, An outbreak vector-host epidemic model with spatial structure: the 2015-2016 zika outbreak in Rio de Janeiro, *Theoret. Biol. Med. Mod.*, **14** (2017), DOI 10.1186/s12976-017-0051.
- [24] K.M. Gamado, G. Streftaris, and S. Zachary, Modelling under-reporting in epidemics, *J. Math. Biol.*, **69** (2014), 737-765.
- [25] N. Grassly and C. Fraser, Seasonal infectious disease epidemiology, *Proc. Roy. Soc. Lond. B: Biol. Sci.*, **273** (2006), 2541-2550.
- [26] K.P. Hadeler, Parameter identification in epidemic models, *Math. Biosci.*, **229** (2011), 185-189.
- [27] K.P. Hadeler, Parameter estimation in epidemic models: simplified formulas, *Can. Appl. Math. Q.*, **19** (2011), 343-356.
- [28] H.W. Hethcote, Qualitative analyses of communicable disease models, *Math. Biosci.*, **28** (1976), 335-356.

- [29] 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.
- [30] H.W. Hethcote, The mathematics of infectious diseases, *SIAM Review*, **42(4)** (2000), 599-653.
- [31] 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.
- [32] 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.
- [33] M. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*, Princeton University Press, Princeton, 2007
- [34] W.O. Kermack and A.G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond. A*, **115** (1927), 700-721.
- [35] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics: II, *Proc. Roy. Soc. Lond. A* **138** (1932), 55-83.
- [36] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics: III, *Proc. Roy. Soc. Lond. A* **141** (1933), 94-112.
- [37] A. Lange, Reconstruction of disease transmission rates: Applications to measles, dengue, and influenza, *J. Theor. Biol.* **400** (2016), 138-153.
- [38] F.C.K. Li, B.C.K. Choi, T. Sly, and A.W.P. Pak, Finding the real case-fatality rate of H5N1 avian influenza, *J. Epid. Comm. Health* **10** (2008), 555-559.
- [39] J. Li and Y. Lou, Characteristics of an epidemic outbreak with a large initial infection size, *J. Biol. Dyn.* **10** (2016), 366-378.
- [40] G. Macdonald, *The Epidemiology and Control of Malaria*, in *Epidemics*, Oxford University Press, London, 1957.
- [41] 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.
- [42] P. Magal, O. Seydi, and G. Webb, Final size of an epidemic for a two-group SIR model, *SIAM J. Appl. Math.* **76**, (2016), 2042-2059.
- [43] P. Magal, G. Webb, and Y. Wu, Spatial Spread of Epidemic Diseases in Geographical Settings: Seasonal Influenza Epidemics in Puerto Rico, (*Submitted*).



- [44] A. Mummert, Studying the recovery procedure for the time-dependent transmission rate(s) in epidemic models, *J. Math. Biol.*, **67** (2013), 483-507.
- [45] J.D. Murray, *Mathematical Biology*, Springer, Berlin, 1993.
- [46] 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.
- [47] 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.
- [48] C. Reed, F.J. Angulo, D.L. Swerdlow, M. Lipsitch, M.I. Meltzer, D. Jernigan, and L. Finelli, Estimates of the prevalence of pandemic (H1N1) 2009, United States, April-July 2009, *Emer. Inf. Dis.*, **15**, No.12 (2009).
- [49] L.I.W. Roeger, Z. Feng and C. Castillo-Chavez, Modeling TB and HIV co-infections, *Math. Biosci. Eng.*, **6**(4) (2009), 815-837.
- [50] R. Ross, *The Prevention of Malaria*, John Murray, London, 1910.
- [51] S.S. Shrestha, D.L. Swerdlow, R.H. Borse, V.S. Prabhu, L. Finelli, C.Y. Atkins, K. Owusu-Edusei, B. Bell, P.S. Mead, M. Biggerstaff, L. Brammer, H. Davidson D. Jernigan, M.A. Jhung, L.A. Kamimoto, T.L. Merlin, M. Nowell, S.C. Redd, C. Reed, A. Schuchat, and M.I. Meltzer, Estimating the burden of 2009 pandemic influenza A (H1N1) in the United States (April 2009-April 2010), *Clin. Infect. Dis.*, **Jan 1;52 Suppl 1:S75-82** (2011).
- [52] T. Stadler, D. Kühnert, S. Bonhoeffer and A.J. Drummond, Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus, *PNAS*, **Jan 2;110** (2013).
- [53] D. Smith and L. Moore, The SIR model for spread of disease - background: Hong Kong flu, *J. Online Math. Appl.*, **Dec.** (2004).
- [54] H.R. Thieme, *Mathematics in Population Biology*, Princeton University Press, Princeton, 2003.
- [55] J.A. Thompson, On the epidemiology of plague, *J. Hygiene*, **6** (1906), 537-569.
- [56] P. Van den Driessche and J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.* **180** (2002) 29-48.
- [57] [www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Estadisticas Influenza/Informe Influenza Semana 26-2017.pdf](http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Estadisticas%20Influenza/Informe%20Influenza%20Semana%2026-2017.pdf), (2017).
- [58] [www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Estadisticas Influenza/Informe Influenza Semana 39-2015.pdf](http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Estadisticas%20Influenza/Informe%20Influenza%20Semana%2039-2015.pdf), (2015).

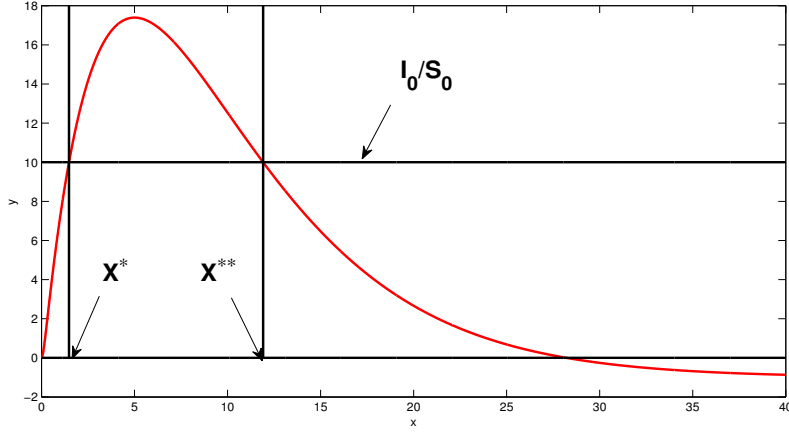


Figure 1: The graph of  $F(X)$  (red), and the two solutions  $X^*$  and  $X^{**}$  of  $F(X) = I_0/S_0$ . One of the two solutions provides a correct numerical simulation of (2.9), (2.10) for the reported case data and their turning point  $tp$ .

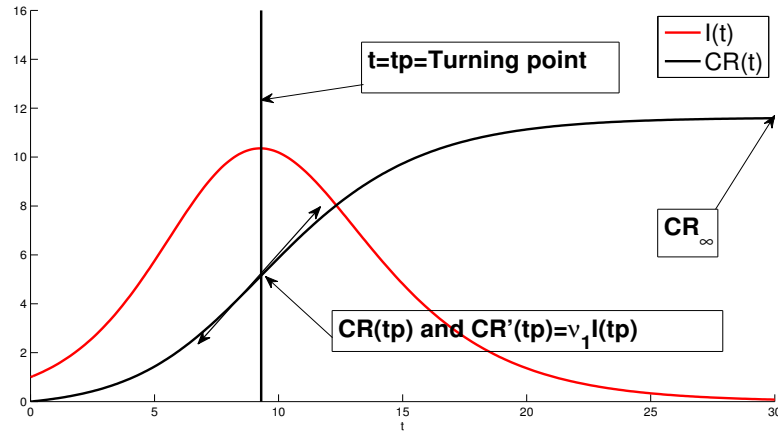


Figure 2: The graph of all cases (both reported and unreported)  $I(t)$  (red) and the graph of cumulative reported cases  $CR(t)$  (black), and their relationship at the turning point  $tp$ .  $I'(tp) = 0$  and  $CR''(tp) = 0$ .

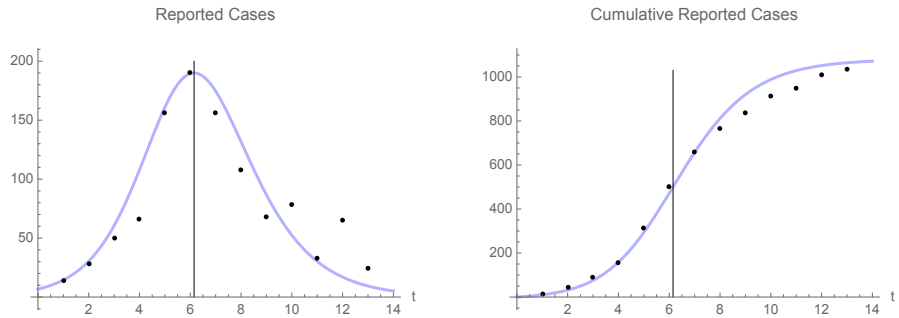


Figure 3: Hong Kong influenza epidemic in New York City in 1968-1969. The weekly reported mortality case data and cumulative reported case data (black dots), and the model output graphs of  $\nu_1 I(t)$  and  $CR(t)$  (blue).

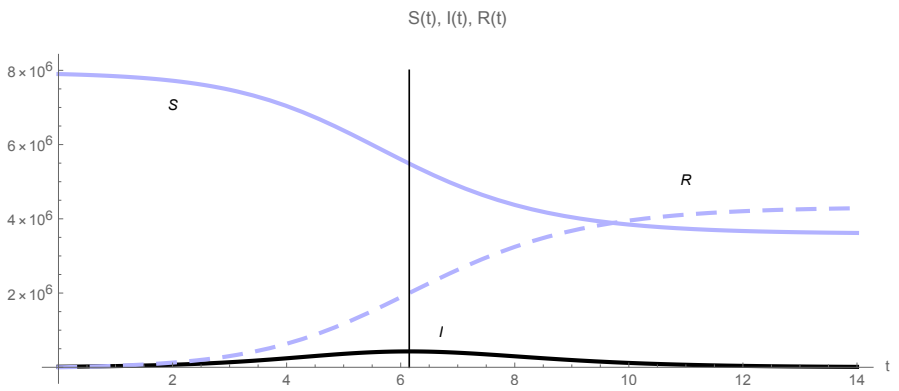


Figure 4: Hong Kong influenza epidemic in New York City in 1968-1969. The model output graphs of  $S(t)$  (blue),  $I(t)$  (black), and  $R(t)$  (dashed). The vertical lines are at the epidemic turning point  $tp \approx 6.15$  weeks.

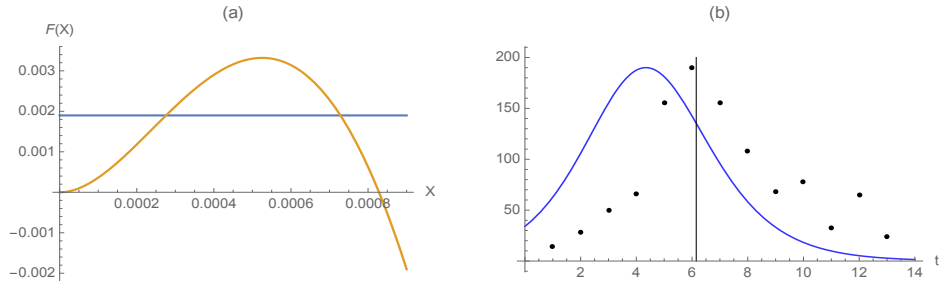


Figure 5: *Hong Kong influenza epidemic in New York City in 1968-1969. (a) The graphs of  $F(X)$  (yellow) and  $I_0/S_0$  (blue). There exists two solutions  $X = 0.000276$  and  $X = 0.000728$  of  $F(X) = I_0/S_0$ . The solution corresponding to the data is  $X = 0.000728$ . (b) The graph of the model output of the cumulative weekly reported case data for the solution  $X = 0.000276$ , which does not match the epidemic data.*

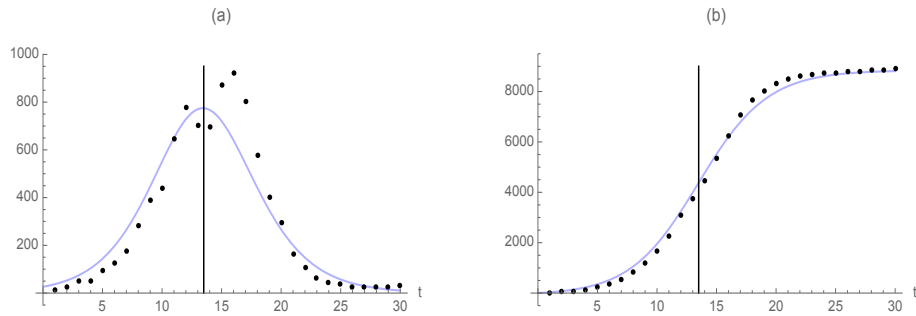


Figure 6: *Bubonic plague epidemic in Bombay, India in 1906. (a) The weekly reported mortality case data (black dots) and model output graph of  $\nu_1 I(t)$  (blue). (b) The weekly reported cumulative mortality case data (black dots) and model output graph of  $CR(t)$  (blue). The vertical lines are at the epidemic turning point  $t_p \approx 13.5$  weeks.*

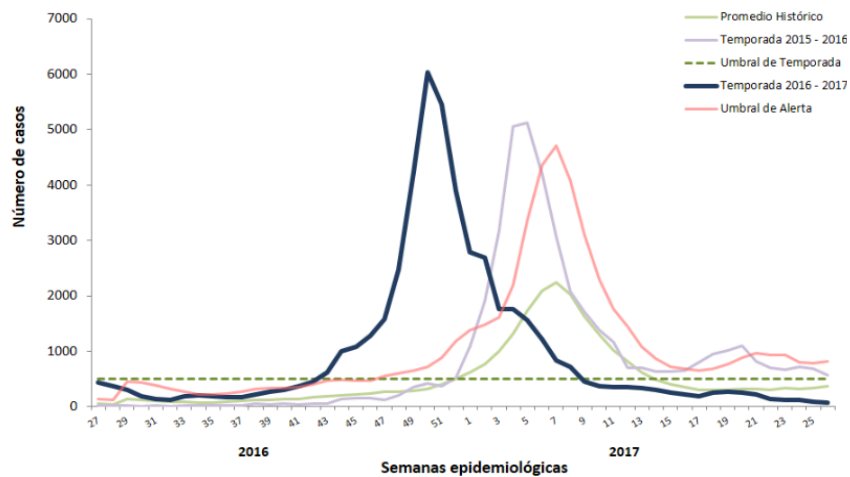


Figure 7: Weekly reported cases of seasonal influenza epidemics in Puerto Rico in the 2015-2016 and 2016-2017 seasons. The graph of reported cases in 2015-2016 (yellow) has two peaks. The graph of reported cases in 2016-2017 (black) has only one peak and satisfies  $CR(tp)/CR_\infty < 1/2$ .

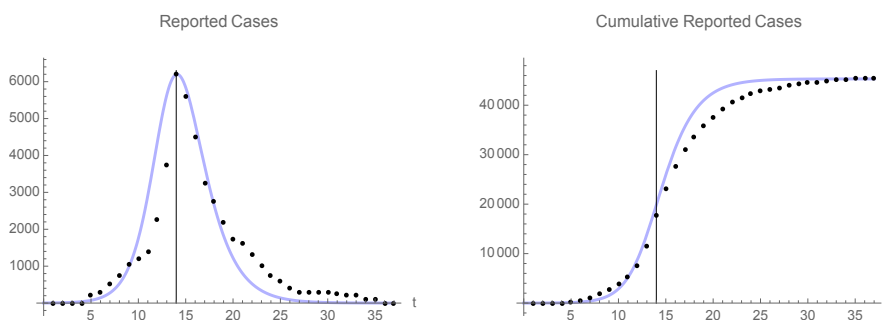


Figure 8: Seasonal influenza epidemic in Puerto Rico in 2016-2017. The weekly reported mortality case data and cumulative reported case data (black dots), and the model output graphs of  $\nu_1 I(t)$  and  $CR(t)$  (blue). The vertical lines are at the epidemic turning point  $tp \approx 14$  weeks.

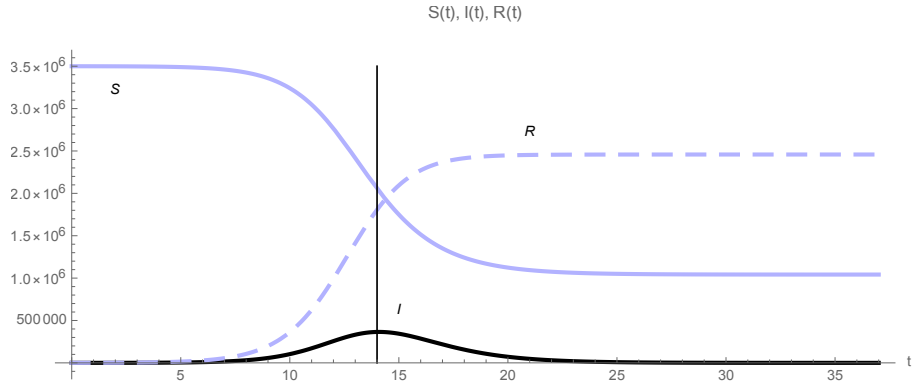


Figure 9: Seasonal influenza epidemic in Puerto Rico in 2016-2017. The model output graphs of  $S(t)$  (blue),  $I(t)$  (black), and  $R(t)$  (dashed).

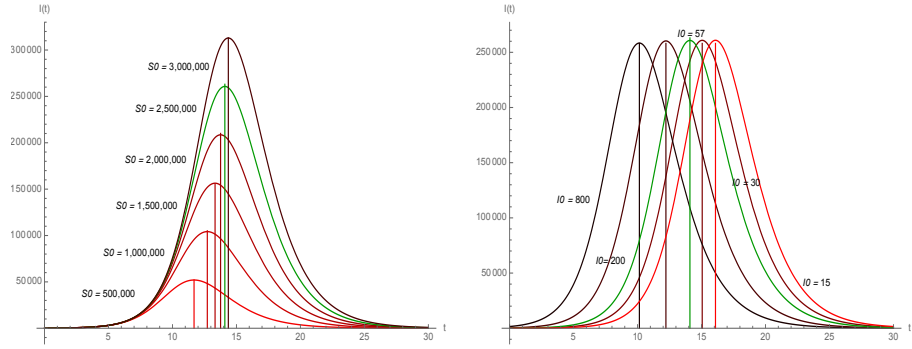


Figure 10: The relationship of  $I_0/S_0$  for the simulation of the 2016-2017 seasonal influenza epidemic in Puerto Rico with model (2.1), (2.2). The algorithm yields the correct turning point  $tp = 14$  for the data with  $I_0/S_0 = 2.28 \times 10^{-5}$ . Left side:  $I_0 = 57$ ,  $tp$  increases linearly with  $\log(S_0)$  for  $I_0/S_0 = 2.28 \times 10^{-5}$ . Right side:  $S_0 = 2,500,000$ ,  $tp$  decreases linearly with  $\log(I_0)$  for  $I_0/S_0 = 2.28 \times 10^{-5}$ . The green graphs correspond to  $S_0 = 2,500,000$  and  $I_0 = 57$ . For  $cI_0/cS_0$ ,  $c \geq 1$ ,  $R_0 \approx 1.73$  independently of  $c$ , and  $\nu_1/\nu_2 \approx 38.7c - 1.0$ .

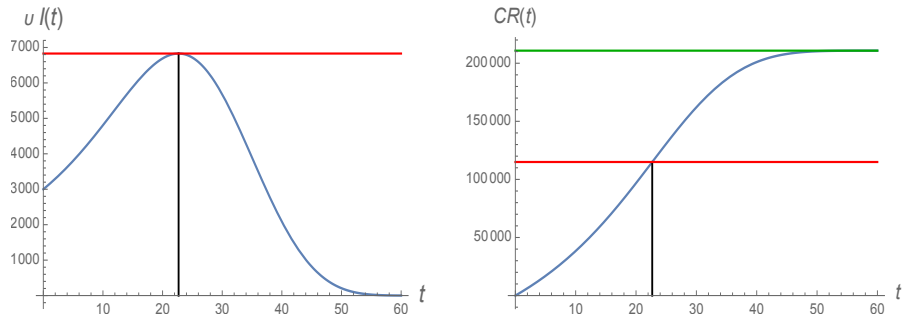


Figure 11: The general incidence model (5.6), (5.7) with  $\tau = 0.14 \times 10^{-6}$ ,  $\nu = 3.0$ ,  $p = 1.09$ ,  $q = .1$ ,  $S_0 = 3,500,000$ ,  $I_0 = 1,000$ .  $\nu$  corresponds to all cases being reported.  $\mathbb{R}_0 = 1.105$ ,  $tp = 22.74$ ,  $\nu I(tp) = 6,829$ ,  $CR(tp) = 114,984$ ,  $CR_\infty = 210,828$ . The cumulative reported cases do not satisfy  $CR(tp)/CR_\infty < 1/2$ .

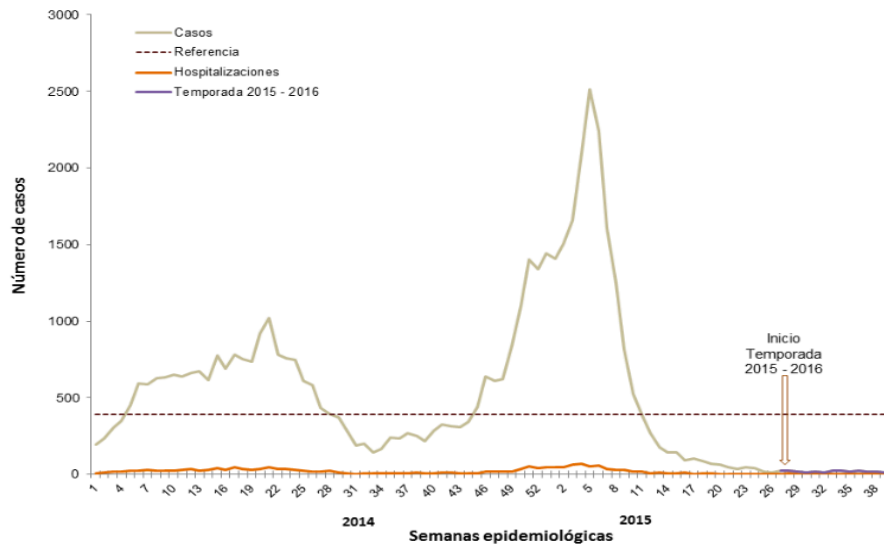


Figure 12: Weekly reported cases of seasonal influenza epidemics in Puerto Rico in the 2013-2014 and 2014-2015 seasons. The cumulative reported cases (corresponding to the areas below the graphs) do not satisfy  $CR(tp)/CR_\infty < 1/2$  for either season.