Modeling Super-spreading Events for Infectious Diseases: Case Study SARS

Thembinkosi Mkhatshwa   Anna Mummert

Abstract—Super-spreading events for infectious diseases occur when some infected individuals infect more than the average number of secondary cases. Several super-spreading individuals have been identified for the 2003 outbreak of severe acute respiratory syndrome (SARS). We develop a model for super-spreading events of infectious diseases, which is based on the outbreak of SARS. Using this model we describe two methods for estimating the parameters of the model, which we demonstrate with the small-scale SARS outbreak at the Amoy Gardens, Hong Kong, and the large-scale outbreak in the entire Hong Kong Special Administrative Region. One method is based on parameters calculated for the classical susceptible - infected - removed (SIR) disease model. The second is based on parameter estimates found in the literature. Using the parameters calculated for the SIR model, our model predicts an outcome similar to that for the SIR model. On the other hand, using parameter estimates from SARS literature our model predicts a much more serious epidemic.

Index Terms—modeling, infectious diseases, super-spreading events, severe acute respiratory syndrome

I. INTRODUCTION

SEVERE acute respiratory syndrome (SARS) is a highly contagious respiratory disease which is caused by the SARS Coronavirus. It is a serious form of pneumonia, resulting in acute respiratory distress and sometimes death. SARS emerged in China late 2002 and quickly spread to 32 countries causing more than 774 deaths and 8098 infections worldwide. One of the intriguing characteristics of the SARS epidemic was the occurrence of super-spreading events. Super-spreading events for a specific infectious disease occur when certain infected individuals, called super-spreaders, infect more than the average number of secondary cases.

According to the U. S. Centers for Disease Control and Prevention, a person is a super-spreader if they cause more than 10 secondary infections. Such super-spreading individuals have been identified in the SARS outbreak and they are thought to have caused most of the secondary infections. For example, in Singapore, about 80% of infections have been attributed to only 5 super-spreading individuals ([1]). One extreme super-spreading individual in Hong Kong caused more than 100 secondary infections ([2]). To contrast, in Singapore, most individuals caused 0 secondary infections ([3]).

One differential equation model for super-spreading individuals was proposed before the recent SARS outbreak, by Kemper ([4]). He presents a modified susceptible - infected - removed (SIR) disease model to capture the effect of the super-spreading individuals in which the infected individuals are split into two different infected classes

Manuscript submitted June 8, 2010; revised February 18, 2011.

T. Mkhatshwa, e-mail: tpsmjc@yahoo.com, tel: 865-805-4292.

A. Mummert, corresponding author, Marshall University, Mathematics Department, Huntington, WV 25755, USA, tel/fax: 304-696-3041/4646 e-mail: mummerta@marshall.edu.

with different transmission rates. The super-spreaders have a higher transmission rate, meaning more of their contacts with susceptible individuals result in a new infection than those of the regular infected individuals. This was the first model designed specifically to address the effect of super-spreading individuals on the course of infectious disease epidemics.

Due to the extreme influence of super-spreading individuals during the SARS outbreak, many mathematical epidemiologists developed models for the spread of SARS, which included the occurrence of the super-spreading events[1]. It is interesting to note how each incorporated the super-spreading events into their model. First are models where the super-spreading individuals have a higher transmission rate than the regular infected individuals, as in Kemper ([4]). For example, Masuda, et al. ([3]) developed a contact network model which had different transmission rates for regular infected individuals and for super-spreading individuals. Second are models in which parameters are taken from probability distributions, making the super-spreading individuals naturally appear as the right-hand tail of the distributions. Lloyd-Smith, et al. ([6]) developed a stochastic compartment model and Fang, et al. ([7]) developed a spacial lattice combined with a deterministic compartment model wherein the individual reproduction numbers are drawn from a continuous probability distribution.

The goal of this paper is to capture the effect of the super-spreading individuals using a modification of the classical SIR disease model. The modification is inspired by Li, et al. ([8]) who determined that for SARS "the daily infection rate did not correlate with the daily total number of symptomatic cases but with the daily number of symptomatic cases who were not admitted to a hospital within 4 days of the onset of symptoms." This means that the number of infected individuals is closely associated with the number of individuals who remain outside of isolation longer than most other infected individuals. These individuals remaining outside of isolation longer than normal are related to the disease severity. They have more contacts with susceptible individuals and thus more chances to spread the disease, in other words, they are the super-spreaders. In our model, we split the infected individuals into two classes, with two different "removal" rates; these two rates determine how long an infected individual remains outside of isolation.

In this paper, we modify the classical SIR disease model to capture the effect of a super-spreading event, using the idea that super-spreading individuals stay out of isolation longer than individuals who are not super-spreading. A description of the $SIPR$ model, the model equations, and some basic model properties are given in Section II. We demonstrate

[1]Many models were developed to try and mathematically capture the SARS outbreak. Interested readers are directed to Bauch, et al. ([5]) for an overview of SARS models.

the model using a small-scale and a large-scale outbreak of SARS in Section III. Our discussion follows in Section IV.

A. Infectious Disease Modeling

One of the earliest mathematical models to study infectious diseases is the Reed-Frost model developed in the early 1900s. Another early model and one of the most well known disease models is the susceptible-infected-recovered (SIR) model proposed and studied by Kermack and McKendrick in the 1920s (details of this model are given in Section I-A1). Though this later model is quite simple, it continues to be widely used to study infectious disease outbreaks. For example, Ching, et al. ([9]) use the classical SIR model to study the SARS outbreak.

Since the pioneering work in mathematical modeling of infectious diseases many SIR model modifications and new models have been developed and studied; also, many new techniques to study models have been developed. New models and modifications allow for the study of additional disease outbreak details, for example, the difference in disease spread in different age groups. The study of models has been enhanced by new mathematical analysis techniques, as well as new numerical techniques. For an overview of some new models, and analytical and numerical techniques, readers are directed to Anderson and May ([10]).

In addition to population-level infectious disease models such as the classical SIR model, a wide variety of in-host models have been developed to study the details of a particular disease within an individual. For example, Lo ([11]) studies the CD4-cell counts of HIV infected patients, and Aggarwala ([12]) presents a system of four differential equations for the in-host progression of HIV into AIDS.

The study of mathematical models has been greatly enhanced by the advent of and improvement in computer technology. Improved computing power allows researchers to perform many numerical simulations, which are practically required if one wishes to study systems with a large number of equations, systems of partial differential equations, agent-based models, models with delays, and optimal control strategies. Roy and Mondal ([13]) present an in-host modified SIR model of an infectious disease (pathogen) and study both delay and optimal control. Also, improved computing power allows researchers to fit the parameters of a proposed model to data using sophisticated statistical techniques requiring many numerical simulations, as demonstrated by Kalhori, Nasehi, and Zeng ([14]) for a model of tuberculosis treatment.

By no means do we claim that these short paragraphs are a comprehensive review of mathematical modeling of infectious disease outbreaks. Interested readers will find models of almost any variety covering any topic of their interest related to disease outbreaks. We only wish to provide readers a sense of the current trends in the broad domain of infectious disease modeling.

1) The SIR model: We conclude the introduction with a brief review of the classical SIR disease model. A more detailed review is given in Murray ([15]), and in Ching, et al. ([9]) and the references given there.

The classical SIR disease model splits a fixed-size population, N, into three distinct classes: the susceptible in-
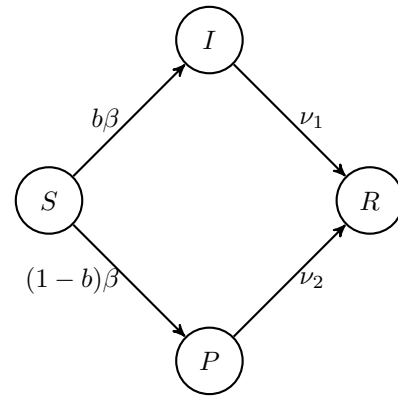

Fig. 1. $SIPR$ model schematic, where S are the susceptible individuals, I are the regular infected individuals, P are the super-spreaders, R are the removed individuals, β is the transmission rate, b is the probability that a new infection will be a regular infected person, $1 - b$ is the probability that a new infection will be a super-spreading individual, ν_1 is the removal rate for a regular infected individual, and ν_2 is the removal rate for a super-spreading individual.

dividuals, S, those do not have the disease and can become infected; the infected individuals, I, those who have the disease and can infect susceptible individuals; and the removed individuals, R, those who have recovered, died, or moved into isolation. Individuals in the removed class gain permanent immunity and remain in the R class forever. Schematically, the SIR model is

$$S \longrightarrow I \longrightarrow R$$

Individuals in the population are assumed to homogeneously mix. Contacts between the susceptible and infected individuals result in a new infected individual at a rate proportional to the number of susceptible and infected individuals. Infected individuals are removed to class R at a rate proportional to the number of infected individuals. The system of nonlinear ordinary differential equations describing the SIR model is

$$\begin{aligned}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \nu I \\
\frac{dR}{dt} &= \nu I
\end{aligned}$$

where β is the transmission rate and ν is the removal[2] rate.

II. THE $SIPR$ MODEL

We describe the $SIPR$ model, a modification of the classical SIR model, which captures the effect of super-spreading individuals. Schematically, the model is given in Figure 1. While the basic SIR model has one class of infected individuals, the $SIPR$ model has a second class of infected individuals, the super-spreaders, denoted by the variable P. In the $SIPR$ model we divide the population, of fixed size N, into four groups, namely the susceptible individuals, S, the regular infected individuals, I, the super-spreaders, P, and the removed individuals, R.

---

[2]The R class is often called the recovered class containing those who have recovered from the disease. In this situation, all individuals recover from the disease; no individuals die or move into isolation. The parameter ν is correspondingly called the recovery rate.

A susceptible individual can become infected through contact with either a regular infected individual or a super-spreading individual. Then with probability b the new infected individuals become regular infected individuals (move to class I) and with probability $1 - b$ the new infected individuals become super-spreading individuals (move to class P). We assume most new infections are regular infected individuals, that is, $b > 1-b$. We assume that the two infected classes, I and P, have the same transmission rate, β.

To capture the effect of the super-spreading individuals, we use the idea that individuals who are super-spreaders stay out of isolation longer than the regular infected individuals. The regular infected individuals stay out of isolation (the R class) for $1/\nu_1$ days and the super-spreading individuals stay out of isolation for $1/\nu_2$ days, with $1/\nu_1 < 1/\nu_2$. Therefore, the model has two distinct removal rates namely, ν_1 and ν_2, corresponding to regularly infected individuals, I, and super-spreading events, P, respectively. When a person is removed to the R class there is no possibility of becoming susceptible again, but rather they recover and gain permanent immunity, or die; in either case, they remain in the removed compartment forever.

Based on the previous descriptions and assumptions we formulate a system of four ordinary differential equations for the $SIPR$ model.

$$\begin{aligned} \frac{dS}{dt} &= -\beta(I + P)S \\ \frac{dI}{dt} &= b\beta(I + P)S - \nu_1 I \\ \frac{dP}{dt} &= (1 - b)\beta(I + P)S - \nu_2 P \\ \frac{dR}{dt} &= \nu_1 I + \nu_2 P \end{aligned} \quad (1)$$

subject to the following initial conditions

$$S(0) = N - I_0 - P_0 = S_0,$$

$$I(0) = I_0, \quad P(0) = P_0, \quad \text{and} \quad R(0) = 0.$$

The $SIPR$ model given by the system of equations (1) has the following properties.

- The $SIPR$ model has a unique global solution.
- The components of the solution, $S(t), I(t), P(t)$, and $R(t)$, of the $SIPR$ model are non-negative and bounded by N for all time, $t \geq 0$.
- The $SIPR$ model has equilibrium points $(N, 0, 0, 0)$, $(S^*, 0, 0, R^*)$, for any $0 < S^* < S_0$, and $(0, 0, 0, N)$.
- The individual reproduction number is

$$R_0 = \frac{b\beta S_0}{\nu_1} + \frac{(1 - b)\beta S_0}{\nu_2}.$$

The properties of the $SIPR$ model are, in general, the natural modification of the corresponding properties of the classical SIR model; proofs of these properties follow in Appendix A. It is interesting to note that the R_0 for the $SIPR$ model is the R_0 for the corresponding SIR model for each of the two infected classes, I and P, multiplied by the probability of a new infected individual becoming an I or a P, respectively.

The $SIPR$ model can be used to analyze any infectious disease where super-spreading events have been identified. As a particular example, super-spreading events have been identified in one outbreak of measles as described in Paunio, et al. ([16]).

TABLE I
ESTIMATED PARAMETERS FOR THE SIR MODEL AND THE SARS OUTBREAK IN THE AMOY GARDENS, HONG KONG (FROM ([9])), AND IN THE HONG KONG SPECIAL ADMINISTRATIVE REGION FOR BOTH FEBRUARY 21 - MARCH 21, 2003 AND MARCH 29 - JUNE 12, 2003.

Parameter	Amoy Gardens	Hong Kong Feb 21 - Mar 21	Hong Kong Mar 29 - Jun 12
β	1.4850/N	0.2586/N	0.1351/N
ν	0.9750	0.0821	0.1923
R_0	1.5154	3.1511	0.7025

III. THE SARS EPIDEMIC

We use the $SIPR$ model to study the spread of SARS on a small scale, in the Amoy Gardens apartment complex in Hong Kong, and on a large scale, in the entire Hong Kong Special Administrative Region. In both cases, we fit the model to the data using two parameter estimation methods. We begin with a description of the estimation procedures.

The general parameter estimation procedure was as follows. The $SIPR$ system of equations (1) was solved repeatedly with parameter sets taken from the allowable range of possible parameters, and the least squares error between the cumulative number of cases of the solution and the actual data was computed. The least squares error was minimized. All computations were done with MATLAB using the function fminsearchbnd.

In parameter estimation Method 1, we used the parameters estimated with a fit of the classical SIR model to the data. For the data on the entire Hong Kong Special Administrative Region, we were unable to find SIR parameters. Therefore, we began with fitting the classical SIR model to the data. A summary of the SIR model parameters is given in Table I. We use the transmission rate of the SIR model as the transmission rate of the $SIPR$ model. The transmission rate, the removal rate, and the initial number of susceptible individuals are used to determine the basic reproduction number R_0. We estimate parameters b, ν_1, and ν_2 so that together they have R_0 as calculated, and they satisfy $b > 1-b$ and $\nu_1 > \nu_2$.

In parameter estimation Method 2, bounds for the parameters in the $SIPR$ model were determined from the literature. Then, using these bounds, we estimate the parameters that give the best fit to the data. Details and references for these bounds are given in the next paragraph and in Subsections III-A and III-B.

For estimation Method 2, we make some assumptions regarding the parameters that apply to both the small- and large-scale outbreaks. In situations with super-spreading individuals, it is assumed that most infected individuals are not super-spreaders, which is the case for the spread of SARS. In fact, Masuda, et al. ([3]) state that for the SARS outbreak in Singapore 80% of infected individuals infected no one else. This leads us to set a lower bound on the probability of becoming a regular infected individual of 0.8, and we have $0.8 < b < 1$. For SARS, after the onset of symptoms regular infected individuals stayed out of isolation between 3 and 5 days ([17]); we assume the average and set $\nu_1 = 1/4$. We assume that on average super-spreading individuals moved into isolation 10 days after they became infectious, that is, $\nu_2 = 1/10$. Finally, many researchers have shown that the

individual reproduction number, R_0, for the SARS outbreak is between 1.5 and 4 ([5]). We assume this region for R_0, which gives a constraint on the parameters β, b, ν_1, and ν_2.

A. The SARS epidemic - small scale

We use the $SIPR$ model to study the spread of SARS at the high rise apartment building Amoy Gardens, Block E, in Hong Kong. We assume that the cumulative number of confirmed cases of SARS in the Amoy Gardens, summarized in Table II, are all from Block E, and that Block E contains a total of 792 individuals, as in Ching, et al. ([9]).

For parameter estimation Method 1, we use the parameters determined by Ching, et al. ([9]). The fixed and estimated parameters of the $SIPR$ model are given in Table III. The resulting number of confirmed cases of SARS are presented in Table II.

For parameter estimation Method 2, we survey literature to determine appropriate bounds on all of the $SIPR$ model equations. We assume that the total number of infected individuals at $t = 0$ is 4, as in Ching, et al. ([9]). We conservatively assume $0 < \beta < 0.01$. The bounds and constraints for each parameter as well as the estimated parameter values are given in Table III. The resulting number of confirmed cases of SARS are presented in Table II.

B. SARS epidemic - large scale

We use the $SIPR$ model to study the spread of SARS in the entire Hong Kong Special Administrative Region. We use the cumulative number of confirmed cases of SARS in Hong Kong as reported by the World Health Organization[3] (WHO). The first case of SARS in Hong Kong appeared on February 21, 2003. The WHO began daily reporting of SARS cases on March 17, 2003, and June 12, 2003 shows the last additional confirmed case, to a total of 1755 confirmed cases. In the parameter estimations, we use all of this data, and we summarize the data in Table IV.

To begin, we identify three distinct time periods of the SARS epidemic in Hong Kong. The first time period is February 21 – March 21, 2003, the second is March 22 – March 28, 2003, and the third is March 29 – June 12, 2003. The first reported case of SARS in Hong Kong occurred on February 21, 2003, when a SARS infected doctor arrived from China. From this date until March 21, 2003, very little was known regarding SARS and few precautions were taken by individuals to prevent the spread of infection. On March 21, Hong Kong officials decided to begin daily reports to the public regarding the threat of SARS infection. As a result of these warnings people began to protect themselves from infection and the dynamics of the disease transmission changed. Due to the continued spread of SARS and the increasing awareness of the seriousness of infection, on March 29, all Hong Kong schools were suspended, again changing the disease dynamics. Classes resumed for all students by May 19 and the last reported case of SARS in Hong Kong occurred on June 12, 2003[4].

---

[3]The WHO numbers were accessed on September 29, 2010, on-line at http://www.who.int/csr/sars/country/en/index.html.

[4]A chronology of the SARS epidemic in Hong Kong can be found in Appendix III of the final report by the SARS Expert Committee of the Hong Kong Special Administrative Region, accessed on October 15, 2010, on-line at http://www.sars-expertcom.gov.hk/

We examine the first and third periods only, using the $SIPR$ model. The first time period corresponds to quick spreading of the disease and we show this region has an R_0 value above 1. The third period corresponds to the end of the disease and correspondingly has an R_0 value below 1.

In 2003, the Hong Kong area had a total population of 6.803 million people. The number of susceptible individuals in Hong Kong during the SARS epidemic is not the entire population; the number of susceptible individuals must be approximated. Katriel and Stone ([18]) give a formula to estimate the percent of the population that is susceptible during an epidemic. The number who are susceptible can be computed using the R_0 value and the percent of the population who became infected. Riley, et al. ([2]) determine the R_0 value specifically for the SARS outbreak in Hong Kong is in the region 2.2 – 3.7. Using this, we compute that the number of susceptible individuals falls in the range 1.8 million – 3.1 million, where the smaller population correspond to larger R_0 values. We assume a large R_0 value, and, therefore we use the lower value, $N = 1.8$ million.

For parameter estimation Method 1, we use the parameters determined by fitting the data sets to the classical SIR model. The fixed and estimated parameters of the $SIPR$ model are given in Tables V and VI, corresponding to time period one and three, respectively. The resulting number of confirmed cases of SARS are presented in Table IV.

For parameter estimation Method 2, we survey literature to determine appropriate bounds on all of the $SIPR$ model equations. We assume that the total number of infected individuals at $t = 0$ is 1 for time period one, corresponding to the index case, and 470 for time period three[5], corresponding to the number of confirmed cases on March 29, 2003. The bounds and constraints for each parameter as well as the estimated parameter values are given in Tables V and VI. The resulting number of confirmed cases of SARS are presented in Table IV.

C. Summary Results

We summarize the infections for each of the data sets and their fitted parameters in Table VII. Specifically, the table contains the total number of individuals who became infected, the total number who became infected through contact with a regular infected (I) and super-spreading individuals (P), the total number of regular infected (I) and super-spreading individuals (P), and the individual reproduction numbers for the I and P classes (as SIR models).

IV. DISCUSSION

We have presented a modification of the classical SIR disease model that captures the effect of super-spreading individuals on an infectious disease epidemic. Using an idea from the progression of the SARS outbreak, we distinguish the regular infected individuals from the super-spreading individuals by how long they remain outside of isolation; the super-spreading individuals spend longer outside of isolation than most infected individuals. The model was fit to data

---

[5]Some of these 470 confirmed cases would have already been removed from the disease transmission dynamics via recovery, death, or removal to isolation.

TABLE II
CUMULATIVE CONFIRMED CASES OF SARS IN AMOY GARDENS, BLOCK E, MARCH 2003 (FROM CHING, ET AL. ([9])). PREDICTED CASES BASED ON PARAMETER ESTIMATION METHODS 1 AND 2.

Day	March 26	March 27	March 28	March 29	March 30
Confirmed	7	22	56	78	114
Method 1	12.2	26.2	48.2	80.3	123.8
Method 2	9.1	19.1	37.4	69.7	122.3

TABLE III
ESTIMATED PARAMETERS FOR THE $SIPR$ MODEL AND THE SARS OUTBREAK IN THE AMOY GARDENS, HONG KONG, MARCH 2003, USING METHODS 1 AND 2. FIXED PARAMETERS WERE FOUND IN CHING ET AL. ([9]).

Parameter	Dimension	Method 1	Fixed or Estimated	Method 2	Bound or Constraint
S_0	people	788	Fixed	788	S_0=788
I_0	people	0.4	Estimated	0.0	$I_0 + P_0 = 4$
P_0	people	3.6	Fixed by I_0	4.0	$I_0 + P_0 = 4$
β	days$^{-1}\times$ people^{-1}	1.4850/N	Fixed	0.8728/N	$0 < \beta < 0.01$
b		0.9315	Estimated	0.9994	$0.8 < b < 1$
ν_1	days^{-1}	0.9997	Estimated	0.2500	$\nu_1 = 0.25$
ν_2	days^{-1}	0.7294	Fixed by b, ν_1, and R_0	0.1000	$\nu_2 = 0.1$
R_0		1.5154	Fixed	3.4765	$1.5 < R_0 < 4$

TABLE IV
CUMULATIVE CONFIRMED CASES OF SARS IN HONG KONG. PREDICTED CASES BASED ON PARAMETER ESTIMATION FOR THE CLASSICAL SIR MODEL, AND THE $SIPR$ MODEL USING METHODS 1 AND 2.

Day	Feb 21	Mar 17	19	21	29	Apr 12	26	May 10	24	Jun 6	20
Confirmed	1	95	150	203	470	1108	1527	1674	1724	1750	1755
SIR	1	100.9	143.8	204.9	470.0	1081.5	1355.8	1478.9	1534.0	1557.6	1569.3
Method 1	1	100.3	142.9	203.6	470.0	776.1	983.6	1124.8	1221.0	1282.6	1328.4
Method 2	1	103.8	144.9	202.2	470.0	1156.5	1486.1	1644.4	1720.5	1755.1	1773.6

TABLE V
ESTIMATED PARAMETERS FOR THE $SIPR$ MODEL AND THE SARS OUTBREAK FOR FEBRUARY 21 - MARCH 21, 2003, IN HONG KONG, USING METHODS 1 AND 2.

Parameter	Dimension	Method 1	Fixed or Estimated	Method 2	Bound or Constraint
S_0	people	1,799,999	Fixed	1,799,999	S_0=N-1
I_0	people	1.0	Estimated	1.0	$I_0 + P_0 = 1$
P_0	people	0.0	Fixed by I_0	0.0	$I_0 + P_0 = 1$
β	days$^{-1}\times$ people^{-1}	0.2586/N	Fixed	0.3738/N	$0 < \beta < 1/N$
b		0.6489	Estimated	0.8000	$0.8 < b < 1$
ν_1	days^{-1}	0.0836	Estimated	0.2500	$\nu_1 = .25$
ν_2	days^{-1}	0.0794	Fixed by b, ν_1, and R_0	0.1000	$\nu_2 = .1$
R_0		3.1511	Fixed	1.9438	$1.5 < R_0 < 4$

TABLE VI
ESTIMATED PARAMETERS FOR THE $SIPR$ MODEL AND THE SARS OUTBREAK FOR MARCH 29 - JUNE 20, 2003 IN HONG KONG, USING METHODS 1 AND 2.

Parameter	Dimension	Method 1	Fixed or Estimated	Method 2	Bound or Constraint
S_0	people	1,799,530	Fixed	1,799,530	S_0=N-470
I_0	people	277.2	Estimated	311.6	$I_0 + P_0 = 470$
P_0	people	192.8	Fixed by I_0	158.4	$I_0 + P_0 = 470$
β	days$^{-1}\times$ people^{-1}	0.1351/N	Fixed	0.1473/N	$0 < \beta < 1/N$
b		0.5237	Estimated	0.8910	$0.8 < b < 1$
ν_1	days^{-1}	25.6988	Estimated	0.2500	$\nu_1 = .25$
ν_2	days^{-1}	0.0920	Fixed by b, ν_1, and R_0	0.1000	$\nu_2 = .1$
R_0		0.7025	Fixed	0.6851	$0 < R_0 < 1$

TABLE VII

SUMMARY OF THE OUTBREAKS FOR EACH OF THE DATA SETS AND THEIR FITTED PARAMETERS: THE TOTAL NUMBER OF INDIVIDUALS WHO BECAME INFECTED; THE TOTAL NUMBER WHO BECAME INFECTED THROUGH CONTACT WITH A REGULAR INFECTED (I) AND SUPER-SPREADING INDIVIDUALS (P); THE TOTAL NUMBER OF REGULAR INFECTED (I) AND SUPER-SPREADING INDIVIDUALS (P); AND THE INDIVIDUAL REPRODUCTION NUMBERS FOR THE I AND P CLASSES (AS SIR MODELS).

	Amoy	Gardens		HK	Period 1		HK	Period 3	
	SIR	M 1	M 2	SIR	M 1	M 2	SIR	M 1	M 2
Total Infected	477.3	479.1	765.8	1,709,770	1,709,769	1,405,410	1578.9	1426.1	1791.5
Contact with I	473.3	426.1	742.1	1,709,769	1,097,524	912,542	1108.9	4.1	876.6
Contact with P		49.0	19.7		612,244	492,867		952.0	444.9
Total I	477.3	443.0	761.4	1,709,770	1,109,516	1,124,365	1578.9	777.9	1489.2
Total P		36.1	4.4		600,253	281,045		648.2	302.3
R_0 for I	1.5154	1.4779	3.4735	3.1511	3.0944	1.4953	0.7025	0.0053	0.5889
R_0 for P		2.0255	8.6838		3.2560	3.7382		1.4691	1.4721

from the SARS epidemic, using two different parameter estimation methods.

Parameter estimation Method 1, used parameters estimated for the classical SIR model. The resulting parameters show only a slight super-spreading behavior in all cases studied. For Amoy Gardens and the short-term outbreak in Hong Kong, the I and P classes are not distinguished by how long each stays outside of isolation. They are also not distinguished by their individual reproduction numbers. In each case, the two recovery rates are similar to each other, and similar to the recovery rate of the SIR model estimated parameters. It is not surprising, therefore, that the final number of infected individuals, given by the final number of individuals in the removed class, in the two models are within just a few individuals.

The fitted parameters for the long-term outbreak in Hong Kong using parameter estimation Method 1 indicate that regular infected individuals move so quickly into isolation that we can disregard their influence on the disease spread. As evidence, over the entire course of the disease, they collectively only infect a total of 4.4 individuals. Due to the incredibly short time spent in the infected class, the individual reproductive number in this case is very low.

Though the outbreak in the entire Hong Kong Special Administrative Region was split into three time periods corresponding to three different disease dynamics, the fit parameters do not match the third time period data set as well as one might hope. This is noticed for the classical SIR model, and, since the SIR parameters are used for estimation Method 1, we also see this in the fit parameters for Method 1. One possible fix for this problem is to split the third time period into other periods in which the outbreak has common dynamics.

Parameter estimation Method 2, used research to set a priori bounds on the parameters of the $SIPR$ model. In every case, the resulting parameters show super-spreading behavior, for example, the two infected classes are distinguished by their individual reproduction rates. All three cases show that, on average, each regular infected individual infects less than 1 other individual, while each super-spreading individual infects more than 1 other individual.

Considering the specific case of the outbreak at the Amoy Gardens apartment complex, Method 2 predicts that, without any other intervention, almost all of the residents of Block E, will become infected. In this case, both the regular infected and super-spreading individuals have basic reproduction numbers larger than 1. It is clear from the dire outcome predicted for the residents of the Amoy Gardens, Block E, that super-spreading individuals must be brought into isolation as quickly as possible.

For the long-term outbreak in Hong Kong (in the third time period), the overall individual reproductive number is less than 1, which matches the notion that individuals were taking precautions to protect themselves, and the disease spread was slowing down. Using the $SIPR$ model, we see that the regular infected individuals do have an individual reproductive number less than 1, however, the super-spreading individuals have a reproductive number larger than 1. Again, we see that it is imperative that the super-spreading individuals be brought into isolation as quickly as possible.

Finally, it is evident from the short-term outbreak in Hong Kong (in the first time period), that without any control measures the spread of SARS in Hong Kong would have been extreme. The model predicts that more than 1.7 million residents would have been infected over the course of the disease outbreak. Correspondingly, they would have experienced a large disease related mortality. The death rate for SARS is estimated to be around 10%, and so, Hong Kong would have lost an estimated 170,000 residents. Noting that the final death toll in Hong Kong was (though tragic) only 299 people, the control measures put in place by the government of Hong Kong saved thousands of lives.

The $SIPR$ model is versatile; it can be used to examine an outbreak of any disease known to have super-spreading individuals, measles for example (see ([16])). On the other hand, the model was built using the idea that super-spreading individuals stay out of isolation longer than regular infected individuals, as during the SARS epidemic. (There are documented cases of SARS infected individuals violating strict isolation mandates[6].) For diseases where the super-spreading behavior is a result of differing transmission rates, one should use the Kemper model ([4]).

Super-spreading individuals for infectious diseases pose a serious threat to public health. The $SIPR$ model clearly demonstrates that infectious individuals must be removed from interactions with the susceptible individuals as quickly as possible to decrease the seriousness of an infectious disease epidemic.

[6]See for example "SARS Epidemic Worsens in Taiwan", by Shu Shin Luh, in the *Washington Post*, May 15, 2003.

APPENDIX A

PROPERTIES OF THE $SIPR$ MODEL

We give a brief justification for each of the properties of the $SIPR$ model. As mentioned in Section II, each of these properties is analogous to that of the classical SIR model. Therefore, details of the proofs follow exactly as for the SIR model.

1) Continuity and (partial) differentiability of the equations in the system of differential equations, along with the fact that each of the solutions is bounded, guarantees a unique global solution.

2) Since $\dfrac{dR}{dt} > 0$ for all t, the number of removed individuals will be non-negative. Since $\dfrac{dS}{dt} = 0$ when $S = 0$, the number of susceptible individuals will be non-negative.

To verify that the number of regular infected individuals is always positive, consider the boundary case when $I = 0$, and all of the other populations are non-negative. In this case $\dfrac{dI}{dt} = b\beta PS \geq 0$; that is, the number of regular infected individuals will not decrease to become negative. The verification for the number of super-spreading individuals is similar.

When the four equations are added together we see that the total change in the population size is 0, and so the population is constant, say N. Therefore, since each population is non-negative, we see that each is bounded between 0 and N.

3) Setting each of the equations in the system of differential equations equal to zero and solving for the values of S, I, P, and R gives the family of equilibrium points.

4) To determine the individual reproductive number R_0, one may compute the largest eigenvalue of the next generation matrix

$$\begin{bmatrix} \dfrac{b\beta S_0}{\nu_1} & \dfrac{b\beta S_0}{\nu_2} \\[2mm] \dfrac{(1-b)\beta S_0}{\nu_1} & \dfrac{(1-b)\beta S_0}{\nu_2} \end{bmatrix}$$

ACKNOWLEDGMENT

REFERENCES

[1] Y. S. Leo, M. Chen, B. H. Heng, and C. C. Lee, "Severe acute respiratory syndrome — Singapore, 2003," *MMWR*, vol. 52, no. 18, pp. 405–411, May 9, 2003.

[2] S. Riley, C. Fraser, C. A. Donnelly, and et al., "Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions," *Science*, vol. 300, pp. 1961–1966, 2003.

[3] N. Masuda, N. Konno, and K. Aihara, "Transmission of severe acute respiratory syndrome in dynamical small-world networks," *Phys Rev E*, vol. 69, pp. 1608–1610, 2004.

[4] J. T. Kemper, "On the identification of superspreaders for infectious disease," *Math. Biosci.*, vol. 48, no. 1-2, pp. 111–127, 1980.

[5] C. T. Bauch, J. O. Lloyd-Smith, M. P. Coffee, and A. P. Galvani, "Dynamically modeling SARS and other newly emerging respiratory illnesses: Past, present, and future," *Epidemiology*, vol. 16, no. 6, pp. 791–801, November 2005.

[6] J. O. Llyod-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, "Superspreading and the effect of individual variation on disease emergence," *Nature*, vol. 438, pp. 355–359, November 17, 2005.

[7] H. Fang, J. Chen, J. Hu, and L. X. Xu, "On the origin of the super-spreading events in the SARS epidemic," *Europhysics Letters*, vol. 68, no. 1, pp. 147–152, 2004.

[8] Y. Li, I. T. S. Yu, P. Xu, and et al.

[9] C. W., S. K. Chung, and T. Ng, "On modeling SARS in Hong Kong," *IAENG International Journal of Applied Mathematics*, vol. 13, no. 1, pp. 1–7, 2003.

[10] R. M. Anderson and R. M. May, *Infectious Diseases of Humans*. Oxford University Press, 1991.

[11] C. F. Lo, "First passage time density for the disease progression of HIV infected patients," *Engineering Letters*, vol. 13, no. 3, pp. 320–325, 2006.

[12] B. D. Aggarwala, "On a four state model for development to AIDS," *Engineering Letters*, vol. 13, no. 3, pp. 308–319, 2006.

[13] P. K. Roy and J. Mondal, "Host pathogen interactions: Insight of delay response recovery and optimal control in disease pathogenesis," *Engineering Letters*, vol. 18, no. 4, pp. 406–421, 2010.

[14] S. R. N. Kalhori, M. Nasehi, and X. Zeng, "A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion," *IAENG International Journal of Applied Mathematics*, vol. 40, no. 2, pp. 102–107, 2010.

[15] J. D. Murray, *Mathematical Biology I. An Introduction*, 3rd ed. Springer, 2002.

[16] M. Paunio, H. Peltola, M. Valle, and et al., "Explosive school-based measles outbreak: Intense exposure may have resulted in high risk, even among revaccinees," *Am J Epidemiol*, vol. 148, no. 11, pp. 1103–1110, 1998.

[17] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L. Ho, T. Thach, P. Chau, K. Chan, T. Lam, L. Tse, T. Tsang, S. Liu, J. H. B. Kong, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson, "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong," *The Lancet*, vol. 361, pp. 1761–1766, May 24, 2003.

[18] G. Katriel and L. Stone, "Pandemic dynamics and the breakdown of herd immunity," *PLoS ONE*, vol. 5, no. 3, p. e9565, 2010.