Assessing the levels of unreporting in the 2009 A-H1N1 influenza epidemic in Lima, Peru

Diego Chowell[1], Baltazar Espinoza[2], Oscar Patterson[3], David Requena[4], Michelle Salas[5], Karen Ríos-Soto[6]
[1] Mathematisch Instituut, Universiteit Utrecht, [2] School of Sciences, Universidad de Colima, [3] Applied Mathematics for the Life and Social Sciences, Arizona State University, [4] Group of Natural Structures and Theoretical Research, Universidad Nacional Mayor de San Marcos, [5] College of Arts and Sciences, University of Portland, [6] Department of Mathematical Science, University of Puerto Rico at Mayagüez

Abstract

Mathematical models can provide insights on how future epidemics may behave. Evaluation and implementation of public health strategies can be more accurate when reliable data is used to estimate parameters. However, not all cases are reported, and the levels of uncertainty generated by the gap between the number of reported cases and the actual number of cases has not been studied in detail. We evaluate the impact of non-reported cases in the calculation of the final epidemic size and the effect of different control measures on reducing the attack rate. A system of non-linear ordinary differential equations is constructed to model the spread of influenza. The final size relation for the total number of infected individuals and the proportion of total reported cases are calculated. Both relations are used to generate an expression that helps us quantify the level of non-reporting. To illustrate our results, we consider the case of the 2009 A-H1N1 influenza outbreak in Lima, Peru. Assuming different scenarios of reporting, we estimate key parameters using the data from the initial exponential phase, from which we conclude that no more than 30% of the actual cases were reported. We also perform Monte Carlo simulations to quantify the uncertainty of the control reproductive number to model parameters. Via numerical simulations, we study the impact of different values of the per-capita isolation rates on the final epidemic size. Furthermore, we explore the effects of social distancing varying the time when the intervention policy is applied and different levels of reduction in the transmission rate.

B Introduction

There are between three and five million reported cases of influenza each year leading to 250,000-500,000 deaths worldwide [20]. However, these numbers are small in comparison to the Spanish influenza pandemic of 1918-1919 where it was estimated that 1/3 of the total population got infected and 50 million deaths occurred, 10% of the total world population at that time [3]. It is thought that this strain of influenza was so deadly because the virus may have been novel to humans. In the United States, a yearly flu vaccine is administered to keep the spread of seasonal influenza under control [21]. However, this vaccine only provides protection against the three strains which are predicted to be most common for that particular year [21]. For example, in 2011-2012 the vaccine will protect against an influenza A-H3N2 virus, an influenza B virus and the H1N1 virus that emerged in 2009 to cause a pandemic [21]. The consequences could be drastic if an influenza strain or subtype which is completely novel or radically different from those in the vaccine becomes prevalent.

Fears of consequences of a new infectious disease or even re-emergent, have generated an interest in mathematical epidemiology modeling [1]. Recent examples are the Severe Acute Respiratory Syndrome (SARS) epidemic in 2003 and the H1N1 pandemic in 2009 [28]. Mathematical models for the spread of influenza can provide insight of how future influenza epidemics with similar epidemiological characteristics may behave. This is extremely important, because quantification of disease transmission are quite difficult.

Epidemiological data such as incidence, prevalence and number of deaths among others, are highly valuable in estimating parameters of a model. Some of these parameters can be the recovery rates, incubation periods, and deaths rates associated to the disease [5]. Having estimates for parameter values, such as isolation, vaccination and transmission rates, are also beneficial as they allow for comparison of control strategies and how effective they are. Models can provide a curve for the cumulative number of cases. However, when trying to fit these curves to these data, inconsistencies might occur. Some of these inconsistencies might be attributed to the model not properly describing the dynamics of the particular disease, the data only quantifying the reported cases, delays in reporting, and incomplete data [1].

There are a number of reasons why accurate reporting is difficult. First, not all regions of a country may be equipped to diagnose a disease [14]. For example, in rural areas in China, clinicians may not have the skills or the equipment needed to correctly diagnose a disease [14]. Also, because of the large number of patients a clinician may have, they may be far too busy to accurately report all confirmed cases of a disease or patients might no even get tested [14]. Furthermore, a nation may choose to non-report cases due to the fear that tourism and trade will drop [23]. In addition, people of lower socioeconomic status often do not seek medical attention or they show only non-severe symptoms, further leading to non-reporting the total number of infected cases [14].

C The influenza A-H1N1 in Lima, Peru

During the 2009 influenza pandemic in Peru, there were four main surveillance systems consisting primarily of sentinel surveillance of influenza-like illnesses with virological surveillance of influenza and other respiratory viruses. The other three are sentinel surveillance of severe acute respiratory infections and associated deaths, surveillance of acute respiratory infections in children under the age of five years and pneumonia in all age groups, and case and cluster surveillance [2].

The first confirmed case of pandemic A-H1N1 influenza in Peru was diagnosed in Lima on May 9, 2009 [2]. In Lima, which is the focus of the current study, the number of reported cases was 2989, and the peak of the disease was reached on June 22, 2009. This peak corresponds to the point in time when the number of new reported cases is the highest, beyond this point, the number of new reported cases starts to decrease. Figure 2 shows the number of new reported/confirmed cases per unit time in Lima from May 1, 2009 to December 31, 2009. However, the most reliable data is up to July 11 because on that date, a policy was put in place. This policy changed the way data were collected: while during the initial phase all individuals with severe symptoms were being tested, after the policy only individuals in risk group who presented severe symptoms were tested (i.e., pregnant women, seniors, children and individuals with pre-existing respiratory symptoms) [**?**]. Also only a portion of individuals exhibiting influenza-like illnesses were tested for A-H1N1 [**?**].

This work focuses on the gap between the total number of cases of influenza and the reported cases. Since reported A-H1N1 individuals were isolated in Peru, a model which takes isolation into account is proposed in Section 3. In Section 4, basic reproductive numbers corresponding to the model with and without isolation are discussed. In Section 5, calculations for the final epidemic size and the total number of reported cases are shown. These computations lead to equations for the level of reporting and non-reporting. In Section 6, parameter values are estimated, and a relation between the final epidemic size proportion and the percentage of reporting is obtained. Section 7 numerically investigates the effects on the final epidemic size of varying the isolation rates. In Section 8, also numerically, we asses the impact on the final size of different social distancing alternatives. In section 9, uncertainty analysis is conducted on the control basic reproductive number and the level of unreporting. Finally in Section 10

we introduce an optimization problem where per-capita isolation rates are optimized under a budget constraint aimed at controlling the disease via isolation efforts.

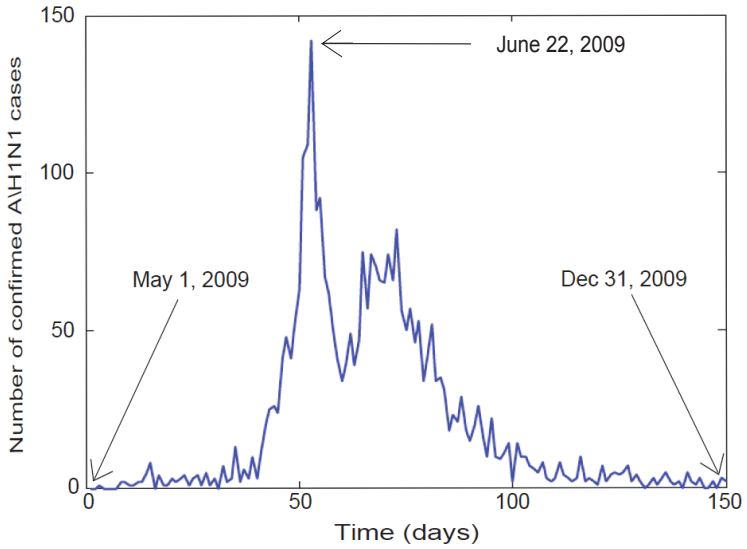


Figure 2: Number of new confirmed cases in Lima, Peru per unit time.

D The Model

The total population is stratified according to their epidemiological states: Susceptible individuals (S), Infected (I_1 and I_2), Isolated (J) and Recovered (R) individuals. Susceptibles (S) acquire an infection via contacts with individuals in the I_1 or I_2 class at a per-capita transmission rate β. The class I_1 consists of individuals with less severe symptoms whose infectiousness is reduced by a factor δ. Individuals in class I_1 can either develop severe symptoms entering class I_2 at a per-capita rate κ or they recover (R) at a per-capita rate γ_1. Severely infected individuals (I_2) recover at a per-capita rate γ_2. Sick, individuals are isolated (J) from I_1 and I_2 at per-capita rates α_1 and α_2, respectively. In order to quantify the proportion of reported individuals, they must be kept in separate compartments than the rest of the population. The parameters of the model and their definition are provided in Table 8.
The model considers a single outbreak of influenza and demographic dynamics is not included. The disease is caused by one strain of influenza transmitted through contact with infectious individuals. It is assumed that individuals who recover develop total immunity against the circulating strain. The fraction of infectious individuals who transferred to the isolated class are assumed to be perfectly isolated from the rest of the population. Hence, we use a modified standard incidence: $N - J$ is the interacting population instead of N. We assume proportionate mixing between individuals, that is, all individuals have an equal probability of interacting with each other. Symptoms are assumed to appear soon after an individual becomes infected. The class J represents those isolated and in this case reported cases. The

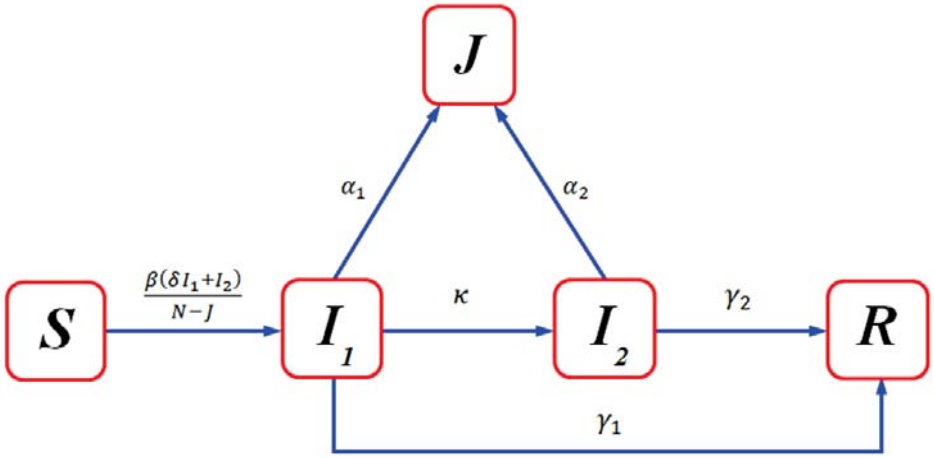


Figure 3: Compartmental diagram: susceptible individuals (S), individuals with less severe flu (I_1), individuals with severe flu (I_2), recovered (R) and isolated (J) individuals.

system of non-linear ordinary differential equations is given by

$$\frac{dS}{dt} = -\beta S\left(\frac{\delta I_1 + I_2}{N-J}\right), \tag{2}$$

$$\frac{dI_1}{dt} = \beta S\left(\frac{\delta I_1 + I_2}{N-J}\right) - (\alpha_1 + \kappa + \gamma_1)I_1, \tag{3}$$

$$\frac{dI_2}{dt} = \kappa I_1 - (\alpha_2 + \gamma_2)I_2, \tag{4}$$

$$\frac{dJ}{dt} = \alpha_1 I_1 + \alpha_2 I_2, \tag{5}$$

$$\frac{dR}{dt} = \gamma_1 I_1 + \gamma_2 I_2, \tag{6}$$

$$N = S + I_1 + I_2 + J + R. \tag{7}$$

Table 8: Parameters and their definition

Parameters	Definition
β	Transmission rate
κ	Rate of development of severe symptoms
γ_1	Recovery rate of individuals with less severe symptoms
γ_2	Recovery rate of individuals with severe symptoms
α_1	Isolation rate for those with less severe symptoms
α_2	Isolation rate for those with severe symptoms
δ	Factor by which the infectivity of the less severe class is reduced

The level of non-reporting in epidemiology may be significant in the case of influenza given that

asymptomatic individuals are not accounted for and some symptomatics show only non-severe symptoms. Hence, the quantification of the infected/recovered population presents a prolonged challenge. Through our model we find an analytical expression for the level of non-reported infected individuals. To find this quantity we will first calculate the final epidemic size and then compute the reported final size. Using these quantities we can have an idea of the magnitude of the non-reporting and the role of the isolation parameters α_1 and α_2.
Under the assumptions of our model, the non-reported cases happen if infected individuals (I_1 or I_2) recover (R) without being isolated (J). We consider two types of infections given that individuals with severe influenza symptoms are more likely to seek medical attention and get reported than individuals with less severe symptoms; hence, we assume that $\alpha_2 > \alpha_1$.

E Basic reproductive numbers

In our model susceptible individuals can become infected due to interactions with either less severe or severe infected individuals. Therefore, we have two contributions to the basic reproductive number, R_0. Typically, R_0 is the average number of secondary infections caused by a single infectious individual introduced into a entirely susceptible population. This term describes the initial phase of the epidemic when the dynamics is determined only by "natural" forces, that is, when no human intervention such as isolation or quarantine are in place, [6]. This scenario can be accommodated to our model by making $\alpha_1 = \alpha_2 = 0$, that is, the isolation process is not active. The parameter β is the transmission rate of the disease and it is also the force of infection of those infected with severe symptoms (I_2). The force of infection of the less severly infected is $\delta\beta$ where $\delta \in [0, 1]$ represents a reduction factor in the infectiousness of individuals with less severe symptoms (I_1). The average infection period for I_1 is $\frac{1}{\kappa + \gamma_1}$, while for I_2 it is $\frac{1}{\gamma_2}$. We must also consider $\frac{\kappa}{\kappa + \gamma_1}$, the proportion of individuals with non-severe symptoms that progress to a more severe stage of the disease. Therefore, R_0 is given by

$$R_0 = \frac{\beta\delta}{\kappa + \gamma_1} + \left(\frac{\kappa}{\kappa + \gamma_1}\right)\left(\frac{\beta}{\gamma_2}\right). \tag{8}$$

The control reproduction number, R_c, describes the number of secondary infections generated during an epidemic outbreak by a typical infected individual when control measures are in place [6], such as isoltion The expression for R_C is given by

$$R_C = \frac{\beta\delta}{\alpha_1 + \kappa + \gamma_1} + \left(\frac{\kappa}{\alpha_1 + \kappa + \gamma_1}\right)\left(\frac{\beta}{\gamma_2 + \alpha_2}\right). \tag{9}$$

The term $\frac{1}{\alpha_1 + \kappa + \gamma_1}$ is the average time an individual spends in class I_1. Hence

$$R_{C1} = \frac{\beta\delta}{\alpha_1 + \kappa + \gamma_1}, \tag{10}$$

is the contribution to R_C by the less severe infected individuals. The term $\frac{1}{\alpha_2 + \gamma_2}$ is the average time an individual spends in class I_2; $\frac{\kappa}{\alpha_1 + \gamma_2 + \kappa}$ is the fraction of individuals reaching infection class I_1 that progress to infection class I_2. Therefore,

$$R_{C2} = \frac{\beta\kappa}{(\alpha_1 + \kappa + \gamma_1)(\alpha_2 + \gamma_2)}, \tag{11}$$

is the contribution to R_C by severe infected individuals.

F Unreported and final epidemic size

Consider a closed population and suppose that it is completely free from a certain organism causing a disease. Assume that in one way or another, the disease is introduced in at least one host. Therefore, the number of susceptible individuals can only decrease and hence it must have a limit when time tends to infinity [**?**]. We may ask the following questions: Will this limit be zero? Or will some fraction of the population escape from getting infected? What proportion of the population will ultimately have experienced infection? How does this fraction depend on the parameters of the model proposed? These questions were posed by Kermack and McKendrick in 1927. In this section we will try to address some of these questions for our model. We calculate and analyze the final epidemic size following the methodology by Brauer ([4,6]) to quantify the non-reported cases (U_∞), in Eq.(**??**), and the role of isolation for the final epidemic of size (Y) and the final reported cases (J_∞).

F.1 Final epidemic size calculation

The final epidemic size is considered to be the final number of infected individuals during an epidemic outbreak. We proceed by getting the expression for the severe infected individuals as a function of time, $I_2(t)$, using the fact that

$$I_2' = -S' - I_1' - (\alpha_1+\gamma_1)I_1 - (\alpha_2+\gamma_2)I_2, \tag{12}$$

where I_1 and I_1' are fixed. To solve Eq. (12) for I_2, we us an integrating factor $\mu = e^{\int(\alpha_2+\gamma_2)dt}$. Knowing that $I_2(0) = 0$ (initially there were no severly infected), we obtain

$$\begin{aligned}\left(I_2 e^{(\alpha_2+\gamma_2)t}\right)' &= \left(-S' - I_1' - (\alpha_1+\gamma_1)I_1\right)e^{(\alpha_2+\gamma_2)t},\\ \implies I_2(t) &= \int_0^t \left[-S' - I_1' - (\alpha_1+\gamma_1)I_1\right]e^{(\alpha_2+\gamma_2)(\tau-t)}d\tau.\end{aligned} \tag{13}$$

Similarly, we obtain an equation for the infected individuals with less severe symptoms as a function of time, given by:

$$I_1(t) = \int_0^t -S' e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)}d\tau. \tag{14}$$

To find the final size relation we divide Eq.(2) by S and replacing Eqs. (13) and (14) we get

$$-\frac{S'(t)}{S(t)} = \frac{\beta(\delta I_1 + I_2)}{N-J}, \tag{15}$$

$$\begin{aligned}\implies \ln\frac{N}{S_\infty} &= \frac{\beta}{N}\int_0^\infty\int_0^\infty -S'(\tau)\left(\delta e^{-(\alpha_1+\gamma_1+\kappa)u} + e^{-(\alpha_2+\gamma_2)u}\right)dud\tau\\ &\quad + \frac{\beta}{N}\int_0^\infty\int_0^\infty \left(-I_1'(\tau) - (\alpha_1+\gamma_1)I_1(\tau)\right)e^{-(\alpha_2+\gamma_2)u}dud\tau.\end{aligned} \tag{16}$$

Further, solving the integrals in Eq.(16), we obtain

$$\ln\frac{N}{S_\infty} = \frac{\beta}{N}(N-S_\infty)\left(\frac{\delta}{\alpha_1+\gamma_1+\kappa} + \frac{1}{\alpha_2+\gamma_2}\right) - \frac{\beta(\alpha_1+\gamma_1)}{N(\alpha_2+\gamma_2)}\int_0^\infty I_1(\tau)d\tau. \tag{17}$$

The typical final size relation is obtained by the substitution of Eq.(14) into Eq.(17), given by

$$\begin{aligned}
\ln\frac{N}{S_\infty} &= \left(1-\frac{S_\infty}{N}\right)\left(\frac{\beta\delta}{\alpha_1+\kappa+\gamma_1}\frac{\beta}{\alpha_2+\gamma_2}\right) \\
&\quad -\left(\frac{\beta}{\alpha_2+\gamma_2}\right)\left(\frac{\alpha_1+\gamma_1}{N}\right)\int_0^\infty\int_0^\tau -S'(s)e^{(\alpha_1+\gamma_1+\kappa)(s-\tau)}dsd\tau, \\
&= \left(1-\frac{S_\infty}{N}\right)\left[\frac{\beta\delta}{\alpha_1+\kappa+\gamma_1}+\frac{\beta}{\alpha_2+\gamma_2}\left(\frac{\kappa}{\alpha_1+\kappa+\gamma_1}\right)\right], \\
\ln\frac{N}{S_\infty} &= \left(1-\frac{S_\infty}{N}\right)R_C. \qquad (18)
\end{aligned}$$

Equation (18) is the typical final size relation [4,6]. If we let $s_\infty = \frac{S_\infty}{N}$ as the proportion of the final susceptible size Eq. (18) yields

$$\ln(s_\infty) = (s_\infty - 1)R_C, \qquad (19)$$

where $0 < s_\infty < 1$ and we have changed the argument in the logarithm and hence the sign of the right hand term. Here $1 - s_\infty$ represents the number of disease cases over the course of the epidemic. Let $y = \frac{Y}{N}$ be the final epidemic size proportion, where Y is the final epidemic size. Notice that $y \neq \lim_{t\to\infty}\frac{I(t)}{N}$ (since $\lim_{t\to\infty}\frac{I(t)}{N} = 0$) and y is the proportion of cumulative number of infected individuals during the whole epidemic, whereas $s_\infty = \lim_{t\to\infty}\frac{S(t)}{N}$. Also notice that for a single outbreak (no death) $y = (1 - s_\infty)$. With that we get a proportional final size relation (from Eq. (19))

$$y = 1 - e^{-yR_C}. \qquad (20)$$

We know that the final epidemic size is directly related to R_C. This can be proven by rearranging terms in Eq.(20) as follows

$$R_C = \frac{1}{y}\ln(1-y),$$

taking derivative with respect to y we obtain

$$\frac{dR_C}{dy} = \frac{1}{y^2}\ln(1-y) + \left(\frac{1}{y}\right)\left(\frac{1}{1-y}\right) > 0.$$

If we plot $\frac{dR_C}{dy}$ as a function of $y : 0 < y < 1$, we get that for all $0 < y < 1$, $\frac{dR_C}{dy} > 0$. That is, the basic reproductive number and the final epidemic size are directly related in such interval, as shown in the Figure 4.

F.2 Final reported cases (J_∞)

From the equation of the rate of change of isolated/reported individuals, that is, Eq.(6), and with initial condition $J(0) = 0$, we have

$$\int_0^\infty J'dt = \int_0^\infty (\alpha_1 I_1 + \alpha_2 I_2)\, dt. \qquad (21)$$

Therefore, replacing the equations for the infected individuals as a function of time, that is, Eq. 13 and

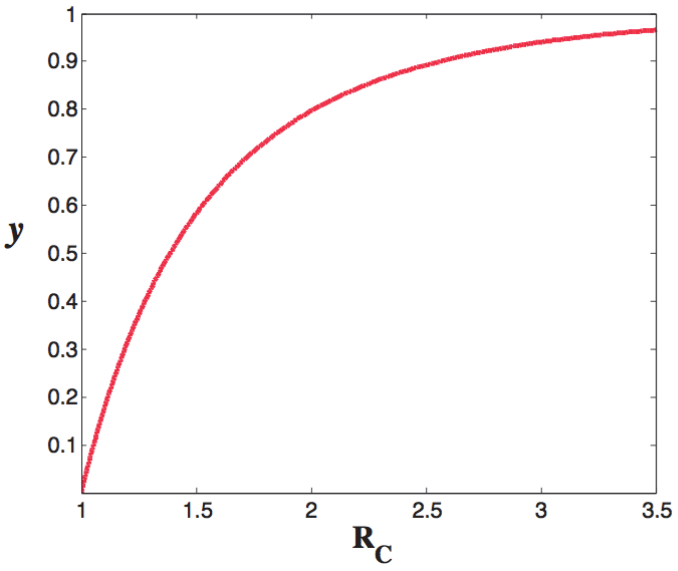


Figure 4: Relation between R_C and the proportional final size relation. As shown in the graph, $R_C = 1.5$ corresponds to the fact that 58.28% got infected.

Eq. 14 we have,

$$\begin{aligned} J_\infty &= \alpha_1 \int_0^\infty \int_0^t -S' e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} d\tau dt \\ &\quad +\alpha_2 \int_0^\infty \int_0^t \left[-S' - I_1' - (\alpha_1+\gamma_1) I_1\right] e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt, \\ J_\infty &= (N - S_\infty) \left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right]. \end{aligned} \tag{22}$$

Dividing by the total population N, and defining j_∞ as the final size proportion of reported cases, Eq. (22) can be expressed as

$$\begin{aligned} j_\infty &= (1 - s_\infty) \left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right], \\ &= y \left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right]. \end{aligned}$$

Then, we obtain an expression for the level of unreporting, which is given by

$$\begin{aligned} u_\infty &= 1 - \frac{j_\infty}{y}, \\ &= 1 - \left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right], \end{aligned} \tag{23}$$

where $\frac{j_\infty}{y}$ is the level of reporting.

Notice that the transmission rate (β) and the reduction of infectivity by less severe individuals (δ) do not appear in this expression. The reason is that the fraction of reported cases does not measure the rate of inflow to the infected classes, connected to β and δ. The fraction of reported cases rather quantifies the progression from infected individuals to isolation and recovered R (non-reported). An upper bound was found for the level of reporting (see Appendix B).
In Eq. (23) the term

$$\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa},$$

is the proportion of non-severe cases that were isolated. The term

$$\left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right),$$

represents the contribution to isolation by severely infected individuals (I_2), that is, given by the product of

$$\frac{\kappa}{\alpha_1+\gamma_1+\kappa},$$

which is the fraction of infected from I_1 develops severe symptoms, and the term

$$\frac{\alpha_2}{\alpha_2+\gamma_2},$$

is the proportion of individuals who were in the more severe infected clas I_2 and become isolated.

G Parameter estimation

The use of mathematical modeling to interpret disease outbreak data has provided many insights into epidemiology, particularly in the context of emerging and re-emerging infectious diseases. In many situations, the basic reproductive number R_0, governs the probability of the occurrence of a major outbreak, the typical size of the resulting outbreak and the rigor of control measures needed to mitigate an outbreak. Estimates of R_0 which we refer to as $\hat{R}_0$ can be obtained by substituting the corresponding individual parameter estimates into the analytical formula of R_0. Model parameters can be estimated using least-square fitting of the model solution to the observed data. The optimal set of parameters best fits the epidemic data by minimizing the sum of the squared differences between the observed data and the model solution. Previous work using parameter estimation methods has been effective to estimate parameters that are not measurable. In the work of Sutton [29] she uses parameter estimation on the effect of vaccination in pneumococcal infection data in Australia from 2002 - 2004. Parameter selection methods in inverse problem formulation has been studied by H. T. Banks and Ariel Cintrón-Arias [30]. They discuss methods for a priori selection of parameters to be estimated with inverse problem formulations such as the Maximum Likelihood, Ordinary and Generalized Least Squares for mathematical models with a large number of parameters. They specifically illustrate their ideas within a host model for HIV dynamics which has been successfully validated with clinical data obtained from the Massachusetts General Hospital in 2007.
In this work we were able to estimate some parameters of our model with our data set by using Ordinary Least Squares (OLS), see [5]. This technique assumes that the epidemiological system is described by some underlying dynamical model with some set of parameters, known as the true parameters, but that the observed data arises from some noise of the output of this system (that is, observational errors). We write the true parameter set as the p-element vector Θ_0, where some of these parameters may be initial conditions of the dynamic model if one or more of these are unknown. The n observations of the system, Y_1, Y_2,..., Y_n, are made at times t_1, t_2,..., t_n. Then the mathematical model is written as

$$Y_i = M(t_i;\Theta_0) + E_i, \quad i = 1, ..., n, \tag{24}$$

where $M(t_i;\Theta_0)$ corresponds to the observed solution for the mathematical model at the i^{th} observation time for a particular vector of p parameters $\Theta \in R^p$. The term Θ_0 represents the parameters which can be perceive from the observations $\{Y_i\}_{i=1}^n$ referred as the "true" parameters in the literature. The terms E_i are random variables which can represent observation or measurement errors, such as "system fluctuations" or other phenomena that cause observations not to be precise.
The appropriate estimation procedure depends on the properties of the errors E_i. We assume that the errors have the following form

$$E_i = M(t_i;\Theta_0)^{\xi}\epsilon_i, \tag{25}$$

where the value ξ determines the method that will be use to estimate your parameters. When $\xi = 0$ we use the ordinary least squares method, in this case $E_i = \epsilon_i$, and it is assumed that the noise *variance* is independent of the magnitude of the predicted value of the observations. For $\xi = 1$ the noise *standard deviation* is assumed to be scale linearly with M and instead of OLS we need to use the generalized least squares method (GLS). Finally, for $\xi = 1/2$ the noise *standard deviation* scales linearly with M (known as Poisson noise) and the GLS method is used to estimate the parameters. The ϵ_i are assumed to be independent, identically distributed random variables (*i.i.d.*) with zero mean and (finite) variance σ_0^2. The random variables Y_i have means given by $E[Y_i] = M(t_i;\Theta_0)$ and variances $\mathrm{Var}[Y_i] = M(t_i;\Theta_0)^{2\xi}\sigma_0^2$, see [19]. We assume that the noise variance in our data in independent and consider $\xi = 0$, hence we use OLS.
The Ordinary Least Squares estimator Θ_{OLS} is a random variable obtained by considering the objective function

$$J(Y|\Theta) = \sum_{i=1}^{n} w_i(Y_i - M(t_i\Theta))^2, \tag{26}$$

where the weights w_i are given by

$$w_i = \frac{1}{M(t_i;\Theta)^{2\xi}}. \tag{27}$$

Since we consider $\xi = 0$, then $w_i = 1$ for all i, and in this case the optimal estimator is given by

$$\Theta_{\mathrm{OLS}} = \arg\min_{\Theta} \quad J(Y|\Theta_o). \tag{28}$$

Parameter Estimation for our model. In order to estimate a set parameters, namely, β and α_2, that are specific to the epidemic we are studying here, we fix all the other parameters in our model. We have fixed the relative measure of infectiousness for the class I_1 to $\delta = 0.4$, as done by [7]. The value for the rate of developing severe symptoms is taken to be $\kappa = \dfrac{1}{3}$ (3 days) as in [25]. For the parameter estimation, we let $\alpha_1 = 0$ to be consistent with our data set, since our data corresponds to the infectious people with severe symptoms (from the class I_2). We have also fixed the recovery rate of individuals with severe symptoms to $\gamma_2 = \dfrac{1}{5}$ (5 days), as seen in [**?**]. We obtained the value of γ_1 from the equation that measures the level of reporting, that is,

$$\frac{j_\infty}{y} = \frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right). \tag{29}$$

Let r be the percentage of reported cases where $\frac{r}{100} = \frac{j_\infty}{y}$(see Eq.(23)). Then, replacing $\alpha_1 = 0$, $\kappa = 1/3$ and $\gamma_2 = 1/5$, we get

$$\frac{r}{100} = \left(\frac{\frac{1}{3}}{\gamma_1+\frac{1}{3}}\right)\left(\frac{\alpha_2}{\alpha_2+\frac{1}{5}}\right). \tag{30}$$

Solving for γ_1, we obtain

$$\gamma_1 = \frac{(500-5r)\,\alpha_2 - r}{3r\,(5\alpha_2+1)}. \tag{31}$$

Therefore we have an expression for γ_1 in terms of α_2 and r provided that $\gamma_1 > 0$ if $\alpha_2 > \frac{r}{5(100-r)}$.

The model parameters $\Theta = (\beta, \alpha_2)$ are fitted by OLS to the initial phase of the incidence of reported cases (20 epidemic days of data, from May 30 to June 20). May 30 was assumed to be the starting point of exponential growth of the epidemic outbreak in Lima. We also assumed that the exponential growth phase continued through June 30 (see Figure 2). We implemented a MATLAB code to estimate our parameters given a set of initial conditions using 20 iterations with a tolerance of 10^{-5}. For different percentages of reporting, $r = 10k\%$, with $k \in \{1, \cdots, 6\}$, we estimate the set of parameters Θ (for $7 \leqslant k \leqslant 10$, the value of γ_1 becomes negative). Given that we do not know what the percentage of reporting was, we explore different parameter values for some of these percentages of reporting. We fixed the level of reporting r to a given value $r \in [10, 60]$ and by means of parameter estimation we obtain values of $\alpha_2 = \alpha_2(r), \beta = \beta(r)$ and $\gamma_1 = \gamma_1(\alpha_2, r)$ as seen in Eq. (31). Figure 5 shows the fitting of the model to the epidemic data using $r = 30\%$. The best approximation of the data that can be obtained in the initial phase of the epidemic are for values of r up to 60%. We also estimated the control reproductive number, R_C, and the final epidemic size proportion, y, for different values of r (see Table 1).

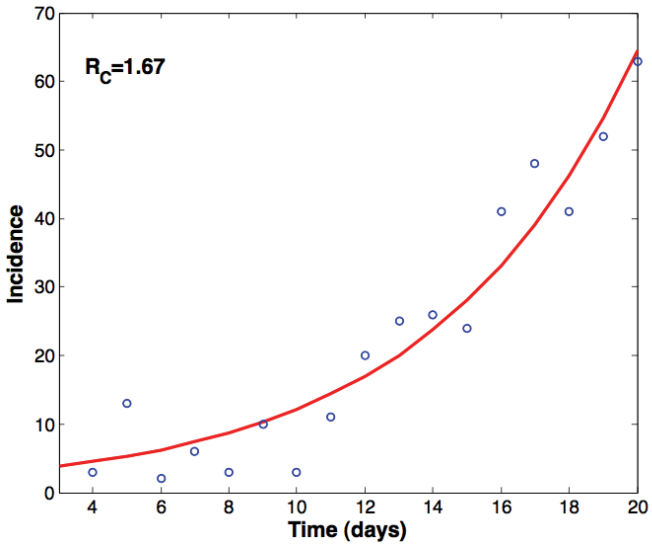


Figure 5: Model fit obtained using $r = 30\%$ and 20 epidemic days of data, from May 30 to June 20, of the 2009 A-H1N1 influenza epidemic in Lima. The plot corresponds to the initial phase of the incidence number of reported cases.

Table 9: Estimates for the set of parameters $\Theta = (\beta, \alpha_2)$, R_C and ξ

$r(\%)$	β	α_2	γ_1	R_C	ξ
10	1.0529	0.0522	0.3566	2.6262	0.9080
20	1.0481	0.1714	0.4358	1.7684	0.7200
30	0.8953	0.2471	0.2807	1.6701	0.6775
40	0.7803	0.2981	0.1654	1.6731	0.6787
50	0.6971	0.3348	0.0840	1.7092	0.6953
60	0.6352	0.3628	0.0248	1.7599	0.7004

Residual plots. As we discussed earlier, the form of the error assumed in our statistical model determines the estimator used. The true form of the error is typically unknown, and there is no way to definitively determine this, but there are two residual tests that can support the chosen assumptions, or indicate that the assumptions may be unreliable [5]. After the estimation procedure is completed for a given set of data, one can plot the residuals $r_i = y_i - M(t_i, \Theta)$ versus time t_i or versus $M(t_i, \Theta)$. If the errors are in fact independent of time *i.i.d.*, a plot of r_i versus t_i would be randomly distributed. Also, if we have assumed constant variance, or that the errors do not depend on the model values $M(t_i, \Theta)$, then a plot of r_i versus $M(t_i, \Theta)$ should also show a random pattern (for more examples and a discussion about error structures, see [5]). Figure 6 shows the residual plot from the best fit solution for the parameters of our data. This plot suggests that it is reasonable to assume constant variance among our observations, providing support for the statistical model underlying the parameter estimation procedure.

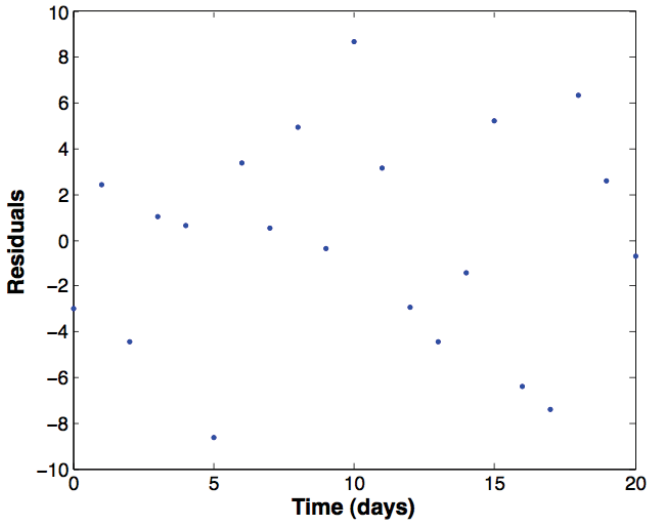


Figure 6: Residuals vs. time. This plot suggests that correct assumptions have been made for the parameter estimation.

G.1 Relationship between the final epidemic size proportion (y) and the percentage of reported cases (r)

We defined the percentage of reported cases, r, as $\frac{r}{100} = \frac{j_\infty}{y}$ where the level of reporting, j_∞, is a constant given by the data. Therefore, the final epidemic size proportion y is inversely proportional to r and given by

$$y(r) = \frac{(100 j_\infty)}{r}$$

Figure 7 provides values of the final epidemic size proportion as a function of reporting (empirically). Different levels of reporting percentage only makes sense for $r < 30$ since after that value of r the graph changes its monotonicity. Looking at this graph of different y values produced with different percentages of reporting, we see that after 30%, y begins to increase. This should not be the case, because in this model, reporting an individual is the same as isolating that individual. Therefore, the more reporting there is, the lower the final size of the epidemic should be. Furthermore, when assuming the data captured more than 30 percent of the actual cases, the estimated parameter values are inconsistent with

what is known to be biologically true. This means we cannot assume the data captured more than 30% of the actual cases of influenza. This is a very important conclusion of this work since it suggests that during the influenza pandemic in Lima, Peru, no more than 30% of the cases were reported.

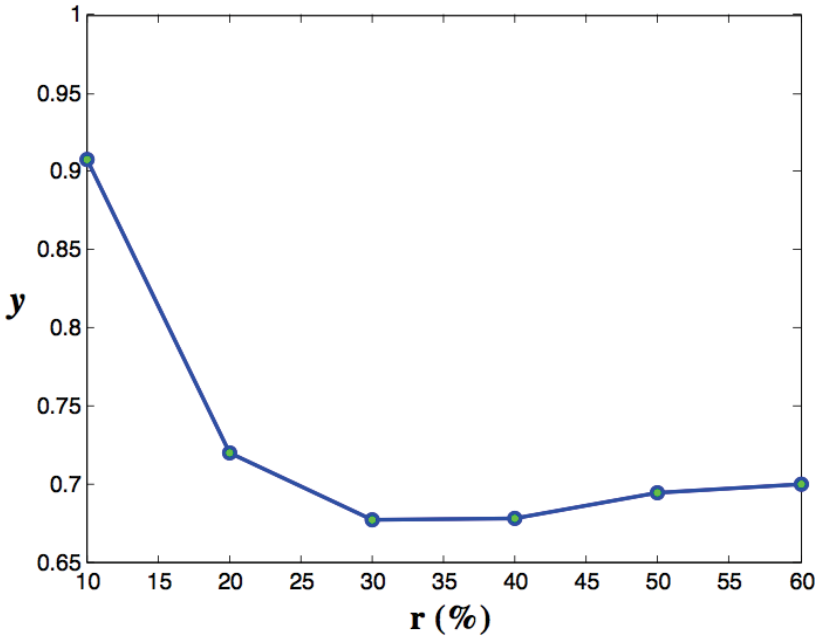


Figure 7: Final epidemic size proportion as a function of the percentage of reporting cases.

H Numerical simulations: role of isolation

In this section is presented the numerical solutions of the model in Equations (2)-(7). We explore dynamical properties of the model that are more difficult to ponder analytically. We make special emphasis in the role of each of the per-capita isolation rates α_1 and α_2 and the effect of social distancing in the final epidemic size.

H.1 The role of α_1

In Lima, Peru only some severe cases were diagnosed and isolated. However we may also introduce public health policies to account for the isolation of the less severe infected individuals. We explore the effects of early diagnose and isolation in the containment of the epidemic by letting $\alpha_1 > 0$. Figure 8 shows that increasing the rate of isolation of less severe cases (I_1) reduces and delay the peak of the epidemic. Comparing Figures 9 and 10, we note that α_1 has a higher impact than α_2 in the final epidemic size reduction and in delaying the epidemic peak time. Furthermore, for smaller values of α_1 we can achieve similar reductions in the final epidemic size than with α_2. For example, a value of $\alpha_1 = 0.3$ reduces the final epidemic size by 80% where as the same value for α_2 reduces the final size by 45%. These simulations show that α_1 is almost two times more effective than α_2 in reducing the total number of infected.

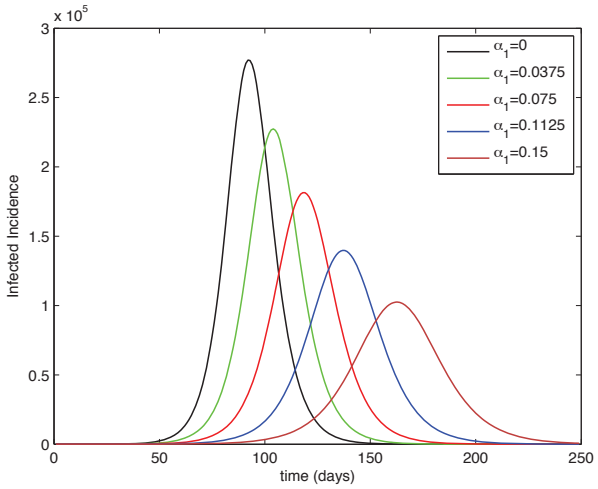


Figure 8: Daily incidence. $\beta = 0.9, \delta = 0.4, \gamma_1 = 0.43, \kappa = 1/3, \gamma_2 = 1/5, \alpha_2 = 0.18$

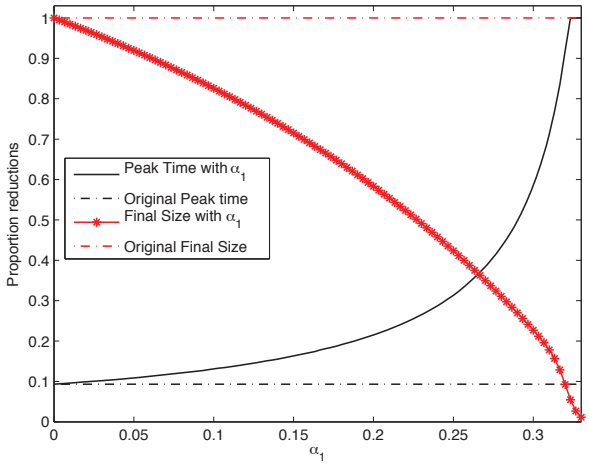


Figure 9: Final size reduction and epidemic's peak delay for α_1 with $\beta = 0.9, \delta = 0.4, \gamma_1 = 0.43, \kappa = 1/3, \gamma_2 = 1/5, \alpha_2 = 0.18$.

H.2 Infection level dependent isolation rate α_2

We assume that the number of individuals that are being isolated with severe infection (I_2) depends on the actual number of infected people at a given moment of time. The term $-\alpha_2 I_2$ in Eq. (4) conveys this idea by stating that the rate of change (outflow to the isolation class) of infected individuals is linear in I_2 with a constant rate α_2. The aim here is to propose and numerically assess the effects of non-constant isolation rates, a rate depending on the number of severely infected individuals. As the epidemic increases it is reasonable to believe that the number of cases being isolated grows as well, not only due to

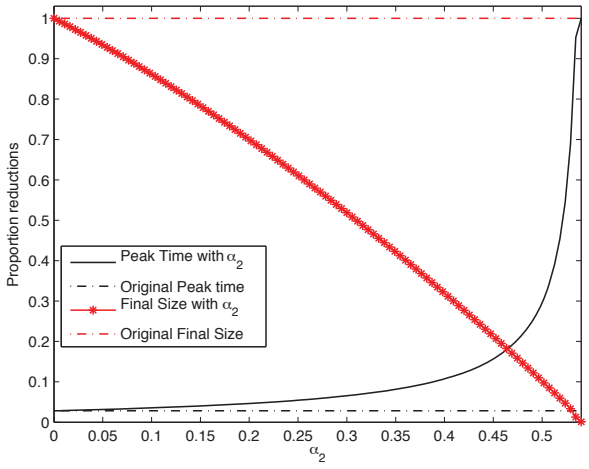


Figure 10: Final size reduction and epidemic's peak delay for α_2 with $\beta = 0.9, \delta = 0.4, \gamma_1 = 0.43, \kappa = 1/3, \gamma_2 = 1/5, \alpha_1 = 0$

an increase of infected individuals, but also as a result of more efficient and intense control measures being put into action to seize the epidemic. However, due to limited resources, the isolation capacity will reach a maximum, namely, α_2^m. A functional dependence that can satisfy all of these conditions is given by a sigmoid function of I_2 provided by

$$\alpha_2(I_2) = 2\alpha_2^m \left(\frac{1}{1 + e^{-aI_2}} - \frac{1}{2} \right). \tag{32}$$

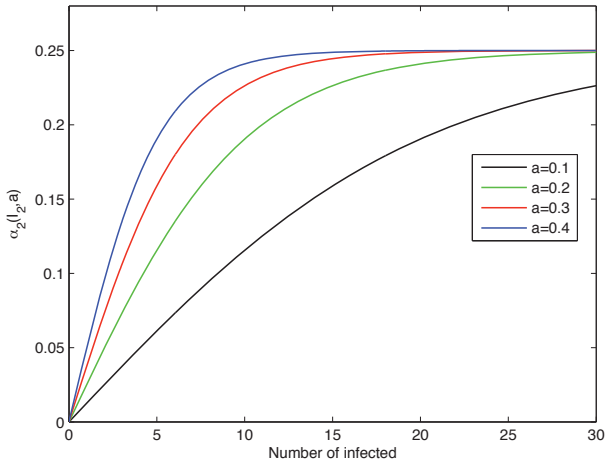


Figure 11: Per-capita isolation rate as a function of severely infected individuals.

Eq. (32) considers the parameter a as a saturation "speed" of the per-capita isolation rate. It is reasonable to conclude that when a gets larger (see Figure 11), α_2 reaches its maximum faster. Note that $\alpha_2(0) = 0$ and $\alpha_2(\infty) = \alpha_2^m$. Therefore, the parameter a is a measure of how fast the saturation value α_2^m is reached. The parameter a can also be viewed as a measure of *how fast* the isolation system reacts to the epidemic: for large a the response to the epidemic is faster.
Figure 12 and 13 show that the final size and the epidemic peak notably changes when a varies. For these graphs we use the parameter values obtained in Section 7 (of the parameter estimation) that is, $\alpha_2^m = 0.25, \beta = 0.9, \delta = 0.4, \gamma_1 = 0.44, \kappa = 1/3, \gamma_2 = 1/5, \alpha_1 = 0$. Moreover, for larger values of a the final size and the delays in the peak are reduced. This suggests that responding faster to the epidemic could reduce the infection level in the population and could also, due to the induced delay, provide a period of time to implement other types of interventions like vaccination campaigns.

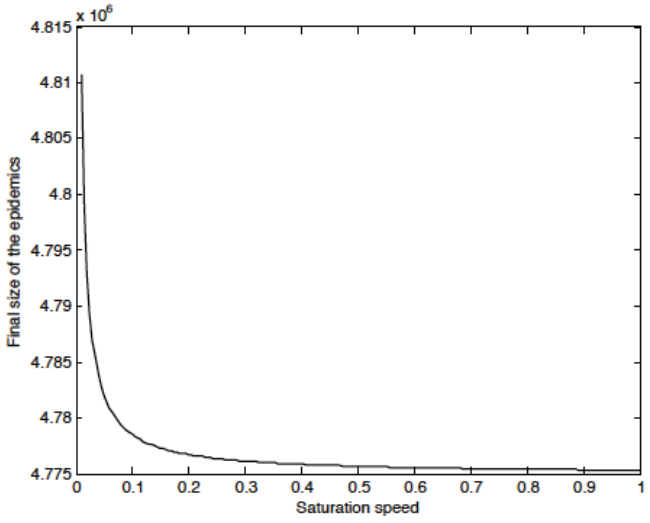


Figure 12: The final size gets reduced as the saturation parameter a increases.

I Numerical simulations: role of temporal social distancing

Social distancing is a public health intervention aimed at reducing the transmission and mitigate disease burden by limiting contact between infectious and non-infectious individuals within the population. Examples of these measures can include school closings, the cancellation of large public gatherings, and the use of facemasks [8].
Recent observational studies support the implementation of school closure interventions to achieve reductions in influenza transmission rates, [8, 22]. In the context of a pandemic, this can be particularly useful to gain time until biomedical resources (vaccines, anti-virals) become available, and to relieve the burden on health care institutions due to a reduced surge of influenza patients. The theory behind the mitigative effect of school closings is that school age children have high contact rates, and tend to be more susceptible to influenza infection than other age groups, [8]. Therefore, reducing transmission in school children may reduce the attack rates of influenza in all age groups.
In general, social distancing is represented in epidemiological models by decreasing the transmission rate (β). There are control measures, such as school closure, that can be incorporated to the model by

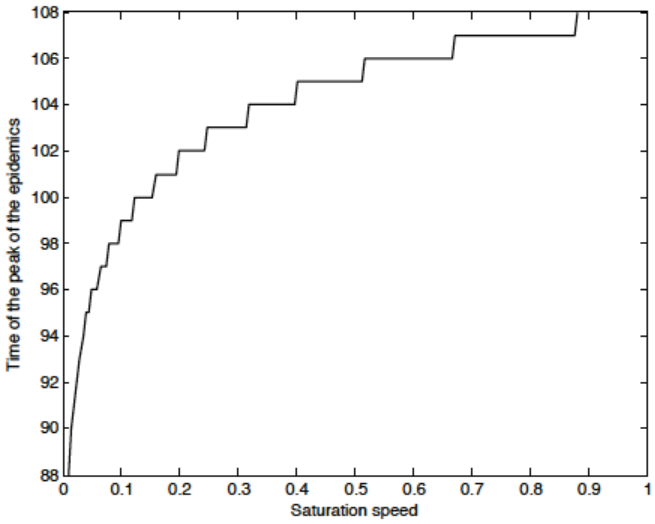
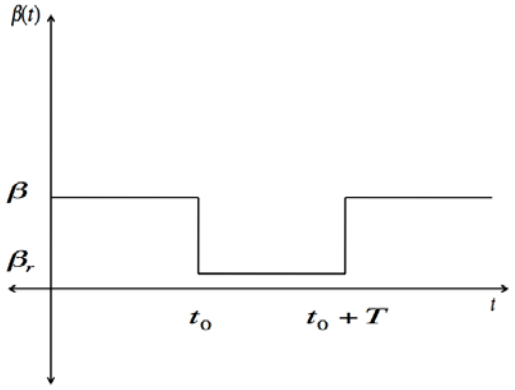


Figure 13: The peak of the epidemics gets delayed with a larger saturation parameter a.

reducing this value *temporarily*, that is for a certain time period $T = 14$ days [18]. Hence, we consider a transmission rate as a function of time , i.e., $\beta = \beta(t)$ (see Figure I) where,
$\beta(t) = \beta_r$, for $0 < t_0 \leq t \leq t_0 + T$
β, otherwise.
The parameter β_r is the reduced force of infection due to social distancing measures (naturally $\beta_r < \beta$) and t_0 is the time at which the social distancing measure is implemented. We numerically investigate the effect of different alternatives when implementing social distancing, namely, timing and reduction level of the control measures. The results presented in this section related to the final size of the epidemic were obtained from simulations that were run long enough (1000 days or more) to ensure all infected individuals were taken into account, either getting isolated or recovered, making sure that the final size of the epidemic was properly computed.



I.1 Social distancing timing

In this section, we investigate the time to apply the social distancing, that is varying the time to introduce the intervention, for fixed values of β and β_r. The social distancing is applied for a fixed period of time of 2 weeks, that is, $T = 14$ days. Figure 14 shows that the daily incidence curves markedly changes whenever the social distancing measure is enforced. It can be seen that during this point of time the social distancing decrease the rate of increasing the incidence. Then, after the intervention is over, the curve tries to take its original course. If the intervention if done just before the peak time without social distancing, that is for values of $t_0 < 75$, the effect decreases the maximum incidence, furthermore it decreases the final size of the epidemic. This behavior can also be explain by Figure 15.

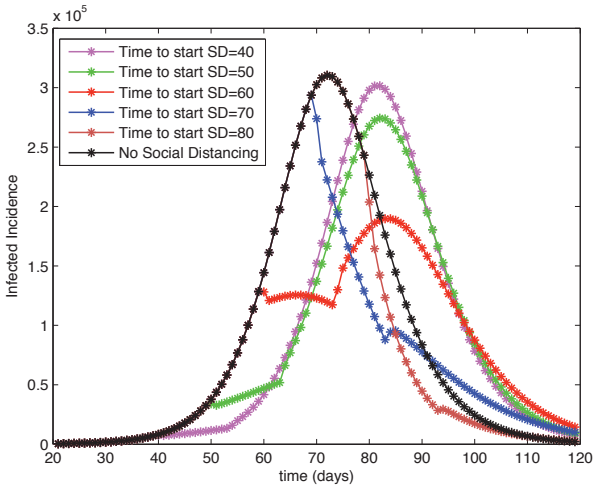


Figure 14: Daily Incidence. $\beta = 0.9, \beta_r = 0.7, \delta = 0.4, \gamma_1 = 0.43, \kappa = 1/3, \gamma_2 = 1/5, \alpha_1 = 0, \alpha_2 = 0.18, T = 14$.

Figure 15 shows a plot of the reduction proportion of the a peak time (time from the onset of the epidemics to the highest peak) and the final size of the epidemic by varying the time of intervention for the social distancing. These quantities are plotted as proportions, that is we normalized the final epidemic size to better capture the effects of social distancing, i.e. each value of the final size at each t_0 was divided by the maximum final epidemic size for all values of t_0.
The same was done for the time of the peak. For example, if $t_0 = 60$ the final size of the epidemic will be reduced by 10%, where as if $t_0 = 40$ the time for the peak delay can be reduced by approximately 12%. Remarkably, there is a time t_0^{min} for which the final size is minimum, that is, for $t_0^{min} \approx 70$. If the social distancing is implemented after or before this optimal time, the effectiveness of the social distancing in lowering the final size diminishes.
In Figure 15 we see a sharp decline in the peak time curve right before the time $t_0^{min} \approx 70$ when the minimum final size is reached. This abrupt shift in the peak time is because at a particular time $60 \leq t_0^s \leq 65$ (s for shift) there are two local maximum with the same daily incidence. For $t_0 << t_0^s$ the only benefit we see is to delay the time of the epidemic. However, for $t_0 >> t_0^s$, the social distancing has merely no effect on the dynamics of the epidemic.

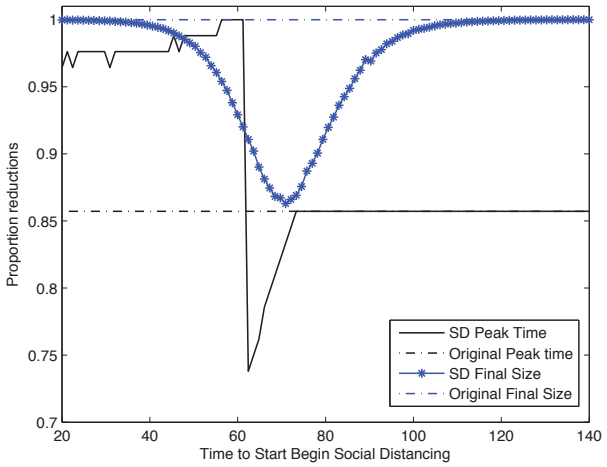


Figure 15: Peak time and final size.

I.2 Varying the social distancing level

In this section we vary level of social distancing by means of temporarily reducing the transmission rate β with different intensities. We ponder the effects of different values of β_r (recall Eq.(I)) on the global dynamics of the epidemics, specially in the epidemic's final size. Figure 16 plots daily incidence for different values of β_r, all applied at 65 days since the beginning of the epidemics and for a period T of two weeks.

Figure 16: Daily incidence. $\beta = 0.9, \delta = 0.4, \gamma_1 = 0.43, \kappa = 1/3, \gamma_2 = 1/5, \alpha_1 = 0, \alpha_2 = 0.18, t_0 = 65, T = 14$.

Figure 16 shows that the peak is delayed when reducing the transmission rate β. However, the size of the peaks begin to increase as well. Hence, the impact on the final size that has reducing β remains unclear, given that the final size is proportional to the area under the incidence curves. Figures 17 and 18 will help to clarify this situation.
In Figure 17, we plot the final size proportion as a function of the reduced β, that is β_r. The final size will reach a minimum value when $\beta_r^o \approx 0.6$ (see Figure 17). Thus, there is an optimal value β_r^o such that the epidemic will have a minimum impact in the susceptible population. Decreasing the force of infection too much (though not enough to make $\beta_r = 0$) could have a negative effect on the population, compared to what could have happened if $\beta_r \approx \beta_r^o$. This can be explained in the following way: by reducing β too much, we are leaving a large pool of susceptible in the population that are not getting infected in the first smaller wave that forms when SD is implemented (see figure above). At the same time, by not making $\beta_r = 0$ for a time long enough so that all infected can recover, we are ensuring that the disease is still prevalent in the population. Therefore, for a large reduction in β (but not down to zero) the effect of SD is almost equivalent to just delay the peak of the incidence curve. Of course, such is still a positive result, given the time is being provided to implement other alternatives into halting the course of the disease. In Figure 18 we can verify how reducing the force of infection delays the peak of the epidemics.

Figure 17: Final Size reduction. $\beta = 0.9, \delta = 0.4, \gamma_1 = 0.43, \kappa = 1/3, \gamma_2 = 1/5, \alpha_1 = 0, \alpha_2 = 0.18, t_0 = 65, T = 14$.

Figure 18: Delay of the peak of the epidemics.

J Uncertainty analysis for R_C

We performed an uncertainty analysis on the control reproductive number R_C by varying the parameter values. We used Monte Carlo simulations (simple random sampling) to quantify the uncertainty of R_C when the model parameters are randomly distributed. A probability density function (PDF) was assigned to parameters $(\beta, \delta, \kappa, \alpha_2, \gamma_1, \gamma_2)$. We sampled this set of parameters 10^3 times for different values of r, holding $\alpha_1 = 0$, $\kappa = \frac{1}{3}, \delta = \frac{2}{5}$ and $\gamma_2 = \frac{1}{5}$ fixed and assuming that β has an exponential distribution, δ a uniform distribution and α_1, α_2, γ_1 and γ_2 a gamma distribution. The parameter value of κ was considered constant. Then, we computed R_C from each set. The distribution of R_C lies in the mean range $(1.67, 2.62)$, depending on the percentage of reporting (see Figure 20). Each distribution of R_C is characterized by its mean, standard deviation and median (see Table 10).

Table 10: Results of uncertainty for R_C for each percentage of reporting.

$r(\%)$	Mean of R_C	Standard Deviation of R_C	Median of R_C
10	2.68	3.32	1.56
20	1.78	3.50	1.61
30	1.67	2.04	1.02

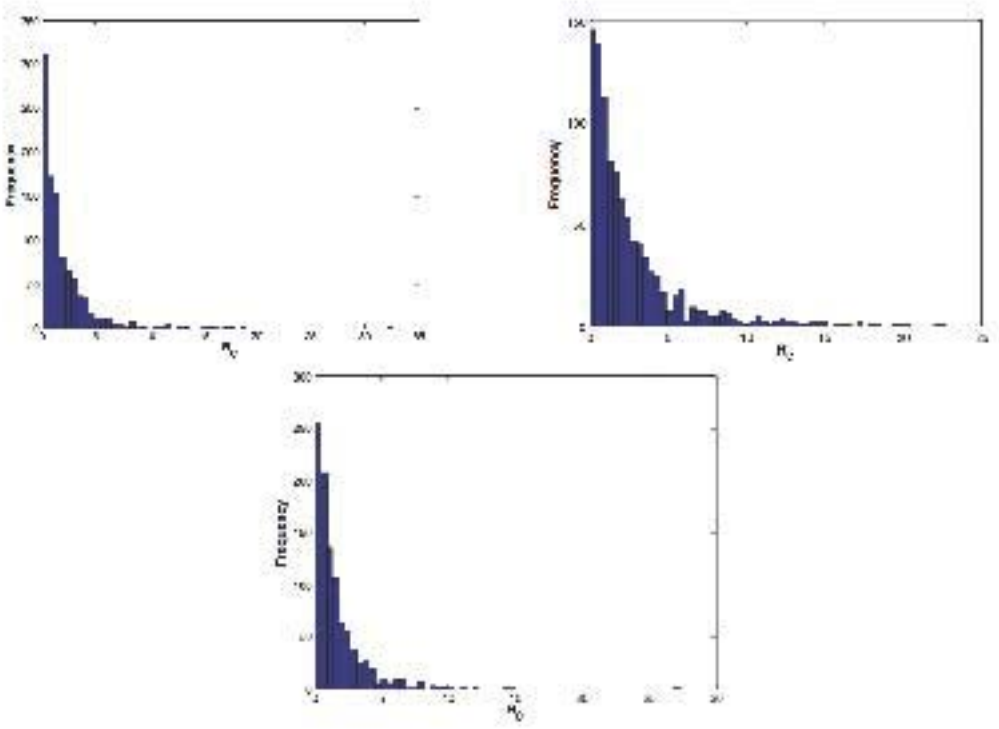


Figure 19: Histograms for R_C, for each percetange of reporting, the mean of R_C is 2.68, 1.78, 1.67, for $r = 10, 20, 30\%$, respectively.

K Optimizing α_1 and α_2

Next we address the question of what are the optimal values of α_1 and α_2 so that the final epidemic size, given by Eq.(20), is minimum under the constraint

$$B = c_1\alpha_1 + c_2\alpha_2. \tag{33}$$

This relation represents a cost-dependent isolation per-capita rate. The value B could represent the"budget" per unit time aimed at controlling the disease via isolation efforts, i.e., its units can be seen as [cost unit]/[day]. The parameters c_1 and c_2 are the respective costs of the per-capita isolation rates from I_1 and I_2 to J. We reasonably assume that $c_1 > c_2$ since detecting earlier cases could be clinically more difficult and also socially costly. The first reason comes from the fact that detecting the disease in those who do not present clear or severe symptoms is medically more expensive due to an increase in testing runs and the medical and technical staff needed to carry them out. The second reason stems from the fact that, if a patient do not present severe symptoms, i.e., that person may not be sick from influenza, and it is expensive to isolate such individuals.
We also know from the previous section that the final epidemic size is proportional to R_C. Hence, if we want to find an optimal relation between α_1 and α_2 to minimize the final size y is equivalent to minimize

R_C with respect to those parameters. We would then have the following optimization problem

$$\begin{aligned} min\left[R_C(\alpha_1,\alpha_2)\right] &= min\left[\beta\left(\frac{\delta}{\kappa+\gamma_1+\alpha_1}+\frac{\kappa}{(\kappa+\gamma_1+\alpha_1)(\gamma_2+\alpha_2)}\right)\right] \\ \text{such that} \quad 0 &= b-\alpha_1-c\alpha_2 \\ 0 &< \alpha_1 \\ 0 &< \alpha_2 \end{aligned}$$

where we have introduced the unitless quantity c, relative cost, given by $c = c_2/c_1 < 1$, and the parameter b defined as B/c_1, with units of [1/day]. We have naturally constrained α_1 and α_2 to be strictly positive. Our Lagrange function, with multipliers λ, μ_1 and μ_2 is given by

$$\Lambda(\alpha_1,\alpha_2,\lambda) = \beta\left(\frac{\delta}{\kappa+\gamma_1+\alpha_1}+\frac{\kappa}{(\kappa+\gamma_1+\alpha_1)(\gamma_2+\alpha_2)}\right) - \lambda\left[b-\alpha_1-c\alpha_2\right] - \mu_1\alpha_1 - \mu_2\alpha_2,$$

and we need to solve $\nabla\Lambda_{(\alpha_1,\alpha_2,\lambda,\mu_1,\mu_2)} = 0$, that is

$$\begin{aligned} \frac{\partial\Lambda}{\partial\alpha_1} &= -\frac{\beta\delta}{(\gamma_1+\alpha_1+\kappa)^2} - \frac{\beta\kappa}{(\gamma_1+\alpha_1+\kappa)^2(\alpha_2+\gamma_2)} + \lambda + \mu_1 = 0, \\ \frac{\partial\Lambda}{\partial\alpha_2} &= -\frac{\beta\kappa}{(\gamma_1+\alpha_1+\kappa)(\alpha_2+\gamma_2)^2} + \lambda c + \mu_2 = 0, \\ \frac{\partial\Lambda}{\partial\lambda} &= b-\alpha_1-c\alpha_2 = 0, \\ \mu_1\alpha_1 &= 0, \\ \mu_2\alpha_2 &= 0. \end{aligned}$$

Since α_1 and α_2 can not be zero, the solution for those two conditions is $\mu_1 = \mu_2 = 0$. Isolating λ and α_1 we obtain

$$\frac{\beta\delta}{(\gamma_1+\alpha_1+\kappa)^2} + \frac{\beta\kappa}{(\gamma_1+\alpha_1+\kappa)^2(\alpha_2+\gamma_2)} = \frac{\frac{\beta\kappa}{c}}{(\gamma_1+\alpha_1+\kappa)(\alpha_2+\gamma_2)^2},$$
$$\alpha_1 = b - c\alpha_2,$$

which leads to the following quadratic equation in α_2

$$(b-c\alpha_2+\gamma_1+\kappa)\frac{\kappa}{c} = (\alpha_2+\gamma_2)^2\delta + \kappa(\alpha_2+\gamma_2),$$

that has a pair of solutions given by

$$\alpha_2^o = \frac{-\kappa c - \delta c\gamma_2 \pm \sqrt{\kappa c(\kappa c + \delta c\gamma_2 + b\delta + \delta\gamma_1 + \delta\kappa)}}{\delta c}$$

We take the biologically significant one (discard the negative solution) for the optimal value of α_2 and its corresponding solution for α_1 as given below

$$\alpha_2^o = -\frac{k}{\delta} - \gamma_2 + \sqrt{\frac{\kappa}{\delta}\left(\frac{\kappa}{\delta}+\gamma_2+\frac{b+\gamma_1+\kappa}{c}\right)}, \tag{34}$$

$$\alpha_1^o = b + c\left[\frac{\kappa}{\delta}+\gamma_2-\sqrt{\frac{\kappa}{\delta}\left(\frac{\kappa}{\delta}+\gamma_2+\frac{b+\gamma_1+\kappa}{c}\right)}\right]. \tag{35}$$

Notice that we have not checked for the positivity of the above solutions. Moreover, there is an infinite number of parameter sets such that α_1^o or α_2^o can be negative. The expressions above are to be taken only for those parameter sets that make both α_1 and α_2 simultaneously positive.

So far we have proven necessary conditions for $R_C(\alpha_1^o, \alpha_2^o)$ to be a minimum or R_C. Next we show sufficiency conditions to guarantee that $\forall(\alpha_1, \alpha_2) \in \mathbb{R}^+, R_C(\alpha_1, \alpha_2) \geq R_C(\alpha_1^o, \alpha_2^o)$. To prove that a multivariable function has a minimum at a given critical point of its domain when there are no constraints, the Hessian matrix's determinant and trace at that point have to be positive definite. These two conditions, along with the fact that the Hessian is a symmetric matrix, would guarantee that the eigenvalues of the Hessian are positive, hence, at the critical point R_C has a minimum. We rewrite R_C, for the sake of simplicity, as follows

$$R_C(\alpha_1, \alpha_2) = \frac{A}{(A_1 + \alpha_1)} + \frac{B}{(A_1 + \alpha_1)(\alpha_2 + B_1)},$$

where the new parameters are given by $A = \beta\delta, A_1 = \gamma_1 + \kappa, B = \beta\kappa, B_1 = \gamma_2$. Notice that they are all positive, given their biological significance. It can be easily shown that the determinant of the Hessian of R_C is given by

$$\begin{aligned} H_{R_C} &= \begin{vmatrix} \frac{\partial^2 R_C}{\partial \alpha_1^2} & \frac{\partial^2 R_C}{\partial \alpha_1 \partial \alpha_2} \\ \frac{\partial^2 R_C}{\partial \alpha_2 \partial \alpha_1} & \frac{\partial^2 R_C}{\partial \alpha_2^2} \end{vmatrix}, \\ &= \frac{B(4A\alpha_2 + 4AB_1 + 3B)}{(A_1 + \alpha_1)^4(\alpha_2 + B_1)^4}, \end{aligned}$$

which is always positive for any positive value of α_2, given that all the other parameters are also positive. The trace, in general, is given by

$$\frac{2(BA_1^2 + 2BA_1\alpha_1 + B(\alpha_1)^2 + A(\alpha_2)^3 + 3A(\alpha_2)^2 B_1 + 3A\alpha_2 B_1^2 + AB_1^3 + B(\alpha_2)^2 + 2B\alpha_2 B_1 + BB_1^2)}{(A_1 + \alpha_1)^3(\alpha_2 + B_1)^3}$$

This trace is positive for all $\alpha_2^o > 0$. Therefore, for any critical point $\alpha_2^o > 0$, and with no regard for the constraint α_1^o and α_2^o are subject to, we can assure that at (α_1^o, α_2^o), R_C has a minimum in $\mathbb{R}^{2+}$. Therefore, we have shown the existence of an analytical expression for the optimal values of a cost-dependent relation of the per-capita isolation rates. That is, knowing the parameter values and the relative cost of each of the isolation rates, we can compute precise values for these per-capita isolation rates so that the epidemics takes a lesser toll on the population.

L Conclusions

We considered the case of the 2009 A-H1N1 influenza outbreak in Lima, Peru and have studied the impact of the unreported cases in the calculation of the final epidemic size. To accomplish this, a system of non-linear ordinary differential equations was constructed. The final size relations for the total number of infected individuals and for the total number of reported cases were computed. These size relations were used to obtain an expression for the level of unreporting.

Estimating model parameters via ordinary least-squares fitting of the model solution to the observed data we have measured the reliability of this parameter estimation under different scenarios of reporting, concluding that we cannot assume the data captured more than 30% of the actual cases of influenza during the initial phase of the epidemic.

An uncertainty analysis was conducted on R_C in order to determine its range of values, depending on the amount of variation in the parameter values. We found that the distribution of R_C lies in the mean range (1.67, 2.62), depending on the percentage of reporting that was assumed. As the percentage of reporting increases, the mean of R_C decreases, which is due to the fact that a larger percentage of individuals are being isolated.

In addition, a sensitivity analysis was also performed for the control reproduction number and for the percentage of non-reporting. We found that R_C is most sensitive to changes in the transmission rate,

followed by α_2. However, as the percentage of reporting is increased, the influence of α_2 on R_C is decreased. The R_C is directly related to the final epidemic size, then, for a control policy it is a better choice to decrease the transmission rate (social distancing), instead of increasing the isolation rate. However, if we focus on the non-reporting, the isolation rate is the better choice for increasing the reporting level.

Isolation of infectious individuals is a strategy used to contain a disease. Usually, only individuals who exhibit severe symptoms are isolated. Looking at the effects that isolating individuals with less severe symptoms would have on the incidence, we were able to see how effective early isolation is. By isolating individuals with less severe symptoms, the size of the peak can be greatly reduced. Also, early isolation (α_1) is more effective than late isolation (α_2) at delaying the time of the peak.

When we investigated an isolation rate that depends on the amount of infected individuals, we could study the effects that the speed at which the isolation system reacts to the epidemic would have on the final epidemic size and on the time of the peak. We found that increasing the speed at which the isolation system responds to the epidemic would always lead to a reduction in the final size of the epidemic, and would also always lead to a delay in the time of the peak.

In order to control an epidemic, temporary social distancing measures are often put into effect. These measures lead to a decrease in the transmission rate, β, since the amount of contacts between infectious individuals and susceptible is reduced. Via numerical simulations we the existence of an optimal value that β should be reduced to in order to obtain the smallest final epidemic size. Furthermore, the timing of the reduction is also important in terms of the social distancing effectiveness. Even though there is an optimal time and transmission rate reduction so that the final size is minimum, social distancing always result in a delay of the peak of the epidemic.

Given that isolating individuals with less severe symptoms is more effective in halting the epidemic, and since it is reasonable to assume that doing so is more expensive, socially and economically, than isolating individuals with severe symptoms, using an optimization problem we found analytical expressions for the two optimal isolation rates as a function of the model parameters and their relative cost.

M Acknowledgments

Special thanks to Dr. Carlos Castillo-Chavez for the opportunity to participate in MTBI. Thanks to Mayteé Cruz-Aponte for all her valuable input and Dr. Luis Melara for his feedback.

This research was conducted in the Mathematical and Theoretical Biology Institute (MTBI) at the Mathematical, Computational and Modeling Sciences Center (MCMSC). This project has been partially supported by grants from the National Science Foundation (NSF - Grant DMPS-0838705), the National Security Agency (NSA - Grant H98230-11-1-0211), the Alfred P. Sloan Foundation and the Office of the Provost of Arizona State University.

Appendix

A Calculations of the unreported and final epidemic size

A.1 Final epidemic size calculation

Calculations of the final epidemic size [4, 6] are shown below. We proceed getting the expression for $I_2(t)$ using the fact that

$$I_2' = -S' - I_1' - (\alpha_1 + \gamma_1)I_1 - (\alpha_2 + \gamma_2)I_2, \tag{36}$$

where I_1 and I_1' are fixed. To solve Eq. (12) for I_2, we us an integrating factor $\mu = e^{\int(\alpha_2+\gamma_2)dt}$. Knowing that $I_2(0) = 0$ (initially no severe infection), we obtain

$$\begin{aligned}\left(I_2 e^{(\alpha_2+\gamma_2)t}\right)' &= \left(-S' - I_1' - (\alpha_1+\gamma_1)I_1\right) e^{(\alpha_2+\gamma_2)t}, \\ \implies I_2(t) &= \int_0^t \left[-S' - I_1' - (\alpha_1+\gamma_1)I_1\right] e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau. \end{aligned} \tag{37}$$

Similarly, we solve the following for $I_1(t)$:

$$I_1' = -S' - (\alpha_1+\gamma_1+\kappa)I_1, \tag{38}$$

using an integrating factor $\mu = e^{\int(\alpha_1+\gamma_1+\kappa)dt}$ and with initial condition $I_1(0) = 0$. This gives

$$\begin{aligned}\left(I_1 e^{(\alpha_1+\gamma_1+\kappa)t}\right)' &= -S' e^{(\alpha_1+\gamma_1+\kappa)t}, \\ \implies I_1(t) &= \int_0^t -S' e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} d\tau. \end{aligned} \tag{39}$$

To find the final size equation we divide Eq.(2) by S and use Eqs. (13) and (14) to get

$$-\frac{S'(t)}{S(t)} = \frac{\beta(\delta I_1 + I_2)}{N-J}, \tag{40}$$

$$\begin{aligned}-\frac{S'(t)}{S(t)} &= \frac{\beta}{N-J}\delta \int_0^t -S' e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} d\tau \\ &\quad + \frac{\beta}{N-J}\int_0^t \left(-S' - I_1' - (\alpha_1+\gamma_1)I_1\right) e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau. \end{aligned} \tag{41}$$

To simplify the following calculation, we consider the number of isolated individuals to be negligible compared to the total population, then $(N-J) \approx N$. Integrating with respect to time from 0 to ∞, we obtain:

$$\begin{aligned}\ln\frac{N}{S_\infty} &= \frac{\beta}{N}\delta \int_0^\infty \int_0^t -S'(\tau) e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} d\tau dt \\ &\quad + \frac{\beta}{N}\int_0^\infty \int_0^t \left(-S'(\tau) - I_1'(\tau) - (\alpha_1+\gamma_1)I_1(\tau)\right) e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt, \end{aligned} \tag{42}$$

$$\begin{aligned} &= \frac{\beta}{N}\int_0^\infty \int_0^t -S'(\tau)\left(\delta e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} + e^{(\alpha_2+\gamma_2)(\tau-t)}\right) d\tau dt \\ &\quad + \frac{\beta}{N}\int_0^\infty \int_0^t \left(-I_1'(\tau) - (\alpha_1+\gamma_1)I_1(\tau)\right) e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt, \end{aligned} \tag{43}$$

$$\begin{aligned}\ln\frac{N}{S_\infty} &= \frac{\beta}{N}\int_0^\infty \int_\tau^\infty -S'(\tau)\left(\delta e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} + e^{(\alpha_2+\gamma_2)(\tau-t)}\right) dt d\tau \\ &\quad + \frac{\beta}{N}\int_0^\infty \int_\tau^\infty \left(-I_1'(\tau) - (\alpha_1+\gamma_1)I_1(\tau)\right) e^{(\alpha_2+\gamma_2)(\tau-t)} dt d\tau. \end{aligned} \tag{44}$$

Where the last step involved a change in order of integration. Letting $u = -\tau + t$ gives

$$\begin{aligned}\ln\frac{N}{S_\infty} &= \frac{\beta}{N}\int_0^\infty \int_0^\infty -S'(\tau)\left(\delta e^{-(\alpha_1+\gamma_1+\kappa)u} + e^{-(\alpha_2+\gamma_2)u}\right) du d\tau \\ &\quad + \frac{\beta}{N}\int_0^\infty \int_0^\infty \left(-I_1'(\tau) - (\alpha_1+\gamma_1)I_1(\tau)\right) e^{-(\alpha_2+\gamma_2)u} du d\tau. \end{aligned} \tag{45}$$

We evaluate the integral in the first term and obtain

$$\begin{aligned}
&\int_0^\infty \int_0^\infty -S'(\tau)\left(\delta e^{-(\alpha_1+\gamma_1+\kappa)u} + e^{-(\alpha_2+\gamma_2)u}\right) du d\tau \\
&= \int_0^\infty -S'(\tau)d\tau \left(\int_0^\infty \left(\delta e^{-(\alpha_1+\gamma_1+\kappa)u} + e^{-(\alpha_2+\gamma_2)u}\right) du\right), \\
&= \int_0^\infty -S'(s)d\tau \left(\frac{\delta}{\alpha_1+\gamma_1+\kappa} + \frac{1}{\alpha_2+\gamma_2}\right), \\
&= (N - S_\infty)\left(\frac{\delta}{\alpha_1+\gamma_1+\kappa} + \frac{1}{\alpha_2+\gamma_2}\right). \qquad (46)
\end{aligned}$$

Substituting Eq. (**??**) into Eq. (16), we obtain

$$\begin{aligned}
\ln\frac{N}{S_\infty} &= \frac{\beta}{N}(N - S_\infty)\left(\frac{\delta}{\alpha_1+\gamma_1+\kappa} + \frac{1}{\alpha_2+\gamma_2}\right) - \frac{\beta}{N}\int_0^\infty \int_0^\infty -I_1'(\tau)e^{-(\alpha_2+\gamma_2)u} du d\tau \\
&\quad -\frac{\beta}{N}(\alpha_1+\gamma_1)\int_0^\infty \int_0^\infty I_1(\tau)e^{-(\alpha_2+\gamma_2)u} du d\tau.
\end{aligned}$$

Then, using

$$\int_0^\infty -I_1'(\tau)d\tau = I_1(\infty) - I_1(0) = 0,$$

gives

$$\ln\frac{N}{S_\infty} = \frac{\beta}{N}(N - S_\infty)\left(\frac{\delta}{\alpha_1+\gamma_1+\kappa} + \frac{1}{\alpha_2+\gamma_2}\right) - \frac{\beta(\alpha_1+\gamma_1)}{N(\alpha_2+\gamma_2)}\int_0^\infty I_1(\tau)d\tau. \qquad (47)$$

To evaluate the integral in Eq. 17 we use Eq.(14) and obtain

$$\begin{aligned}
\ln\frac{N}{S_\infty} &= \left(1 - \frac{S_\infty}{N}\right)\left(\frac{\beta\delta}{\alpha_1+\kappa+\gamma_1}\frac{\beta}{\alpha_2+\gamma_2}\right) \\
&\quad -\left(\frac{\beta}{\alpha_2+\gamma_2}\right)\left(\frac{\alpha_1+\gamma_1}{N}\right)\int_0^\infty \int_0^\tau -S'(s)e^{(\alpha_1+\gamma_1+\kappa)(s-\tau)} ds d\tau, \\
&= \left(1 - \frac{S_\infty}{N}\right)\left(\frac{\beta\delta}{\alpha_1+\kappa+\gamma_1} + \frac{\beta}{\alpha_2+\gamma_2}\right) \\
&\quad -\left(\frac{\beta}{\alpha_2+\gamma_2}\right)\left(\frac{\alpha_1+\gamma_1}{N}\right)\left[(N - S_\infty)\left(\frac{1}{\alpha_1+\kappa+\gamma_1}\right)\right], \\
&= \left(1 - \frac{S_\infty}{N}\right)\left[\frac{\beta\delta}{\alpha_1+\kappa+\gamma_1} + \frac{\beta}{\alpha_2+\gamma_2}\left(\frac{\kappa}{\alpha_1+\kappa+\gamma_1}\right)\right], \\
\ln\frac{N}{S_\infty} &= \left(1 - \frac{S_\infty}{N}\right)R_C. \qquad (48)
\end{aligned}$$

Equation (18) is the typical final size relation [4,6]. If we let $s_\infty = \frac{S_\infty}{N}$ as the proportion of the final susceptible size Eq. (18) yields

$$\ln(s_\infty) = (s_\infty - 1)R_C, \qquad (49)$$

where $0 < s_\infty < 1$ and we have changed the argument in the logarithm and hence the sign of the right hand term. Here $1 - s_\infty$ represents the number of disease cases over the course of the epidemic.

A.2 Final reported cases (J_∞)

From Eq.(6), and with initial condition $J(0) = 0$, we have

$$\int_0^\infty J' dt = \int_0^\infty (\alpha_1 I_1 + \alpha_2 I_2)\, dt. \tag{50}$$

Therefore, by Eqs.(13)-(14) we have,

$$\begin{aligned}
J_\infty &= \alpha_1 \int_0^\infty \int_0^t -S' e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} d\tau dt \\
&\quad +\alpha_2 \int_0^\infty \int_0^t \left[-S' - I_1' - (\alpha_1+\gamma_1) I_1\right] e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt, \\
&= \alpha_1 \int_0^\infty \int_0^t -S' e^{(\alpha_1+\gamma_1+\kappa)(\tau-t)} d\tau dt + \alpha_2 \int_0^\infty \int_0^t -S' e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt \\
&\quad +\alpha_2 \int_0^\infty \int_0^t -I_1' e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt \\
&\quad +\alpha_2(\alpha_1+\gamma_1) \int_0^\infty \int_0^t -I_1 e^{(\alpha_2+\gamma_2)(\tau-t)} d\tau dt.
\end{aligned} \tag{51}$$

The four integrals in Eq.(**??**) further simplify (as done to find the final size relation), providing the final number of reported cases, given by

$$\begin{aligned}
J_\infty &= \frac{\alpha_1(N-S_\infty)}{\alpha_1+\gamma_1+\kappa} + \frac{\alpha_2(N-S_\infty)}{\alpha_2+\gamma_2} - \left(\frac{\alpha_2(\alpha_1+\gamma_1)}{\alpha_2+\gamma_2}\right)\left(\frac{N-S_\infty}{\alpha_1+\gamma_1+\kappa}\right), \\
&= (N-S_\infty)\left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \frac{\alpha_2}{\alpha_2+\gamma_2} - \left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\left(\frac{\alpha_1+\gamma_1}{\alpha_1+\gamma_1+\kappa}\right)\right], \\
&= (N-S_\infty)\left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \frac{\alpha_2}{\alpha_2+\gamma_2}\left(1 - \frac{\alpha_1+\gamma_1}{\alpha_1+\gamma_1+\kappa}\right)\right], \\
J_\infty &= (N-S_\infty)\left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right].
\end{aligned} \tag{52}$$

Dividing by the total population N, and defining j_∞ as the final size proportion of reported cases, Eq. (22) can be expressed as

$$\begin{aligned}
j_\infty &= (1-s_\infty)\left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right], \\
&= y\left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right].
\end{aligned}$$

We obtain an equivalent expression for the level of non-reporting given by

$$\begin{aligned}
u_\infty &= 1 - \frac{j_\infty}{y}, \\
&= 1 - \left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right],
\end{aligned} \tag{53}$$

where $\frac{j_\infty}{y}$ is the level of the final size proportion of reported cases.

B Bounds for Levels of reporting and unreporting

B.1 An upper for the level of reporting

Equation (**??**) provides the level of the final size proportion of reported cases. We further proceed to analyze this expression in order to provide a bound to $\frac{j_\infty}{y}$. Under different levels of infection for an influenza outbreak, the rate of progression from class I_1 to I_2 is positive, that is $\kappa > 0$ and parameters for the recovery and isolated cases are $\alpha_1, \alpha_2, \gamma_1, \gamma_2 \geq 0$, we get the following expressions

1. $\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} \leq \frac{\alpha_1}{\alpha_1+\kappa}$, the probability of isolation from class I_1 without recovery is larger than with recovery,
2. $\frac{\kappa}{\alpha_1+\gamma_1+\kappa} \leq \frac{\kappa}{\alpha_1+\kappa}$, the probability of progression to I_1 from I_2 is longer without recovery than with recovery,
3. $\frac{\alpha_2}{\alpha_2+\gamma_2} \leq \frac{\alpha_2}{\alpha_2} = 1$, for individuals with more severe infections I_2 the probability of becoming isolated is always one provided no recovery.

Then,

$$\begin{aligned} \frac{j_\infty}{y} &= \frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} + \left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right), \\ &\leq \frac{\alpha_1}{\alpha_1+\kappa} + \frac{\kappa}{\alpha_1+\kappa} = 1. \end{aligned} \tag{54}$$

From Eq. (54) we get that $\frac{j_\infty}{y} = 1$ if $\gamma_1 = 0$ and $\gamma_2 = 0$, which means that under no recovery every infected individual gets reported.
Under a typical scenario of influenza epidemiological surveillance (data) only individuals with severe symptoms are being isolated ($\alpha_1 = 0$), then Eq.(**??**) becomes

$$\frac{j_\infty}{y} = \left(\frac{\kappa}{\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right), \tag{55}$$

thus we can bound Eq.(55) with $\gamma_1 > 0$, obtaining

$$\frac{j_\infty}{y} = \left(\frac{\kappa}{\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right) < \left(\frac{\kappa}{\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right), \tag{56}$$

which implies

$$\frac{j_\infty}{y} < \frac{\alpha_2}{\alpha_2+\gamma_2}, \tag{57}$$

under $\alpha_1 = 0$, that is when less severe infections are not reported. Equation (57) provides an upper bound for the reporting level. Since in a real epidemic scenario we cannot control the value of γ_2, this upper bound only depends on the isolation level of people with severe symptoms (α_2), for fixed values of γ_2 .

B.1.1 Asymptotic behavior of the non-reporting u_∞

The level of non-reporting $u_\infty \to l$, $l < \infty$ when $\alpha_2 \to \infty$, but there is no asymptote for α_1. The explanation for this is connected to their "position" in the chain of the infection-isolation events. The parameter α_1 acts on those just infected, or, in the first chain of events, and for large α_1 the infected will go from I_1 to J at a much higher rate than to R or I_2 that is, if

$$\alpha_1 \gg \{\gamma_1, \kappa\} \implies \frac{\alpha_1}{\alpha_1+\gamma_1+\kappa} \gg \left\{\frac{\gamma_1}{\alpha_1+\gamma_1+\kappa}, \frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right\}. \tag{58}$$

From the total proportion of non-reported relation, Eq.(23), we have $\lim_{\alpha_1 \to \infty} u_\infty = 0$. This suggests that under perfect isolation of less severe infected individuals the non-reported will be minimal. The vast majority of less severely infected individuals will be reported. In this case, the model behaves like a classic SIR with the R class representing the isolated class.
For α_2 we have that, given that it acts on I_2, there is a whole infectious compartment, I_1, that can escape from this isolation process. From Eq.(23) we then have

$$\begin{aligned} \lim_{\alpha_2 \to \infty} u_\infty &= 1 - \left(\frac{\alpha_1}{\alpha_1 + \gamma_1 + \kappa} + \frac{\kappa}{\alpha_1 + \gamma_1 + \kappa} \right), \\ &= \frac{\gamma_1}{\alpha_1 + \gamma_1 + \kappa}. \end{aligned}$$

that is, under perfect isolation of severe infected individuals (I_2), individuals with less severe infections (I_1) will still recover.

C Fitting of the model to data for different percentage of reporting (all data)

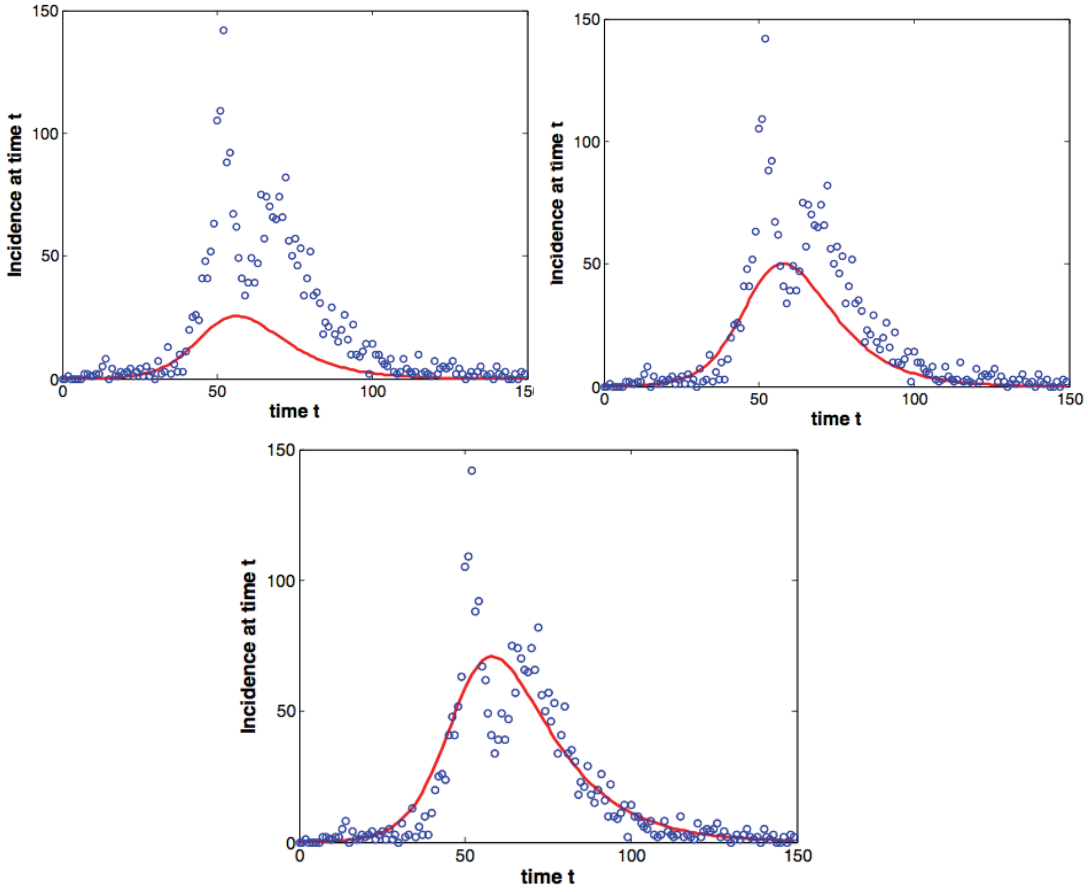


Figure 20: Model fit to all the data from Lima, Peru of the incidence number of reported cases varying the percentage of reporting (r), that is, $r = 10\%, 20\%$ and 30%, respectively.

C.1 Fitting of the model to data for different percentage of reporting (initial phase of the epidemic)

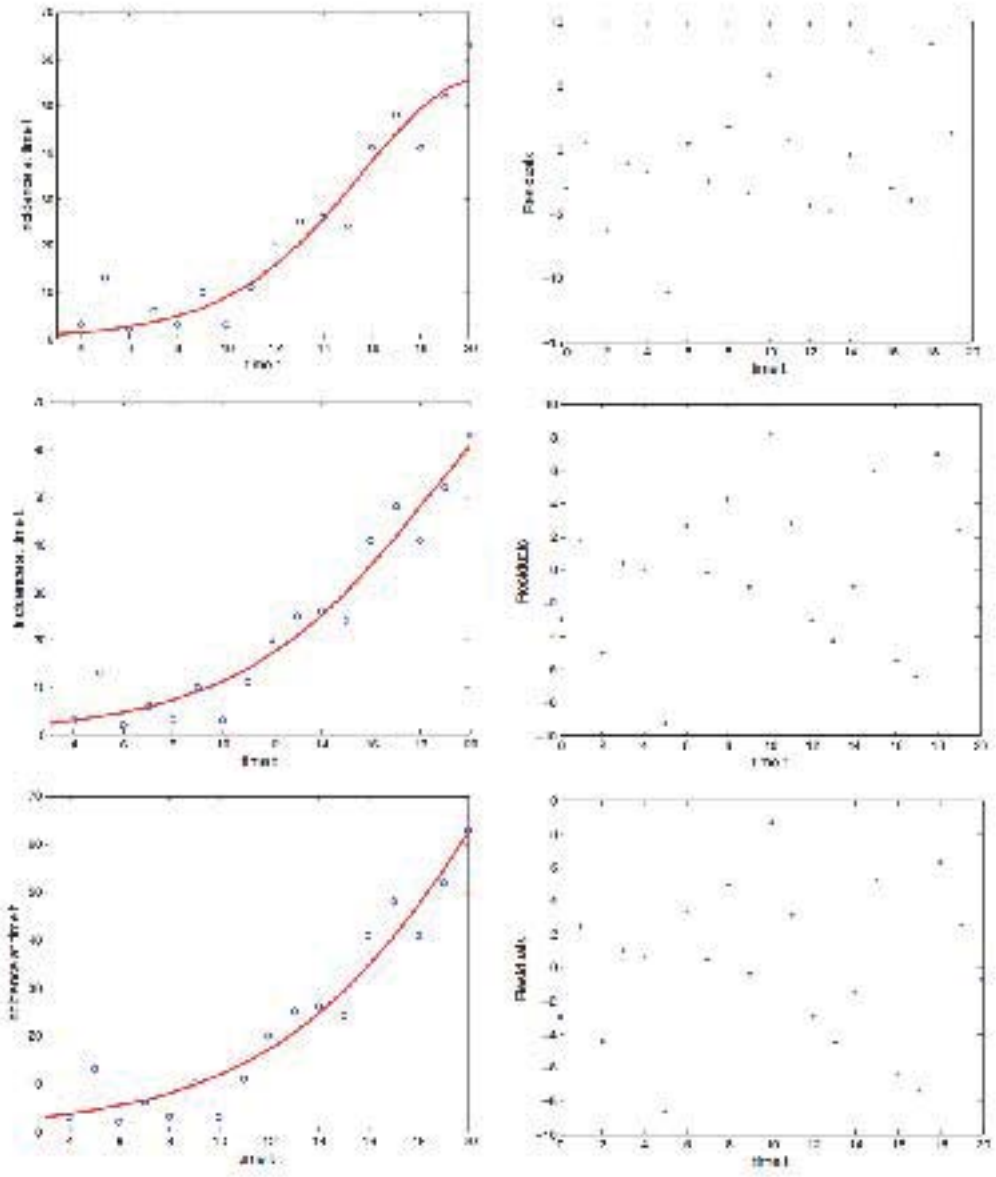


Figure 21: Model fit to the initial phase of the incidence number of reported cases varying the percentage of reporting (r), that is, $r = 10\%, 20\%$ and 30%, respectively.

D Sensitivity analysis

D.1 Sensitivity analysis for R_C

To explore the sensitivity of R_C to the variability of the parameters of the model, we let λ represent any of the parameter values that define R_C in our model. Considering a small perturbation of λ by $\Delta\lambda$, R_C will be perturbed by ΔR_C as well. The normalized sensitivity index S_λ is the ratio of the corresponding normalized changes. We define the sensitivity index for parameter λ as

$$S_\lambda^{R_C} = \frac{\frac{\Delta R_C}{R_C}}{\frac{\Delta\lambda}{\lambda}} = \left(\frac{\lambda}{R_C}\right)\left(\frac{\partial R_C}{\partial\lambda}\right).$$

Since the control reproductive number for the model is given by

$$R_C = \beta\left(\frac{\delta}{\kappa+\gamma_1+\alpha_1} + \frac{\kappa}{(\kappa+\gamma_1+\alpha_1)(\gamma_2+\alpha_2,)}\right),$$

we calculate the sensitivity indices

$$\begin{aligned}
S_\beta^{R_C} &= \frac{\beta}{R_C}\left(\frac{\partial R_C}{\partial\beta}\right) = 1,\\
S_\delta^{R_C} &= \frac{\delta}{R_C}\left(\frac{\partial R_C}{\partial\delta}\right) = \frac{\delta(\alpha_2+\gamma_2)}{\delta\alpha_2+\delta\gamma_2+\kappa},\\
S_\kappa^{R_C} &= \frac{\kappa}{R_C}\left(\frac{\partial R_C}{\partial\kappa}\right) = \frac{\kappa(-\delta\alpha_2-\delta\gamma_2+\alpha_1+\gamma_1)}{(\alpha_1+\kappa+\gamma_1)(\delta\alpha_2+\delta\gamma_2+\kappa)},\\
S_{\alpha_2}^{R_C} &= \frac{\alpha_2}{R_C}\left(\frac{\partial R_C}{\partial\alpha_2}\right) = \frac{-\alpha_2\kappa}{(\alpha_2+\gamma_2)(\delta\alpha_2+\delta\gamma_2+\kappa)},\\
S_{\gamma_1}^{R_C} &= \frac{\gamma_1}{R_C}\left(\frac{\partial R_C}{\partial\gamma_1}\right) = \frac{-\gamma_1}{\alpha_1+\kappa+\gamma_1},\\
S_{\gamma_2}^{R_C} &= \frac{\gamma_2}{R_C}\left(\frac{\partial R_C}{\partial\gamma_2}\right) = \frac{-\gamma_2\kappa}{(\alpha_2+\gamma_2)(\delta\alpha_2+\delta\gamma_2+\kappa)}.
\end{aligned}$$

We computed the sensitivity indices for R_C for each percentage of reporting, where parameters $\delta = \frac{2}{5}, \gamma_2 = \frac{1}{5}, \kappa = \frac{1}{3}$, and $\alpha_1 = 0$ are fixed, and parameters β, α_2, and γ_1 are varied according to r, the level of reporting cases (see Table **??** for the different parameter values). The values of the sensitivity indices for R_C are given in Table 11. The transmission rate β is the most influential parameter in decreasing R_C since the sensitivity index is 1. Also, since the isolation rate for those with severe symptoms, α_2, provides possible intervention strategies, we examine how changes to this parameter affect the control reproductive number R_C. For example, for $r = 10\%$, we have $S_{\alpha_2}^{R_C} = -0.159$ which means that a 6.3% increase in α_2 results in 1% decrease in R_C. Whereas, for $r = 30\%$, we have $S_{\alpha_2}^{R_C} = -0.360$ which means that a 2.7% increase in α_2 results in a 1% decrease in R_C.

Table 11: Sensitivity analysis of R_C

Index	10%	20%	30%
$S_{\beta}^{R_C}$	1	1	1
$S_{\delta}^{R_C}$	0.232	0.308	0.349
$S_{\kappa}^{R_C}$	0.285	0.258	0.108
$S_{\alpha_2}^{R_C}$	-0.159	-0.319	-0.360
$S_{\gamma_1}^{R_C}$	-0.517	-0.567	-0.457
$S_{\gamma_2}^{R_C}$	-0.609	-0.372	-0.291

D.2 Sensitivity analysis for u

Since one of the aims of this work is to investigate the impact of non-reporting on the final epidemic size, a sensitivity analysis is also conducted on u.

$$
\begin{aligned}
u &= 1-\left[\frac{\alpha_1}{\alpha_1+\gamma_1+\kappa}+\left(\frac{\kappa}{\alpha_1+\gamma_1+\kappa}\right)\left(\frac{\alpha_2}{\alpha_2+\gamma_2}\right)\right] \\
S_{\kappa}^{u} &= \frac{\kappa}{u}\left(\frac{\partial u}{\partial \kappa}\right)=\frac{\kappa(\alpha_1\gamma_2-\gamma_1\alpha_2)}{(\alpha_1+\kappa+\gamma_1)(\kappa\gamma_2+\gamma_1\alpha_2+\gamma_1\gamma_2)} \\
S_{\alpha_2}^{u} &= \frac{\alpha_2}{u}\left(\frac{\partial u}{\partial \alpha_2}\right)=-\frac{\alpha_2\kappa\gamma_2}{(\alpha_2+\gamma_2)(\kappa\gamma_2+\gamma_1\alpha_2+\gamma_1\gamma_2)} \\
S_{\gamma_1}^{u} &= \frac{\gamma_1}{u}\left(\frac{\partial u}{\partial \gamma_1}\right)=\frac{\gamma_1(\alpha_1\alpha_2+\alpha_1\gamma_2+\kappa\alpha_2)}{(\alpha_1+\kappa+\gamma_1)(\kappa\gamma_2+\gamma_1\alpha_2+\gamma_1\gamma_2)} \\
S_{\gamma_2}^{u} &= \frac{\gamma_2}{u}\left(\frac{\partial u}{\partial \gamma_2}\right)=\frac{\alpha_2\kappa\gamma_2}{(\alpha_2+\gamma_2)(\kappa\gamma_2+\gamma_1\alpha_2+\gamma_1\gamma_2)}
\end{aligned}
$$

Table 12 provides the values for the sensitivity index for u. We vary values of α_2, one of the most influential parameters. However, as the percentage of reporting increases, the sensitivity index of α_2 gets closer to zero, meaning that a higher percent change in α_2 is needed to obtain a 1% decrease in u. For example, when the percentage of reporting is $r = 10\% \Rightarrow S_{\alpha_2}^{u} = -0.779$. This suggest a 1.25% increase in α_2 would result in a 1% decrease in u. When the level of reporting is $30\% \Rightarrow S_{\alpha_2}^{u} = -0.374$, this suggest that as 2.6% increase in α_2 gives a 1% in u.

Table 12: Sensitivity analysis of u

Index	10%	20%	30%
S_{κ}^{u}	-0.057	-0.142	-0.196
$S_{\alpha_2}^{u}$	-0.799	-0.533	-0.374
$S_{\gamma_1}^{u}$	0.057	0.142	0.196
$S_{\gamma_2}^{u}$	0.799	0.533	0.374

Table 12 provides the values for the sensitivity index for u. The parameter whose value can be altered in reality is α_2, which happens to be one of the most influential parameters. However, as the percentage of reporting increases, the sensitivity index of α_2 gets closer to zero, meaning that a higher percent change in α_2 is needed to obtain a 1% decrease in u. For example, when the percentage of reporting is $r = 10\%$, $S^u_{\alpha_2} = -0.779$, meaning a 1.25% increase in α_2 would result in a 1% decrease in u. When the level of reporting is 30%, $S^u_{\alpha_2} = -0.374$, which means that to get a 1% decrease in u, α_2 needs to be increased by 2.6%.

References

[1] Arino, J., Bauch,C., and Brauer, F., et al., *Pandemic influenza: modelling and public health perspectives*, Mathematical Biosciences and Engineering (2011) **8**:1-20.

[2] Gomez, J., Vibud, C., Munayco, C., et al., *Pandemic influenza in a southern hemisphere setting: the experience in Peru from May to September, 2009.*, Euro surveillance bulletin european sur les maladies transmissibles European communicable disease bulletin, (2009) **6**:e21287.

[3] Taubenberger, J. and Morens, D., *1918 Influenza: The mother of all pandemics*, Emerging Infectious Diseases (2006) **17**:15-22.

[4] Brauer, F., Dreissche, P., and Wu, J., *Mathematical Epidemiology* Springer (2008).

[5] Banks, H., Tran, H., *Mathematical and Experimental Modeling of Physical and Biological Processes* Chapman and Hall (2009).

[6] Brauer, F., *Some simple epidemic models*, Mathematical Biosciences and Engineering (2006) **3**:1-15.

[7] Fraser, C., Riley S., Anderson, R. and Ferguson, N., *Factors that make an infectious disease outbreak controllable*, PNAS (2004) **101**:6146-6151.

[8] Cauchemez, S., Ferguson, N. and Wachetel, C., et al., *Closure of schools during an influenza pandemic*, Lancet Infectious Diseases (2009) **9**:473-481.

[9] De Serres, G., Rouleau, I. and Hamelin, M., et al., *Contagiousness period for pandemic (H1N1) 2009*, Emerging Infectious Diseases (2010) **16**:783-788.

[10] Cahill, E., Crandall, R., Rude, L. and Sullivan, A., *Space-time influenza model with demographic, mobility, and vaccine parameters*, PSIpress (2009).

[11] *Morbidity and mortality weekly report*, Centers for Disease Control and Prevention (2010) **59**:162-165.

[12] Lin, F., Muthuraman, K. and Lawley, M., *Lin, F. and Muthuraman, K. and Lawley, M.*, BioMed Central (2010) **10**:1-13.

[13] Wang, X., Yang, P. and Seale, H., et al., *Estimates of the true number of cases of pandemic(H1N1) 2009, Beijing, China*, Emerging Infectious Diseases (2010) **16**:1786-1788.

[14] Yan, W., Zhou, Y., Wei, S. and Zhang, H., *The difficulties of early detection for infectious disease outbreak in China: a qualitative investigation*, Journal of Nanjing Medical University (2008) **22**:66-70.

[15] Albert, R. and Barabasi, A., *Statistical mechanics of complex networks*, Reviews of Modern Physics (2002) **74**:47-97.

[16] Ross, S., *A First Course in Probability*, Prentice Hall (2005).

[17] Watts, D. and Strogatz, S., *Collective dynamics of 'small-world' networks*, Nature (1998) **393**:44-442.

[18] Herrera-Valdez, M., Cruz-Aponte, M. and Castillo-Chavez C., *Multiple outbreaks for the same pandemic: local transportation and social distancing explain the different "waves" of A-H1N1PDM cases observed in Mexico during 2009*, Mathematical Biosciences and Engineering (2011) **8**:21-50.

[19] Capaldi, A., Behrend, S., and Berman, B., *Parameter estimation and uncertainty quantification for an epidemic model*, unpublished (2009).

[20] World Health Organization, *Influenza* http://www.who.int/mediacentre/factsheets/2003/fs211/en/ Accessed: July, 19 2011.

[21] Center for Disease Control, *CDC Says "Take 3" Actions to Fight the Flu*, `http://www.cdc.gov/flu/protect/preventing.htm` Accessed: July 19, 2011.

[22] Centers for Disease Control and Prevention, *Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States-early, targeted, layered use of nonpharmaceutical interventions* (2007).

[23] Nuzzo, J. and Gronvall, G., *Global health security: closing the gaps in responding to infectious disease emergencies*, http://www.upmc-biosecurity.org/website/resources/publications/2011/2011-06-15-closing_-gaps_-inf_-disease.html Accessed: July 21, 2011.

[24] Center for Disease Control, *How the flu virus can change: drift and shift*, `http://www.cdc.gov/flu/about/viruses/change.htm` Accessed: July 21, 2011.

[25] CDC, *Interim guidance on infection control measures for 2009 H1N1 influenza in healthcare settings, including protection of healthcare personnel*, `http://www.cdc.gov/h1n1flu/guidelinesinfectioncontrol.htm` Accessed: July 21, 2011.

[26] *Ordenanza N 111-2009-MDH* `http://www.munihualmay.gob.pe/docs/Ordenanzas/ORDENANZA111-2009.pdf` Accessed: July 27, 2011.

[27] *Municipalidad Distrital de Pimentel* , `http://www.munipimentel.gob.pe/Ordenanzas/OM%20N%20120_M DP.pdf Accessed` : $July 27, 2011. WHO, World now at the start of 2009 influenza pandemic (WHO), $`http://www.who.int/mediacentre/news/state`$ July 30, 2011.$

[28] White, L.F. and Pagano, M., *Reporting errors in infectious disease outbreaks, with an application to Pandemic Influenza A/H1N1*, Epidemiologic Perspectives & Innovations (2010) **7**:12.

[29] Sutton, K.L., Banks, HT and Castillo-Chavez, C., *Inverse Problem Methods as a Public Health Tool in Pneumococcal Vaccination*, Tech. Rep. CRSC-TR08-19, Center for Research in Scientific Computation, North Carolina State University, NC.

[30] Banks, HT, Cintrón-Arias, A. and Kappel, F., *Parameter Selection Methods in Inverse Problem Formulation*,Technical report, Center for Research in Scientific Computation (2010).