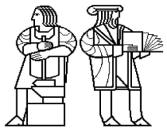


Massachusetts Institute of Technology
Engineering Systems Division

Working Paper Series

ESD-WP-2008-10

REVISITING R_0, THE BASIC REPRODUCTIVE NUMBER FOR PANDEMIC INFLUENZA

Richard C. Larson

Center for Engineering Systems Fundamentals
Engineering Systems Division
Massachusetts Institute of Technology
rclarson@mit.edu

February 2008

Revisiting R_0, the Basic Reproductive Number for Pandemic Influenza

Richard C. Larson
Center for Engineering Systems Fundamentals
Engineering Systems Division
Massachusetts Institute of Technology
February 8, 2008

ABSTRACT

This paper focuses on a fundamental input parameter for most existing mathematical models of pandemic influenza, the 'basic reproductive number R_0,' defined to be *the mean number of new influenza infections created by a newly infected person in a population of all susceptible people*. We argue that R_0 is limited in policy and scientific value as is any single parameter attempting to characterize a complex probabilistic process. In particular, we demonstrate by simple logic that R_0 does not exist as a separate 'constant of a particular influenza,' but rather its value is determined by social context and behavioral patterns as well as by the "physics'' of the influenza virus. To the extent that R_0 is useful, it is best viewed as an output of a modeling analysis, not an input. But with R_0 being the mean of a random variable, much more information is contained in the entire probability distribution. With this view, we show – again by simple arguments – that R_0 can be greater than 1.0 and still, contrary to popular belief, the probability of an exponentially growing pandemic may be arbitrarily small. Finally, we show that attempts to estimate R_0 from data of previous pandemics is fraught with methodological complexities, due primarily to heterogeneities in the population that cause super-spreaders and socially active people to be the first propagators of the disease. Unless one is careful, statistical estimates of R_0 based on early exponential growth of reported cases may be significantly upwardly biased.

Key words: *epidemiology, mathematical modeling, basic reproductive number, statistical estimation, bias*

Revisiting R_0, the Basic Reproductive Number for Pandemic Influenza

Richard C. Larson

Pandemic influenza represents a credible threat to the lives of hundreds of millions of people worldwide. The trigger event will likely be the mutation of a flu virus presently in birds, pigs or other animals, to become human-to-human efficiently transmittable. From the moment of the trigger event, current science suggests that at least six months will be required to develop a safe and effective vaccine, and then only for a small fraction of the planet's inhabitants. For six months or longer, we will all be 'naked' against the flu.

Much 'flu research', including mathematical modeling, suggests that once the flu starts, it will simply run its course -- implying there is not much we can do. It's as if the flu imposes on us a Russian Roulette partially loaded gun, and when nature pulls her trigger there is a given chance that we will be hit, regardless of what else we do. But we do not believe that to be true. Evidence from the 2003 SARS epidemic, recent analyses of the 1918-1919 "Spanish Flu," and our own mathematical modeling suggest that there is much we can do to reduce the likelihood that we as individuals, our friends and family members and our co-workers will become sickened with the flu.

There is a fundamental input constant for most existing mathematical models of the flu, the 'basic reproductive number R_0,' defined to be ***the mean number of new influenza infections created by a newly infected person in a population of all susceptible people.*** Suppose that early in the epidemic I become infected with the flu and that I infect 3 other individuals before I am committed to bed and rest. Suppose you are also infected and that you infect 2 others before you are isolated in bed. Our 'average' 'R_0' in this simple case is (3 + 2)/2 = 2.5. With past pandemic flu's such as in 1918-1919, a typical R_0 across the entire population is estimated to be between 1.8 and 2.5. One can see that if R_0 were greater than one, the epidemic would seem to grow exponentially for a while, until the number of remaining susceptible people drops below some critical point. In many existing mathematical models, after the epidemic starts, R_0 is replaced by $R(t)$, where $R(t)$ is defined to be the mean number of new influenza infections created by a newly infected person at time t in the epidemic, where t is measured in days or in "generations'' of the disease. Due to the fact that a smaller fraction of the population is susceptible to the disease, as it progresses through the population, $R(t)$ should be a decreasing function of t, with its maximum at the start of the disease, at which time $R(0) = R_0$. So, for instance, if R_0 starts the infectious disease with a value $R_0 = 2.0$, the disease multiplication factor $R(t)$ will only become smaller as time progresses.

What is surprising is that only a small value of R_0 is needed to create a pandemic. Intuitively, seeing the "Great Influenza" havoc caused in 1918-1919, one might think that R_0 needs to be 10 or 20! No, averaged across the population, it is less than 3.0, often near 2.0. (Mills *et al.,* 2004) That's the 'good news' as it is so much easier to reduce R_0 from 2.0 to one than from 10 to 1. If one can find a sequence of simple steps to reduce R_0 to be less than 1.0, rather than exponential growth, one then enjoys geometric decay as the disease dies away. Evidence suggests that this is what happened with the eradication of

SARS in Hong Kong and elsewhere in 2003. And, as we will demonstrate, an R_0 equal to 2.0 does not necessarily imply exponential growth.

The importance of R_0 cannot be overstated. It is the primary parameter in most mathematical models of pandemic influenza, and it is central to the modeling of other infectious diseases as well. But from a decision and policy point of view, the dominance of R_0 frames our policy and decision space, unnecessarily narrowly in our opinion. An understanding of the limitations of the R_0 concept as a modeling device for pandemic influenza should open additional more insightful decision and policy alternatives. That is our goal in this paper.

1. Historical Perspective

The concept of using R_0 seems to be accepted in an unchallenged way in the epidemiology community. From Heesterbeek [2002], an expert of the history of R_0, we hear,

> The basic reproductive ratio (or number) R_0 is arguably ***the most important quantity on the study of epidemics*** and notably in comparing population dynamical effects of control strategies. The quantity is defined as the expected number of new cases of an infection caused by a typical infected individual in a population consisting of susceptibles only. In the last 10-15 years R_0 is an ingredient in almost all papers that use some mathematical modeling in studying the spread of infectious agents. (emphasis added)

As described by Heesterbeek [2002], R_0 was created in Germany by demographers in the 1880's and formalized in 1925 to model the progression of a country's population. The original R_0 was defined to be the average number of female offspring born to one female over her entire life. For the year 1879, this number for Germany was estimated by Richard Bockh to be 1.06. The time scale was decades and the system was in approximate equilibrium. With an influenza epidemic, the time scale is in days and weeks and nothing approximating equilibrium exists. To the contrary, the system is characterized by markedly changing parameter values as society copes daily with the influenza's evolution. Over the last three decades, epidemiologists have adopted the R_0 concept and applied it to a variety of diseases, some of which (e.g., malaria) exist in a type of quasi-equilibrium similar to that of population demographics. But the original demographic motivation and near steady state environment supporting R_0 simply do not exist in a dynamic influenza epidemic situation. In summary, R_0 and its successor $R(t)$ as fixed-trajectory concepts in rapidly evolving infectious disease epidemics are of limited value at best.

We often hear epidemiologists attach to an infectious disease a given number for R_0, as if that number characterizes some constant of nature, independent of anything else. One might hear, "Consider an infectious disease with R_0 equal to 3.14159, etc., etc." One mathematical researcher even calls R_0 the "…one parameter that (almost) does it all." (Keeling 2001) Such simplistic statements ignore the contextual social and physical environments in which the disease is developing. These disease environments play a significant role in determining the numerical value for R_0 and for subsequent values of

$R(t)$. Even in demography, where quasi steady-state operation supports use of the R_0 concept, human behavior demonstrates that R_0 is far from an immutable constant. In Germany today, more than a century after the first estimate of Germany's R_0, the current R_0 is estimated to be about 0.70, a 33 percent drop from Bockh's 1879 estimate of 1.06. (note 1) Worldwide, the demography interpretation of R_0 today varies by a factor of seven, from over 3.5 daughters per female (Mali and Niger) to under 0.5 (Hong Kong) (note 2). In demography, we see that the numerical value of R_0 depends strongly on social and environmental context. It is not a constant of nature. So too in infectious disease applications we should expect R_0 to depend on context. In influenza, as in demography, the numerical value of R_0 depends strongly on the societal situation in which it is embedded.

An excellent overview of the history and use of R_0 in deterministic epidemiology modeling is given by Heffernan *et al.* (2005). We are not the first to explore the role of stochastic behavior in modeling pandemics (see, for example, Nasell 1995) nor are we the first to examine how population heterogeneities after disease dynamics (see, for example, Lindholm 2007 and Diekmann *et. al.* 1990). We especially appreciate those using social network modeling to identify high-risk heterogeneous segments of the population (e.g., school children) who when targeted with social distancing may greatly reduce incidence of the disease (see Glass *et. al.* 2006 and Glass *et. al.* 2008). Admittedly, ours are stylistic models but with a probability focus. Our sense is that the simple "thought experiment'' arguments made herein may shed additional insights into the complex problem area of infectious disease progression and control.

We now explore some of the mathematical properties of R_0, an exploration that we hope will demonstrate that the concept must be used with extreme care in complex decision making situations involving epidemics and pandemics.

2. R_0 is the Mean Value of a Random Variable

The consensus definition of R_0 states that it is the mean value of a random variable. As in all probabilistic situations, the mean of a random variable conveys some useful information. But expressing the mean in terms of other more fundamental quantities can yield additional insights.

Suppose I come face to face with N people on a day that I am infectious but asymptomatic. Many people who become infected with the flu have one such day before they feel and appear sick, and not being able to identify these people is what makes eradication of the flu so difficult. Define an 'indicator variable' as follows:

$$X_i = \begin{cases} 1 \text{ if person } i \text{ becomes sick as a result of exposure to me} \\ 0 \text{ if person } i \text{ does not become sick as a result of exposure to me} \end{cases}$$

Now, we let NI be defined to be the number of people I will infect on this day. NI can be written as simply counting the indicator variables,

$$NI = X_1 + X_2 + X_3 + ... = \sum_{i=1}^{N} X_i \quad (1)$$

Suppose for example $N = 50$ and that all X_i's are 0 except for X_0, X_{18} and X_{45}, each being equal to one. In that case, I have infected 3 of the 50 individuals I have came face to face with on this day.

Now, at any given level of intensity of face-to-face contact, there is a probability p that I will pass the infection on to the person I am facing. Using this fact, we can write an expression for the mean number of people I will infect on this day. It is simply the mean of $NI = X_1 + X_2 + X_3 + ... = \sum_{i=1}^{N} X_i$, which equals Np. We thus have a simple expression for R_0, and that is

$$R_0 = Np. \quad (2)$$

As we stated before, for pandemic flu, R_0 appears to many as some constant of nature, such as the gravitational constant = 6.67300 $(10^{-11})m^3\ kg^{-1}\ s^{-2}$. But flu is an infectious respiratory disease, spread by human contacts. Reduce human contacts, and reduce prevalence of the flu. By writing $R_0 = Np$, we have expressed R_0 in terms of two other parameters, each of which we can control to some extent. We have a fighting chance of reducing R_0, perhaps a little, perhaps even to below 1.0, the critical value to assure that the disease dies away rather than grows exponentially. In the sense of this discussion, R_0 does not exist as a separate quantity. It is a function of both the inherent properties of the given virus *and* the population's behavioral responses to it.

How do we control N and p? One reduces N simply by reducing the number of face-to-face contacts we have each day. If a parent is shopping for groceries, rather than following the European tradition of daily shopping, perhaps one switches to weekly shopping, or, better yet, to groceries delivered to one's door. If you manage a team of employees, rather than have face-to-face meetings during a flu emergency, have conference calls instead, with many workers telecommuting. Many companies have already created comprehensive pandemic flu plans that include telecommuting, reduced face-to-face encounters and even minimum desk spacing between workers. The desk spacing idea relates more to the parameter p, the probability that any given face-to-face contact will result in a new infection. How else can we reduce p? Wash hands with hot water and soap several times daily. Do not shake hands during greetings with colleagues. Cough or sneeze into your elbow, not into the open air. Be careful not to touch surfaces that might have recently been contaminated with flu virus. Encourage your city's large employers to stagger work hours so that public transportation subways and busses are less crowded during now-stretched-out rush hours. Even run the subways and busses with windows opened. The key here is that R_0 is a direct function of social context and human behavior, behavior that can be altered to reduce the numerical value of R_0.

Yes, there are limitations to this analysis. The causal model creating infection is more complex than just counting the numbers of face-to-face contacts. One can touch surfaces contaminated minutes or even hours before by individuals who we do not see face to

face. If contaminated hands then touch one's mouth or eyes, infection can result. With SARS (Severe Acute Respiratory Syndrome), residents of a Hong Kong high-rise apartment complex became infected by a faulty sewage system, again not 'seeing' the infected person responsible for spreading the infection. (note 3) But we believe that a model that counts the number of face-to-face contacts and includes the intensity of these contacts represents a valid primary mechanism for depicting how the disease propagates through the population. Adding complexities such as the two just cited does not alter the main conclusions of our arguments. Our approach is buttressed by findings of others. For instance, Riley at al (2003) credits reduction in the number of face-to-face contacts in Hong Kong as the primary cause for reduction in spread of SARS.

3. Variance of *NI*

The mean of a random variable contains only very aggregate information about it. In fact, as we will show, a very diverse set of probability distributions can give rise to the same mean value. Therefore, we need to look at measures of dispersion about the mean and other properties of the entire probability distribution.

Since the random variable *NI* has a probability distribution, it has a variance as well as a mean R_0,defined as the squared second moment about them mean,

$$VAR[NI] \equiv \sigma_{NI}^2 \equiv E[(NI - E[NI])^2] = E[(NI - R_0)^2].$$

As with all random variables, knowledge of the variance tells us much about the level of uncertainty in the infection process. Let us write Equation (1) again,

$$NI = X_1 + X_2 + X_3 + ... = \sum_{i=1}^{N} X_i .$$

In Sec. 1 above, we found the mean for the random variable *NI*. Since the indicator random variables X_i are mutually independent and identically distributed, we have

$$\sigma_{NI}^2 \equiv N\sigma_{X_i}^2 = Np(1-p), \qquad (3)$$

where $p(1-p)$ is the variance of any given indicator random variable X_i. Notice how the variance is small for small values of p, as would most likely be found in influenza for as random casual face-to-face contact, but the variance builds to a maximum value as p increases up to 50%.

4. Mean and Variance when *N* is a Random Variable

In real life the number of people one has face-to-face contact with on any given day is itself a random quantity. This fact adds more uncertainty to the number of people an infected person will infect.

Let us again return to our first result, Equation (1),

$$NI = X_1 + X_2 + X_3 + ... = \sum_{i=1}^{N} X_i .$$

Recognizing that the number N of people we interact with on any given day is uncertain, we wish to make this more realistic. N too is a random variable. Suppose N has probability mass function $p_N(n) \equiv P\{N = n\}$, with mean $E[N]$ and variance $\sigma_N^2 \equiv E[(N - E[N])^2] = E[N^2] - E[N]^2$. And suppose that all the random variables X_i and N are mutually independent. Under these circumstances it known that

$$\begin{aligned} E[NI] &= E[N]E[X_i] = pE[N] \\ \sigma_{NI}^2 &= E[N]\sigma_{X_i}^2 + (E[X_i])^2\sigma_N^2 = E[N]p(1-p) + p^2\sigma_N^2 \end{aligned} \tag{4}$$

Here we see that the variance of NI has a term equivalent to that found in Eq. (3), namely $E[N]\sigma_{X_i}^2 = E[N]p(1-p)$ and a new, second term, $(E[X_i])^2\sigma_N^2 = p^2\sigma_N^2$. So, randomness in the number of people that patient zero has face-to-face contact with increases the total variance of the number of new infections he causes. But the mean remains unchanged at $E[NI] = E[N]E[X_i] = pE[N]$.

5. Extinction Probabilities

In most mathematical modeling of infectious disease pandemics, it is commonly believed that any value for R_0 greater than 1.0 will yield a pandemic (or epidemic) with near-term exponential increase in number of infected people. Here we show that such an assumption is false. Using an R_0 equal to 2.0, a value commonly associated with pandemic influenza, we provide simple examples showing how many 'patient zeros' may initiate a disease that dies out rapidly in one or two generations. That is, we can have self-extinction of the disease, not as a rare event, but as a common one.

As a simple 'thought experiment,' consider that NI can take on only two possible values, 0 and 4, each with equal likelihood. That is, patient zero will infect zero others with probability 1/2 and exactly 4 others with probability 1/2. Here R_0 is the expected value of NI which is easily computed, $R_0 = (1/2)*0 + (1/2)*4 = 2$. But with probability equal to 1/2, the disease never progresses beyond patient zero. In other words, the probability of self-extinction of the disease is at least 50%, even though R_0 is 2.0.

We can write an equation from which we can compute the exact value for the self-extinction probability, which we will call P_E. For our simple example, we can write

$$P_E = (1/2) + (1/2)P_E^4 \quad .$$

The logic is this: P_E is equal to 1/2, due to the 50% chance that patient zero will infect no others, plus (1/2) times the probability that each of the four people infected under the second possibility for patient zero will themselves spawn an infection process that dies out – each independently and each with probability P_E. The numerical solution to this equation is $P_E = 0.543$. So, we have a feasible situation in which R_0 is 2.0 and yet 54.3%

of the 'epidemics'' die out very quickly on their own. There is no exponential growth, obviously, for such cases.

More generally, suppose that the number of people *NI* infected by patient zero has an arbitrary probability law,

$$p_{NI}(n) \equiv P\{NI = n\} = p_n, \quad n = 0,1,2,...$$

Then, following the logic above, we can write a general equation for the extinction probability,

$$P_E = p_0 + p_1 P_E + p_2 P_E^2 + ... + p_n P_E^n + ... = \sum_{n=0}^{\infty} p_n P_E^n . \tag{5}$$

This turns out to be the discrete or geometric transform of the probability mass function $p_{NI}(n)$, evaluated at $z = P_E$, where the transform is defined

$$p_n^T(z) \equiv \sum_{n=0}^{\infty} p_n z^n \quad \text{for } |z| \le 1.$$

(The function $p_n^T(z)$ is also called the moment generating function for $p_{NI}(n)$.) Thus, to solve for a numerical value for the self-extinction probability P_E, we must solve the functional equation

$$P_E = p_n^T(P_E). \tag{6}$$

A not unlikely distribution for the number of infections caused by face-to-face contacts per day is the Poisson distribution, perhaps reflecting a situation in which the infections are generated by a Poisson process. As an example, retail sales clerks might fall into this group. For a Poisson process with mean λ, the probability mass function and associated geometric transform are, respectively,

$$\begin{aligned} p_n &= \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0,1,2,... \\ p_n^T(z) &= e^{\lambda(z-1)} \quad |z| \le 1 \end{aligned} \tag{7}$$

Applying Eq.(6) in this case, we find P_E by solving the equation,

$$P_E = e^{\lambda(P_E - 1)}.$$

Suppose we have $\lambda = 2$, which corresponds to $R_0 = 2$. Then P_E is the solution to the equation $P_E = e^{2(P_E - 1)}$. The solution is found to be $P_E \approx 0.203$. So, with a Poisson

distributed number of new infections in a day created by an infected and infectious person, a process with $R_0 = 2.0$ yields a self-extinction probability of over 20%. A much larger $R_0 = 3.0$ yields a self-extinction probability of only about 0.06. But a reduction from 2.0 to 1.8 (advocated by some current researchers for pandemic influenza (note 4)) yields $P_E \approx 0.267$.

As a final example of this line of inquiry, let us consider what we believe to be the most realistic model of infection progression early in the disease. That is one in which the number of contacts each day is random and each contact has a given probability *p* of becoming infected. Thus, the number who become infected from patient zero is given by the familiar counting process of Eq. (1),

$$NI = X_1 + X_2 + X_3 + ... = \sum_{i=1}^{N} X_i .$$

If the random variables X_i and N are mutually independent, then it is known that the geometric transform of the probability mass function for the random variable *NI* is

$$p_{NI}^T(z) = p_N^T(p_{X_i}^T(z)). \tag{8}$$

Suppose the number of contacts *N* each day is given by a Poisson random variable with mean λ and each contact results in a new infection with probability *p*. Then,

$$p_N^T(z) = e^{\lambda(z-1)}$$
$$p_{X_i}^T(z) = (1-p) + pz, \text{ both requiring } |z| \le 1.$$

Substituting into Eq.(8), we have

$$p_{NI}^T(z) = e^{\lambda([1-p+pz]-1)} = e^{\lambda p(z-1)}. \tag{9}$$

Comparing Eq.(9) to Eq.(7), we immediately see that *NI* is also Poisson-distributed, but with mean λp. So, for any set of values for λ and *p* such that there product λp=2.0, we have a self-extinction probability of $P_E \approx 0.203$.

The mathematical modeling of self-extinction is not new. In fact one can revert to the classic 1875 paper by Watson and Galton, "On the Probability of the Extinction of Families," to see essentially the same arguments put forward here. Watson and Galton were concerned with the extinction of surnames in Victorian England. With current contemporary estimates of fertility rates, a single adult male having a unique surname has –according to the Watson-Galton model -- a probability of about 0.89 of his surname becoming extinct. (Whittle 1970 and note 5). With an infectious disease starting with a patient zero, the situation is directly analogous to a unique surname in a population. The probability of self-extinction is computed in identical ways in each case. The demographers who introduced the R_0 concept in Germany in the 1880's, notably *after* the Watson-Galton paper, did not have to be concerned with self-extinction due to the millions of residents of Germany. The likelihood of an entire country's own self-extinction is negligibly small. Thus we see another problem carrying the demography-

invented R_0 concept to disease propagation: the appropriate model for disease propagation is the small sample size model of family surname progression, not the large sample size model of a country's population evolution.

6. Estimating R_0

In applying models using R_0, one is faced with the need to estimate numerically R_0 from data of past pandemics. Given the existence of significant self-extinction probabilities, our contention is that any estimate of R_0 from actual non-self-extinction pandemics will in fact be overestimates of the true R_0. As an example, our 'thought experiment' example having a 50% chance of no additional infections by patient zero and a 50% chance of exactly 4 additional infections, a statistical estimate of R_0 would likely only be performed on actual observable pandemics, not ones that die out after one or two generations. The quickly dying pandemics would not be called pandemics. Statisticians may not even see them. Focusing on patient zero, an unbiased estimate of R_0 for our thought experiment for those pandemics that do not die out would be 4.0, not the true 2.0, an overestimate of 100%. Such situations might correspond to patient zero being called a super-spreader, meaning that he or she infects many more people than the average infected person. Depending on the probability laws involved, it may be that many actual observed pandemics require that patient zero be a super-spreader, else the pandemic dies out quickly by self-extinction.

Let us consider the issue more generally. Suppose we define

$$P_E(j) \equiv P\{\text{pandemic becomes extinct at generation } j\}.$$

Following usual logic, we can immediately write

$$P_E(1) = p_0$$
$$P_E(2) = p_1 p_0 + p_2 (p_0)^2 + p_3 (p_0)^3 + \ldots + p_n (p_0)^n + \ldots = p_n^T(p_0) - p_0$$

Evaluating $P_E(j)$ for values of $j>2$ becomes computationally complex. But in practice we do not need that to demonstrate our point. Consider the realistic Poisson model for generation-to-generation propagation of the disease, again setting $\lambda = 2.0$. In this case, we have

$$P_E(1) = p_0 = e^{-2} \approx 0.135$$
$$P_E(2) = p_n^T(p_0) - p_0 = e^{\lambda(p_0 - 1)} - p_0 \approx e^{2(0.135-1)} - 0.135$$
$$= e^{2(0.135-1)} - 0.135 = e^{-1.73} - 0.135 \approx 0.177 - 0.135 = 0.042$$

Of the total extinction probability of 0.203, the first and second generations give a summed extinction probability of 0.135 + 0.042 = 0.177, or about 87% of the total extinction probability. That is, if this Poisson process pandemic dies out by itself, 87% of the time it will become extinct in generation 1 or generation 2. An apparently unbiased statistical estimate of the mean number of new infections created by patient zero (i.e.,

generation 1) would be $2.0/(1-0.135) \approx 2.31$, an overestimate of about 15% in contrast to the true R_0 value of 2.0. This may not appear to be large, but when one is dealing with exponential growth with a growth parameter greater than 1.0, a 15% overestimate may be significant from both a prediction and policy point of view.

6.1. Example: A Three-Point Probability Mass Function for *NI*.

As a thought experiment, consider a situation in which the probability law for the number of people *NI* that are infected by patient zero has three possibilities: $NI = 0$, 2 and some larger number, the larger number and the probabilities arranged so that $R_0 = 2.0$. We will select the following:

$$P\{NI = 0\} = \alpha(1-p)$$
$$P\{NI = 2\} = 1-\alpha$$
$$P\{NI = 2/p\} = \alpha p$$

One can easily verify that the three respective probabilities sum to one and that the expected value of *NI* is $E[NI] = 0[\alpha(1-p)] + 2(1-\alpha) + (2/p)\alpha p = 2.0$. This is a probability mass function having parameters α and p. With α =0, we have a strictly deterministic process in which each newly infected person in a population of nearly 100% susceptibles generates precisely 2 additional infections. Over time, this deterministic process will generate an exponential growth curve starting at time $t = 0$, with the exponential factor for generation-to-generation growth being $R_0 = 2.0$. As α increases from 0 towards 1.0, another more divergent process comes into play. This is a process for which a fraction $\alpha(1-p)$ of newly infected persons infect zero others while the fraction αp infect $2/p$ others. Here we see that for small values of the parameter p, most people in this second process infect zero others while a very few (a fraction αp) infect a large number $2/p$. This last category of persons may be called super-spreaders. As α grows towards 1.0, this second process dominates more and more.

Invoking Eq. (5), we see that to solve for the self-extinction probability P_E, we need to solve the functional equation,

$$P_E = \alpha(1-p) + (1-\alpha)P_E^{\,2} + \alpha p(P_E)^{2/p} \qquad (10)$$

where, formally, we require $2/p$ to be integer, since we are dealing with discrete random variables taking on non-negative integer values. Suppose, for example, $\alpha = 1/2$ and $p = 0.1$. Then we have $P_E = 0.45 + 0.5P_E^{\,2} + 0.05(P_E)^{20}$. The third term on the right hand side of the equation is essentially zero, and we are left with solving the quadratic equation $P_E = 0.45 + 0.5P_E^{\,2}$, whose solution is $P_E \approx 0.684$. Here we have a plausible probability distribution of number of people infected by a newly infected person, R_0 =2.0, and yet more than two-thirds of the "pandemics" die out quickly by self-extinction. For these, there is no exponential growth. Applying Eq. (10) parametrically, a family of curves displaying $P_E(\alpha, p)$, that is P_E as a function of α and p, is shown in Figure 1. Note that as α, the fraction of the population that is super-spreaders or non-spreaders, increases, so

too does the self-extinction probability. Note also that the self-extinction probabilities increase at p decreases, meaning that fewer super-spreaders are in the highly variable population but each one is more dangerous, as the number that each infects is equal to $2/p$ – a number that grows very large as p becomes smaller.

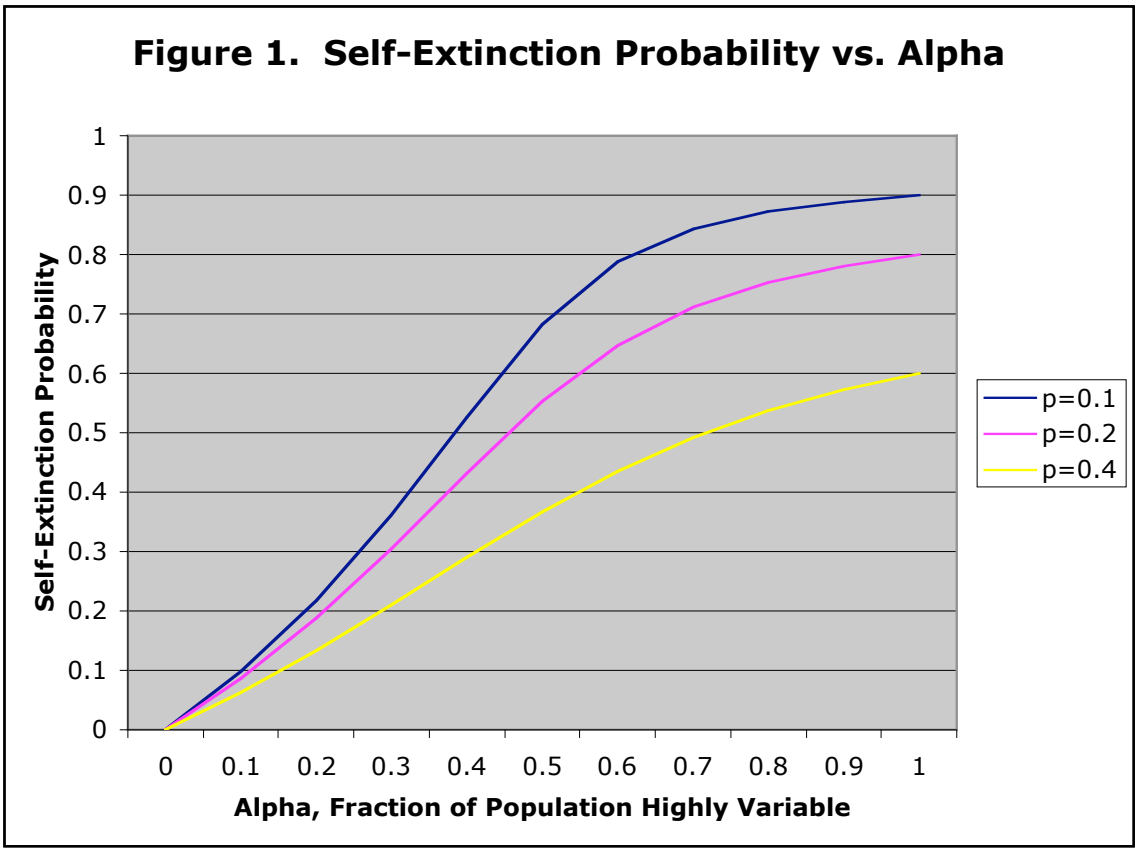
The careful reader may question our modeling as "extreme" using super spreaders. But super-spreaders are known to have started some epidemics. Among these are Mary Mallan (AKA Typhoid Mary) who is said to have infected 47 people, Gaetan Dugas -- a Canadian flight steward who allegedly infected many men with HIV/AIDS virus, and Professor Liu Janlun who allegedly started the SARS epidemic in Hong Kong by infecting people staying on the same floor of Hotel Metropole of Kowloon and his brother-in-law (who subsequently infected up to 79 others). These examples demonstrate that one individual – a 'super-spreader' -- may infect scores of others, far from any 'average' disease reproductive ratio. In fact, without a "start-up" super spreader, many "epidemics" or "pandemics"' may die out by self-extinction in one or two generations. And after the initial super spreader has done his or her deed, the remainder of the disease dynamics proceeds as if the R_0 were the moderate average value, 2.0 in our examples.

There are those who recommend splitting R_0 into two components, one due to super spreaders and the other representing the remainder of the population (see, for example, Riley 2003). But we have shown that in a highly diverse population containing both super spreaders and the great majority who are not super spreaders, once the disease has taken hold with a first "patient zero" super spreader, the chance of subsequent self-extinction is small. And the population will very likely have other super spreaders who will play a key role in subsequent exponential increase of incidence of the disease. Removing the super spreaders from computations may result in finding that R_0 for the non-super-spreaders is less than one, perhaps dramatically so. So, both super spreaders and others must be considered in estimating growth parameters of the disease. But policy and decision options should rightly be focused differently for each group. Ideally, one would like to identify before-the-fact likely super spreaders and reduce or even eliminate their possible contribution to propagation of the disease. If a cause of super spreading is related to large numbers of daily face-to-face intense contacts, that cause can usually be addressed.

6.2. Biasing Effects on Estimates of R_0 Due to Self-Extinction

The effect of self-extinguishing pandemics on statistical estimation of R_0 can be dramatic. Again to illustrate, we use a "thought experiment." Suppose we consider pandemics characterized by Eq. (10) with $\alpha = 1.0$, that is, no newly infected people subsequently infect only two people; they either infect zero additional people or they infect a large number, namely $2/p$. For values of p less than about 0.25, self-extinction only occurs at the first generation. That is, either the pandemic dies immediately with patient zero who infects nobody else, or patient zero is a super-spreader who infects $2/p$ others. If patient zero is a super-spreader, then the likelihood of self-extinction – once $2/p$ people are infected at generation 1 -- is small enough to ignore. After the super-spreader infects $2/p$ additional individuals, the mean value function of the number newly infected people over time grows exponentially, doubling at every generation – due to R_0 being equal to 2.0.

While the variance of the number infected at each generation will be large due to the highly dispersed nature of the probability law for *NI*, the eventual doubling of cases per generation should be seen in the data, perhaps with large deviations from the mean along the way.



We now consider how this affects biasing in the estimates of R_0. Illustrative numerical results are shown in Table 1. Each row of the table represents a generation of the influenza, starting at generation zero and building down the table to generation 10. The generation numbers are shown in the left-most column. In practice, each generation requires roughly three or four days from initial infection until asymptomatic infectivity to isolation, so the table covers the first month or so of the observed (non-self-extinguished) pandemic. The second column depicts deterministic exponential growth in the numbers infected at each generation, assuming that R_0 equals 2.0 and no super-spreaders. This simple exponential growth is often portrayed in epidemiological differential equation models and other strictly deterministic epidemiological models. Columns 3, 5 and 7 depict the mean number of new cases per generation, assuming that generation 1 (after "patient zero" in generation zero) is launched by a super-spreader where the three columns depict super-spreaders infecting $2/p = 8$, 16 and 32 people, respectively, for columns 3, 5 and 7. For these cases, the corresponding values of p are $p = 0.25$, 0.125 and 0.0625, respectively. The most relevant columns are columns 4, 6 and 8. Each of these columns contains estimates of R_0, where the estimate is based on fitting an exponentially increasing curve from 1 (at generation zero) to the number observed infected in generation i, $i = 1, 2, \ldots, 10$. As an example of the calculations, let us

consider column 6, generation 3. The entry in that cell is 4.00. That would be the estimate of R_0 if one attempted to fit an exponentially increasing curve from generation zero to generation 3, where at generation 3 the mean number of newly infected persons is 64 (column 5, generation 3). In order to climb from 1 infected individual to 64 in 3 generations, one needs the exponential sequence 1, 4, 16, 64, where $64 = 4^3$. Thus the generation-to-generation growth factor, otherwise known as R_0, is imputed to be 4.0 in this case. But we know that the underlying value for R_0 is not 4.0, but rather 2.0. If we tried to fit our curve up through the 10^{th} generation, our error would be less, this time estimating R_0 to be 2.46, still a positive biasing error of 23 percent. Note that the smallest bias error in the table is 15 percent (column 4, generation 10). Also note that the biasing error becomes larger as the size of the super-spreader "cohort" becomes greater.

While it is unlikely that any given pandemic will exhibit probabilistic behavior this extreme, the general message remains valid in our opinion. That is (1) pandemics can self-extinguish in the first or second generation; (2) only non-self-extinguishing pandemics are recorded for later statistical analysis; (3) the non-self-extinguishing pandemics are likely to be initiated by one or more super-spreaders or other similar non-representative phenomena, thereby accelerating the growth of the curve of those infected beyond the simple deterministic exponential curve. This early, accelerated growth makes estimating the numerical value of R_0 exceedingly complex. If one is not careful, the early accelerations may result in an estimate of R_0 that is biased towards higher values than that of the true underlying R_0.

Estimates of R0 as a Function of Size of Super-Spreader 1st Generation and Number of Generations Used for Estimate

Gen. #, t	E[NI(t)], Determ.	E[NI(t)], p=.25	Est. R_0 p=.25	E[NI(t)], p=.125	Est. R_0 p=.125	E[NI(t)], p=.0625	Est. R_0 p=.0625
0	1	1	-	1	-	1	-
1	2	8	8.00	16	16.00	32	32.00
2	4	16	4.00	32	5.66	64	8.00
3	8	32	3.17	64	4.00	128	5.04
4	16	64	2.83	128	3.36	256	4.00
5	32	128	2.64	256	3.03	512	3.48
6	64	256	2.52	512	2.83	1024	3.17
7	128	512	2.44	1024	2.69	2048	2.97
8	256	1024	2.38	2048	2.59	4096	2.83
9	512	2048	2.33	4096	2.52	8192	2.72
10	1024	4096	2.30	8192	2.46	16384	2.64

Table 1

6.3. A Heterogeneous Population Mixing Model

Many mathematical models of infectious disease progression assume a homogeneous population, in which essentially we all act statistically as identical clones. But real societies are quite heterogeneous, in many ways. Here we focus on one source of

variability across the population, a source that affects directly the spread of infectious diseases, and ultimately can cause troubles again in estimating R_0.

For illustrative purposes, we allow for two types of persons in the susceptible population: Highly socially active persons and persons whose social activity is Low. We assume that 'social activity' refers to frequency of human contacts, not necessarily in purely social situations but often in professional and day-to-day living situations. A retail store clerk has many human contacts per day and is thus socially active. A novelist working from an office at home has many fewer social contacts on a typical day and is thus characterized by a Low level of social activity. We assume that social contacts occur as a homogenous Poisson process, with rate parameters defining the level of social activity. In particular, define

λ_H = Poisson rate of social contacts per day of a High Activity person
λ_L = Poisson rate of social contacts per day of a Low Activity person
n_H = initial population of High Activity persons
n_L = initial population of Low Activity persons
$n_H + n_L$ = total population

We like to think of human interactions, generated by Poisson processes, in terms of some physical model that we can visualize. Suppose that each time a person interacts with another she leaves a slip on the ground, labeled H or L, depending on whether she is High Activity or Low Activity, respectively. Each interaction provides 2 slips, one from each of the two people who interacted. In this simple example, we ignore higher-level interactions of three of more people simultaneously (the 'three body problem"!) At the end of day we can sample interactions by randomly 'picking up pairs of slips.' As perceived by those who are interacting, there are on average $n_H\lambda_H$ interactions of High Activity people during the day and $n_L\lambda_L$ interactions of Low Activity people during the day. But just as in the clinking of wine glasses, one must divide by two to count the number of clinks – since it takes two to clink, and it takes two to interact. Thus the mean total number of person-to-person interactions during a day is $(n_H\lambda_H + n_L\lambda_L)/2$. This result clearly generalizes to any number of activity-level categories.

To obtain some comfort and familiarity with this heterogeneous mixing model, consider a random person, G, High Activity or Low Activity. The next interaction of G with another will be with a High Activity person with probability $n_H\lambda_H/(n_H\lambda_H + n_L\lambda_L)$. That is simply because the fraction of interactions that represent High Activity interactions is $n_H\lambda_H/(n_H\lambda_H + n_L\lambda_L)$. If High Activity interactions comprise 90 percent of all interactions, then G's next interaction will be with a high activity person with probability 0.90. The next interaction of G will be with a Low Activity person with probability $n_L\lambda_L/(n_H\lambda_H + n_L\lambda_L)$. Now consider a randomly selected interaction pair, with persons G1 and G2. The next G1-G2 interaction is likely to be High Activity with another High Activity with probability $[n_H\lambda_H/(n_H\lambda_H + n_L\lambda_L)]^2$. Recall that each person who interacts with another figuratively leaves a slip of paper on the ground at the point of interaction. Think of this result as picking up two paired slips of paper from the ground. An interaction between an inactive person and another inactive person occurs with probability $[n_L\lambda_L/(n_H\lambda_H +$

$n_L\lambda_L)]^2$. The fraction of interactions that involve both a High Activity person and a person of Low Activity is $2[n_H\lambda_H/(n_H\lambda_H + n_L\lambda_L)]\ [n_L\lambda_L/(n_H\lambda_H + n_L\lambda_L)]$. These results also generalize to more than two activity-level categories.

Now let's examine R_0 in this context of heterogeneous activity levels. We can write R_0 as follows:

$R_0 \equiv E[NI]$ = Expected number of new infections generated by a person randomly selected from the population and who is infectious, assuming a population of all susceptables.

That is, R_0 is the basic reproductive number when averaged over all people in the population, those Highly Active and those with Low Activity. We can express R_0 as conditional expectations, conditioned on activity level,

$$\begin{aligned} R_0 &= E[NI \mid \text{High Activity Person}](n_H/\{n_H + n_L\}) + \\ &\quad E[NI \mid \text{Low Activity Person}](n_L/\{n_H + n_L\}) \\ &= [R_0 \mid \text{High Activity}](n_H/\{n_H + n_L\}) + [R_0 \mid \text{Low Activity}](n_L/\{n_H + n_L\}). \end{aligned}$$

If each interaction yields infection with probability p, then we can write,

$$R_0 = p\lambda_H(n_H/\{n_H + n_L\}) + p\lambda_L(n_L/\{n_H + n_L\}).$$

As an example, suppose $p = 0.04,\ \lambda_H = 100,\ \lambda_L = 10,\ n_H = n_L$. Then we have $R_0 = 0.04(100)(1/2) + 0.04(10)(1/2) = 2.0 + 0.20 = 2.2$. Yet the R_0 for High Activity people is 4.0 and the R_0 for Low Activity people is 0.40. In the initial days and perhaps weeks of the pandemic, the huge majority of new infections will be caused by High Activity people, and the exponential growth rate of the disease will be seen as close to 4.0 new infections per generation, not 2.2. Eventually, due to their becoming infected early in the pandemic, the High Activity people will leave the population of susceptibles much faster than the Low Activity people, and R_0 will soon and very naturally drop to a value less than 1.0, meaning geometric dying out of the infectious disease. Any statistical estimate of R_0 based on curve fitting to the exponential growth during the early days and weeks of the pandemic will most likely result in significant upward biasing of the estimate of R_0.

This potentially significant upward biasing is not limited to one type of heterogeneity. In addition to heterogeneity in social activity, there are multiple other sources of heterogeneity, including a person's susceptibility to becoming infected and an infected person's level of infectivity. The consequences of these additional sources of heterogeneity are explored with simple spreadsheet models in Larson (2007) and its appendix. Each type of heterogeneity adds complexity to estimating R_0 from data of past pandemics, and each causes the early days of the pandemic to be uncharacteristically influenced by extremes in the distributions, extremes in levels of social contacts, in

propensity to become infected and in level of infectivity to others. These same models are extended to include spatial heterogeneities in Nigmatulina and Larson (2007).

Other researchers are now noticing the many limitations of R_0 as well. M. Lipsitch, *et. al.*(2003), in analyzing SARS, stated, "Future work should certainly focus on quantifying transmission and other epidemiological parameters in a variety of circumstances… to construct more detailed models of transmission that realistically incorporate **the effects of heterogeneities** in specific settings." (emphasis added) Both he and Riley (2003) went far beyond the usual R_0 concept to explain what was happening in the Hong Kong population.

At time of this writing we have not explored in detail the statistical estimates of others when attempting to estimate a numerical value for R_0 from past pandemics. (Note 6.) But, interestingly, if upward biasing is found due to insufficient consideration of the effects of heterogeneity, than the oft-quoted estimated low values for R_0 for the Great Influenza of 1918-1919 may in fact be overestimates, and the true value of R_0 may be less, perhaps significantly less, perhaps quite close to 1.0. That question is the subject of on-going research.

7. Why a Constant?

One might ask why epidemiologists and health policy analysts enjoy treating R_0 as a given constant, a type of constant of nature, rather than considering its entire probability distribution and its emergent evolution over the course of a pandemic. At this time, we can only speculate on the answer. We list here our hypotheses:

1) **Simplicity.** Dealing only with the mean of a random variable is considerably simpler than dealing with its entire probability law.

2) **Tipping Point.** Treated as a constant, a value of R_0 greater than 1.0 will yield an exponentially growing disease, whereas any value less than 1.0 suggests geometric decay of the disease. Thus, with elegant simplicity, one has an intuitive tipping point that is easy to understand.

3) **Analytical Tractability.** A constant unchanging R_0 is a convenient parameter to insert into (what appear to be) fancy coupled, differential equations with constant coefficients. These equations are used to project the time evolution of the disease. Those who are mathematically inclined can prove theorems about these equations, but similar theorems are very difficult and sometimes impossible to prove with more complicated, more realistic models. The reader may rightfully ask, "Which is better? …proving theorems about highly stylized and unrealistic models or gaining decision insights by reviewing numerical results from more complex models?"

4) **Fixed Value Makes it a Medical Concern.** If R_0 is considered to have a given fixed value, say R_0=2.00, then the time-evolution of the disease is certain and all

decisions deal with health care system responses to those who become ill. It's as if the citizenry, members of the susceptible class, are totally helpless. The medical issues are numerous: surge capacity of hospitals, distribution of anti-virals, stocking of respirators, delaying elective surgeries, etc. If R_0 were treated as the movable quantity that it truly is, then the importance of human behavioral patterns involving social distancing and hygienic activities would come into play. These non-pharmaceutical interventions (NPI's) are not the focus of the medical industry and establishment. They are rooted in other disciplines and industries.

One of our key points is that treating R_0 as the total definition of the problem sharply delimits the policies available to treat the threat of the disease.

8. Summary.

We have focused on R_0, the 'basic reproductive number,' a fundamental input constant for most existing mathematical models of pandemic influenza as well as many other infectious diseases. R_0 is the mean of a random variable, defined as *the mean number of new influenza infections created by a newly infected person in a population of all susceptible people*. In many circles R_0 has taken on almost sacred significance.

We have argued that R_0 is limited in policy and scientific value as is any single parameter attempting to characterize a complex probabilistic process. If one accepts the assumption that the act of becoming infected is related directly to the number and intensity of daily human contacts, we demonstrated by simple logic that R_0 does not exist as a separate 'constant of a particular influenza,' but rather its value is determined by social context and behavioral patterns as well as by the "physics'' of the influenza virus. For an infected individual circulating among a population of susceptibles, R_0 can be expressed as the product of the mean number of human contacts per day and the probability that any random human contact will result in infection. Since both of these quantities are under our control to a limited extent, R_0 is best viewed as an output of a modeling analysis, not an input.

Much useful information is contained in the entire probability distribution that has R_0 as its mean. With this view, we showed by simple arguments that R_0 can be greater than 1.0 and still, contrary to popular belief, the probability of an exponentially growing pandemic may be arbitrarily small. This counterintuitive result is a direct consequence of the stochasticity of the system. Finally, we showed that attempts to estimate R_0 from data of previous pandemics is fraught with methodological complexities, due primarily to heterogeneities in the population that cause super-spreaders and/or socially active people to be the first propagators of the disease. Unless one is careful, statistical estimates of R_0 based on early exponential growth of reported cases may be significantly upwardly biased.

Acknowledgements.
The author would like to thank for helpful discussions his MIT colleagues, Dr. Stan Finkelstein, Ms. Karima Nigmatulina and Mr. Katsunobu Sasanuma. The Sloan Foundation of New York supports this research.

Notes:

1. http://www.eu-cu.com/germany.htm, accessed July 29, 2006.
2. People Statistics> Total fertility rate (most recent) by country http://www.nationmaster.com/graph/peo_tot_fer_rat-people-total-fertility-rate, accessed December 2, 2007.
3. "Hong Kong's health secretary, Dr. Yeoh Eng-kiong, said … that sewage is responsible for SARS infecting 321 people at the Amoy Gardens high-rise apartment complex in Hong Kong. She said ventilation fans sucked particles of the virus out of a faulty sewage system and spread it throughout the building." http://www.wired.com/medtech/health/news/2003/04/58534 1/26/08.
4. Hollingsworth TD, Ferguson NM, Anderson RM. Frequent travelers and rate of spread of epidemics. Emerg Infect Dis [serial on the Internet]. 2007 Sep [*9*]. Available from http://www.cdc.gov/EID/content/13/9/1288.htm
5. See also http://homepages.newnet.co.uk/dance/webpjd/offstats/profiling.htm 1/26/08.
6. Lipsitch, *et.al*. estimate R_0 for the 1918 influenza pandemic not with curve fitting but by using 'excess deaths' and then inserting the data into standard S-I-R models for disease progression.

References

Diekmann, O., A. P. Heesterbeek, J. A. J. Metz, "On the Definition and computation of the basic reproductive ration R_0 in models for infectious diseases in heterogeneous populations," Journal of Mathematical Biology, Vol. 28, pp. 365-382, 1990.

Glass R.J., Glass L.M., Beyeler W.E., Min H.J., "Targeted social distancing design for pandemic influenza." *Emerg Infect Dis* [serial on the Internet]. 2006 Nov [*cited Feb. 6, 2008*]. Available from http://www.cdc.gov/ncidod/EID/vol12no11/06-0255.htm

Glass, L. M., R. J. Glass, "Social contact networks for the spread of pandemic influenza in children and teenagers," to appear *BMC Public Health*, 2008.

Heesterbeek JAP, "A Brief History of R_0 and a Recipe for its Calculation," *Acta Biotheor*. 2002; 50:189-204.

Heffernan, J. M., R.J. Smith and L.M. Wahl, "Perspectives on the basic reproductive ratio," *Journal of The Royal Society Interface*, Vol. 2, No. 4, 22 Sept. 2005, pp. 281 – 293.

Keeling, Matt. ""The Mathematics of Diseases," *+plus magazine,* March 2001, http://plus.maths.org/issue14/features/diseases/2pdf/index.html/op.pdf (accessed December 28, 2007)

Larson, R. C. "Simple Models of Influenza Progression within a Heterogeneous Population," *Operations Research*, Vol. 55, No. 3, pp. 399-412, 2007.

Lindholm, Mathis, "Stochastic epidemic models for endemic diseases: the effect of population heterogeneities," Research Report 2007:10, Licentiate thesis, Mathematical Statistics, Stockholm University, Stockholm, Sweden, ISSN 1650-0377.

Lipsitch, M., C. E. Mills, J. Robins, "Estimates of the Basic Reproductive Number for 1918 Pandemic Influenza in the United States -- Implications for Policy," 2007, published only on line for MIDAS, Models of Infectious Disease Agent Study, at http://www.ghsi.ca/documents/Lipsitch_et_al_Submitted%2020050916.pdf (accessed Feb. 8, 2008).

Lipsitch, Marc, *et. al.* Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *ScienceExpress*, http://www.sciencexpress.org, 23 May 2003 Page 1/10.1126/Science.1086616.

Mills, C.E., J. M. Robins and M. Lipsitch, "Transmissibility of 1918 pandemic influenza." *Nature* **432**, 904-906. 2004.

Nasell, I. "The threshold concept in stochastic epidemic and endemic models," In *Epidemic Models: their structure and relations to data* (ed. D. Mollison), pp. 71-83. Cambridge University Press, 1995.

Nigmatulina, Karima R. and Richard C. Larson, "Stopping Pandemic Flu: Government and Community Interventions in a Multi-Community Model," submitted to *European Journal of Operational Research.* 2007.

Riley, Steven, *et. al.* Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *ScienceExpress*, http://www.sciencexpress.org, 23 May 2003 Page 1/10.1126/Science.1086478.

Watson, H. W., Francis Galton, "On the Probability of the Extinction of Families," *The Journal of the Anthropological Institute of Great Britain and Ireland*, Vol. 4, 1875 (1875), pp. 138-144.

Peter Whittle, *Probability*. John Wiley, 1970, ISBN 0-471-01657-8, pp 124-125.