Disease Spread as a Function of Socioeconomic Status in Manhattan

Julie Blackwood[1], Carlos Chiquete[2],

Russell Latterman[3], Stephen Small[4]

[1] Department of Mathematics and Statistics
Rochester Institute of Technology, Rochester, NY 14623
[2] Department of Mathematics,
University of Arizona, Tucson, AZ 85719
[3] Department of Mathematics and Statistics,
Arizona State University, Tempe, AZ 85281
[4] Department of Mathematics
Norfolk State University, Norfolk, VA 23504

August 5, 2005

Abstract

When studying the spread of disease, it is imperative to consider variable factors such as population distributions, the interactions between differing populations, and socioeconomic factors. We use a network model of interacting nodes with contact rates dependent on population size and socioeconomic status to explore the disease spread across the twelve districts of Manhattan, New York City. Influenza was chosen as an example due to its short infection period and negligible disease-related deaths in comparison to prevalence levels. Since transmission occurs primarily through casual contact, proportionate mixing is incorporated in the model. Numerical simulation of the model and sensitivity analysis of its parameters are then used to identify critical factors, dependent on socioeconomic status, responsible for the severity of the epidemic. Vaccination strategies are also implemented to explore what methods will have the greatest effects on the dynamics of the model.

1 Introduction

As we head into the 21st century, health care is one of the most prominent topics of discussion in the United States. Taking proper measures to ensure that everyone in the country receives adequate health care is a major concern, and Congress has been petitioned to ensure the Health Care Bill of Rights is not ignored. The bill would provide universal coverage, cost control, and prescription drug coverage [12]. These are all reasonable items that many Americans would like to receive, yet even today, many disparities persist concerning health care in this country.

Forty-one million Americans in the United States do not have health care. Surprisingly, eighty-two percent of the uninsured population come from working families [11]. However, there are various factors affecting these numbers. Availability of health care can often be linked to socioeconomic issues of differing neighborhoods and families. Ethnicity, age, income, and education play vital parts in the average amount of health care a person in a certain area receives.

There are some recurring trends among certain families and their ability to receive health care coverage. One dominant trend is that families with higher incomes are healthier since health care coverage is more readily available to them. In contrast, health coverage is scarce for families with median or below median income. The financial status of a family increasingly determines how much health care they receive. For low-income families, health care often becomes a passive proposition. This passivity is manifested in reduced vaccination rates for influenza.

Vaccines are a means to decrease the probability of infection in many populations. A person is supposed to receive a flu shot every year, however that does not often occur. Different people have varying reasons for refraining from flu shots. Five common reasons for missing an annual flu vaccination among African-Americans are: not knowing it was needed (20.6 %), thinking the shot could cause the flu (18.4 %), thinking the shot has adverse effects (15 %), thinking the shot could prevent the flu (14.5 %), and simply forgetting to get the vaccination (12.6 %) [10]. However, educating people about the vaccine for influenza may increase the probability of receiving a vaccination and thus decrease the annual flu epidemic. Educational methods may include giving general information on flu vaccinations in order to correct many misconceptions and to inform people about locations where vaccines can be received.

Every year millions of Americans are infected with the Influenza virus, precipitating a set back in the economy due to lost work days and lowered work efficiency. Population density and socioeconomic status have significant effects on outbreak. Data shows that flu affects poor districts disproportionately with respect to national averages [14] [3]. Vaccination is a safe and effective method of prevention, but the majority of the population does not get vaccinated [5]. In response to this, the Center for Disease Control (CDC) has pushed to increase vaccination rates in the US [11].

Data is published annually that can attest to the discrepancies in health care and socioeconomic status [14]. The availability of information on Manhattan allows exploration concerning the differences in socioeconomic status between communities. A major factor for using New York City in this model is that New York City has a wide income gap and is representative of many different levels of economic status. New York City's five boroughs, in particular the Manhattan borough, exhibit many more disparities in income, population, and health status than those prevalent in the United States.

Given the disparities in socioeconomic status that characterize the borough of Manhattan, the wealth of data, and the CDC's willingness to improve vaccination rates in the United States, we develop a mathematical model of the spread of the flu throughout the twelve densely populated, socioeconomically diverse districts of Manhattan, New York City. The model will attempt to optimize the vaccination increases that may result from increased education efforts. This subject does not lack in importance since the impacy of the flu cannot be understated in terms of cost both human and economic.

2 Model Formulation

We develop a model of Manhattan, one of New York City's five boroughs. Manhattan is composed of twelve districts, each of which we consider to be a subpopulation of the city. Figure 1 displays the divisions of each borough as well as the districts within them according to the New York City Department of City Planning [14]. Interactions occur between the people of Manhattan having important effects on the dynamics of the spread of a disease when it is introduced into areas of Manhattan, thus making this a metapopulation model [9].

Each district is represented as a node within a network; there are twelve nodes that are all interconnected. Our model is compartmental with four classes: susceptible, latent, infective and recovered (SLIR) and the spread of disease, in our case influenza, is affected by the interactions and activity levels within districts. Susceptible are those at risk of contracting the disease, latent are those who are infected but not yet infective, infective are those able to pass on the disease and recovered are those who have recovered from the flu. The model consists of four difference

Figure 1: Boroughs of New York City; our focus is on Manhattan

equations for each of Manhattan's districts and is given by

$$\begin{align} S_{i_{t+1}} &= S_{i_t} G(I_{(j=1,..,12)_t}) \tag{1} \\ L_{i_{t+1}} &= (1 - G(I_{(j=1,..,12)_t}))S_{i_t} + \lambda L_t \tag{2} \\ I_{i_{t+1}} &= (1-\lambda)L_{i_t} + \omega_i I_{i_t}, \tag{3} \\ R_{i_{t+1}} &= (1-\omega_i)I_{i_t} + R_{i_t} \tag{4} \end{align}$$

where $1 \leq i \leq 12$. Since we are only observing short time periods, we assume that in the time course of the simulations there are no births, deaths and the proportion of people who move into and out of a particular district is insignificant. Note that $(S_i + L_i + I_i + R_i)_{t+1} = (S_i + L_i + I_i + R_i)_t = N_i = K_i$, where N_i is the population at district i and K_i is a constant.

Here, $G(I_{(j=1,..,12)_t})$ is the probability of an individual not becoming infected from t to $t+1$. For homogeneously mixed populations, $G(I_{(j=1,..,12)_t})$ will be taken to be of the form $\sum_{j=1}^{12} e^{-\alpha_j \frac{I_{j_t}}{N_j}}$, where $\frac{I_{j_t}}{N_j}$ is the proportion of infected individuals in district j and α_j denotes the force of infection. We assume that the force of infection is higher among more populous districts and also in a more impoverished neighborhood, given that poor neighborhoods tend to have more community housing (such as apartments or public housing); thus α_j is district-dependent. In Table 1, we define the characteristics of each of the districts, including population size and poverty level.

District	Population Size	Proportion of Population in Poverty
1	34420	0.1313
2	93119	0.1086
3	164407	0.2826
4	87479	0.1490
5	44028	0.1116
6	136156	0.0974
7	207699	0.1090
8	217063	0.0663
9	111724	0.3050
10	107109	0.3608
11	117743	0.3727
12	208414	0.3070

Table 1: District characteristics (data is taken from the New York City Department of City Planning)

Letting $\alpha_j \frac{I_{jt}}{N_j} = x_{jt}$, where x_{jt} is the proportion of contacts that are infective, and assuming that $0 < x_{jt} << 1$, finding the Taylor expansion about x_{jt} gives,

$$e^{-x_{jt}} = (1 - x_{jt}) + h.o.t.$$

where h.o.t. refers to higher order terms of order x_{jt}^2 and above. Since x_{jt} is the probability of becoming infected from an individual in district j, then $1 - x_{jt}$ approximates the probability of not becoming infected. Writing the difference equations in this form avoids any negative solutions; all parameters are between 0 and 1.

Now we must take into account interactions between districts. We consider proportionate mixing which is represented by

$$\bar{P}_j = \frac{C_j N_j}{\sum_{m=1}^{12} C_m N_m}, \tag{5}$$

where for some district i, each $\bar{P}_j$ represents the proportion of total contacts per individual made in district j and C_j is the probability of contacting a certain person in district j [7]. The contact probability, C_j, is approximated by equating it to the proportion of the population in district j. Therefore it is more likely to contact a person from a more populous district. Note that $\sum_{j=1}^{12} \bar{P}_j = 1$. Combining the probability of not becoming infected given that one has come in contact with an

Parameter	Definition
λ	Probability of remaining latent
ω_i	Probability of remaining infectious
α_j	Force of transmission in district j
$\bar{P}_j$	Proportional mixing matrix
C_j	Probabilities of contact
v_i	Proportion of vaccinated individuals in district i

Table 2: Parameters

infected individual and proportionate mixing, we have

$$G(I_{(j=1,..,12)_t}) \equiv \sum_{j=1}^{12} \bar{P}_j e^{-\alpha_j \frac{I_{jt}}{N_j}}. \tag{6}$$

Also, in Equations (2) and (3), λ represents the probability of remaining in the latent stage of influenza. Similarly, the probability of remaining infected is given by ω_i, which is district-dependent because in a more poor district, access to antiviral drugs are not as readily available.

Finally, in New York City, a proportion of susceptibles will receive vaccinations. To account for this it is assumed that people who are vaccinated receive the vaccine prior to the peak flu season, which is where we begin our analysis. At the initial time, a fraction of the susceptible class of each district i, given by v_i, is immediately moved into the recovered class for the purpose of taking them out of the group at risk of becoming infected. Thus, the initial conditions for the susceptible class are given by $S_{i_0} = (1 - v_i)(N_i - I_{i_0} - L_{i_0})$. The methods by which these initial vaccinated proportions are determined can be found in the Parameter Fitting Section and Table 2 lists the parameter definitions.

3 Parameter Fitting

To determine our model's parameter values, United States time series data on influenza is used to approximate the time series data for each of the twelve districts. Data available from the Center for Disease Control (CDC) gives the weekly admission rates to the hospital as a result of influenza [3]. From this information, we find the proportion of people in the United States in the hospital for influenza during a given week and assume these weekly percentages also hold in Manhattan (see figure 2). By finding the average total number of people admitted to the hospital during the influenza season as well as the total number of influenza cases in the United

States during a given year, it is found that there are approximately 71 cases of influenza for every one person admitted to the hospital.

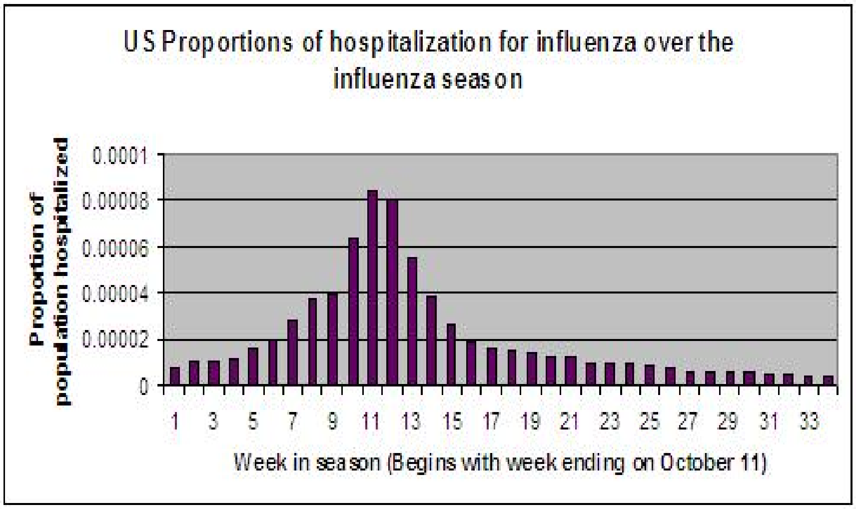

Figure 2: Time series data for the United States taken from the Centers for Disease Control

This information is used to approximate the expected proportion of total influenza cases in the United States during a given week. Since we are assuming that the districts of Manhattan follow these same patterns, we need information on admission rates to the hospital for each district throughout the year. Our information on the twelve districts is taken from the New York City District Profiles from the CDC and although it gives information on the population size of each as well as the proportion of people living in poverty, it does not give hospitalization rates [10]. However, the DOHMH gives this information but their division of Manhattan has only ten districts and the divisions are significantly different than ours [15].

To solve this problem and find the rates for our districts, we compared the hospitalization rates to the proportion of people living in poverty within the ten district data and found that there is a correlation coefficient r of 0.80697 (see figure 3). Because of the significant correlation, we assume that hospitalization rates and level of poverty in an area are correlated to one another by the linear regression line

$$y = 0.0006x - 0.0586$$

where y is the admission rate to hospital (per 100,000 people) and x is the proportion of the population in poverty. Using this regression line, we calculate the admission rates to the hospital (related to influenza) based on the levels of poverty for each of the twelve districts.

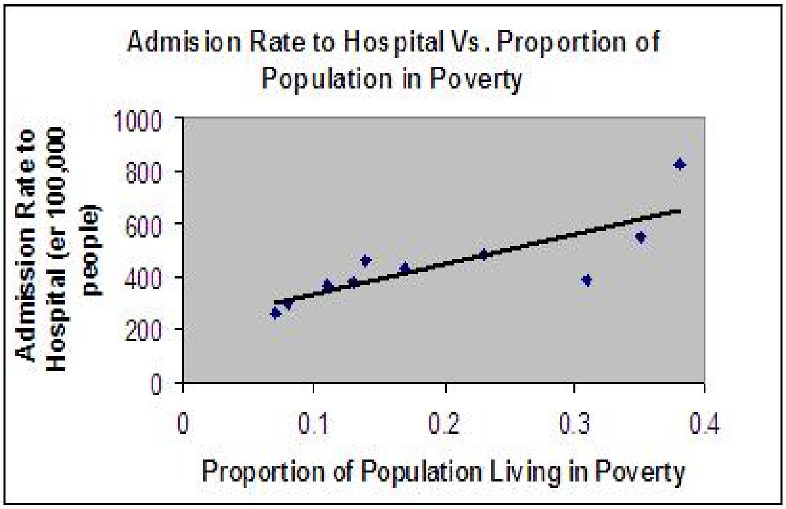

Figure 3: Regression for the admission rate to the hospital for influenza vs. the proportion of population living in poverty in the districts of Manhattan. Data is taken from the New York City Department of Health and Mental Hygiene.

From these calculated rates, we approximate the total number of hospitalizations for each week based on the hospitalization rates and the time series data for the United States. Using the hospitalization rates we calculate the number of hospitalized individuals during a given week in each district and from that the corresponding total number of infected individuals (there are 71 infected for every one hospitalized individual) are estimated.

Once these time series data are obtained, we fit the parameters γ_i, ω_i, λ as well as the initial conditions for the latent and infectious classes. The initial conditions for the recovered class are approximated according to the percentage of vaccinations made in each district; the number of vaccinated individuals are considered to be the initial conditions for the recovered class.

In our model we assume that people who are vaccinated receive vaccinations prior to the beginning of the flu season, which is true in most cases. Since flu vaccination data is not given specifically for each district of the Manhattan borough we used a weighted method based on age distribution data from the various districts. First, we use data from the National Health Interview Survey (NHIS) to find what percentage of people get flu vaccinations [5]. The breakdown of the statistics are age related; the age groups include 18 to 49, 50 to 65, and 65 and above. We use the age breakdowns for the districts of Manhattan and relate them to the national survey, assuming that the vaccination proportions in Manhattan are comparable to those in the United States. For each respective age group we multiply the population of the Manhattan district by the national percentage corresponding to the age group vaccinated. Since the NHIS has no available data for people under age 18, we assume that 25 % of that population is vaccinated. We then take a weighted average of the percentage of vaccinated with respect to age for each district and find, in total, what percent of the population receive vaccinations. This percentage of people is then removed from the susceptible class and placed directly into the recovered class.

The latent period of influenza spans from one to four days and the infectious period lasts for up to seven days. Thus in fitting these parameters, we constrain λ to this information and allow ω_i to correspond to the likelihood of remaining infectious for 4 to 7 days. The probability of becoming infected given that a susceptible individual contacts an infectious individual can be anywhere between 0.2 and 0.9, so we allow this parameter to range between these values [6]. Since α_j is also a probability, it is allowed to range from 0 to 1 for each j and is dependent on the population size and poverty level of each district j.

After performing parameter fits, they were implemented in a program constructed in MATLAB. A sample of one such fit can be seen in figure 4. Our main concern in determining the parameters is that the peak of the influenza season has an approximate match to the actual data. Our focus lays in the peak levels of influenza activity, subsequent influenza cases are less important for our purposes–thus, the tails of the fits are inconsequential. There are, however, instances where the peak of the fitted curve does not match the maximum number of infected individuals. This can be attributed to the fact that our data is approximated, so there is room for error in finding a best fit curve although all peaks are within a certain tolerance.

4 Analysis

4.1 Equilibrium Points

The equilibrium points that occur in our difference equation system depend on the following conditions derived from equations (1-4).

$$S_{i_\infty} = G(I_{(j=1,..,12)_{infty}})S_{i_\infty} \tag{7}$$

$$L_{i_\infty} = (1 - G(I_{(j=1,..,12)_{infty}}))S_{i_\infty} + \lambda L_{i_\infty} \tag{8}$$

$$I_{i_\infty} = (1 - \lambda)L_{i_\infty} + \omega_i I_{i_\infty} \tag{9}$$

$$R_{i_\infty} = (1 - \omega_i)I_{i_\infty} + R_{i_\infty} \tag{10}$$

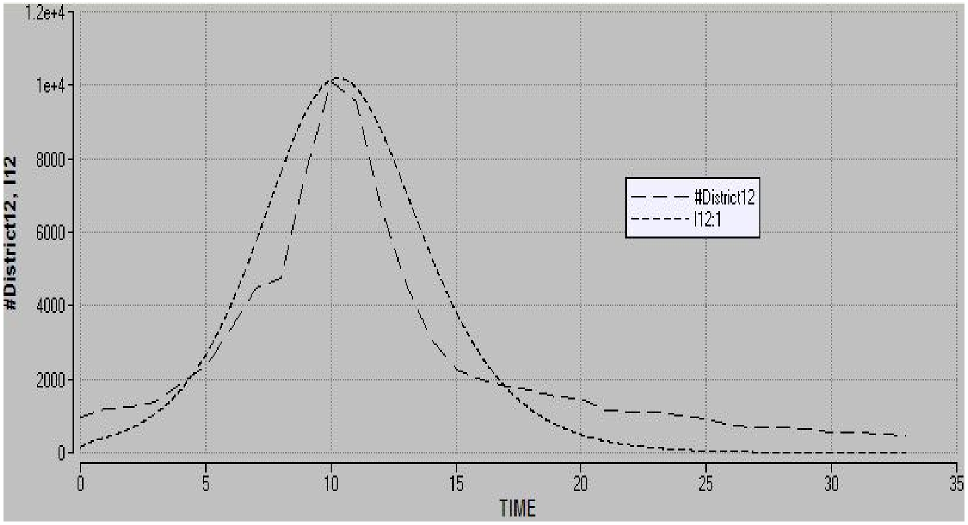

Figure 4: Parameter fit for District 12 where the vertical axis represents the number of infected individuals at time t, the smooth curve represents the fitted data, and the jagged curve represents the time series data obtained from the DOHMH

where S_{i_∞}, L_{i_∞}, I_{i_∞}, and R_{i_∞} refer to the equilibrium values of susceptible, latent, infective, and recovered, respectively.

The condition for S_{i_∞} suggests that either $G(I_{(j=1,..,12)_\infty}) = 1$ or $S_{i_\infty} = 0$. If we consider the former then, from eq. 6,

$$G(I_{(j=1,..,12)_{infty}}) = \sum_{j=1}^{12} \bar{P}_j e^{-\alpha_j \frac{I_{j_{infty}}}{N_j}}.$$

Clearly, in our case $G(I_{(j=1,..,12)_\infty}) = 1$ requires that $I_{j_{infty}} = 0$ for all i.

Using this last result, Equation 9 gives that,

$$(1 - \lambda)L_{i_\infty} + \omega_i(0) = 0$$

So $L_{i_\infty} = 0$ for all i since $1 - \lambda \neq 0$. Finally, using these results, Equation 10 becomes,

$$R_{i_\infty} = (1 - \omega_i)(0) + R_{i_\infty}$$

And so the results is that $R_{i_\infty} = R_{i_\infty}$ for all i.

Therefore, the equilibrium state only has people in the susceptible and recovered classes and is disease free. We note that this equilibrium is an element of a line of equilibria since they are constrained by initial conditions satisfying $S_{i_\infty} + R_{i_\infty} = N_i$. The equilibrium line holds for all districts in the model.

A possible interpretation of the equilibrium point is that the infective die out and the susceptible class retains individuals that eventually stay in the class permanently since the infection term is zero as there are no infected, i.e. $G(I_{i=1...12_{infty}}) = 1$. If the proportion $G(I_{(j=1,..,12)_{infty}}) = 1$, this implies that the probability that you will remain in S_i is 1, therefore an individual is bound to remain susceptible. Individuals who are infective and latent must enter the recovered class since everyone eventually recovers absent a death rate. Therefore this equilibrium is designated as the Disease Free Equilibrium (DFE),

$$E_{DFE} = (N_{i=(1,..,12)} - R_{(i=1,..,12)_\infty}, 0...0, 0...0, R_{(i=1,..,12)_\infty}).$$

The remaining condition for S_{i_∞} suggests that $S_{i_\infty} = 0$ for all i. The second condition becomes,

$$L_{i_\infty} = \lambda L_{i_\infty}.$$

Again, $\lambda \neq 0$ and so necessarily $L_{i_\infty} = 0$ for all i. Similarly, $I_{i_\infty} = 0$ for all i. Finally,

$$R_{i_\infty} = R_{i_\infty}$$

The above condition holds for all i. This appears to be a special case of the first equilibrium found, i.e. $S_{i_\infty} = 0$ and so $R_{i_\infty} = N_i$ for all i.

A possible interpretation of this point is that the susceptible class has no members at the initial time. This suggests that everybody has been vaccinated.

4.2 The Basic Reproductive Number

The *basic reproductive number*, R_0, represents the average number of secondary infections resulting from the introduction of one new infectious individual into a population. For our discrete model we use the Jury Criteria which give conditions that guarantee that the system is locally asymptotically stable at the DFE. We then find a condition that must hold to guarantee the criteria is satisfied for each individual node.

We first linearize the equations $L_{i_{t+1}}$ and $I_{i_{t+1}}$ for each individual district assuming that there are no interactions between districts and that $S_{i_\infty} \approx \gamma_i N_i$ where $\gamma_i = 1 - v_i$ (i.e. γ_i is the proportion of the susceptible class who are not vaccinated):

$$J_{(L_{i_t}, I_{i_t})} = \begin{bmatrix} \lambda & \gamma_i \alpha_i e^{-\alpha_i \frac{I_{i_t}}{N_i}} \\ 1-\lambda & \omega_i \end{bmatrix}. \tag{11}$$

Calculating this at the disease free equilibrium produces the 2×2 matrix

$$J_{(0,0)} = \begin{bmatrix} \lambda & \gamma_i \alpha_i \\ 1-\lambda & \omega_i \end{bmatrix} \tag{12}$$

By the Jury Criteria, the following conditions need to be satisfied:

$$a_1 + a_2 + 1 > 0 \tag{13}$$
$$1 - a_1 + a_2 > 0 \tag{14}$$
$$1 - a_2 > 0 \tag{15}$$

where a_1 is the negative trace, $-(\lambda+\omega_i)$, and a_2 is the determinant, $\lambda\omega_i - (1-\lambda)\gamma_i\alpha_i$ [2]. The only condition required for the above three conditions to be satisfied is that $\alpha_i \frac{\gamma_i}{1-\omega_i} < 1$ and therefore for every district i,

$$R_{0_i} = \alpha_i \frac{\gamma_i}{1-\omega_i}$$

where $\frac{1}{1-\omega_i}$ is the average amount of time spent in the infectious stage and if $R_{0_i} < 1$, the disease free equilibrium of district i is locally asymptotically stable. Thus we now have the basic reproductive number for each district, assuming that there are no interactions between districts.

Based on prior work on mixing, we suspect that the value of $R_{0_{total}}$, the basic reproductive number is representative of the entire system, is

$$R_{0_{total}} = \sum_{i=1}^{12} \left[\gamma_i \frac{\alpha_i}{1-\omega_i} \frac{N_i}{\sum_{k=1}^{12} N_k} \right],$$

That is, the weighted average of each value of R_{0_i} according to their population size.

4.3 Sensitivity Analysis

The solutions that we found depend on various parameters including force of infection (α_i's), initial unvaccinated proportions (γ_i's), recovery probability $(1-\omega_i)$, and the latent-infective transition probability $(1-\lambda)$. This last transition probability, $(1-\lambda)$, is the same for all districts.

4.3.1 Sensitivity of $R_{0_{total}}$

We previously found that

$$R_{0_{total}} = \sum_{i=1}^{12} \left[\gamma_i \frac{\alpha_i}{1-\omega_i} \frac{N_i}{\sum_{k=1}^{12} N_k} \right].$$

and all parameters appearing in $R_{0_{total}}$ are district-dependent; α_i is dependent on population size as well as the level of poverty within a district, γ_i is dependent on the age breakdown of each district and ω_i is again dependent on the level of poverty in district i.

We find the sensitivity indexes for each of these parameters which allows us to analyze which parameters have the greatest effects on the dynamics of the system given a slight change in the parameter value. For a parameter p, the sensitivity index is given by

$$S_p = \frac{\partial J}{\partial p}\frac{p}{J} \tag{16}$$

where $J(u)$ is a function dependent on u (and thus depends on some parameter p), ∂p represents a perturbation on p and ∂J is the resulting change in J from the perturbation [1].

Finding the sensitivity indexes of the α_i's with respect to $R_{0_{total}}$ requires us to first find $\frac{\partial R_{0_{total}}}{\partial \alpha_i}$. To make this clearer, we expand $R_{0_{total}}$:

$$R_{0_{total}} = \frac{\alpha_1\gamma_1}{1-\omega_1}\frac{N_1}{N_{total}} + \frac{\alpha_2\gamma_2}{1-\omega_2}\frac{N_2}{N_{total}} + \cdots + \frac{\alpha_{12}\gamma_{12}}{1-\omega_{12}}\frac{N_{12}}{N_{total}}$$

where $N_{total} = \sum_{k=1}^{12} N_k$. It is now clear that for $1 \leq i \leq 12$,

$$\frac{\partial R_{0_{total}}}{\partial \alpha_i} = \frac{\gamma_i}{1-\omega_i}\frac{N_i}{N_{total}}. \tag{17}$$

From Equation 16 we now multiply Equation 17 by $\frac{\alpha_i}{R_{0_{total}}}$ which gives us the sensitivity index of each α_i,

$$S_{\alpha_i} = \frac{\alpha_i\gamma_i}{R_{0_{total}}(1-\omega_i)}\frac{N_i}{N_{total}}.$$

Since $S_{\alpha_i} > 0$ for all i, an increase in α_i will result in an increase in $R_{0_{total}}$; if the force of transmission is increased then the peak of the epidemic will increase. Substituting in the respective parameter values we find that districts with the highest sensitivity of α_i are predominantly those districts that have the highest proportion of people living in poverty (percent in poverty is >25 %) as well as the districts with the largest population. In particular, α_{12} has the largest sensitivity index. Using the same methods, we find the sensitivity indexes for the γ_i's, which are identical to S_{α_i} for all i:

$$S_{\gamma_i} = S_{\alpha_i} = \frac{\alpha_i\gamma_i}{R_{0_{total}}(1-\omega_i)}\frac{N_i}{N_{total}}.$$

Similarly, since $S_{\gamma_i} > 0$ for all i, an increase in the proportion of non-vaccinated individuals will result in an increase in the peak of the epidemic.

For the sensitivity of the ω_i's we find

$$S_{\omega_i} = \frac{\omega_i \alpha_i \gamma_i}{R_{0_{total}}(1-\omega_i)^2} \frac{N_i}{N_{total}}.$$

Again since $S_{\omega_i} > 0$ for all i an increase in ω_i will result in an increase in $R_{0_{total}}$. We find that the districts with the largest sensitivity indexes of ω_i are again those with more poverty and larger population sizes. Again, District 12 has the largest sensitivity index for this parameter. Also, within a district i, ω_i is always the most sensitive parameter (greater than that of the index of γ_i and α_i). This implies that $R_{0_{total}}$ is affected the most by changes in the parameters in the districts with the most poverty and largest populations. Clearly, changes in the amount of time spent in the infectious stage will have the greatest effects on $R_{0_{total}}$ and in turn has the greatest effects on the dynamics of the influenza epidemic.

4.3.2 Numerical Sensitivity Analysis

Using Berkeley-Madonna curve-fit program we used numerical methods to explore the dependence of the numerical solution on the the parameters of the model. The plots are produced using Berkeley-Madonna's numerical sensitivity function. This program produces the numerical partial derivative ($\frac{\partial I_{i_t}}{\partial p}$)of the infective number for some district over any particular parameter (p). These derivatives are then multiplied by $\frac{p}{I_{i_t}}$ in order to obtain,

$$S_p = \frac{\partial I_{i_t}}{\partial p} \frac{p}{I_{i_t}}.$$

This is the definition of the sensitivity index according to Equation (16). These were calculated over time and plotted for various parameters.

The results of the sensitivity analysis plots suggest that a perturbation in ω_i has the largest effect on the number of infective at any one time. Figure 5 shows the sensitivity of the solution with respect to various parameters including various force of infection parameters (α_i's) and proportions of vaccinated individuals as well as λ and ω_i over time. The figure is representative of the structure of all the infected classes across the districts. In this case the graph shows the sensitivity parameter of infective in District 4 as given by equation (16).

In this case ω_i initially dominates the plot. The α_i's and to a greater extent λ controls the behavior as well. It is worth noting that λ affects the solution negatively in the short run but gains prominence and a positive change in λ increases the function after about $t = 12$ weeks. This positive change in λ delays the transition of latent individuals into the infected class initially since the latent individuals stay

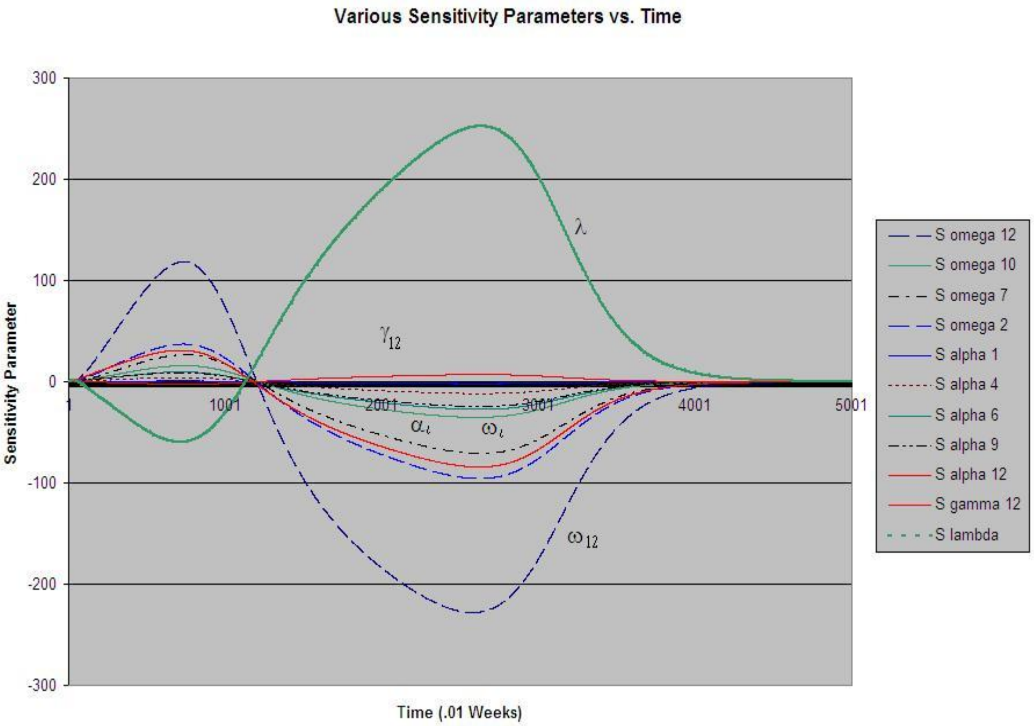

Figure 5: The sensitivity indices plot for the parameters including γ_{12}, and various ω_i's, α_i's and λ over time. In this case, the sensitivity indices depend on time. After $t = 10$ weeks, the number of infective go to zero sharply.

.

latent longer. Eventually they leave the latent class and enter the infected. This has an effect of broadening the epidemic but lessening the peak of the outbreak initially.

Additionally, the various other parameters have a much smaller effect on the infected class in District 4 than either λ or ω_{12}. The α_i's increase the infected class since higher contact among the districts clearly increases the epidemic.

Although λ and the ω_i's affect the number greatly it is impractical to consider education programs to change these parameters. For our purposes, it is far more practical to consider changing the proportion of individuals who are vaccinated in the various districts. This can perhaps be achieved in the near future.

Figure 6 displays the sensitivity parameters for the non-vaccinated proportions of the populations as a function of time. In this case, District 11 is studied although the trend holds for all other districts. Initially we can see an increase in vaccination, i.e. a decrease in γ_i causes a decrease in the number of individuals in the infective class. This decrease then causes an increase later on since it broadens the epidemic

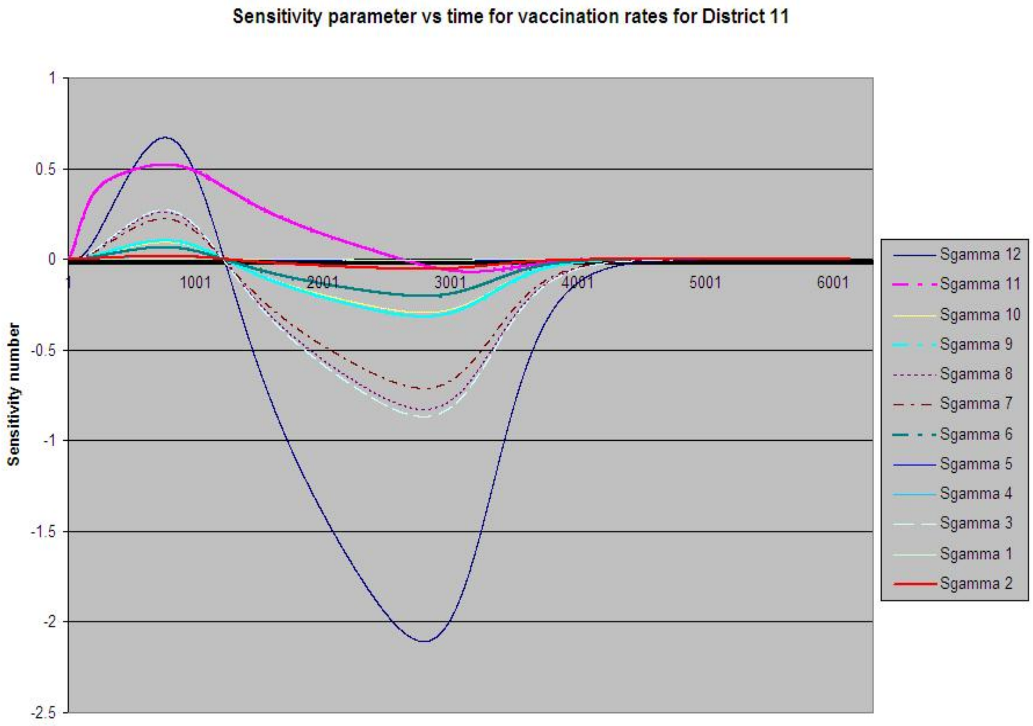

Figure 6: The sensitivity index plot for the γ_i's over time.

slightly. Additionally, we see that $S_{\gamma_{12}}$ is the largest sensitivity index. This is due to both the population as well as the activity level of the district which is higher than others. District 12 has a high index of poverty and population and therefore the γ_{12} parameter affects all of the other districts immensely.

5 Computer Methods and Simulations

We write our own programs to run simulations using MATLAB's object oriented programming language. Our main program, which we call "I vs T" is created to model the spread of the flu throughout the twelve districts of Manhattan, NY. The model calculates the number of susceptible, latently infected, infective, and recovered individuals for each district. Varying contact probabilities are estimated using the method of proportionate mixing. Other parameters are estimated in Berkeley-Madonna using data from the Center of Disease Control (CDC) [4]. By manipulation of the program structure itself, different scenarios are simulated. Using the fitted parameters, we simulate different vaccination strategies and derive possible recommendations.

5.1 Other Custom-Made Programs

We write several other programs to help us analyze our basic reproductive number and to observe the effects that different vaccination scenarios have on its value. We've written a program called "InfectionVSR0" which produces a plot of the total population of infected individuals versus increasing values of our estimated reproductive number (see Figure 8). The results support our assumption that this estimated $R_{0_{total}}$ is close to the "actual" $R_{0_{total}}$. Another program, called "RandomFit," randomly distributes 100,000 additional vaccinations to different districts. Hundreds of trials are run to find progressively better values for $R_{0_{total}}$, producing a visual representation (see Figure 9). A similar program, called "avgrandom" also performs operations to find the most efficient method of vaccination using uniform random numbers weighted by population and poverty level.

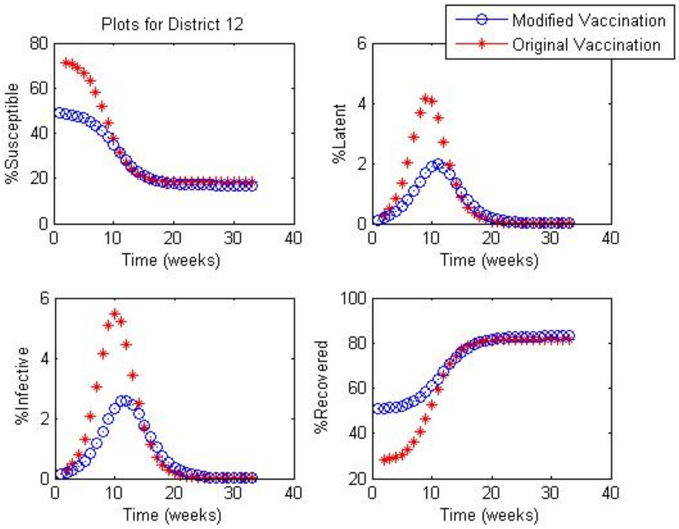

Figure 7: The is an example of a predicted flu outbreak in District 12. A simulation using original vaccination data is compared to a simulation with an additional 25% of the population vaccinated.

We discover that $R_{0_{total}}$ is lowest when all vaccinations are concentrated in the low-income districts. After thousands of computer simulated trials, the values for the smallest $R_{0_{total}}$ converge to the value produced when all vaccinations are placed in District 12. This simply means that District 12 has the greatest affect on the spread of disease; this observation is also supported through our sensitivity analysis which shows that District 12 has the greatest effect on $R_{0_{total}}$. This also makes

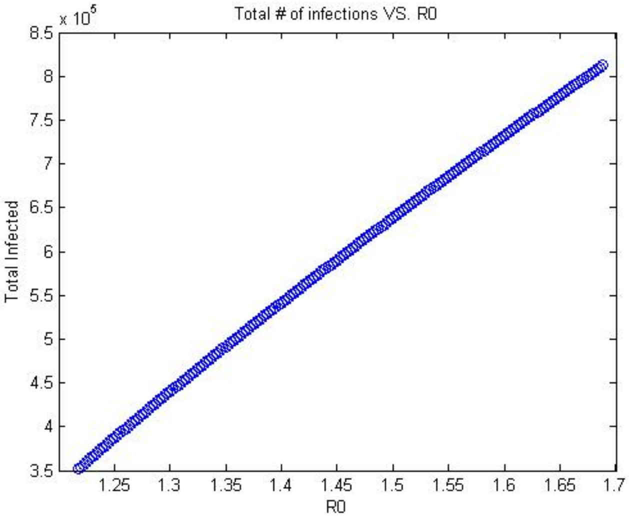

Figure 8: Infection vs. $R_{0_{total}}$

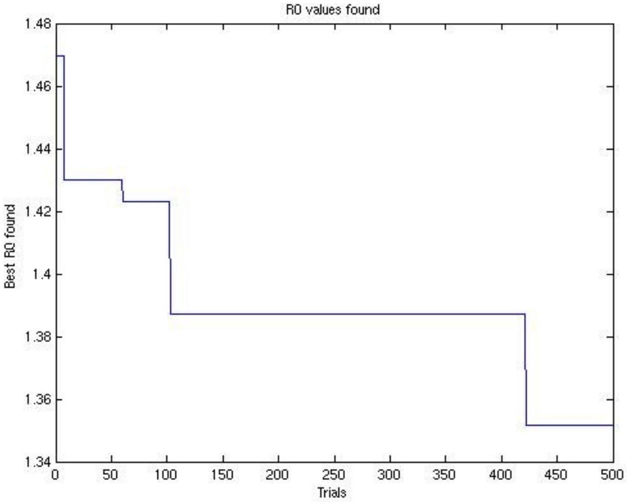

Figure 9: Progression of discovered $R_{0_{total}}$ values (values decrease with simulations)

sense from a practical point of view since District 12 has one of the highest poverty levels as well as the the second largest population. From these findings we develop a suggestion for future vaccination that places emphasis on the low-income districts. The following quantifies these results.

5.2 Optimizing Improvements in Vaccination Proportions

Using age-related data and census information, we estimate the proportion of vaccinated individuals within the twelve districts. These are referred to as the baseline vaccination proportions. It is then postulated that for a subsequent year there is an improvement of 6.5% in the number of people vaccinated before the flu seasons which translates into approximately 100,000 more vaccinations. It is assumed that a targeted education program could produce these increases in the various districts. These increases translate into a decrease in the γ_i's (non-vaccinated proportion of district i). Simulation were carried out to maximize the reduction in infected as a result of differing vaccination strategies. The following simulations attempt to isolate the factors that will lead to an optimized vaccination proportion. Initially, we test the extent to which random vaccination across the districts reduces the number of infected. We then compare the results of differing vaccination procedures including vaccinating predominantly poor districts or vaccinating only in large population districts. We compare these to the initial trivial case where we vaccinate randomly.

5.2.1 Random Uniform Distribution

Using MATLAB, we create a program to randomly distribute the number of vaccinations to each district. This process. The randomization increases the vaccination proportions for all districts in comparison to the baseline proportions. This is done using a uniform distribution that is normalized in order for the sum of the proportions to equal to one. The program then averages over 1000 simulations the percent decrease in infections using these uniform random distributions. This average was found to be 16.8%.

Figure 10 is a typical result for a uniformly random distribution of vaccination proportions per district. It displays the baseline vaccination proportions, the new proportions and the percent reduction of infected people in each district after 33 weeks (after the epidemic has died). The total reduction is also calculated and included in the figure as well as R_0 and effectivity (the number of infections avoided as a consequence of vaccinating one person above the baseline levels).

5.2.2 Uniform Distribution among the 'Rich'

There are two general classes in Manhattan. There is a gulf between the districts in terms of poverty. Harlem and its environs are generally substantially less affluent than other neighborhoods in Manhattan. It is natural to separate these into two camps. We test uniform random vaccination only among the richer districts and allow the same 100,000 additional vaccinations to be distributed between these dis-

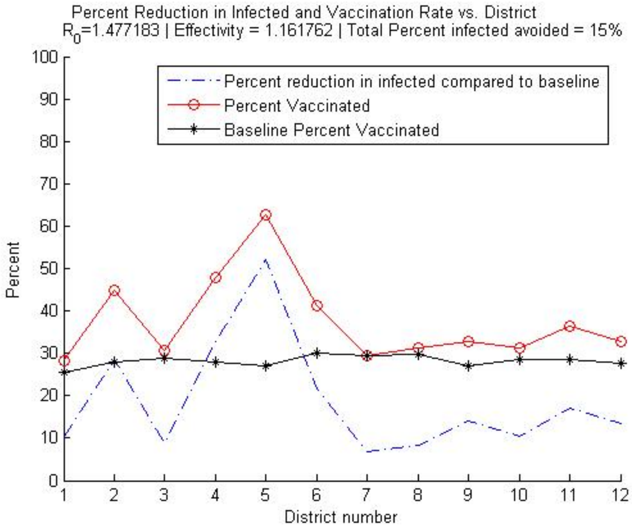

Figure 10: Result of random uniform distribution for new vaccinations

tricts (1, 2, 4, 5, 6, 7, and 8). Figure 11 displays the same information as in Figure 10. This was done in order to isolate the particular vaccination proportion that will have the largest effect on the number of infected. In this case we isolate the cases to the rich districts.

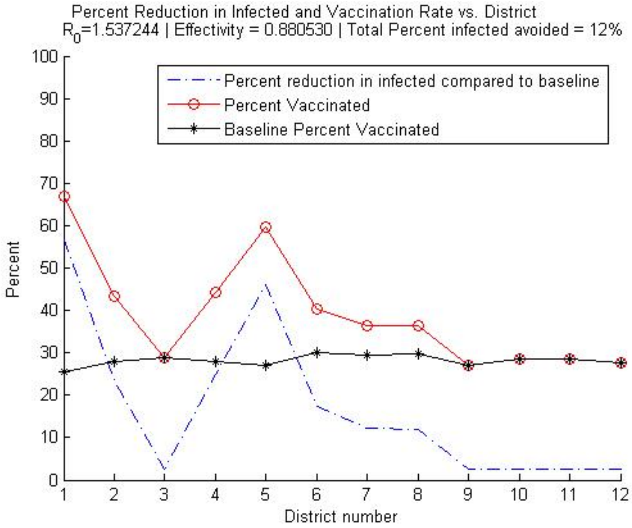

Figure 11: Result of random uniform distribution for only the 'rich' districts

We find that the typical result for this configuration of vaccination proportions yields a reduction in infection of about 12% on the average. Comparing this to the uniform distribution among all districts (16% reduction), this result demonstrates that vaccinating only rich districts lessens the efficacy of the extra vaccinations with respect to random vaccination. Given later results, this amounts to stating that since no one in the poorer districts were vaccinated results in decreased vaccination efficiency with respect to random variance in vaccination proportions. Vaccinating among the rich districts is less effective than vaccinating randomly.

5.2.3 Uniform Distribution among the ‘Poor’

The same procedure as in the ‘rich’ case is used instead for the ‘poor’ districts. These districts are 3, 9, 10, 11 and 12. The vaccination proportions are then varied randomly using the usual uniform distribution. Figure 12 displays the results of a typical simulation.

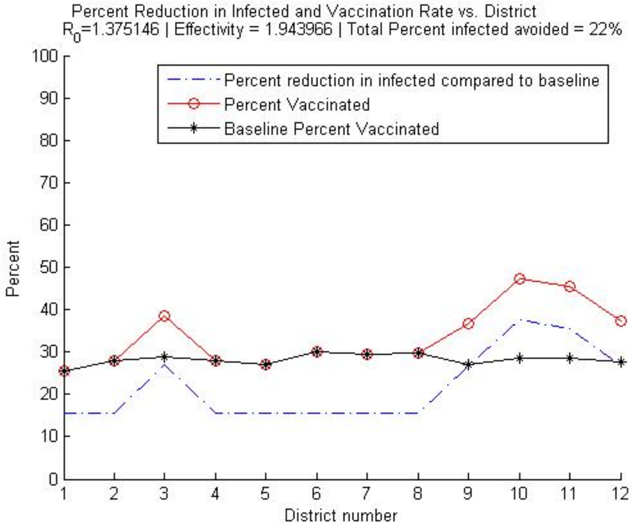

Figure 12: Result of random uniform distribution for only the ’poor’ districts

Using this distribution only among the poor classes we find that the average reduction in infected is 22% which is an improvement in efficacy over both of the other two cases (uniform random distribution among the rich and among all districts). This result suggests that the differences in force of infection has a large effect upon the behavior of the system as a whole since a perturbation in the proportion of vaccinated people among the poor classes has a much larger affect on the rich districts than vice-versa.

5.2.4 Finding the Optimal Vaccination Strategy

If there is an optimum vaccination strategy as suggested by the above results, minimizing $R_{0_{total}}$ over large number of iterations will reveal a pronounced pattern.

Therefore, we produce 100,000 different sets of vaccination proportions for the 12 districts. A program calculates $R_{0_{total}}$ for each of these random proportions and keeps the set of proportions that minimizes $R_{0_{total}}$. The results of this process appear in Figure 13.

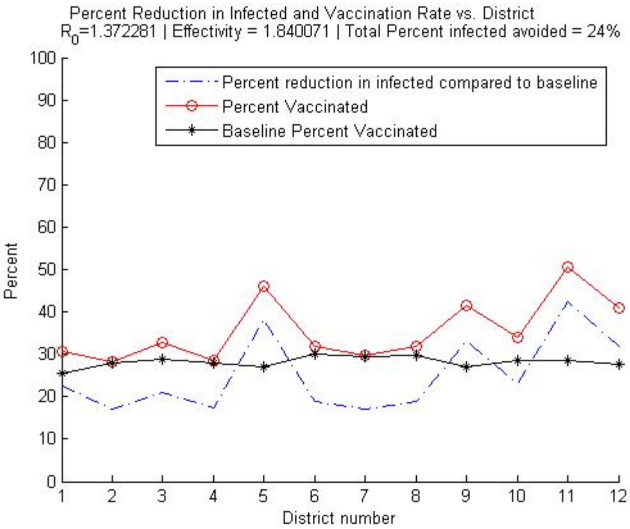

Figure 13: Result of iterated random vaccination proportions

The results show marked increases in all ‘poor’ districts and smaller changes in the ‘rich’ districts with the exception of District 5. The reduction in infected is 24%, an improvement over uniform random vaccination in poor districts. The difference lies in that inferior sets of proportions are discarded and seem to converge to a behavior as shown in Figure 13. This process was repeated various times and produced the same general behavior.

The results produced suggest that the ’poor’ districts have an inordinate effect on their richer neighbors. In order to narrow down the reason for this effect, we produced vaccination percentages weighted by population so as to favor larger populations no matter if they were ’rich’ or ’poor’. As before, 100,000 weighted random vaccination proportions were produced. The results appear in Figure 14.

The results clearly shows that favoring higher populations produces the same result as in Figure 12. Even though Districts 7 and 8 are two of the largest populations, the additional vaccination in those districts is minimal. On the contrary, the vaccination increases largely occur in the poor districts. The favoring of ’poor’ districts by the optimizing algorithm isn’t solely due to population size alone as shown in these results. In addition, the percent reduction in infected improves to 27%.

We have shown that vaccinating rich districts while ignoring the ’poor’ districts

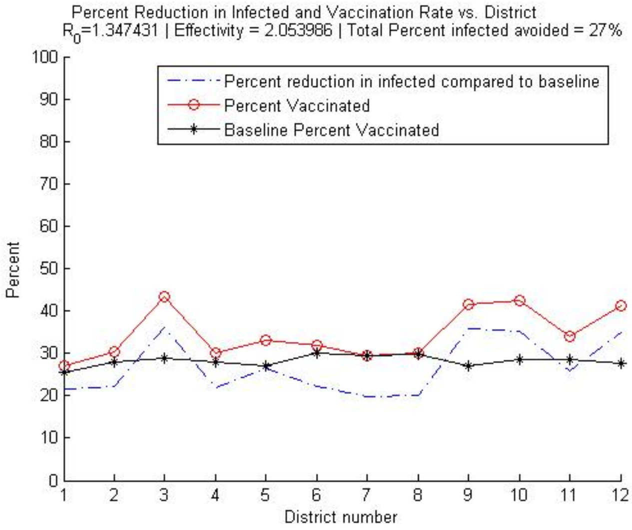

Figure 14: Result of random uniform distribution weighted by population size

reduces the effectivity. Our iterative procedure favors vaccinating the poor districts disproportionately. Motivated by the high sensitivity parameter for γ_{12} as shown in the sensitivity analysis section, we allow vaccination to occur only in poor districts and we weigh the proportions for population size. As usual, we produce $100,000$ such sets of proportions. The results are shown in Figure 15.

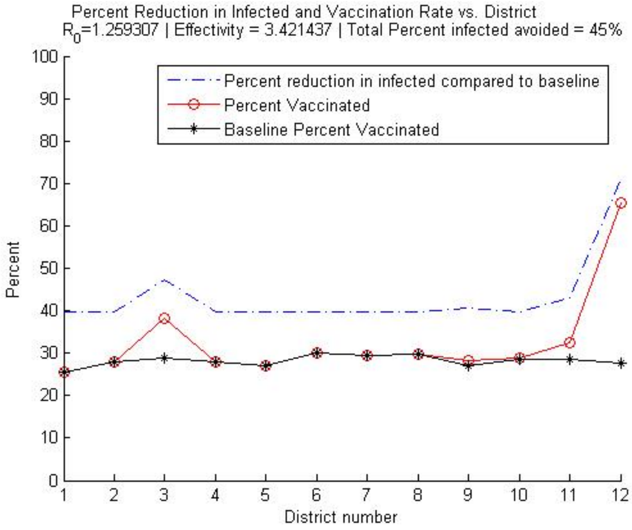

Figure 15: Result of random uniform distribution weighted by population size restricted to poor districts

The results show the overwhelming effect of favoring District 12 over the others. The reduction in infected (45%) is a large improvement compared to previous simulations, including those that only uniformly varied among the poor districts. Concentrating additional vaccination in District 12 has the largest effect on $R_{0_{total}}$, and these results show that it has the largest effect on reducing infected globally, the ultimate goal of any vaccination strategy. However, poverty is not a defining characteristic in and of itself for the large effect observed from large variation in γ_{12}. Figure 16 shows a uniform random distribution for all districts weighed by poverty. Since the highest poverty districts (11 and 10) aren't favored by the optimization procedure, we conclude that poverty in and of itself does not guarantee that particular district will be vaccinated to a large degree in the optimized case.

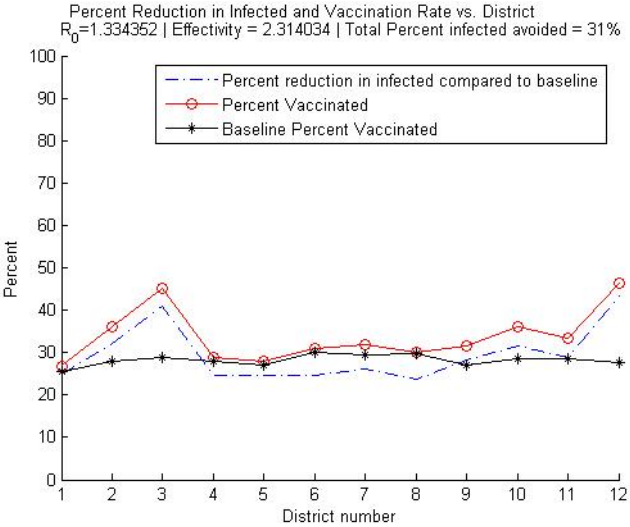

Figure 16: Result of random uniform distribution weighted by poverty level

6 Results

6.1 Optimized Vaccination Proportions

The results exhibited in Figure 16 point to the fact that although District 12 and 3 do not have the highest poverty levels, the optimized vaccination proportions favor these districts over the higher poverty level of Districts 10 and 11 notwithstanding the proportions are weighed by poverty level. These results as well as those presented in Figure 14 show that both population and poverty level in *conjunction* determine the optimized proportion of vaccinations. Districts 7 and 8 have the highest population and yet their effect on reducing the number of infected is minimal in comparison to

District 12 and even the other 'poor' districts. A possible explanation for this result lays in that District 12 has increased activity level, i.e. greater force of infection and consequently more contacts. This is in addition to the fact that an individual is more likely to meet a person from District 12 simply by the size of its population than a person from District 2 or 1, i.e. proportionate mixing. Therefore, if it is less likely that a person from District 12 is infected due to vaccination, then your chances of an individual getting infected from any other district is lessened as a result. It happens that District 12 has a large population (a high proportion of contacts are with individuals from District 12) and higher force of infection. Since the epidemic spreads faster among poor districts, this engenders many secondary cases in other districts both poor and rich. Since District 12 affects the spread so immensely, a dramatic reduction in susceptible individuals within that district due to vaccination, reduces the magnitude of the spread of the disease.

Consequently, District 12 is unique in that it has a very high population as well as high poverty levels. In our model, high poverty is distinguished in comparison with other districts in the higher force of infection (α_i) in those districts. Simulation shows that the combination of higher contact probabilities and the size of the population of District 12 (District 3 to a much lesser extent) produce the large global effect that reduces the extent of infection among all districts.

This last result explains the lack of effect of vaccinating the richer classes and other poor districts as well, since their contribution to the G term is much smaller than District 12's, i.e. the *amplitudes* for these districts are small. Any increase in vaccination in those districts is largely offset by the interaction between the infected and susceptible in District 12.

Sensitivity analysis suggests that variations of γ_{12} have a larger effect on $R_{0_{total}}$ and in consequence the size of the epidemic in question. As shown before, the sensitivity parameter of γ_{12} has the largest effect on $R_{0_{total}}$ and therefore can reduce $R_{0_{total}}$ most efficiently. Since R_0 is a measure of the strength of the epidemic it is not difficult to see that it would be most efficient to reduce γ_{12} by vaccinating more people in District 12. This reduction in $R_{0_{total}}$ then results in a reduction in the total infected, a key result of the simulations.

6.2 The Healthy People Initiative

Our results show a strong correlation between poverty and the spread of disease. Therefore, we suggest that future vaccination efforts consider this to prevent flu outbreak in a more efficient manner. An example of how current efforts could be improved is the Healthy People Initiative. The plan recommends vaccination for 60% of people aged 18-65 and 90% of people 65 and over by the year 2010 [4]. Our model demonstrates that in theory, it should be possible to double the effectiveness of the above goal simply by focusing on the low-income districts. Although it will require greater distribution of free flu vaccines in low-income areas, studies show that benefits outweigh the cost [13].

According to our simulations results, if vaccinations are simply distributed evenly throughout Manhattan according to the Healthy People Initiative, only 18% of the total population will be infected by the flu during a season similar to the one in our model (see Figure 17). However, if vaccinations are distributed systematically with emphasis on the low-income districts (3, 9, 10, 11, and 12), only $9-10\%$ of the total population will become infected (see Figure 18). This result is surprising considering the fact that there is no difference in total number of individuals vaccinated. The systematic plan is to provide free vaccinations to all low income districts and to provide public education such as fliers, TV commercials, and door to door information handouts in an effort to increase people's willingness to receive vaccinations. With a simple even distribution, the low income districts end up increasing their vaccination rates by 21, 21, 18, 18, and 19%, respectively. Systematic vaccinations could result in a 31, 31, 27, 27, and 34% increase in vaccination proportions with slightly lower increases in the remaining 6 districts. As our simulations and sensitivity analysis suggest, we placed the most weight on District 12. With this distribution we find that the overall effectiveness of vaccination is nearly doubled. Our findings suggest that an effective strategy for reducing flu outbreaks should consider increased efforts in less receptive environments, i.e. low-income areas.

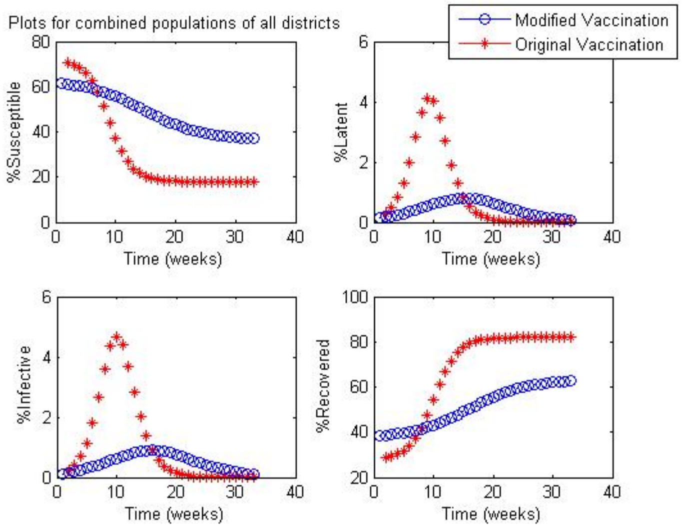

Figure 17: Prediction of the results of the Health Care Initiative using even distributions of vaccination proportions

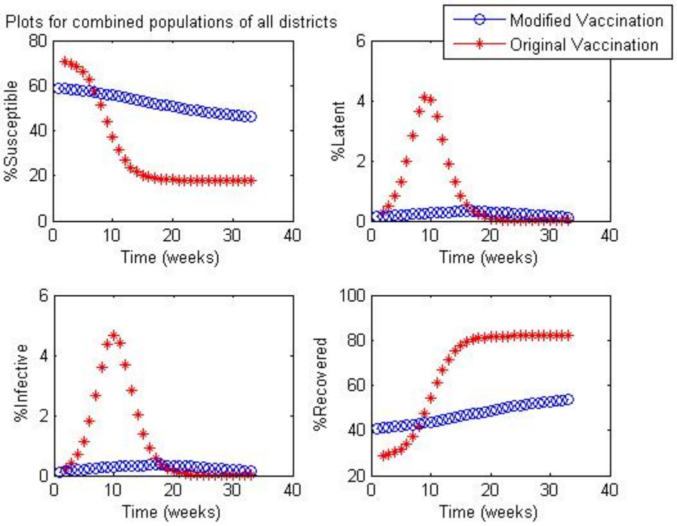

Figure 18: Prediction of the results of the Health Care Initiative using systematic distributions of vaccination proportions

7 Discussion

Through the development of a mathematical model, the effects of the influenza epidemic within the districts of Manhattan is shown using different scenarios. Our study primarily focuses on the socioeconomic status of districts within Manhattan. It is clear that areas with higher levels of poverty need to be targeted when implementing methods of reducing the number of influenza cases during a given year. For one, the size of the peak of the flu season is substantially greater in poor districts in comparison with those having more wealthy inhabitants. Additionally, finding the optimal strategy for distributing more vaccinations throughout Manhattan demonstrates that the greatest reduction in the overall epidemic occurs when the number of potentially infected individuals in a more impoverished district is reduced. The vaccination targets from the CDC's Healthy People 2010 Initiative, if feasible, will reduce infections substantially according to our model. We expect that the CDC's efforts will produce improvements among all income levels especially the most in need.

Often, people do not receive vaccinations because they do not find them necessary or simply do not know about potential resources for finding a free flu vaccination. Educating and encouraging influenza vaccination for people within Manhattan, in particular individuals living in the more impoverished areas, will significantly reduce the size of the annual peak of the influenza season.

There are, however, several ways in which to refine our model to make it more realistic. The model can be extended to include natural birth and death rates, which may or may not show significant differences within our model. However, if the length of time studied is increased to include multiple strains of influenza, the inclusion of these rates may have an impact on the realism of the results. Other future modifications include the addition of the effects of tourism; Manhattan is a location attracting many visitors and the disease brought in by these individuals is bound to have an effect on the dynamics of the epidemic and would give us means to include many external factors.

Another modification that can be made to the model is that New York City is comprised of five boroughs–not only Manhattan. There are substantial interactions between all of these boroughs. Our model considers these interactions to be negligible, although we expect this to be a oversimplification. A final future modification is to use an alternative method of mixing within the districts. Rather than using proportionate mixing, a more realistic mixing matrix may be implemented in which the interactions between individuals is dependent on where people spend most of their time. For instance, a larger portion of an individual's time is spent in his or her own district as well as districts where they work. This modification may produce more accurate predictions of the spread of influenza between districts.

8 Acknowledgements

We thank Karyn Sutton, Xiaohong Wang, Carlos Castillo-Chavez, Christopher Kribs-Zaleta, Baojun Song, Anuj Mubayi, Gerardo Chowell-Puente and the rest of the MTBI community for their support. This research has been partially supported by grants from the National Security Agency, the National Science Foundation, the T Division of Los Alamos National Lab (LANL), the Sloan Foundation, and the Office of the Provost of Arizona State University. The authors are solely responsible for the views and opinions expressed in this research; it does not necessarily reflect the ideas and/or opinions of the funding agencies, Arizona State University, or LANL.

Carlos Chiquete Valle would like to acknowledge his grandfather, Miguel Valle Montijo. This work is a tribute to him and his family. Q.D.E.P. *Tata*.

References

[1] Arriola, Leon, and Hyman, M. Forward and Adjoint Sensitivity Applications in Dynamical Systems, Linear Algebra and Optimization. 2005.

[2] Brauer, Fred, Castillo-Chavez, Carlos. *Mathematical Models in Population Biology and Epidemiology.* Springer-Verlag. New York: 2001.

[3] Centers for Disease Control. “2003-2004 Influenza Season Summary.” Atlanta, 29 June 2004.
http://www.cdc.gov/flu/weekly/weeklyarchives2003-2004/

[4] Centers for Disease Control and Prevention. “Healthy People 2010.” 2005.
http://www.healthypeople.gov/

[5] Centers for Disease Control and Prevention. “Self-Reported Influenza Vaccination Coverage Trends 1989-2003.” Atlanta, 2004.
http://www.gov/flu/professionals/vaccination/pdf/vaccinetrend.pdf

[6] Centers for Disease Control. “Key facts about Influenza and the Influenza Vaccine.” Atlanta, 14 July 2005.
http://www.cdc.gov/flu/keyfacts.htm

[7] Chowell, G., Castillo-Chavez, C. 2003. Worst-Case Scenarios and Epidemics. In: T Banks and C. Castillo-Chavez (eds.), Bioterrorism: Mathematical and Modeling Approaches in Homeland Security. *SIAM’s series Frontiers in Applied Mathematics.*

[8] Fiscal Policy Institute. “Pulling Apart in New York: An Analysis of Income Trends in New York State and New York City.” New York: 1 January 2000.
http://www.fiscalpolicy.org/pullingapart.htm

[9] Hanski, I, and Gilpin, M. “Metapopulation dynamics: Brief history and conceptual domain.” Biological Journal of the Linnean Society [BIOL. J. LINN. SOC.]. Vol. 42, no. 1-2, pp. 3-16. 1991.

[10] Harvard University Gazette. “Flu Vaccination Rates Lower Among African-Americans”. Cambridge, 2001.
http://www.news.harvard.edu/gazette/2001/09.27/09-flushots.html

[11] Health Care Initiative. “The Facts Concerning Health Care.” Pittsburgh, 14 July 2005.
http://www.uswa.org/uswa/program/content/407.php

[12] Health Care Initiative. “The Health Care Bill of Rights.” Pittsburgh, 14 July 2005.
http://www.uswa.org/uswa/program/content/408.php

[13] Innovations Report. “Stanford researcher shows flu shot benefits outweigh costs in healthy young adults.” 20 August 2002.
http://mednews.stanford.edu/releases/2002/august/fluvaccine.html

[14] New York City Department of City Planning. “Community District Profiles.” New York City, 2005.
http://www.nyc.gov/html/dcp/html/lucds/cdstart.html

[15] The New York City Department of Health and Mental Hygeine. “NYC Community Health Profiles.” New York City, 2005.
http://www.nyc.gov/html/doh/html/data/data.shtml

[16] The New York City Department of Heath and Mental Hygiene. “Who Should Recieve the Flu Vaccine?” New York City, 2005.
http://www.nyc.gov/html/doh/html/imm/imminflu.shtml