It's Worth a Shot: Urban Density, Endogenous Vaccination Decisions, and Dynamics of Infectious Disease

Andrew Souther[1], Myong-Hun Chang[2] and Troy Tassier[3]

Abstract:

We develop an agent-based model of vaccination decisions across a spatial network model with urban and rural regions. In the model, agents make rational decisions to vaccinate or not, based on the relative private costs of vaccinations and infections as well as an estimated probability of infection if not vaccinated. The model is a methodological advance in that it provides an economic rational for traditional threshold models of vaccine decision making that are commonly used in agent based network models of vaccine choice. We find that increased connections between agents that are created in more dense economic regions create increased vaccine usage within the model. This finding adds to the more commonly discussed socio-economic reasons for higher levels of vaccinations in urban areas compared to rural areas.

Keywords: Economic Epidemiology, Urban Density, Vaccinations, Agent-based Modelling

[1] Honors Program FCRH, Fordham University, Bronx, NY 10458; e-mail: asouther@fordham.edu

[2] Department of Economics, Cleveland State University, Cleveland , OH 44115; e-mail: m.chang@csuohio.edu

[3] Department of Economics, Fordham University, Bronx, NY 1048; e-mail: tassier@fordham.edu

1. Introduction

Rates of vaccination differ greatly between urban and rural areas in many parts of the world. Lack of access to health care, poverty, and socio-economic status are seen as the primary reason for this difference in many areas (Hinman and McKinlay 2001). In this paper we depart from these factors and consider that urban density may lead to a higher risk of contracting an infectious disease. If people in dense urban areas have more contacts with other individuals, then potentially, they are at greater risk of contracting an infectious disease. We model urban density as a difference in the number of contacts with other agents. We then model the vaccine decision of agents in terms of a rational cost benefit calculation which includes an agent estimated probability of infection, a private cost of vaccination, and a private cost of infection. The benefits of a vaccine are the avoidance of infection when a vaccine is chosen. Our modeling choice corresponds to a growing economic epidemiology literature incorporating endogenous rational choice into models of vaccine decision making. (Brito et. al. 1991, Kremer 1996, Geofard and Philipson 1996 and 1997, Bauch and Earn 2004, Boulier et. al. 2007, Goyal and Vigier 2010, Galeotti and Rogers 2013, Tassier et. al. 2015)

1.1 Background

Influenza causes high rates of morbidity and mortality throughout the world. In the U.S. alone, influenza accounts for 30,000-40,000 deaths, and 200,000-500,000 hospitalizations in a typical year (CDC 2020a). The influenza vaccine is the most common public health tool used to fight influenza. Unfortunately the observed rates of influenza vaccination fall short of levels needed to achieve herd immunity. In the U.S. only about one-half of individuals received an

influenza vaccine during the 2019-2020 influenza season (CDC 2020b), while approximately three-quarters are needed to achieve herd immunity (Plans Rubio 2012).

Further, rates of influenza vaccination are not uniform across the country. At the state level, rates of influenza coverage vary from 41% (Idaho) to 56% (Rhode Island and Massachusetts). At finer levels of resolution, vaccine coverage levels vary more greatly; some counties have rates below 10% while others are in the mid-60% range (County Health Rankings 2019).

Most investigations of vaccine behavior are taken up outside of a traditional economic model of rational choice. Most concentrate on demographic factors that affect access to healthcare or on the growing global anti-vaccine movement that has been perpetuated by many well-known public figures (Olpinski, 2012). In general, researchers find that higher socio-economic status results in higher levels of influenza vaccination (Lucyk et. al. 2019). This extends to differences across race as well (CDC 2020b and Hebert et. al. 2005). Much of the vaccination rate literature concentrates on the vaccination status of children and the Measles Mumps and Rubella (MMR) vaccine. Some empirical models identify “hot spots” where vaccination rates are particularly low (Olive et. al. 2018; Wallinga et. al. 2005; Carrel and Bitterman, 2015; Lieu et. al. 2015; Omar et. al. 2008; Salman et. al. 1999). Others examine the effect of income, religion, race, and other demographic factors on vaccine coverage. Williams et. al. (1995) and Luman (1995) find that poverty is associated with under-immunization of children (children that have some vaccines but fall short of recommended medical guidelines). Smith et. al. (2004) finds a distinction between under-vaccinated children and unvaccinated children

(children that have no vaccines at all). Under-vaccinated children tend to come from families that are black, lack college degrees, and are in poverty. Unvaccinated children are more commonly from families that are white, have college degrees, high incomes, and high degrees of concern for vaccine safety. These families also tend to be highly clustered geographically. Richards et. al. (2013) find lower vaccine use as a function of census track income. Other models consider the effect of varying rates of vaccine coverage on the spread of infectious disease (Atwell et. al. 2013; Feikin et. al. 2000; Salmon et. al. 1999). Vaccine safety is a commonly mentioned reason by parents for lack of vaccine use in their children (Allred et. al. 2005; Freed et. al. 2010; Gust et. al. 2004; Salmon et. al. 2005).

More closely related to this paper is a literature that examines differences in rural versus urban vaccination rates. Urban areas have consistently been found to have higher rates of influenza vaccinations than rural areas, differing by 10 to 20 percent (Zhai et. al. 2020 and O'Leary 2015). One reason for this difference is lack of easy access to healthcare and poor vaccine distribution in rural areas compared to more urban areas (Bennett et. al. 2011.)

This has led to a discussion in academics and the general press concerning these localized communities with low vaccination rates and the accompanying increase in risk of infection (Feikin et. al. 2000; Atwell et. al. 2013, Wallinga et al 2005, Chang and Tassier, 2019). Some of this research is conducted using agent-based computational models (Bausal 2006, Barrat et. al. 2010, Del Valle et. al. 2013, Vardavas and Marcum 2013, Tassier et. al. 2017, Chang and Tassier 2019). Other models within economics consider analytical network based vaccine choice at a group level (Galeotti and Rogers 2010 and Goyal and Vignier 2014).

In this paper we combine several features of these economic and agent-based models and

examine how vaccine behavior may differ across urban and rural areas. First, we use an explicit network model implemented in an agent-based framework. Second, agents in the model make rational choices given their estimation of the risk of infection in their region and associated private costs of infection and vaccination. The agents choose to vaccinate when vaccination maximizes expected utility and forgo a vaccination when it does not. By varying costs and network structure we view the incidence of an infectious disease across time and the corresponding endogenous choice to vaccinate made by agents in the model.

In the following sections of the paper, we demonstrate how county level vaccinations differ across the United States as a function of population density using data from the CDC as well as other publicly available sources. We then discuss a basic strategic model of vaccine choice which we embed into an agent-based epidemiological model. The epidemiological model is a commonly used SEIR model (Anderson and May 1992), which will be discussed in full later in the paper. This model can be used to represent a common infectious disease such as influenza, measles, or the current novel coronavirus pandemic (once an effective vaccine is widely available). After describing the experimental design of the agent-based model we run instances of the model with varying costs and network structures and discuss the results.

2. Empirical Justification

As an empirical justification for our model, consider the yearly influenza vaccinations at the county level across the United States. The data are taken from the 2019 County Health Rankings website which is maintained by the Robert Wood Johnson Foundation.[4] The data are

[4] The data can be found at the following URL: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation

compiled from various US government sources and published yearly. It includes data on a host of health and demographic data aggregated to county level percentages.

Influenza vaccinations as well as other vaccinations vary greatly over geographic space. As an example, in the data, influenza vaccinations vary between about 10% and 65% across counties in the United States. Figure 1 displays the percentage of county residents that are vaccinated for influenza versus the natural log of population density along with the linear trend line. As one can see there is a strong positive correlation between population density and influenza vaccination rates.

Going a step further, we control for other potential explanatory variables and perform a linear regression model of the vaccination rate on density and several additional demographic variables that are commonly used to explain differences in vaccine coverage. The results indicate that density remains a key determinant of county level vaccination rates when these demographic variables are added as controls.[5] This suggests that, in addition to socio-demographic factors and lack of access to health care in rural areas, dense urban areas have a direct role in increasing vaccination coverage. Likely, this is due to an increase in the risk of infection. Below we model density as an increase in agent contacts and view the effect that various model parameters have on vaccination choice and on the rates of infection in urban versus rural areas.

[5] Note that we are not attempting to develop a formal model or to precisely identify (in an econometric sense) effects of these variables on vaccination choice with this regression. Instead, we are simply showing that urban density is likely one factor relevant in the decision to be vaccinated. We leave formal econometric modeling of this issue to other research.

3. Model

3.1 Overview

Many infectious diseases follow an SEIR compartmental design. This acronym refers to the various states or compartments in which agents may find themselves within the model: Susceptible (to being infected), Exposed (infected but not yet able to infect others), Infectious (both infected and able to infect others), and Recovered (no longer infected nor able to infect others).

As a general description of our model, consider a population composed of N agents. Let S_t, E_t I_t, and R_t denote the number of agents in each of the states described above in period t. We implement an SEIR model over a course of seasons, denoted by *s*, which may be thought of as a calendar year. Each season is divided into *t* discrete time periods (days). Each of these *N* agents has contacts with other agents in the population in each time period, *t*. These contacts compose a graph denoted as Γ. We assume that Γ is fixed throughout the model such that each agent has the same contacts in each day and in each season. We vary Γ across replications of an experiment and as a parameter of the model across agent-based experiments.

In each period of the model, each agent contacts the agents to whom she is directly connected in Γ. If an agent is susceptible and contacts an infectious agent in state I then the susceptible agent moves to state E with probability α. Once in state E, the agent moves to the infectious state with probability ϵ (in each time period). The agent then remains in state I until recovering with probability ρ (in each time period). Thus the expected duration in state E is $1/\epsilon$ and in state R is $1/\rho$. Once recovered, the agent becomes and remains immune for the duration of

the epidemic season. Our model is based on an infectious disease such as influenza which evolves each season of infection. Thus in the next influenza season each agent enters the model season in the susceptible state, S, regardless of her state at the end of the previous season.

3.2 Networks, Γ

Each agent lives in a geographic region. There are two types of regions in the model, rural and urban. Regions of each type vary by population and density. Regions are composed of a set of agents of size N_u or N_r, which is the population of the region. Each region varies by land size so that different regions have different densities. Greater density per area results in more contacts for an agent living in a region with a smaller land area (holding population constant.) This results in a larger number of daily contacts for agents in more dense regions. For simplicity we assume that agents live in only two types of regions: urban areas (those with high density) or rural areas (those with low density). Agents in urban areas have γ_u contacts and agents in rural areas have γ_r contacts, $\gamma_u \geq \gamma_r$. Thus differing levels of density appear in the model directly as differences in the number of contacts. We vary the number of urban and rural contacts as well as the population size of urban and rural areas as parameters of the model.

Each agent has some contacts with members of his own region and some contacts with members of other regions. We refer to the percentage of same-region contacts as *h,* a homophily parameter (McPherson et. al. 2001 and Moody 2001). h will be varied across agent-based experiments. This implementation of regions and networks is a version of a stochastic block model (Abbe 2018) which has been used elsewhere to study externalities in infectious disease models (Chang and Tassier 2019).

3.3 Vaccinations

We allow agents the opportunity to vaccinate at the start of a season. We assume that vaccines are fully effective.[6] The vaccination decision is based on a best response to the epidemic outcome in the previous season. Let agents in the model have a base utility level U each season. If an agent is infected during the course of a season, the agent pays a utility cost, C_i. If the agent chooses to be vaccinated it pays a cost C_v. The agent weighs these costs versus a probability of being infected. If the agent believes there is a sufficiently low probability of infection, and the cost of vaccination, compared to the cost of infection, is relatively high, then the agent should forgo a vaccine and the associated cost. If, however, the agent believes that the likelihood of infection is sufficiently high and the cost of a vaccine is relatively low, compared to the cost of infection, then the agent should choose to be vaccinated. Specifically, the agent considers the following inequality:

$$U - C_v > U - \pi_{r,s-1} C_i \qquad (1)$$

where $\pi_{r,s-1}$ is the probability that an unvaccinated agent, in the agent's own region r, would have been infected in the previous season of the model. The left side is the base utility level less the cost of the vaccine. The right side is the base utility level less the probability of infection multiplied by the cost of infection. If the left side is larger than the right side, the agent vaccinates, else the agent forgoes vaccination. Note that this equation only considers the last season in the model and the region in which the agent lives. Other methods of updating can be

[6] If we assume that vaccinations are less than 100% effective, that simply changes the probability of infection calculation below and rescales the vaccination choice. But, it does not change the underlying implications of the model in any significant manner.

considered. For instance one could use a weighted average of infection probabilities in past seasons with more recent seasons weighted more heavily than seasons further in the past. One could also consider other sets of agents, such as all regions, or a subset of nearby regions, or simply the agents to whom an agent is connected. We leave these additional possibilities for exploration in future work. Equation 1 simplifies to yield the following decision rule for agents, vaccinate if:

$$C_v < \pi_{r,s-1} C_i \qquad (2)$$

or vaccinate if:

$$\pi_{r,s-1} > \frac{C_v}{C_i} \qquad (3)$$

$\pi_{r,s-1}$ is decreasing in the number of other agents in the model who choose to be vaccinated. Thus when there is a sufficiently large epidemic in the last season, all agents would choose to be vaccinated in the model; this would then result in no new infections in the next season. With no infections, all agents would then forgo vaccination in the following season. This cycle would potentially continue indefinitely. However, given that the probability of infection decreases with the number of vaccinated individuals in the population, the overall population may evolve to a stable configuration in which the vaccination rate leads to a probability of infection equal to the ratio of the cost of vaccination to the cost of infection.

$$\pi_{r,s-1} = \frac{C_v}{C_i} \qquad (4)$$

Thus agents should vaccinate up until the probability of infection is equal to the ratio of the cost of vaccination to the cost of infection. In our model, we approximate achieving this result by allowing only a small percentage of agents to update their choice each period. Specifically, let δ be the probability with which an agent is given an opportunity to update his vaccine choice. Then with probability δ the agent chooses to vaccinate, or not, based on equation 3. And, with probability $1 - \delta$, the agent keeps her vaccination status the same as it was in season s-1. Thus, agents have persistence in vaccine choice and only δ percent consider a different strategy than was chosen in the previous period.[7]

3.1 Agent Based Experiment Design

We implement this model in an agent-based framework.[8] All experiments have 1,000 agents allocated across 10 regions; 5 of these regions are labelled urban (dense) and 5 are labelled rural (non-dense). Our primary interest in the experiments will be the differences in vaccination rates and number of infections across these two types of regions. We run a set of experiments to examine the effect of different parameter choices. Each computational experiment (parameter combination) is run for 25 seasons and replicated ten times. We present the results in tables that report the average outcome across the final 5 seasons of each of the 10 replications for each experiment.

---

[7] While we do not implement a formal game theoretic model, our implementation is similar in spirit to an evolutionary attainment of a mixed strategy Nash equilibrium described by Oechssler (1997).

[8] Our agent-based simulation was implemented using the Python package *mTree*. Currently under development in the Computational Economics Lab (CeLab) at George Mason University, *mTree* is a package for designing agent-based simulations and online behavioral experiments in one unified framework.

3.1.1 Implementation

We implement our model in an agent-based framework. Agent based models have been recommended as ideally suited to investigate issue of public health in heterogeneous agent environments (Epstein 2009).

To begin each experiment we create 10 regions with 5 designated as urban and 5 designated as rural. We distribute 1,000 agents across this space. In all instances of the model considered here each rural (urban) region has the same number of agents and contacts as all other rural (urban) regions. (As an example, each of the 5 rural areas may contain 80 agents and each of the 5 urban areas may contain 120.) We then initialize the network contacts of each agent according to the parameters of a particular computational experiment. Each agent within an urban region is assigned γ_u contacts and each agent within a rural region is assigned γ_r contacts. We assign these contacts using a matching algorithm as follows. Agents are chosen in turn. Each time an agent is chosen, the algorithm considers whether she already has a full set of contacts equal to the contact rate for her region. If this agent does not have a full set of contacts, she is repeatedly matched with other agents until she reaches that number. These other agents are chosen as a function of the homophily parameter, h. With probability h, each new contact is chosen from within the agent's own region; and with probability (1-h) the new contact is chosen from the other 9 regions. As a requirement, each of these other agents must also lack a full set of contacts. This process continues until all agents have a full set of contacts.

Theoretically, this algorithm may fail to find a full set of contacts for agents matched late in the process. This failure tends to affect 0-2 agents out of 1000 in each replication. However,

agents are matched in a random order to ensure that these occasional failures are also randomly distributed. Three example graphs are shown in Figure 2, 3, and 4 for values of h={0.79, 0.89., 0.99}. In practice, this algorithm generates networks with true rates of same-region contacts that closely match the desired parameter h. Very high values for h (such as 0.99) tend to be biased downwards by about 0.01; lower values for h (such as .79) are biased upwards by the same amount. Urban areas end up with slightly higher homophily than rural areas (on average, a difference of 0.02).

Once all agents are assigned a set of contacts, we begin the vaccination and epidemic process with season 1. In season 1, as an initialization, we vaccinate 15% of agents chosen randomly from the entire population; we then seed 10 agents (chosen randomly from across the entire unvaccinated population) as infected. The epidemic continues until no agents remain in the exposed or infectious state. Once this happens we have reached an absorbing state with each agent remaining either in the susceptible state or having moved through the infection process and arrived in the recovered state. We use the number of agents in the recovered state as our measure of the size of the epidemic. This completes season 1.

In each season s after season 1, we update the vaccination status of each agent. We begin this process by setting the season s vaccination status of each agent equal to the same status as the agent had in season s-1. We then give 8% of the agents a chance to change to a different vaccination status in season s. We do this as follows: if an agent is selected for updating, she compares her estimated probability of infection to the ratio of the cost of vaccination to the cost of infection. For the estimated probability of infection, the agent uses the number of infected

agents in her home region divided by the number of unvaccinated agents in her home region. We define this as $\pi^{e}_{r,s-1}$. This is an estimate of the probability that the agent would have been infected if she had not been vaccinated in the previous season, season *s-1*. If this estimate of infection probability is greater than the ratio $\frac{C_v}{C_i}$ then the agent chooses to be vaccinated in season *s*.

$$\pi^{e}_{r,s-1} > \frac{C_v}{C_i} \qquad (5)$$

If not, the agent forgoes vaccination in season *s*. Any agents who is not given an opportunity to update her vaccination status will remain with the same vaccination status in season *s* as she had in season *s-1*. Note that the reason that we do not allow all agents to update vaccination status at one time is that it would create cycles and would not settle to an equilibrium state. In seasons where there was a large outbreak all agents would choose to be vaccinated and the following season would have no infections. Following the season with no infections, all agents would forgo vaccination. Then, eventually, a sufficiently large outbreak would occur and all agents would again subsequently choose to be vaccinated. Thus the system would oscillate between all agents and no agents being vaccinated. Allowing a small but non-negligible fraction of agents to update their vaccination status in each season allows us to approximate a steady state of the model described above.

Once the vaccination decision process has been completed we again initialize a set of 10 seed infections from among the unvaccinated agents and run the SEIR model across the network.

We repeat this process until we have completed 25 seasons. We show below that this is a sufficient number of season to reach a steady state.[9]

We repeat this process 10 times for each parameter configuration that we investigate and call this set of 25 seasons replicated 10 times a computational experiment. We choose 25 seasons because we find this to be sufficient time for the model to reach steady state. All experiments reach a steady state in vaccinations and infections by about the 10th season. Figure 5a is a plot of average vaccination rates over time for a computational experiment using our base parameters. Figure 5b is an equivalent plot of average infection rates.When each experiment is complete, we calculate the average number of infections and the average vaccination rate over the final 5 seasons for the urban and rural areas separately. Comparisons of these averages over the final 5 seasons are the primary results of our experiments.

4. Results

We use the following parameters for the SEIR model throughout the paper:

Transmission rate (α): 0.03

Recovery rate (ρ): 0.08

Probability of updating the vaccine choice (δ): 0.08

Exposed to Infected transition rate (ϵ): 0.33

We have performed additional experiments with other values and there are not significant qualitative differences in the results when we do so.

We begin with a base implementation of 1,000 agents distributed across 10 regions (5 urban and 5 rural) with 80 agents per rural region and 120 per urban region. The urban regions

[9] As will be shown in graphs below, we reach a state with relatively constant behavior. Of course, since the model is stochastic, there will always be some amount of fluctuations within the model.

are considered to be more dense and each agent has 12 contacts; the rural regions are less dense and each agent has 8 contacts. We set cost of infection, C_i, to 3 and cost of vaccination C_v to 1. The homophily parameter is set to 0.89 We use these values in the following experiments unless otherwise indicated. Figure 6 displays a typical season of the model when there are no vaccinations. In the figure you see the typical characteristic shape of an SEIR epidemic model. We now begin describing the results of various combinations of parameter configurations. Our results take the form of comparative statics exercises where we vary one parameter at a time across a range of reasonable values and discuss the effects of these changes.

4.1 Experiment 1: Heterogeneity in Homophily

Our first experiment considers the effect of varying geographic homophily, h. We consider 3 values, $h=\{0.79, 0.89, 0.99\}$. We display example networks which are characteristic of these values in Figures 2-4. For $h= 0.99$ the regions remain distinct. Lowering h to 0.79 creates a great amount over overlap between regions and, we expect, makes it much easier for an infectious disease to spread. h = 0.89 is an intermediate case with a mixing of regions but each region still holding a distinct cluster of agents. Note also that even though regions become less distinct as h decreases there remains assortative mixing in that urban to urban contacts are more common than urban to rural contacts. Results are reported in Table 2. One would expect that less homophily in networks would lead to a more widespread epidemics for a given level of *exogenous* vaccinations (Chang and Tassier 2019). However, we see that this does not develop in our model because of the *endogenous* vaccination choice of agents. Agents vary their vaccine choice behavior and achieve a relatively constant level of infections in the model, 15-16% in the

rural regions and 18-19% in the urban regions. Agents achieve these consistent levels of infection by varying their vaccination behavior. As the amount of homopjhily decreases (leading to more contacts between regions) we see that more agents choose to be vaccinated; this is the response of agents to greater infectious disease risk. For the rural areas, vaccination rates increase from 28% to 39% to 40% and for the urban areas vaccination rates increase from 43% to 53% to 55%. Note that most of the increase occurs as we move from 99% in-region contacts to 89% in-region contacts. Similar to Watts & Strogatz (1998) network characteristics change rapidly with only a small amount of between-region contacts. The other thing to note from this experiment is that the urban vaccination rates are significantly larger than the rural vaccination rates. This is a direct responses to the larger number of contacts (a proxy for urban density) leading to an easier spread of an infectious disease and an increase in vaccination rates to offset the added risk.

4.2 Experiment 2: Heterogeneity in Contact Rate

As our next computational experiment we consider the effect of further increases in density, again through an increase in the contact rate. We hold our base parameters constant and change the contacts rates. In this experiment the contact parameters in urban and rural regions takes on one of three pairs of values: $\gamma_u, \gamma_r = \{14,6\}$, $\{12, 8\}$, or $\{10,10\}$. Thus we are changing the difference in urban and regional contacts from a large difference, to an intermediate difference, to no difference. In the case where both the urban and rural contacts are equal to 10 there is no difference between the urban and rural regions and thus we report them as one set of agents. Thus it provides a benchmark for comparison to the other two parameter settings. The

results are provided in Table 3. As one can view in the table, as contacts increase in the urban regions there is a minimal change in the rate of infection. This is achieved through a substantial increase in the vaccination rate. When each region has equal numbers of contacts about 47% of the population is vaccinated. This number increases to 54% and 57% in the urban regions as contacts increase from 10 to 12 and to 14. The magnitude of the decrease in infection rates in the rural areas is much larger and the decrease in vaccination rates in the rural areas is also much larger (compared to the urban regions). Again, with 10 contacts in all regions, the vaccination rate is about 47%. The rate decreases to 39% with 8 contacts and to 25% with 6 contacts. The net decrease in rural areas is 22% and the net increase in urban areas is 10%.

4.3 Experiment 3: Ratio of Cost of Vaccination to Cost of Infection

We again return to benchmark levels of contacts, 12 in urban regions and 8 in rural regions and now vary the cost of infection between 2, 3, and 4. We hold the cost of vaccination constant at the baseline value of 1. We present results in Table 4. As expected, as the cost of infection decreases, the vaccination percentages decrease and the level of infections increases. Interestingly, we see a much larger decrease in the rural areas than we see in the urban areas. The decrease in vaccination coverage the urban areas between C_i of 4 and 2 is only about 2.5% compared to a drop of 6.5% in the rural areas. Interestingly, the decrease in vaccination coverage in rural regions in nearly identical in magnitude to the increase in rate of infection. However, for urban regions the decrease in vaccination rate is much smaller than the increase in rate of infection. We believe that this suggests interesting externalities are at play between the region types. We expect to further explore these relationships in future versions of the model.

We also calculate the total cost *per agent* on average in each region which is a measure of welfare. We do so by multiplying the cost of infection by the percentage of agents infected in each region and adding this value to the percentage of agents who vaccinate (cost of vaccination being equal to 1.) We can then compare the total decrease in utility per agent resulting from infections and vaccinations, on average, in each type of region.

When viewing the results one sees that the total cost difference between urban and rural regions is relatively constant for values of C_i of 3 and 4. Urban agents have costs of about 0.165 more than rural agents for both C_i of 3 and 4. Rural agents pay a lower cost and thus have a higher level of utility on average compared to the urban agents. However, at C_i=2 we find that rural agents expand on this cost savings to a difference of about 0.185 less than urban agents. This occurs because, as mentioned above, rural agents respond much more strongly in their vaccine choice to the decrease in cost.

4.4 Effect of Population Size

Finally, we look at the vaccination rates for different sized urban and rural regions for each of the three contact rate pairs discussed above. We present the results in Table 5. For all contact rates we observe negligible changes in the vaccination rates of the agents as we vary the population size of the regions. This suggests that the effect of urban areas on vaccine choice in this model occurs because of increased density (increased contact with others) and not aggregate population size differences within the regions.

5. Conclusion

We create an agent based model of endogenous vaccine choice to examine the effect of increased density in urban areas on rates of infection and on rates of vaccine coverage. We find that urban agents in the model vaccinate at higher rates than rural areas in predictable ways. Even small differences in urban density can lead to relatively large differences in vaccination rates. Further even small rates of across region contacts have large effects on the behavior of agents in the model.

The model used in the paper provides a methodological advance in the use of a threshold vaccine decision that is based on economic principles of choice. This model can be expanded in many ways to incorporate different beliefs of an agent's perception (or estimation) of infection risk or beliefs of vaccine cost and risk. We plan to incorporate additional levels of agent heterogeneity into future research using this model.

References:

Abbe, E. (2018). Community detection and stochastic block models: Recent Developments. *Journal of Machine Learning Research*, *18*, 1–86.

Anderson, Roy M., and Robert M. May. 1992. *Infectious diseases of humans: dynamics and control.* Vol. 26, Wiley Online Library.

Bansal, S., Pourbohloul, B., & Meyers, L. A. (2006). A comparative analysis of influenza vaccination programs. *PLoS Medicine*. https://doi.org/10.1371/journal.pmed.0030387.

Barrat, A., Berthelemy, M., & Vespignani, A. (2010). *Dynamical processes on complex network*. New York: Cambridge University Press.

Bauch, Chris T., and David J. D. Earn. 2004. "Vaccination and the theory of games."

Proceedings of the National Academy of Sciences of the United States of America, 101(36): 13391–4.

Bennett, Kevin J. Bennett, Chaiporn Pumkam, Janice C. Probst. 2011." Rural–urban differences in the location of influenza vaccine administration." *Vaccine*, 29 (35): 5970-5977. https://doi.org/10.1016/j.vaccine.2011.06.038

Brito, Dagobert L., Eytan Sheshinski, and Michael D. Intriligator. 1991. "Externalities and compulsary vaccinations." *Journal of Public Economics*, 45(1): 69–90.

Boulier, Bryan L., Tejwant S. Datta, and Robert S. Goldfarb. 2007. "Vaccination externalities." *The BE Journal of Economic Analysis & Policy*, 7(1): 23.

CDC 2020a. https://www.cdc.gov/flu/about/burden/index.html

CDC 2020b. "Flu Vaccination Coverage, United States, 2019–20 Influenza Season" https://www.cdc.gov/flu/fluvaxview/coverage-1920estimates.htm

Chang, Myong-Hun and Tassier, Troy. 2019. "Spatially Heterogeneous Vaccine Coverage and Externalities in a Computational Model of Epidemics," *Computational Economics*, https://doi.org/10.1007/s10614-019-09918-7

County Health Rankings 2020: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation

Del Valle, S. Y., Mniszewski, S. M., & Hyman, J. M. (2013). Modeling the impact of behavior changes on the spread of pandemic influenza. In Manfredi & d'Onofrio (Eds.), *Modeling the interplay between human behavior and the spread of infectious diseases*. New York: Springer.

Epstein, J. M. (2009). Modelling to contain epidemics. *Nature*, *460*(6), 687.

Eubank, S., Hasan Guclu, V. S., Kumar, A., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*, 180–184.

Geoffard, Pierre-Yves, and Tomas Philipson. 1996. "Rational epidemics and their public control." *International Economic Review*, 37(3): 603-624.

Geoffard, Pierre-Yves, and Tomas Philipson. 1997. "Disease eradication: private versus public vaccination." *The American Economic Review*, 87(1): 222–230.

Germann, T., Kadau, K., Longini, I. M., & Macken, C. A. (2006). "Mitigation strategies for pandemic influenza in the United States." *PNAS*, *103*(15), 5935–5940.

Goyal, Sanjeev and Adrien Vigier. 2014. "Attack, Defence, and Contagion in Networks." *The Review of Economic Studies*, vol. 81, (289): 1518–1542.

Hebert et. al. 2005. "The Causes of Racial and Ethnic Differences in Influenza Vaccination Rates

among Elderly Medicare Beneficiaries." *Health Services Research,* 40(2): 517–538. doi: 10.1111/j.1475-6773.2005.00370.x

Hinman, Alan R. and Mark A. McKinlay. 2015. Immunization equity, *Vaccine*, 33 (4):D72-D77. https://doi.org/10.1016/j.vaccine.2015.09.033.

Kremer, Michael. 1996. "Integrating behavioral choice into epidemiological models of AIDS." *The Quarterly Journal of Economics*, 111(2): 549.

Lucyk et. al. 2019. "The association between influenza vaccination and socioeconomic status in high income countries varies by the measure used: a systematic review." *BMC Medical research Methodology*, 19: 153. doi: 10.1186/s12874-019-0801-1

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology*, 27(3):415–444.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America1." *American Journal of Sociology*, 107(3): 679–716.

O'Leary ST, Barnard J, et. al. 2015. "Urban and Rural Differences in Parental Attitudes About Influenza Vaccination and Vaccine Delivery Models." *Journal of Rural Health,* 31(4):421-30. doi: 10.1111/jrh.12119.

Jörg Oechssler, An Evolutionary Interpretation of Mixed-Strategy Equilibria, Games and Economic Behavior, Volume 21, Issues 1–2, 1997, Pages 203-237, https://doi.org/10.1006/game.1997.0550.

Plans-Rubió, Pedro. 2012. "The vaccination coverage required to establish herd immunity against influenza viruses." *Preventive Medicine*, 55 (1): 72-77. https://doi.org/10.1016/j.ypmed.2012.02.015.

Rogers, Brian and Andrea Galeotti Strategic immunization and Group Structure, *American Economic Journal: Microeconomics*, 5(2), May 2013: 1-32).

Smith, D. J., Forrest, S., Ackley, D. H., & Perelson, A. S. (1999). Variable efficacy of repeated annual influenza vaccination. *PNAS*, *96*(24), 140001–140006.

Stroud, P., Del Valle, S., Sydoriak, S., Riese, J., & Mniszewski, S. 2007. Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation*, *10*(4), 9.

Tassier, Troy, Polgreen, Philip and Segre, Alberto. 2015. Vaccination Games with Peer Effects in a Heterogeneous Hospital Worker Population," *Administrative Sciences*, 5: 2-26.

Tassier, T., Polgreen, P., & Segre, A. (2017). Network Position and Healthcare Worker Infections. *Journal of Economic Interaction and Coordination*, *12*, 277–307.

Vardavas, R., & Marcum, C. S. (2013). Modelling influenza vaccination behavior via inductive reasoning games. In Manfredi & d'Onofrio (Eds.), *Modeling the interplay between human behavior and the spread of infectious diseases*. New York: Springer.

Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (1998). https://doi.org/10.1038/30918

Zhai et. al. 2020. "Rural, urban and suburban differences in influenza vaccination coverage among children." *Vaccine,* 38 (48): 7596-7602. DOI: 10.1016/j.vaccine.2020.10.030

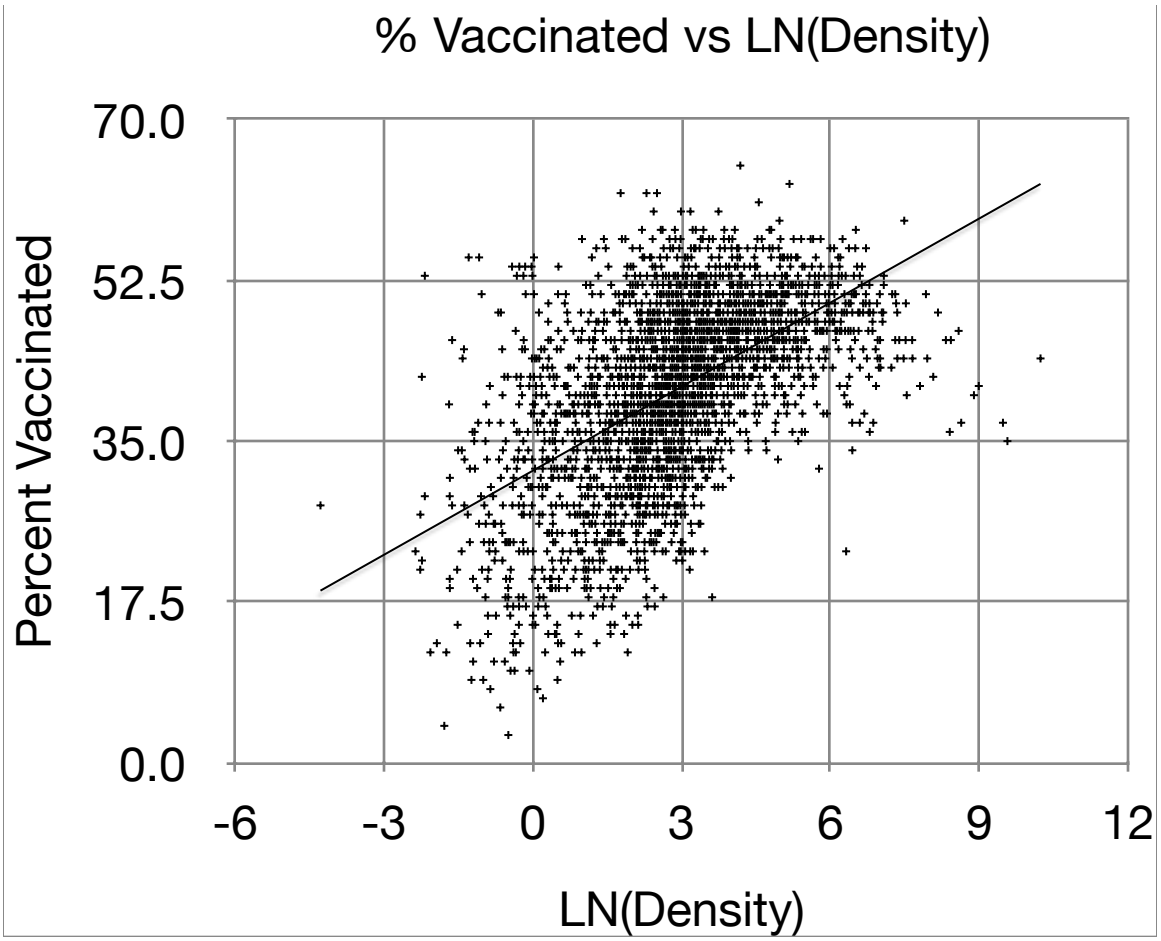

Figure 1: Plot of County Level Influenza Vaccinations as a function of LN(population density)

Figure 2: Example network with h=0.99.

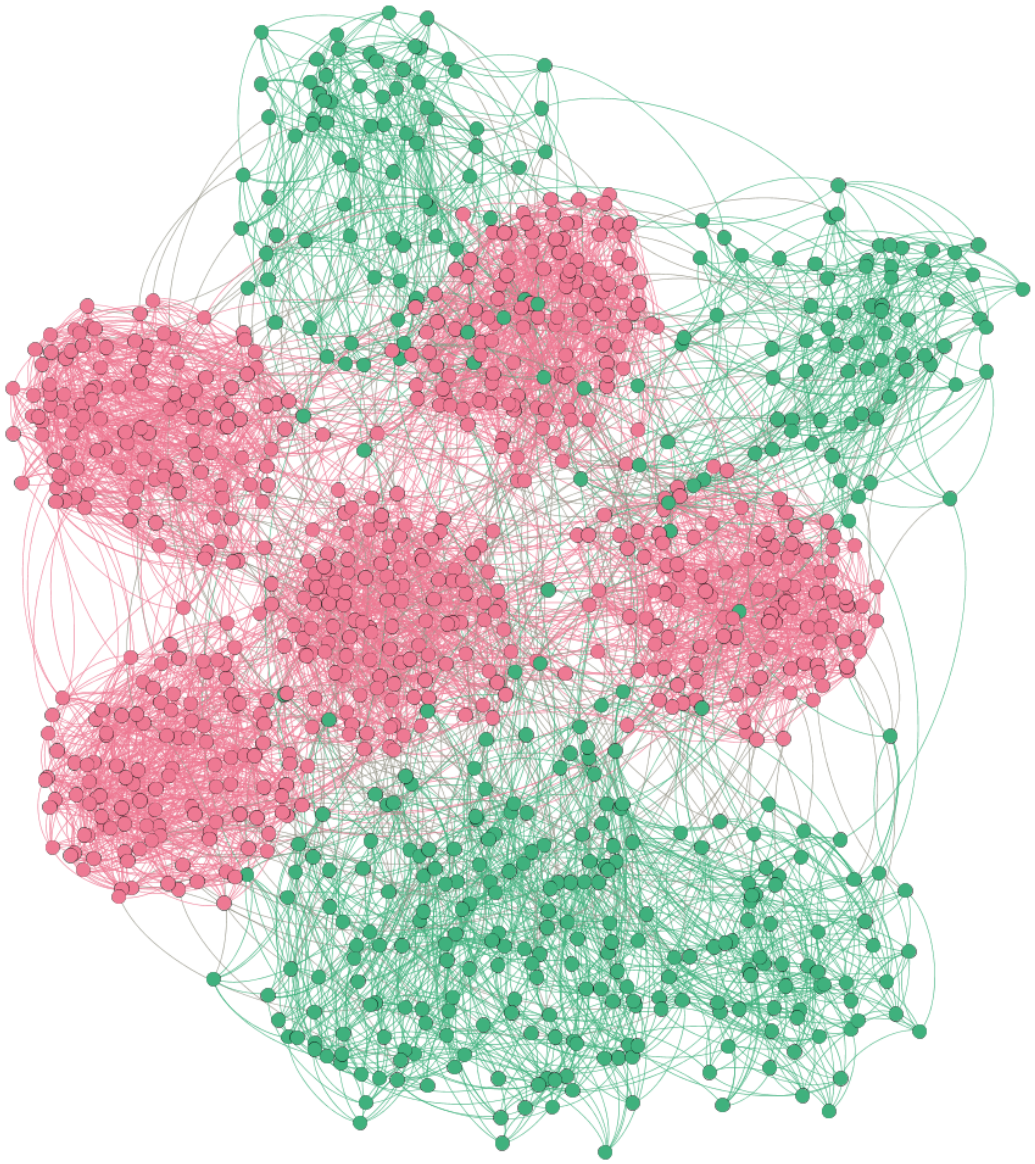

Figure 3: Example network with h=0.89.

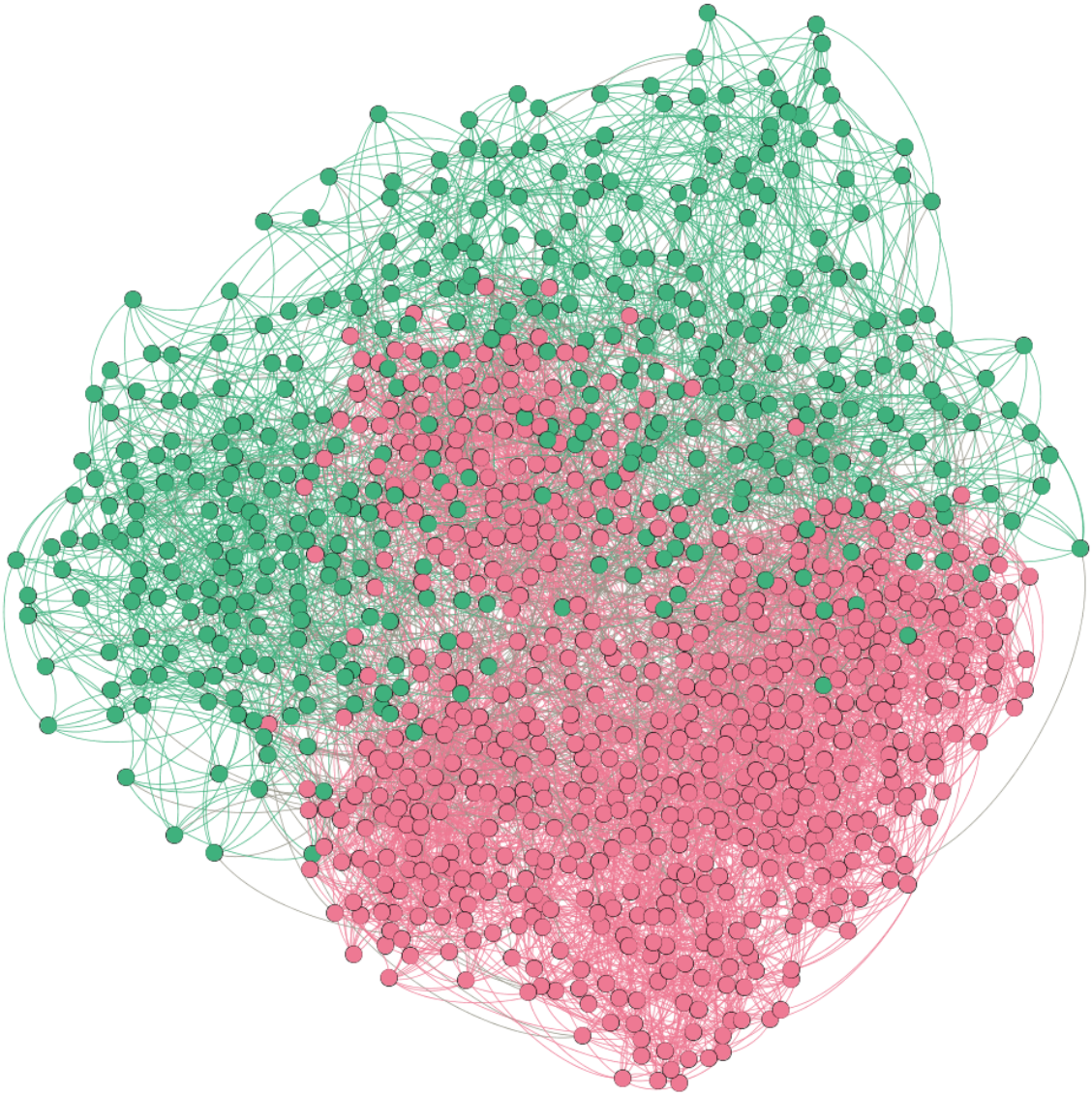

Figure 4: Example network with h=0.79.

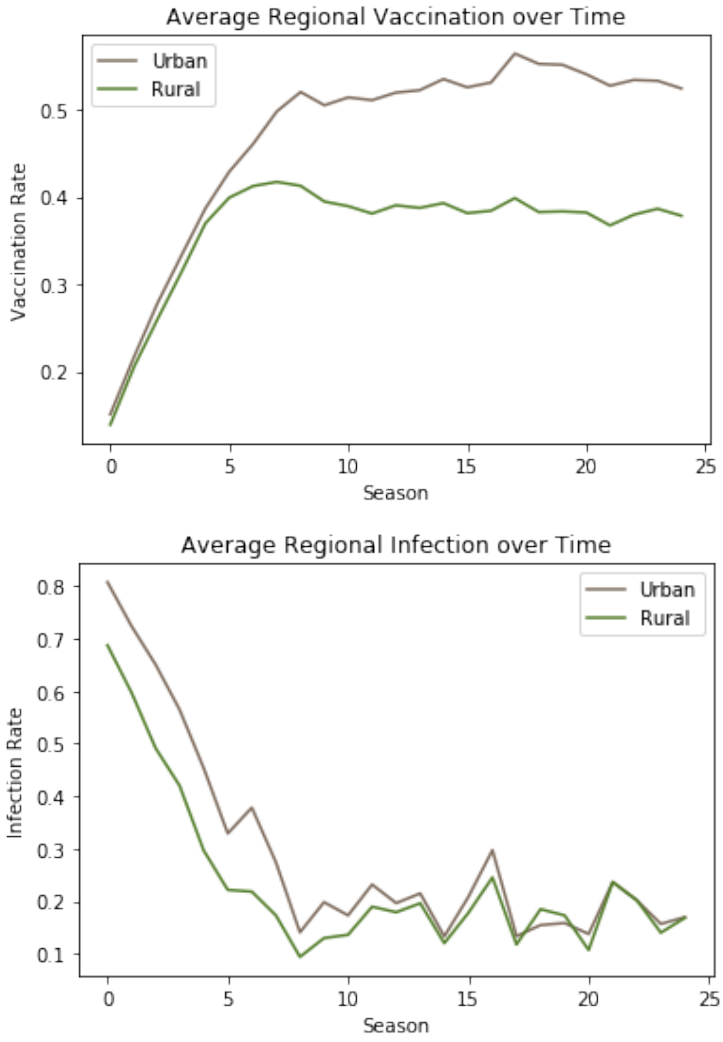

Figure 5: A set of 25 seasons showing that we reach steady state vaccinations by season 20. Then we use the last 5 seasons to estimate steady state behavior for a season.

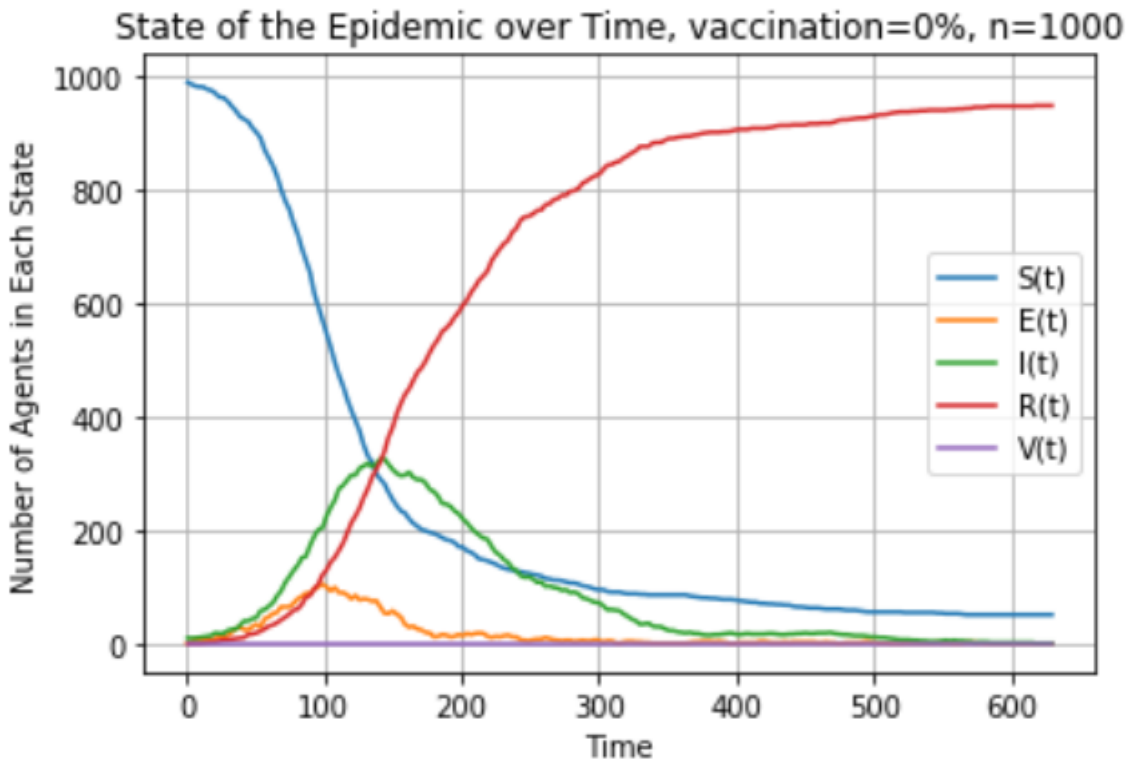

Figure 6: Example SEIR season with no vaccinations.

Table 1: Regression results for the county level vaccination percentage as a function of commonly used explanatory variables.

Regression Results for County Level Percentage of Residents Vaccinated for Influenza		
Independent Variable	Coefficient	t Stat
Intercept	-13.2	-0.95
LN(Density)	2.80	29.5
LN(Med Income)	3.63	2.99
Percent in Poverty	-0.12	-2.78
Percent White	0.08	7.65
Percent Black	0.03	2.40
Percent Some College	0.04	2.49
Percent Uninsured	-0.08	-2.03

Table 2: The effect of homophily.

Experiment 1: Epidemics with heterogeneity in homophily (proportion of contacts inside one's region)

Homophily	Region Contact Rate	Mean (SD) Percent Infected	Mean (SD) Percent Vaccinated
0.79	8	16.22 (4.68)	40.42 (5.75)
	12	18.04 (5.52)	54.78 (4.61)
0.89	8	16.80 (5.91)	39.3 (6.21)
	12	18.64 (6.24)	52.82 (5.43)
0.99	8	15.26 (8.40)	27.88 (7.12)
	12	18.78 (9.82)	43.14 (6.10)

Notes: We consider data at the hub-level. Each replication of the experiment contains 10 regions, where 5 of those regions have a contact rate of 8 and the remaining 5 have 12. For each level of homophily, we run 10 replications.We collect steady-state values of infection and vaccination by averaging these values for each region over the last five seasons of a replication.

Table 3: The effect of changes in contact rate.

Experiment 2: Epidemics with heterogeneity in contact rate

Treatment	Region Contact Rate	Mean (SD) Percent Infected	Mean (SD) Percent Vaccinated
A	6	14.78 (5.47)	25.44 (5.87)
	14	18.94 (5.94)	56.84 (4.71)
B	8	16.80 (5.91)	38.30 (6.21)
	12	18.64 (6.24)	52.82 (5.43)
C	10	18.6 (5.53)	46.81 (5.17)

Notes: We consider data at the region-level. For each treatment, we run 10 replications. We collect steady-state values of infection and vaccination by averaging these values for each region over the last five seasons of a replication.

Table 4: The effect of changes in the cost of infection.

Experiment 3: Epidemics with heterogeneity in the cost of infection

Cost of Infection	Region Contact Rate	Mean (SD) Percent Infected	Mean (SD) Percent Vaccinated	Cost Per Agent
2	8	23.90 (8.42)	32.94 (4.53)	0.81
	12	24.06 (6.38)	51.16 (4.7)	0.99
3	8	17.94 (6.54)	36.32 (7.67)	0.90
	12	17.68 (6.24)	53.60 (4.23)	1.06
4	8	16.38 (6.08)	39.46 (6.11)	1.05
	12	16.92 (5.76)	53.76 (4.55)	1.21

Notes: We consider data at the region-level. Each replication of the experiment contains 10 regions, where 5 of those regions have a contact rate of 8 and the remaining 5 have 12. For each cost of infection, we run 10 replications. We collect steady-state values of infection and vaccination by averaging these values for each region over the last five seasons of a replication.

Table 5: The effect of population size.

Experiment 4: The effect of population size and contact rates on region vaccination rate

		Population Size Treatment I	Population Size Treatment II	Population Size Treatment III
		Urban: 60, Rural: 140	Urban: 80, Rural: 120	All Hubs: 100
Contact Rate Treatment	Region Contact Rate			
A	6	25.44 (5.87)	24.98 (7.04)	24.56 (4.96)
	14	56.84 (4.71)	57.7 (3.90)	56.02 (5.21)
B	8	38.94 (6.01)	38.20 (7.18)	38.98 (4.98)
	12	54.04 (5.21)	53.98 (5.41)	51.52 (5.71)
C	10	46.91 (5.17)	44.82 (6.38)	45.90 (5.53)

Notes: We consider data at the region-level. Standard deviations are presented in parentheses. For each contact rate treatment and each population size treatment, we run 10 replications. Each replication contains 10 regions. We collect steady-state vaccination rates by averaging these values for each region over the last five seasons of a replication.

`