REGULAR ARTICLE

It's worth a shot: urban density, endogenous vaccination decisions, and dynamics of infectious disease

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

Abstract

We develop an agent-based model of vaccine decisions across a heterogeneous 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 rationale for traditional threshold models of vaccine decision-making that are commonly used in agent-based network models of vaccine choice. In the model, more dense urban regions have more connections between agents than less dense rural regions. Higher density leads to higher levels of vaccine usage and lower rates of infection in urban regions within the model. This finding adds to the more commonly discussed socio-economic reasons for higher levels of vaccination usage in urban areas compared to rural areas. In addition to this direct contribution, the paper emphasizes the importance of endogenous decision-making in models of epidemiology. For instance, we find that networks that lead to larger epidemics in exogenous vaccination models lead to smaller epidemics in our model because agents use vaccinations to offset the additional risk introduced by these network structures. Endogenous agent responses to risk need to be incorporated into theoretical and empirical models of economic epidemiology.

✉ Troy Tassier
  tassier@fordham.edu

  Andrew Souther
  asouther@fordham.edu

  Myong-Hun Chang
  m.chang@csuohio.edu

[1] Honors Program FCRH, Fordham University, Bronx, NY 10458, USA

[2] Department of Economics, Cleveland State University, Cleveland, OH 44115, USA

[3] Department of Economics, Fordham University, Bronx, NY 10458, USA

⌖ Springer

1 Introduction

Rates of vaccination differ greatly between urban and rural areas in many parts of the world. Many factors contribute to the differences such as lack of access to health care, poverty, and socio-economic status (Hinman and McKinlay 2015). In this paper we consider an additional contributing factor: urban density. Higher levels of density in urban areas 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 exposure to infection and thus may be more likely to choose to be vaccinated. We use differences in the number of contacts with other agents as a proxy for differences between urban and rural density. We then model the vaccine decision of agents in terms of a rational cost benefit calculation which includes an agent-estimated probability of infection within a local area, 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, Geoffard and Philipson 1996 and 1997, Bauch and Earn 2004, Boulier et al. 2007, Goyal and Vigier 2014, Galeotti and Rogers 2013, Tassier et al. 2015). We discuss later in the paper how our model differs in significant ways from these previous models.

1.1 Background

Influenza causes high rates of morbidity and mortality throughout the world. In the U.S. alone, between the years of 2010 and 2019, influenza has created an estimated 12,000 to 52,000 deaths each year (CDC 2020a). In addition to mortality, influenza causes 140,000–710,000 hospitalizations in a typical year (Rolfes et al. 2018). The influenza vaccine is the most common public health tool used to fight influenza. Despite its recommendation, in the U.S., only about one-half of individuals received an influenza vaccine during the 2019–2020 influenza season leaving the population short of levels needed to achieve herd immunity (CDC 2020b; Plans-Rubió 2012).

Further, rates of influenza vaccination are not uniform across the country. According to the U.S. Center for Disease Control (CDC) *weekly flu vaccine dashboard*, "Coverage among states and DC as of January 8, 2022 ranges from 28.0 to 72.1%; national coverage is 50.3%" (CDC 2022). There is even greater variation at the county level (data to be discussed below).

Most empirical investigations of vaccine behavior concentrate on factors that affect access to healthcare, demographic variables, or the growing anti-vaccine movement (Ołpiński 2012). In general, researchers find that higher socio-economic status results in higher levels of influenza vaccination (Lucyk et al. 2019). Differences exist across race and ethnicity as well (CDC 2020b and Hebert et al. 2005). Much of the empirical vaccination rate literature concentrates on the vaccination status of children and particularly 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; Omer et al. 2008;

Salmon et al. 2005). Others examine the effect of income, religion, race, and other demographic factors on vaccine coverage. Smith et al. (2004) find a distinction between under-vaccinated children and unvaccinated children. Under-vaccinated children have some vaccines but not all or have started a vaccine schedule but not completed the schedule on time. In contrast to under-vaccinated children, unvaccinated children have parents who reject vaccines. As an example, an under-vaccinated child may have received the first dose of a MMR vaccine but not a second dose, and the child is significantly past the recommended age for the second dose. An unvaccinated child simply has never received a MMR vaccine and is significantly past the recommended age. Smith et al. (2004) find that under-vaccinated children tend to come from families that are black, lack college degrees, and are in poverty. In contrast, 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. 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). 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).

There exists a smaller empirical literature that specifically examines differences in rural versus urban vaccination rates. Urban areas have been found to have higher rates of influenza vaccinations than rural areas, differing by 10 to 20 percent (O'Leary et al. 2015; Zhai et al. 2020). One reason cited 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). To the best of our knowledge there is no published research with formal modeling of urban versus rural vaccination choice. This is a primary goal of this research.

There is an established literature of vaccine behavior within economics and related disciplines. Some of this research is conducted using agent-based computational models with some similarity to the model that we develop (Bansal et al. 2006; Barrat et al. 2010; Del Valle et al. 2013; Vardavas and Marcum 2013; Tassier et al. 2017; Chang and Tassier 2021).

Strands of the agent-based and analytical modeling literature can be divided into several subsets of modeling frameworks. One subset consists of analyzing society level optimal vaccination rates in the presence of vaccine externalities. This research is conducted with populations assumed to be well-mixed (randomly mixed) in standard susceptible-infected-recovered (SIR) or susceptible-infected-susceptible (SIS) models. This subset does not consider heterogenous network structure. Examples of this type of research include Geoffard and Philipson (1996), Gersovitz and Hammer (2003), and Boulier et al. (2007). Other models within economics consider analytical network-based vaccine choice but in a different context than we study. For instance, Galeotti and Rogers (2013) consider individual vaccine choice when there are two groups with the possibility of overlapping network connections between the groups. However, their model incorporates an SIS setting. SIS models typically result in steady state levels of infections making them more tractable analytically. However, many of the most common vaccine preventable diseases (such as influenza or measles) are most commonly modeled within a SIR or SEIR framework as opposed to a SIS framework.

SIS models are commonly used for infectious diseases which do not confer immunity upon recovery such as gonorrhea. Other models incorporate an SIR or SEIR setting and a network structure, but often have a simplified decision-making structure. For instance, Vardavas and Marcum (2013) incorporate a heterogeneous network structure and endogenous decision-making but agents respond to population level epidemic size and individual experience. Each agent has an exogenously given threshold parameter. If the agent was vaccinated in the previous year and the global population has an epidemic size above the pre-assigned threshold, the agent chooses to be vaccinated again; if the global population threshold is not met, the agent forgoes vaccination. If unvaccinated in the previous year, the agent uses her experience of having been infected or not when making a decision. In comparison, the model that we develop below allows agents to respond to infection risk in a local region (as opposed to infection risk in the overall population) and the networks that we consider vary in ways that can be interpreted in a geographic context.

Other models in the agent-based paradigm change network structure variables or policy variables to investigate the effect of these changes on epidemic spread. However, these models do not incorporate endogenous decision-making by the agents. Instead, vaccination rates, parameters that determine network structure, and other relevant policy variables are assigned as parameters of the model. Examples of this literature include Bansal et al. (2006), Del Valle et al. (2013), Tassier et al. (2017), and Chang and Tassier (2021). Finally, Goyal and Vigier (2014) allow an agent to choose the population level network structure within a game theoretic setting before choosing a method of defense to contagion (which could be interpreted as a vaccine). This does not fit our research questions as no one agent in our population has the ability to design a top-down global network structure.

In this paper we combine several features of these economic 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. Agents exist in either a rural (less-dense) or urban (more-dense) environment which we call a region. There does not exist an explicit geography in the model but one could interpret the model as being based on geographic density. Second, as we have seen during the Covid-19 pandemic, often individuals do not act with only global risk in mind. Instead, agents consider local conditions in a state or city, for example, in making judgements of infection risk. In addition, because infection risk can vary by location, it is important to incorporate local differences in risk into an economic model of vaccine decision-making. In our model, agents make rational choices given their estimation of the risk of infection in their local region and compare this risk to private costs of infection and vaccination, which we allow to vary across agents. The agents choose to vaccinate when vaccination maximizes expected utility and forgo a vaccine when it does not. By varying costs and network structure we view the incidence of an infectious disease across time and the corresponding endogenous choices to vaccinate made by agents in the model. We find that incorporating endogenous choice of vaccine behavior is an important element of understanding epidemic outcomes. For instance, we find that changes in network structure that create larger epidemics in exogenous vaccination environments, create smaller epidemics in our endogenous vaccination environment. This occurs because agents offset the increased network risk with higher levels of

vaccine usage. Thus, endogenous behavior is an important component of theoretical analysis and, importantly, needs to be incorporated into policy advice and empirical investigation.

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 Susceptible Exposed Infectious Recovered (SEIR) model (Anderson and May 1992), which will be discussed in full later in the paper. This model can be interpreted as similar to a common infectious disease such as influenza, measles, or the current novel coronavirus pandemic. 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

To begin, we provide an empirical justification for the importance of density in vaccine decision-making. We do this by analyzing a dataset maintained by the Robert Wood Johnson Foundation that contains yearly influenza vaccinations for Medicare enrollees at the county level across the United States in 2019 and 2020. While Medicare enrollees are but one subset of the population, they are an important group because the elderly are most at risk for severe outcomes from influenza infection. Specifically, the data that we analyze can be found at the 2019 and 2020 County Health Rankings website which is maintained by the Robert Wood Johnson Foundation.[1] The data are compiled from various US government sources and published yearly. The dataset includes a wide range of health and demographic data aggregated at the county level. We choose to analyze two years of data due to concerns that vaccination behavior may have changed in 2020 due to the ongoing coronavirus pandemic. Despite our concern, our analysis of the influenza data across counties in these two years yields very consistent results. The county level correlation in vaccination rates between the two years is over 0.96.[2]

To begin the analysis we note that influenza vaccinations as well as other vaccinations vary greatly over geographic space. As an example, in the 2020 data, influenza vaccinations vary between about 5% and 65% across counties in the United States. Figure 1 displays the percentage of county Medicare residents that were vaccinated for influenza in 2020 versus the natural log of population density (measured as population per km^2) along with the linear trend line. As one can see there is a strong positive correlation between population density and influenza vaccination rates.

---

[1] The 2020 data can be found at the following URL: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation and downloaded from 2020 County Health Rankings National Data. The 2019 data can be found at the following URL: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation/national-data-documentation-2010-2019 and downloaded from the 2019 County Health Rankings National Data link. Both websites last accessed February 16, 2022.

[2] We note that a preliminary version of the 2021 data is available. However, this type of data is routinely updated and we are not confident that there will not be significant updates to the 2021 data after the date of writing.
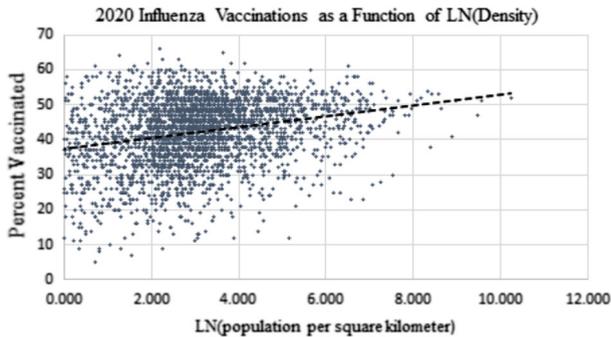
Fig. 1 Plot of County Level Influenza Vaccinations as a function of LN(population density). Population density measured as population per km^2

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

Regression Results for County Level Percentage of Residents Vaccinated for Influenza		
Dependent Variable Independent Variable	2019 Flu Vax Coefficient (t Stat)	2020 Flu Vax Coefficient (t Stat)
Intercept	− 161.0 (− 11.9)	− 171.4 (− 12.6)
LN(Density)	1.09 (12.1)	1.08 (12.0)
LN(Med Income)	17.2 (14.5)	18.2 (15.3)
Percent in Poverty	0.19 (4.29)	0.20 (4.70)
Percent White	0.09 (7.46)	0.09 (7.50)
Percent Black	0.11 (7.37)	0.11 (7.86)
Percent Some College	0.05 (2.89)	0.06 (3.32)
Percent Uninsured	− 0.12 (− 3.55)	− 0.10 (− 2.76)

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 regression results are presented in Table 1 and indicate that density remains a key determinant of county level vaccination rates when these demographic variables are added as controls.[3] This suggests that, in addition to socio-demographic factors (such as income or educational attainment) and lack of access to health care in rural areas, dense urban areas have a direct role in increasing vaccination coverage. Below

---

[3] 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 researchers.

we use different levels of agent contacts as a proxy for differences in density between urban and rural areas. We view the effect that these model parameters have on vaccination choice and on the rates of infection in urban versus rural areas.

3 Model

3.1 Overview

Epidemiologists commonly use compartmental models in the study of infectious disease. In these models, agents are assigned to a compartment based on a present state. The states may include such things as: *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). Agents then transition from state to state within the epidemic model. This is similar to agents in a labor market model transitioning from an employed state to an unemployed state. Epidemic models are commonly referred to by an acronym that lists the possible state transitions in the model. For instance, a model where agents flow from susceptible to exposed to infectious to recovered would be referred to as an *SEIR* compartmental model. This is the structure that we use in the present paper.

We first discuss an overview of our model and then describe each component of the model in more detail below. To begin, we describe the timing of the model. The basic unit of time in the model is a period. A period may be thought of as a day. Within each period agents interact with other agents across a contact graph, Γ, and also transition between states in the *SEIR* epidemic model described below.

A set of periods defines a season. In the initial period of each season, an epidemic is seeded with a set of infectious agents. The epidemic spreads until no agents remain infected. This concludes a season. After each season, agents are given an opportunity to make a vaccination decision. After each agent makes these decisions, a new season begins with a new initial period of seeded infections. Thus, each season may be thought of as a calendar year that starts with vaccination decisions in the fall and concludes after the traditional "flu season".

We repeat this process until we have completed 25 seasons. (We show below that 25 seasons is sufficient to generate steady state behavior in terms of vaccine decision-making.) This set of 25 seasons is called a replication. Throughout a replication all parameters of the model are held constant. For each parameter set of the model that we investigate, we perform 200 replications. In terms of notation, a specific season is denoted by s and a time period is denoted by t.

Next, 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 of a given season. (We suppress indexing each of these variables by s, unless necessary, to simplify notation.) 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. (The construction of Γ is described below.) If an agent is in susceptible state S and contacts an infectious agent in state I then the susceptible agent moves to state E with probability α. With probability $1 - \alpha$ she remains in state S. Once in state E, the agent moves to the

infectious state with probability ϵ (in each time period). Once in state I the agent then remains there until recovering. The agent recovers with probability ρ (in each time period). Thus, the expected duration in state E is $1/\epsilon$ and in state I 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 and does not allow for long-lasting immunity. Thus, in the next influenza season each agent enters the model in the susceptible state, S, regardless of her state at the end of the previous season.

3.2 Networks, Γ

Each agent has a home region. There are two types of regions in the model, urban and rural. Regions of each type vary by population size and density. Urban regions have a population size N_u and rural regions have a population size N_r. Urban regions have greater density than rural regions. Greater density is assumed to result in more contacts for an agent living in a more-dense region. For simplicity we assume that there are only two levels of density which can be thought of interchangeably as the number of contacts. 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 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 regions as parameters of the model.

Each agent has contacts with members of his own region and 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* percent of contacts will be drawn from an agent's own region while (*1-h*) percent will be drawn from other regions. *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 2021). The model also has similarities to the canonical Watts and Strogatz (1998) model of network re-wiring. Our contacts from outside the agent's own region are similar to their non-local rewired contacts. These contacts from outside an agent's own region also greatly shorten the characteristic path length as do the re-wired contacts of the Watts-Strogatz model (Chang and Tassier 2021).

3.3 Vaccinations

We allow agents the opportunity to vaccinate at the start of a season. The vaccination decision rests on the comparison of expected costs computed based on each agent's assessment of the infection risk. Individual agents assess this infection risk using a combination of observed epidemic outcomes from the past seasons. Most importantly, we assume in this paper that agents use the observed epidemic outcomes in their "home" region. This is in contrast to the papers discussed above that commonly consider agents making choices on the basis of "global" epidemic size.

The costs relevant for the vaccination decision are specified as follows. If an agent is infected during the course of a season, the agent pays a cost, C_I. If the agent chooses

to be vaccinated it pays a cost C_V. If vaccinated, we assume that the agent is fully protected from infection.[4] The agent weighs these costs against 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 will vaccinate if and only if:

$$C_V < \pi_{j,s} C_I \tag{1}$$

where $\pi_{j,s}$ is the probability that an unvaccinated agent, in the agent's own region j, will be infected in the current season of the model. Note that the right side of the inequality is the expected cost of infection that follows when the agent is not vaccinated.

Rearranging Eq. (1) we get:

$$\pi_{j,s} > \frac{C_V}{C_I} \tag{2}$$

In Inequality (2), notice that the individual values of C_V and C_I appear as the ratio $\frac{C_V}{C_I}$. This ratio is then compared to the probability of infection. Because only the ratio matters for the decision-making, for the remainder of this paper, we will not refer to the individual values of C_V and C_I; instead, we will refer to the cost ratio $\widetilde{C} \equiv \frac{C_V}{C_I}$. It should also be noted that our analytical interests are restricted to those cases with $C_V < C_I$, because $\pi_{j,s}$ is a probability, taking a value between 0 and 1. If $C_V \geq C_I$, then the agents will never vaccinate in our model, rendering the analysis trivial.

In our model, each agent k draws an individual cost ratio $\widetilde{C}(k)$ from the uniform distribution $[\widetilde{C}_L, \widetilde{C}_H]$, where $\widetilde{C}_L$ and $\widetilde{C}_H$ represent the lower and upper bounds of cost ratios in the model. We assign each agent her own cost ratio to reflect the heterogeneous costs that infection and vaccination impose on different members of a society. In certain experiments, we also allow the distribution to vary across urban and rural regions to reflect potential differences in costs due to different levels of access to vaccinations or different costs of infection.

3.4 Equilibrium (in a simplified setting)

Of course, the *social* aspect of infectious diseases implies that the probability with which any unvaccinated agent gets infected in the upcoming season depends crucially on how many other agents choose to be vaccinated as well as the contact structure in the population, Γ. Since all agents make this vaccination choice simultaneously, none of them can directly calculate this probability in advance.

---

[4] If instead we were to 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 a qualitative manner. For instance, any level of effectiveness can be rescaled by an offsetting change in costs of vaccination and infection.

This creates a strategic, game theoretic situation. In the extreme, if all other agents choose to be vaccinated, a given agent should forgo a vaccine and free-ride on the choices of other agents. If no other agent vaccinates, and these choices create a sufficiently large risk of infection then the agent should vaccinate. This interdependence of choice and potential for free-riding create difficulties in finding outcomes of endogenous vaccination models. Because of this strategic situation and lack of knowledge of other agent choices, a given agent will need to estimate, in some manner, the expected probability of infection for the present period.

For the moment, let us consider a two-agent normal form representation of this situation where there is only one region (again we suppress the season s). For the moment, also assume that all agents have the same costs. Note that the probability of infection for an agent depends on the choice made by the other agent. Let $\underline{\pi}$ denote the infection probability if the other agent vaccinates, while $\overline{\pi}$ denotes the infection probability when the other agent does not vaccinate. We assume: $0 \leq \underline{\pi} < \overline{\pi} \leq 1$. Hence, an agent is more likely to be infected if the other agent is unvaccinated than if he is vaccinated. Based on the infection probabilities and the costs of vaccination and infection, the payoffs to the two agents, facing a choice of strategy from {Vaccinate, Don't Vaccinate}, can be captured in the following payoff matrix:

	Vaccinate	Don't Vaccinate
Vaccinate	$-C_V, -C_V$	$-C_V, -\underline{\pi} C_I$
Don't Vaccinate	$-\underline{\pi} C_I, -C_V$	$-\overline{\pi} C_I, -\overline{\pi} C_I$

Hence, if both agents vaccinate, they both incur the cost of C_V. Alternatively, if neither agent vaccinates, they both incur the expected cost of infection, $\overline{\pi} C_I$. (Note that the probability of infection is high in this case since neither is vaccinated). Finally, if one agent vaccinates while the other agent does not, the vaccinated agent incurs the cost of vaccination, C_V, while the unvaccinated agent incurs the expected cost of infection, $\underline{\pi} C_I$, where the probability of infection is relatively low since the other agent is vaccinated.

Denoting by $\widetilde{C} \equiv \frac{C_V}{C_I}$, there are three possibilities: (1) $\widetilde{C} \leq \underline{\pi}$; (2) $\underline{\pi} < \widetilde{C} < \overline{\pi}$; (3) $\widetilde{C} \geq \overline{\pi}$. The first case is when the cost of vaccination relative to the cost of infection is very low. In this case, it is straightforward that "Vaccinate" is the dominant strategy for both agents; hence, the pure-strategy Nash equilibrium entails both agents vaccinating. In the last case, the cost of vaccination relative to the cost of infection is very high, resulting in "Don't Vaccinate" as the dominant strategy: neither agent vaccinates in equilibrium. Finally, if the cost ratio (between vaccination and infection) takes an intermediate value as in case (2), "Vaccinate" is optimal for only one of the two agents and, hence, both (Vaccinate, Don't Vaccinate) and (Don't Vaccinate, Vaccinate) emerge as pure-strategy Nash equilibria. The main insight from this simple two-agent representation is that the equilibrium vaccination decisions depend on how the infection probability for an agent (which depends on the rival agent's action) compares to the cost ratio, $\widetilde{C}$.

We can generalize the above strategic interaction in the context of a larger population consisting of N agents. Let us retain the assumptions that all agents reside in one single region and have the same costs. Extending the formulation taken in the two-agent example above, let us denote by $\pi(n)$ the probability of an unvaccinated agent getting infected when n other agents are vaccinated, with $0 \leq n < N$. We make the following assumptions on $\pi(n)$:

(A.1) $0 \leq \pi(N-1) < \widetilde{C} < \pi(0) \leq 1$;
(A.2) $\pi'(n) < 0$ for all $0 \leq n \leq N-1$.

(A.1) states that the infection risk for an unvaccinated agent, if everyone else in the population is vaccinated, is sufficiently low to be below $\widetilde{C}$—i.e., $0 \leq \pi(N-1) < \widetilde{C}$, while the infection risk, when no one else in the population is vaccinated, is sufficiently high to be above $\widetilde{C}$—i.e., $\widetilde{C} < \pi(0) \leq 1$. It is also implicit in this assumption that the cost of vaccination cannot exceed the cost of infection since $\widetilde{C} \leq 1$. (A.2) states that the risk of infection for an unvaccinated agent monotonically declines as the number of vaccinated agents in the population increases.

(A.1) and (A.2) together guarantee that the Nash equilibrium number of vaccinated agents remains strictly between zero and N. To see this, consider the vaccination decision for an unvaccinated agent when there are n vaccinated agents in the population. If he chooses to vaccinate, he expects to incur the cost of C_V. Alternatively, if he chooses not to vaccinate, he expects to get infected with the probability of $\pi(n)$, in which case he will incur the cost of C_I, while with the probability of $(1 - \pi(n))$ he remains uninfected and avoids having to pay any cost. Hence, an unvaccinated agent will choose to vaccinate if and only if:

$$C_V < \pi(n) \cdot C_I \text{ or, equivalently } \pi(n) > \frac{C_V}{C_I} \left(\equiv \tilde{C} \right). \tag{3}$$

Otherwise, the agent will choose not to vaccinate. Note from (A.1) that $\pi(0) > \widetilde{C}$, while $\pi(N-1) < \widetilde{C}$. Since $\pi'(n) < 0$ for all n from (A.2), it follows that there exists a unique number, n^*, of vaccinated agents in the population such that $\pi(n^*) = \widetilde{C}$. Furthermore, for all $n < n^*$, there is a positive incentive for an unvaccinated agent to get vaccinated, while for all $n > n^*$ there is a negative incentive to do so. The stable Nash equilibrium number of agents who vaccinate is then defined by n^*.

Note that both of the assumptions (A.1) and (A.2) are required for there to be a unique and strictly interior equilibrium (i.e., $0 < n^* \leq N-1$). Suppose $\pi(0) < \widetilde{C}$, violating (A.1). This is if the infection probability, even when no one else is vaccinated, is extremely low to be below the cost threshold. In this case, the equilibrium entails no agent in the population vaccinating. Conversely, if $\pi(N-1) > \widetilde{C}$, the infection probability, even when everyone else is vaccinated, is sufficiently high that everyone in the population will vaccinate in equilibrium. If these corner solutions were possible, it could have interesting implications for the dynamics of vaccination in the repeated seasons setting of our model. Specifically, suppose that the infection probability in one season is such that no one vaccinates. That outcome will lead everyone in the population to vaccinate in the next season, which would then lead to zero vaccinations in the following season, so on and so forth. What we may observe is the wide

fluctuation in vaccination rates from one season to the next—an alternating sequence of population-wide vaccination and no vaccination. However, under (A.1) and (A.2) we get a strictly interior equilibrium in which there is always a non-empty group of agents who choose not to vaccinate.

While we expect the analytically attained equilibrium solution as described above to be what the population of agents would evolve toward in the long run, the agents in the model, especially with their heterogeneous costs of vaccination and infection along with heterogenous contacts within complex networks, have limited computational capabilities to compute the equilibrium infection probabilities in reaching their vaccination decisions. In addition, agents must coordinate in achieving the stated equilibrium with some agents vaccinating and others not. As such, we assume that the agents use "perceived" infection probabilities that they construct based on their observations from the past seasons. These observations allow agents an implicit form of coordination based on prior vaccination behavior and epidemic outcomes. In particular, we allow each agent to look at the *past rates of infection* in her own region j over all past seasons. Based on this information, the agent calculates an estimated, time discounted, probability of infection based on the infections observed in her home region. Call this estimate $\pi^e_{j,s}$. This is an estimate of the probability that the agent would have been infected if she had not been vaccinated in all the previous seasons. Specifically, it is calculated as follows: the agent calculates a time discounted, weighted average of past infection probabilities of unvaccinated agents in the agent's own region j, considering all seasons prior to the upcoming season s. We define this as $\pi^e_{j,s}$. Each agent calculates this weighted average using the following formula:

$$\pi^e_{j,s} := \frac{\sum_i^{s-1} \delta^{s-1-i} p_{j,i}}{\sum_i^{s-1} \delta^{s-1-i}} \tag{5}$$

where $p_{j,i}$ is the proportion of unvaccinated agents belonging to region j who were infected over the course of season i. δ is a discount rate, $0 \le \delta \le 1$. If the parameter δ is set to 0, notice that an agent would only consider the proportion of unvaccinated agents who were infected in the previous season. In other words, she ignores all information from seasons $1, 2, \ldots, s-2$. If $\delta = 1$, all seasons receive equal weight. For intermediate values of δ, more recent seasons receive greater weight than less recent seasons. We describe outcomes from a range of values for δ below, but in the majority of computational experiments, we set the value of δ to 0.9. This value is sufficiently large to allow convergence to an equilibrium without significant cycles masking individual decision-making.

If this estimate of infection probability is greater than the cost ratio $\widetilde{C}(m)$,

$$\pi^e_{j,s} > \widetilde{C}(m) \tag{6}$$

then an agent m in region j chooses to be vaccinated in season s. If not, the agent forgoes vaccination in season s.

3.5 Agent-based experiment design

We implement this model in an agent-based framework. Agent-based models have been recommended as ideally suited to investigate issues of public health in heterogeneous agent environments (Epstein 2009; Eubank et al. 2004). All experiments have 1000 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 200 times. We show results below to justify that 25 seasons are sufficient for convergence in behavior. (Convergence occurs by about season 10 in all experiments). Because the model is stochastic there are small variations in behavior even once convergence is reached. As such, we use the outcomes of the final five seasons in each replication as a summary statistic for epidemic outcomes. We present these results in tables and figures that report these average outcomes over the 200 replications for each experiment.

3.5.1 Agent-based implementation

To begin each experiment, we create 10 regions with 5 designated as urban and 5 designated as rural. We distribute 1000 agents across these regions. 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 agents). 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 randomly one at a time. 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 she does, a new agent is chosen. If she does not have a full set of contacts, she is repeatedly matched with other agents until she reaches that number. The agents to which she is matched 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.

At the end of this process, it is possible that the algorithm may fail to provide 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 randomly distributed. 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 0.79) are biased upwards by the same amount. Urban areas end up with slightly higher homophily than rural areas
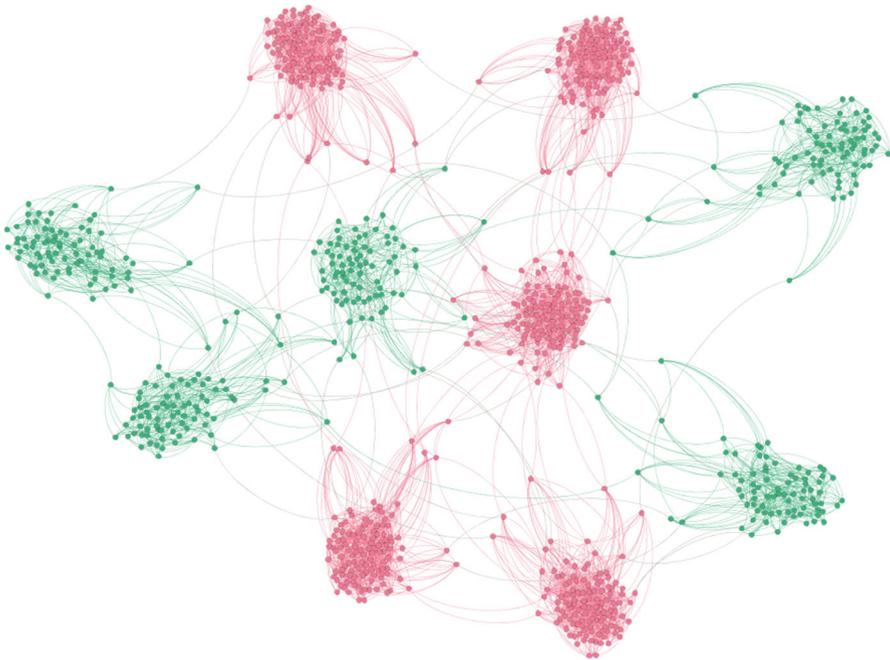
Fig. 2 Example network with $h = 0.99$. Red indicates agent in urban region, green indicates agent in rural region

(on average, a difference of 0.02). Three example networks are shown in Figs. 2, 3, and 4 for values of $h \in \{0.99, 0.89, 0.79\}$.

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. At this time, each agent k also draws her own cost ratio $\widetilde{C}(k) \equiv \frac{C_V}{C_I}$ from the uniform distribution $[\widetilde{C}_L, \widetilde{C}_H]$, where the bounds $\widetilde{C}_L$ and $\widetilde{C}_H$ have been established as an initial parameter of the experiment. After each agent is assigned this cost ratio, it does not change for the remainder of the 25-season replication. The epidemic continues according to the SEIR framework until no agents remain in the exposed or infectious state. Once this happens we have reached an absorbing state where each unvaccinated agent is either in the susceptible state or has 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 do this as follows: each agent compares her estimated probability of infection to her cost ratio. If agent k's estimate of the probability of infection in her home region j, is larger than her cost ratio,
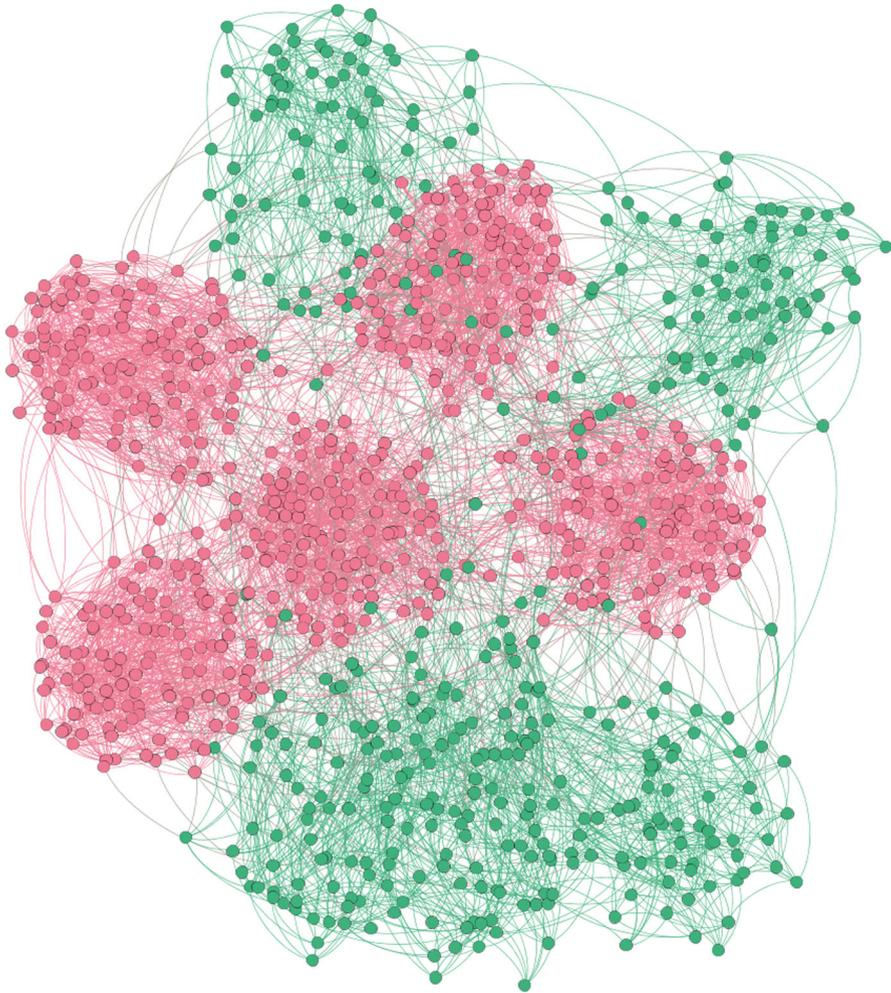
$$\pi_{j,s}^e > \widetilde{C}(k) \tag{6}$$

Fig. 3 Example network with $h = 0.89$. Red indicates agent in urban region, green indicates agent in rural region

then agent k chooses to be vaccinated. If not, the agent forgoes vaccination in season s. Note that each agent in region j has the same estimate of the probability of infection. However, there will be agent level variation in this choice because each agent has an individual level cost ratio. There will be a cost ratio $\widetilde{C}^*$ where all agents in region j with $C(k) < \widetilde{C}^*$ will vaccinate and all agents with $C(k) > \widetilde{C}^*$ will not vaccinate. In this sense the model is similar to Vardavas and Marcum (2013). However, in their model agents respond to global (as opposed to regional) infection rates in the most recent period (as opposed to a discounted history of infection rates). In addition, their vaccination threshold is not developed out of a model of rational decision-making; instead, it is assigned as an exogenously chosen random threshold from a specified distribution.
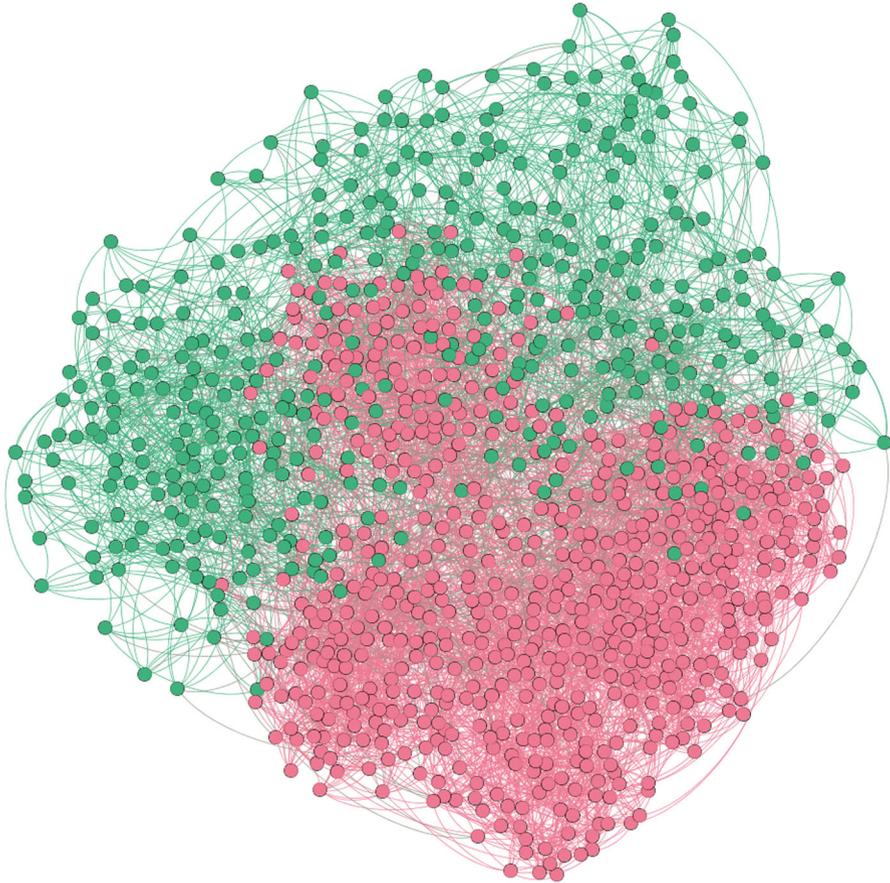
Fig. 4 Example network with $h = 0.79$. Red indicates agent in urban region, green indicates agent in rural region

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 seasons to converge to a steady state.[5]

We repeat this process for 200 replications for each parameter configuration that we investigate. All experiments reach a consistent level of vaccinations and infections by about the 10th season.

---

[5] As will be shown in graphs below, we reach an equilibrium with consistent behavior. Of course, since the model is stochastic, there will always be some amount of fluctuations within the model.

4 Results

We use the following parameters for the SEIR model throughout the paper: Transmission rate $(\alpha) = 0.03$, Recovery Rate $(\rho) = 0.08$, Exposed to Infected Transition Rate $(\epsilon) = 0.33$.

The recovery rate and exposed to infected transition rate were chosen to roughly match state duration values for influenza. We have performed additional experiments with a wide range of parameter values that are not presented here. Changes in these parameter values do not qualitatively alter the results discussed below.

We begin with a base implementation of 1000 agents distributed across 10 regions (5 urban and 5 rural) with 80 agents per rural region and 120 agents per urban region. The urban regions 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 a lower cost ratio bound of 1/4 and an upper bound of 1/2; $\widetilde{C}(k) \equiv \frac{C_V}{C_I}$, is drawn for each agent k from uniform [1/4,1/2]. Because the numerator is the cost of vaccination and the denominator is the cost of infection, you may think of a cost $\widetilde{C}(k)$ of $1/x$ to be a cost of vaccination of 1 and a cost of infection of x. Thus one may think of the cost of vaccination as a "numeraire" cost. The homophily parameter is set to 0.89. In Fig. 5 we vary the discount value, δ, across 0.0 to 1.0 at intervals of 0.1. The figure displays a representative example replication for each value of δ. As can be seen in the figure, smaller values of δ create cyclical best response behavior and larger values of δ allow convergence to a steady state. Recall from above that $\delta = 0$ indicates pure best response to last season's epidemic outcome. As expected, this yields oscillations between no agent vaccinations and all agent vaccinations across the population. From $\delta = 0.1$ to $\delta = 0.5$ this strong cyclical behavior continues with minimal differences in behavior
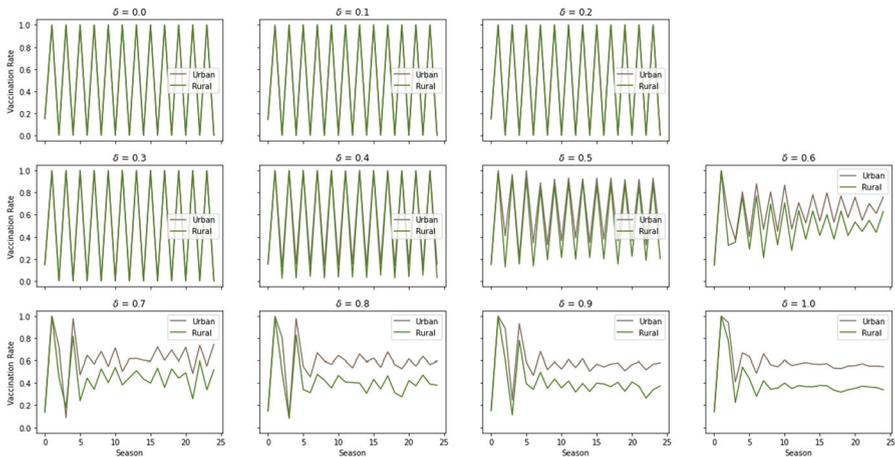

Fig. 5 Vaccination behavior of the model for different values of the discount parameter, δ. $\delta = 0$ indicates pure best response to most recent period and results in persistent cycles. Similar behavior exists for low values of δ. When δ is sufficiently large the model reaches a steady state behavior after a few seasons. A larger value of δ results in quicker convergence to steady state. $\delta = 1$ indicates that all periods of history are equally weighted

between the urban and rural populations. As larger values of δ are implemented, the cycles begin to dampen and the behavior of urban and rural agent populations begins to separate. For $\delta = 0.7$ and larger, we observe distinct behavioral differences between rural and urban agents, and steady state behavior (within a range that shrinks as δ grows) occurs after a sufficient number of seasons. Again, because the model is stochastic, one should never expect to observe pure steady state behavior that never changes over time; there will always be some fluctuations across seasons within an individual replication. Further note that the cycles tend to be correlated between rural and urban regions. This correlation occurs because infections in one region spill over into other regions whenever the regions are interconnected. Thus peaks and valleys in urban regions tend to occur in the same seasons as rural regions. Also note that the model settles into a stable range of vaccine choices between season 5–10 and afterward. We see this consistently in all replications with sufficiently large δ. We choose $\delta = 0.9$ as a baseline discount value that leaves us within this interesting range of stable convergence and distinct urban versus rural behavior.

We consider the values listed above as the base parameters for the analysis of the model. We then perform a series of computational comparative statics experiments where we vary some of these parameters, one at a time, to investigate the effects of these changes.

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. In addition, to better understand the output of the model, we perform a series of 200 replications with our base parameters. For each of the ten regions of each of the 200 replications we collect the infection rate at steady state. This amounts to 2000 individual regional infection rates, in total. We plot the histogram of these infection rates for all regions (top panel) and for urban (middle panel) and rural (bottom panel) regions separately in Fig. 7. As one can see, the histogram has a traditional skewed bell shape. This indicates that the average of the distribution will be an informative statistic to use as
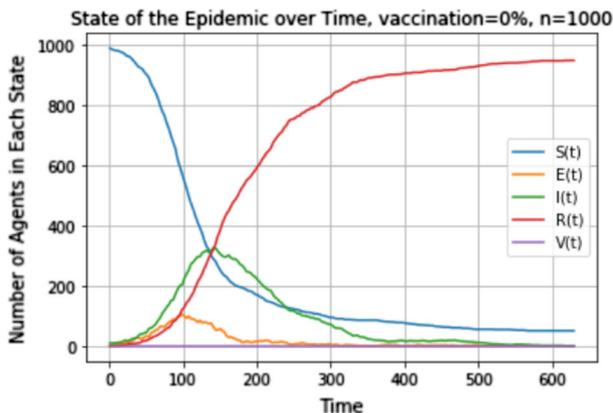

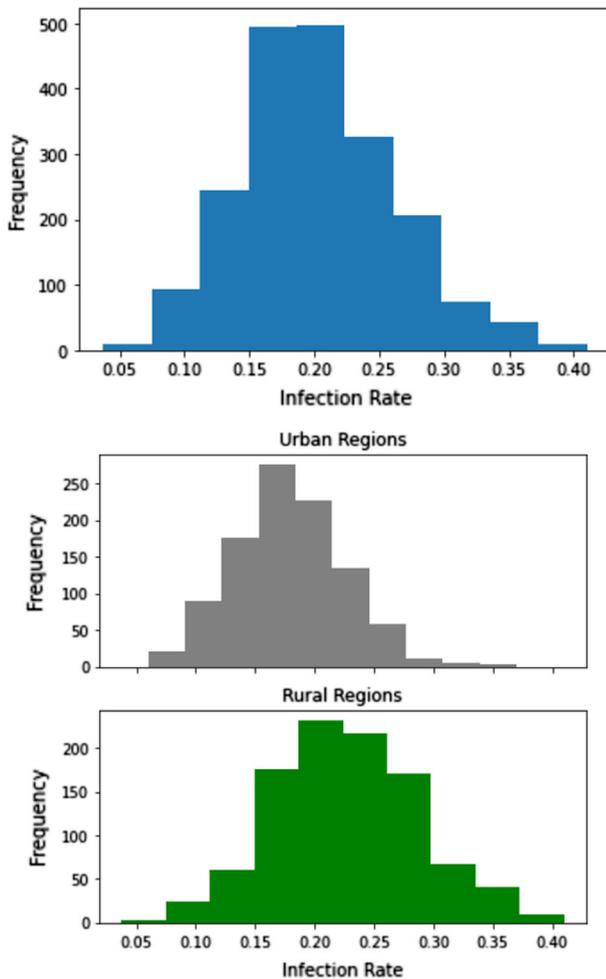**Fig. 6** Example SEIR season with no vaccinations

Fig. 7 Histogram of fraction of agents infected at steady state in each of the ten regions over 200 replications. This results in 2,000 observations of regional infection rates at steady state

a comparison in our comparative static experiments. (If the distribution was bi-modal or had another unusual shape, the average would be less informative).

We now begin describing the results of various combinations of parameter configurations. Again, our results take the form of comparative statics exercises where we vary one parameter at a time across a range of 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, $h \in \{0.79, 0.89, 0.99\}$. As stated earlier, we display example networks which are characteristic of these values in Figs. 2, 3, and 4. For $h = 0.99$ the

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	21.95 2 (5.57)	36.54 (7.95)
	12	17.06 (4.10)	58.26 (5.51)
0.89	8	22.80 (6.16)	35.33 (8.20)
	12	17.90 (4.64)	56.38 (6.18)
0.99	8	25.02 (9.70)	24.63 (11.50)
	12	21.93 (7.38)	46.14 (9.88)

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 level of homophily, we run 200 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

regions remain distinct. Lowering h to 0.79 creates a great amount of 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 and rural to rural 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 larger epidemics for a given level of *exogenous* vaccinations because of the increased regional interconnections leading to a shorter average network distance between agents (Chang and Tassier 2021). However, we see that this does not develop in our model because of the *endogenous* vaccination choice of agents.

With endogenous vaccination, infection rates grow larger, not smaller, as homophily increases. For the two lowest values of homophily displayed, the increases are relatively small, from 22 to 23% for rural regions and from 17 to 18% for urban regions. When we increase the homophily parameter to 0.99 (which results in minimal interaction between regions) we observe the largest values of infections, and larger increases in the levels of infections, 23% to 25% in rural regions and 18% to 22% in urban regions. These increases in rates of infection occur due to decreasing use of vaccinations. Again, for the two lower values of the homophily parameter the difference is small (1–2%) and a much larger difference (about 10%) occurs when homophily becomes more extreme. We also observe that the difference in urban and rural vaccination rates grows larger as homophily increases.

Agents achieve these various levels of infection by varying their vaccination behavior. As the amount of homophily decreases (leading to more contacts across regions) we see that more agents choose to be vaccinated; this is the response of agents to greater infection risk. For the rural areas, as homophily increases, vaccination rates decrease from 37 to 35% to 24% and for the urban areas vaccination rates decrease

from 58 to 56% to 46%. Note that most of the decrease occurs as we move from 89% in region contacts to 99% in region contacts.

That the largest change in behavior occurs for initial decreases in homophily is intuitive in comparison with Watts and Strogatz (1998). In their model, the largest changes in characteristic path length occur for the initial random rewirings of a small number of neighboring links to randomly chosen links. Here we find that the largest change in behavior within our model occurs with initial increases in inter-regional links (decreases in homophily). The fundamental insight is that even a small number of interregional or global contacts can have a large effect on network structure (the Watts and Strogatz model) and subsequently on behavior (our model).

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 response 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.

In summary, the largest takeaway from this experiment is that there is a large difference in epidemic outcomes when vaccine use is endogenously and not exogenously determined. The shorter distance between agents that occurs with lower rates of homophily (again see Chang and Tassier, 2021) leads to larger epidemics for an exogenously given level of vaccinations. However, when we allow agents to respond to this larger risk with an endogenously determined vaccination decision, we observe that infection risk (epidemic size) *decreases* with less homophily (more interregional contacts). Thus, not considering agent responses to mitigate infection risk may lead to incorrect insights about epidemic outcomes in complex environments.

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 vary the contacts rates, γ_u and γ_r. In this experiment the contact parameters in urban and rural regions take 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 (our base parameters), 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 with 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 from 10 to 12 to 14, the rate of infection decreases moderately from 20 to 18% to 17%. This is achieved through a substantial increase in the vaccination rate from 48 to 56% to 62%. The infection rates in the rural areas decrease by a similar magnitude from 24 to 23% to 20% as contacts increase from 6 to 8 to 10. However, these decreases in infection rates occur in tandem with much larger changes in vaccination rates (compared to the urban regions). In rural regions the vaccination rate decreases from 48 to 35% and then to 15% for decreases in contact from 10 to 8 to 6. The net change in vaccination rates in rural areas is 33% and the net change in urban areas is 14%.

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	24.32 (6.61)	14.64 (8.62)
	14	16.67 (4.56)	61.98 (6.02)
B	8	22.80 (6.16)	35.33 (8.20)
	12	17.90 (4.64)	56.38 (6.18)
C	10	20.05 (5.22)	47.96 (7.22)

We consider data at the region-level. For each treatment, we run 200 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

The results of this experiment lead one to hypothesize that the effect of increasing density in urban areas may have a nonlinear effect. Starting from equal contact rates in rural and urban regions within our model, increasing contacts (from 10 to 14 contacts per period) has a much smaller effect on vaccination rates (in urban regions) than decreasing density (from 10 to 6 contacts per period) has on decreasing vaccination rates (in rural regions). Also note that, like the previous experiment, more contacts leads to a lower rate of infection because of the large increase in vaccine use. Again this highlights the importance of models incorporating endogenous behavior. Agents are able to offset increased risk from more contacts with vaccination mitigation measures.

4.3 Experiment 3: ratio of cost of vaccination to cost of infection

We now vary the cost of infection in two different ways in our next two experiments. First, we vary the interval of possible costs for both the urban and rural regions across [1/3,1], [1/4,1/2], and [1/5,1/3]. Recall that the cost ratio is defined as the cost of vaccination over the cost of infection. For any ratio larger than 1, it is never optimal to vaccinate. Thus, the only relevant cost ratios are less than 1. As the cost ratio gets smaller, this indicates that the cost of vaccination is getting smaller, relative to the cost of infection. Again, these costs should be thought of in utility terms where cost includes time to be vaccinated, lost time from work if infected, decreased utility from illness, and other broad items.

We again return to benchmark levels of contacts, 12 in urban regions and 8 in rural regions. We present results of Experiment 3 in Table 4. As expected, as the range of values for the cost ratio decreases, the vaccination rates increase because vaccines become less expensive relative to infections. This increase in vaccine use leads to a decrease in the level of infections. Note that the change in magnitude for infections and vaccinations across these levels of costs ratios is very similar for urban and rural regions.

Table 4 The effect of changes in the cost ratio

Cost Ratio Bounds	Region Contact Rate	Mean (SD) Percent Infected	Mean (SD) Percent Vaccinated
[1/3, 1]	8	30.79 (6.14)	29.55 (6.22)
[1/3, 1]	12	25.93 (4.76)	50.60 (4.62)
[1/4, 1/2]	8	22.80 (6.16)	35.33 (8.20)
[1/4, 1/2]	12	17.90 (4.64)	56.38 (6.18)
[1/5, 1/3]	8	18.48 (6.31)	42.73 (10.20)
[1/5, 1/3]	12	14.41 (5.32)	63.61 (7.87)

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 200 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

4.4 Experiment 4: changing relative costs

We vary cost in a second way, rural cost relative to urban cost. Here we hold the urban cost constant at a range of [1/4,1/2] and allow the rural cost to vary across four different intervals, [1/3,1], [1/4,1/2], [1/5,1/3], and [1/6,1/4]. On average, in the first set of costs rural regions have a larger cost than urban regions, in the second set they are equal (as has been done in the other experiments so far), and in the last two sets the rural cost is less than the urban cost. The other parameters within this experiment follow our base model with 8 contacts in the rural regions and 12 in the urban regions and populations of 80 in the rural regions and 120 in the urban regions. Recall that as the cost ratio increases the relative cost of vaccination relative to infection is increasing. Thus, for the larger costs ratios for the rural regions (Treatment A of Table 5), vaccination rates are lower. As the relative cost of vaccination decreases (relative to the cost of infection), the rural regions increase their rates of vaccination, which has the effect of decreasing the rate of infection in the rural regions. For treatment D, the vaccination rates of urban and rural regions cross and the rural vaccination rate becomes one percent larger than the urban vaccination rate. Yet, the infection rate in the rural region is lower than in the urban region by about 4% due to the lower number of contacts for rural agents. Further, note that as the vaccination rate of the rural region increases (as a function of the decrease in vaccination cost) the vaccination rate in the urban region decreases by a small amount. This is an indication of the externalities present between the urban and rural regions. As agents in the rural regions vaccinate more, the infection rate in those regions decreases, and this offers protection to the agents in urban regions who are connected to rural regions. The decrease in the urban regions (4%) is much smaller than the increase in the rural regions (31%). Finally, note that the decrease in vaccinations in the urban regions is offset by a small 1.5% increase in the infection rate in the urban regions. Once again, this underscores the importance of endogeneity of

Table 5 Changing the cost ratio in rural regions holding the urban regions constant

Experiment 4: Effect of heterogeneity with respect to cost ratio

Treatment	Cost Ratio Bounds	Region Contact Rate	Mean (SD) Percent Infected	Mean (SD) Percent Vaccinated
A	[1/3, 1]	8	32.15 (6.80)	24.95 (6.94)
	[1/4, 1/2]	12	17.10 (4.20)	58.65 (5.67)
B	[1/4, 1/2]	8	22.80 (6.16)	35.33 (8.20)
	[1/4, 1/2]	12	17.90 (4.64)	56.38 (6.18)
C	[1/5, 1/3]	8	17.06 (5.58)	45.92 (9.63)
	[1/4, 1/2]	12	18.66 (4.71)	54.85 (6.42)
D	[1/6, 1/4]	8	14.34 (5.49)	55.66 (11.48)
	[1/4, 1/2]	12	18.68 (4.78)	54.62 (6.50)

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 treatment, we run 200 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

the vaccine decisions that agents make in our model. Changes in vaccination rates in one type of region are a product of infection rates in that region. However, in addition, these infection rates are a product of the interlinked networks across regions and the vaccination decisions and infection outcomes in all regions.

4.5 Experiment 5: effect of population size

Finally, we look at the vaccination rates for different sized urban and rural regions for each of the three urban and rural contact rate pairs discussed above, (14 and 6), (12 and 8), and (10 and 10). The regional population sizes vary across five pairs from 20 rural,180 urban (which is a similar ratio to the mix of urban vs rural population in the United States) to 100 rural, 100 urban (in other words, all regions are equal). In Fig. 8
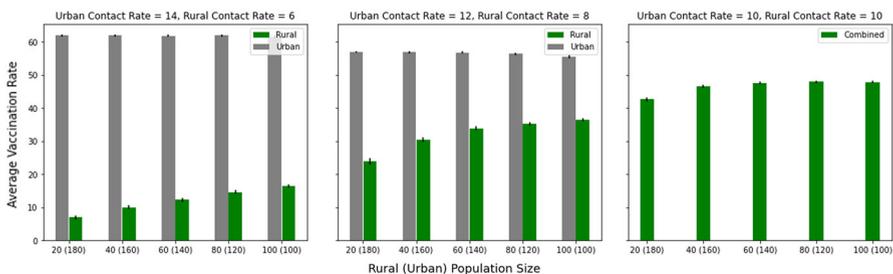

Fig. 8 As relative population size for the rural regions increases (relative to urban population size), the rural regions increase their average vaccination rate. Note that the vaccination rate of urban regions remains constant

we observe that changing the population splits does not change the vaccination rates in the urban regions when there is a difference in contact rates between the urban and rural regions (left and middle panels). However, holding the contact rates constant, rural vaccination rates increase as the number of agents in the rural regions increases (left and middle panels). In the right panel we present the overall vaccination rate when there is no difference in urban and rural contact rates as a benchmark for comparison.

Overall, the results of this experiment suggest that the population split between urban and rural regions is less important than the difference in contact rate for explaining urban vaccination rates. However, there is a significant effect on the rural vaccination rate as the rural areas become larger relative to the urban areas.

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 intuition for changes in epidemic outcomes becomes more difficult in an endogenous vaccination environment. For instance, changes in network structure that lead to larger epidemics in an exogenous vaccination environment lead to smaller epidemics when endogenous vaccine decision-making is incorporated into our model. This is because agents respond to increased risk by choosing to vaccinate more frequently. This suggests that policy prescriptions need to account for endogenous behavior of agents in order to make accurate predictions of policy outcomes. When comparing urban and rural decision-making, we find that urban agents in the model vaccinate at higher rates than rural areas in predictable ways as a function of differences in costs and contacts among agents. Even small differences in urban contacts (density) can lead to relatively large differences in vaccination rates. This is an overlooked reason for differences in vaccination behavior in rural versus urban areas.

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 rational choice. This model can be expanded in many ways to incorporate different beliefs about 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.

Declaration

References

Abbe E (2018) Community detection and stochastic block models: recent developments. J Mach Learn Res 18:1–86

Allred NJ, Shaw KM, Santibanez TA, Rickert DL, Santoli JM (2005) Parental vaccine safety concerns: Results from the National Immunization Survey, 2001–2002. Am J Prev Med 28(2):221–224

Anderson RM, May RM (1992) Infectious diseases of humans: dynamics and control, vol 26. Wiley, New York

Atwell JE, Van Otterloo J, Zipprich J, Winter K, Harriman K, Salmon DA et al (2013) Nonmedical vaccine exemptions and pertussis in California, 2010. Pediatrics 132(4):624–630

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

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

Bauch CT, Earn DJD (2004) Vaccination and the theory of games. Proc Natl Acad Sci USA 101(36):13391–13394

Bennett KJ, Pumkam C, Probst JC (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 DL, Sheshinski E, Intriligator MD (1991) Externalities and compulsary vaccinations. J Public Econ 45(1):69–90

Boulier BL, Datta TS, Goldfarb RS (2007) Vaccination externalities. BE J Econ Anal Policy 7(1):23

Carrel M, Bitterman P (2015) Personal belief exemptions to vaccination in california: a spatial analysis. Pediatrics 136(1):80–88. https://doi.org/10.1542/peds.2015-0831 ((**Epub 2015 Jun 1. PMID: 26034242**))

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

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

CDC 2022. https://www.cdc.gov/flu/fluvaxview/dashboard/vaccination-dashboard.html, last Accessed 10 Feb 2022

Chang M-H, Tassier T (2021) Spatially heterogeneous vaccine coverage and externalities in a computational model of epidemics. Comput Econ 58:27–55. 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 SY, Mniszewski SM, Hyman JM (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. Springer, New York

Epstein JM (2009) Modelling to contain epidemics. Nature 460(6):687

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

Feikin DR, Lezotte DC, Hamman RF, Salmon DA, Chen RT, Hoffman RE (2000) Individual and community risks of measles and pertussis associated with personal exemptions to immunization. JAMA 284(24):3145–3150

Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM (2010) Parental vaccine safety concerns in 2009. Pediatrics 125(4):654–659

Galeotti A, Rogers B (2013) Strategic immunization and group structure. Am Econ J Microecon 5(2):1–32

Geoffard P-Y, Philipson T (1996) Rational epidemics and their public control. Int Econ Rev 37(3):603–624

Geoffard P-Y, Philipson T (1997) Disease eradication: private versus public vaccination. Am Econ Rev 87(1):222–230

Gersovitz M, Hammer JS (2003) Infectious diseases, public policy, and the marriage of economics and epidemiology. World Bank Res Obs 18(2):129–157

Goyal S, Vigier A (2014) Attack, defence, and contagion in networks. Rev Econ Stud 81(289):1518–1542

Gust DA, Strine TW, Maurice E, Smith P, Yusuf H, Wilkinson M et al (2004) Underimmunization among children: effects of vaccine safety concerns on immunization status. Pediatrics 114(1):e16–e22

Hebert PL, Frick KD, Lane RK, McBean AM (2005) The causes of racial and ethnic differences in influenza vaccination rates among elderly medicare beneficiaries. Health Serv Res 40(2):517–538. https://doi.org/10.1111/j.1475-6773.2005.00370.x

Hinman AR, McKinlay MA (2015) Immunization equity. Vaccine 33(4):D72–D77. https://doi.org/10.1016/j.vaccine.2015.09.033

Kremer M (1996) Integrating behavioral choice into epidemiological models of AIDS. Q J Econ 111(2):549

Lieu TA, Ray GT, Klein NP, Chung C, Kulldorff M (2015) Geographic clusters in underimmunization and vaccine refusal. Pediatrics 135(2):280–289. https://doi.org/10.1542/peds.2014-2715 ((**Epub 2015 Jan 19 PMID: 25601971**))

Lucyk K, Simmonds KA, Lorenzetti DL, Drews SJ, Svenson LW, Russell ML (2019) The association between influenza vaccination and socioeconomic status in high income countries varies by the measure used: a systematic review. BMC Med Res Methodol 19:153. https://doi.org/10.1186/s12874-019-0801-1

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. Ann Rev Sociol 27(3):415–444

Moody J (2001) Race, school integration, and friendship segregation in America. Am J Sociol 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. J Rural Health 31(4):421–430. https://doi.org/10.1111/jrh.12119

Olive JK, Hotez PJ, Damania A, Nolan MS (2018) The state of the antivaccine movement in the United States: a focused examination of nonmedical exemptions in states and counties. PLoS Med 15(6):e1002578. https://doi.org/10.1371/journal.pmed.1002578

Ołpiński M (2012) Anti-vaccination movement and parental refusals of immunization of children in USA. Pediatr Pol. https://doi.org/10.1016/j.pepo.2012.05.003

Omer SB, Enger KS, Moulton LH, Halsey NA, Stokley S, Salmon DA (2008) Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. Am J Epidemiol 168(12):1389–1396

Plans-Rubió P (2012) The vaccination coverage required to establish herd immunity against influenza viruses. Prev Med 55(1):72–77. https://doi.org/10.1016/j.ypmed.2012.02.015

Rolfes MA, Foppa IM, Garg S, Flannery B, Brammer L, Singleton JA, Burns E, Jernigan D, Olsen SJ, Bresee J, Reed C (2018) Annual estimates of the burden of seasonal influenza in the United States: a tool for strengthening influenza surveillance and preparedness. Influenza Other Respir Viruses 12(1):132–137. https://doi.org/10.1111/irv.12486

Salmon DA, Moulton LH, Omer SB, deHart MP, Stokley S, Halsey NA (2005) Factors associated with refusal of childhood vaccines among parents of school-aged children: a case-control study. Arch Pediatr Adolesc Med 159(5):470–476

Smith PJ, Chu SY, Barker LE (2004) Children who have received no vaccines: who are they and where do they live? Pediatrics 114(1):187–195

Tassier T, Polgreen P, Segre A (2015) Vaccination games with peer effects in a heterogeneous hospital worker population. Adm Sci 5:2–26

Tassier T, Polgreen P, Segre A (2017) Network position and healthcare worker infections. J Econ Interac Coord 12:277–307

Vardavas R, Marcum CS (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. Springer, New York

Wallinga J, Heijne JC, Kretzschmar M (2005) A measles epidemic threshold in a highly vaccinated population. PLoS Med 2(11):e316 (**Epub 2005 Oct 18**)

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

Zhai Y, Santibanez TA, Khan KE, Srivastav A, Walker TY, Singleton JA (2020) Rural, urban and suburban differences in influenza vaccination coverage among children. Vaccine 38(48):7596–7602. https://doi.org/10.1016/j.vaccine.2020.10.030