



# Spatially Heterogeneous Vaccine Coverage and Externalities in a Computational Model of Epidemics

Myong-Hun Chang<sup>1</sup> · Troy Tassier<sup>2</sup>

Accepted: 21 September 2019 / Published online: 5 October 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

We develop an agent-based computational model of epidemics with varying amounts of heterogeneity in regional vaccination coverage. We examine the severity of the resulting epidemics and the externalities created between high and low vaccination regions. The resulting size of the epidemics depends upon the heterogeneity of coverage, the interaction structure in the population, and how close the population is to herd immunity. Holding overall vaccination levels constant, the mean epidemic size increases and a larger level of vaccinations is needed to reach herd immunity when there is greater variability in the regional vaccination rates as long as the vaccination level is sufficiently high. This relationship amplifies when the contacts of individuals in the model have a greater share of inter-regional linkages. However, when inter-regional linkages and vaccination levels are sufficiently low, the mean epidemic size can decrease when there is greater variability in the regional vaccination rates. Further, we find that the magnitude of externalities between regions with different levels of vaccine coverage depends crucially on how close the population is to herd immunity. If population level vaccination coverage is sufficiently high (low) shifting connections from low coverage regions to high coverage regions will decrease (increase) mean epidemic size.

---

We thank Jason Barr, Marcos Fernandes, Chris Ruebeck, and other participants at the 44th Eastern Economic Association Meetings (Boston) for their comments. The comments of two anonymous referees were extremely helpful. The first author acknowledges the institutional support from the Cleveland State University in the form of a sabbatical leave (Fall 2017) during which period a substantial portion of this research was carried out. The second author acknowledges the support of a Fordham University Faculty Fellowship in carrying out this research.

---

✉ Myong-Hun Chang  
m.chang@csuohio.edu

Troy Tassier  
tassier@fordham.edu

<sup>1</sup> Department of Economics, Cleveland State University, Cleveland, OH 44115, USA

<sup>2</sup> Department of Economics, Fordham University, Bronx, NY 10458, USA

**Keywords** Epidemics · Vaccine coverage · Infectious disease · Influenza · Externality · Agent-based model

## 1 Introduction

Despite the availability of highly effective and inexpensive vaccines, many infectious diseases persist throughout the world. As an example, influenza and pneumonia consistently rank among the top ten causes of death in the U.S. (CDC 2017a) yet coverage rates for these and other vaccines are relatively low. In the US only 40–50% of individuals choose to be vaccinated for influenza each year and slightly more than 20% of adults are vaccinated for pneumonia (CDC 2017b). Measles, Mumps, and Rubella (MMR) vaccination rates are much higher, partly because of mandated vaccination for school age children in many areas. Yet still, between 5 and 10% of parents forgo recommended MMR vaccines for their young children and many seek exemptions to mandatory vaccination requirements (CDC 2017c).

Economic incentives help to explain part of the low vaccination coverage: Given the rarity with which many of these infectious diseases cause death or extreme illness in otherwise healthy people, many judge the risks of forgoing a vaccination to be low and do not bother to be vaccinated even though the vaccines have a minimal monetary cost (Geoffard and Philipson 1997). In many cases it is likely a time cost that proves more prohibitive than the monetary cost which is one reason why many larger firms offer influenza vaccines to be distributed at work facilities. In addition to the cost effects of vaccines, there is a growing anti-vaccine movement which has been perpetuated by many well-known public figures (Olpinski 2012). Despite the lack of scientific evidence to support the claims, or in many cases scientific evidence in opposition to the claims, the anti-vaccine movement remains strong especially in some locales. This vaccine opposition often has a localized distribution that plays out in vaccine denial being especially high in particular localized areas. As a result, researchers have found a great deal of spatial heterogeneity in rates of vaccination coverage (Carrel and Bitterman 2015; May and Silverman 2003; Lieu et al. 2015).

The anti-vaccine movement and spatial heterogeneity in vaccine coverage has sparked a wide spread discussion in academics and the general press.<sup>1</sup> Most of these discussions focus on decreases in overall vaccination rates and the accompanying increase in risk of infection (Feikin et al. 2000; Atwell et al. 2013). Other researchers examine the demographic correlates of vaccine choice and try to identify factors that lead to anti-vaccine sentiment, vaccine refusal, and simply lower levels of vaccination among some groups. Factors investigated include: distrust of vaccine safety; poverty and lack of easy access to healthcare; misunderstanding of both the benefits of vaccines and the potential risks of avoiding vaccination; and demographic factors (McNutt et al. 2016; Birnbaum et al. 2013; Sugarman et al. 2010; Gust et al. 2004). Outside of research in public health, medicine, and public policy, there is a growing interest in economics on the welfare implications and externalities inherent in vaccinations and

<sup>1</sup> For example see Ingraham (2015) or Oster and Kocks (2018) for discussions in the popular press. Or see Buttenheim et al. (2012), or Poland and Jacobson (2001) for discussion in academic journals.

epidemic processes (Gersovitz 2013; Philipson 2000; Chen and Toxvaerd 2014). Most recently, economists have begun to more thoroughly consider the effect of group structure and vaccine choice (Galeotti and Rogers 2013 and 2015). Our research continues to expand this area of interest.

Despite the above discussed interest in spatial vaccine coverage and distribution, there exists a lack of research explicitly exploring this area (apart from interest in a decrease in the overall vaccination rate). In this paper we take observed spatial heterogeneity as given and examine how the spatial heterogeneity itself, holding overall vaccination levels constant, can generate unique features of an epidemic process. These features depend strongly upon the interaction structure within the population. The combination of these two factors is a unique area that has not been explored in the literature to date. We develop an agent-based computational model of an epidemic process that occurs across regions. In the model we create spatial disparities in vaccination rates. The population in our model exists in a discrete set of non-overlapping regions with varying levels of vaccination coverage across the regions. In order to isolate the effect of only spatial heterogeneity we change the variance in the vaccination rates across regions while holding the population level vaccination rate constant. This isolation of the variance apart from changes in overall vaccination levels allows us to more fully explore and understand the implications for spatial differences in vaccine coverage. This is a unique feature of our research which allows us to isolate and examine potential externalities that occur across regions with differing levels of vaccine coverage and with different rates of interaction between the regions.

The most direct result of our model shows that spatial heterogeneity by itself can increase epidemic size (on average) when the population has a sufficiently high rate of vaccine coverage. We find this effect to be most pronounced when agents have more inter-regional connections. These results are consistent with expectations. However, we also find that this spatial heterogeneity can decrease epidemic size in special cases where vaccination levels are sufficiently low and there are sufficiently few (but some) connections between regions. We then use our model to identify inter-regional externalities created by differing vaccination rates as we change the contact structure of the model. When we increase connections between high and low vaccine neighborhoods, as one would expect, we find that unvaccinated individuals in the high vaccine neighborhoods have increased risk of infection (a negative externality), while those in low vaccine neighborhoods have decreased risk of infection (a positive externality). The magnitude of these externalities, relative to each other, determines whether increasing connections between high and low vaccine areas will ultimately increase or decrease the average epidemic size. Holding overall vaccination levels, neighborhood heterogeneity in vaccine coverage, and the percentage of inter-regional contacts constant we find that magnitude of the externalities relative to each other, depends upon the overall level of vaccinations in the population and, specifically, how close the population is to reaching herd immunity. When the overall vaccination rate is sufficiently high, near or above levels that produce herd immunity, allowing more interconnections between the high and low regions can lower epidemic size. Conversely, when vaccine levels are sufficiently below the level needed for herd immunity, allowing more interconnections between high and low level regions significantly increases epidemic size.

All of the results stated above depend on the non-linearity of the relationship between vaccine coverage and epidemic size. But, this relationship depends intricately on both the heterogeneity in vaccine coverage across regions and the interaction structure in the population. The interaction structure in particular creates non-trivial difficulties in analytically isolating the effects that we find. We use an agent-based model to better understand the complex relationships between interaction structure, heterogeneous vaccine coverage, and epidemic size.

In economics, agent based models are most frequently used to incorporate into theoretical models elements of agent heterogeneity, agent adoption or learning, or situations of complex agent interactions. Frequently these models examine the underlying dynamics of the model and often emphasize out of equilibrium dynamics (either as a model result or on the path to an equilibrium). Other uses of agent based models include comparative statics. Agent based models allow one to fluidly change parameter settings and to parsimoniously examine how those changes in parameters result in changes in model outcomes through theoretical “computational experiments” (Miller and Page 2007). The current paper is an example of this feature of agent based modeling.

Because epidemic processes frequently depend on interaction structures that are not tractable analytically, agent-based computational models similar to the one implemented in this paper are used to understand epidemic process (Epstein 2009; Eubank et al. 2004). These models are used in a variety of ways. In some instances, computational models are used to make real-time predictions as an epidemic progresses (Tizzoni et al. 2012). Most, however, attempt to better understand the dynamic processes involved in an epidemic spreading on a complex network and on how behavioral or interaction based changes can modify these processes.<sup>2</sup>

For the purposes of our paper, agent-based computational models are ideal. This modeling paradigm allows us to easily change assumptions about regional vaccination coverage and inter-regional interactions in a precise and controlled manner and to examine the comparative statics properties of changes in vaccine coverage heterogeneity and interaction structure. In doing so we can isolate the effect of these variables and gain a greater understanding of their effect.

The most unique contribution of our paper is the explicit study of vaccine coverage heterogeneity apart from changes in overall vaccination coverage. As described above, researchers have observed variance of regional vaccination rates. The results presented in this paper demonstrate that this phenomenon by itself, apart from changes in overall vaccination rates, has significant importance for understanding the economics of disease transmission.

To better understand the degree of variation in vaccination rates, next we provide an example of geographic vaccination coverage data. This will make clear how wide ranging vaccination coverage rates are across regions of the U.S. Following this example we describe our formal model that will be used to examine the effect of this heterogeneity. In Sect. 4, we describe the set of computational experiments used to isolate and understand the effects of vaccination coverage heterogeneity. We describe

---

<sup>2</sup> Examples include: Barrat et al. (2010), Vardavas and Marcum (2013), Del Valle et al. (2013) Tassier et al. (2017), Bansal et al. (2006), Germann et al. (2006), Smith et al. (1999), Stroud et al. (2007).

**Table 1** Influenza vaccine claims at county level

State	Mean	Min county	Max county	SD
California	44.43	16.86	53.08	9.02
Texas	46.98	8.58	60.62	11.06
Florida	46.27	20.64	59.05	8.40
New York	50.59	38.77	60.66	4.08

our results in Sect. 5. Particularly we show the changes in epidemic size that result from changes in the spatial distribution of vaccine coverage and how this interacts with changes in the number of inter-regional connections in the contact structure of the model. We also show the magnitude of positive and negative externalities between high and low vaccination regions. The magnitude of these externalities, relative to each other, depend crucially on how close the population is to reaching herd immunity and on the interaction structure of the population.

## 2 Influenza Medicare Claims Data

To give the reader a sense for the magnitude of vaccine coverage differences across geography, consider influenza vaccination rates among Medicare fee-for-service beneficiaries during the 2016–2017 influenza season. This data is publicly available at the state, county, and zip code level, on the U.S. Health and Human Services website for every “flu season” since 2012–2013. Of course, this data does not cover the entire U.S. population, but, it does provide a detailed look at the degree of heterogeneity in influenza coverage for this sample of the population. In addition, the overall vaccination level of this population during the 2016–2017 influenza season is 48.2%. This rate is similar to the CDC estimate for the overall U.S. population, 46.8%.<sup>3</sup>

In this data the state level vaccination coverage ranges from 34.98% (Alaska) to 56.88% (Delaware). The median states, Utah and South Carolina, have vaccination coverage slightly below 49%. Looking at a finer level of granularity, consider the 4 most populated states in the U.S.: California, Texas, Florida, and New York. The mean, minimum, maximum, and standard deviation at the county level are given for each of these states in Table 1. As one can see in the table, there are large differences across counties within these states. The difference between the minimum and maximum coverage county is about 21% in NY, 36% in CA, 38% in FL, and 52% in TX. The differences in these states are typical for most states in the country.<sup>4</sup>

Moving to an even finer level of geographic area yields a similar outcome in terms of coverage heterogeneity. In Table 2, we report the coverage statistics at the zip code level for the counties containing five of the major cities in the U.S. As can be seen in the table, again, there is a great deal of heterogeneity within these counties at zip code

<sup>3</sup> Again, other examples of spatial heterogeneity are given in the research cited above.

<sup>4</sup> We note, there are a few examples of states with more uniform coverage. Specifically, Delaware and Connecticut both have relatively uniform coverage at the state level and are the two highest coverage states. However, they are also states with a very small number of counties, 3 and 8 respectively. The third highest coverage state, Wisconsin, has a coverage range similar to the states reported in the table.

**Table 2** Influenza vaccine claims at zip code level

County (major city)	Mean	Min zip code	Max zip code	SD
Dade (Miami)	25.09	7.48	50.13	8.99
New York (NY City)	44.71	21.58	68.75	11.24
Cook (Chicago)	47.61	22.86	68.48	10.16
Harris (Houston)	43.85	19.27	66.67	8.65
King (Seattle)	53.39	28.57	70.37	7.88

level and also across these particular counties, (the Miami, Dade County area having much lower coverage than the Seattle, King County area).

What these two tables demonstrate is that there is large variability in vaccine coverage at county and local (zip code) levels. These are geographic levels that are likely to have a good number of cross-regional interactions such that an epidemic can spread across them. Most specifically the differences between high and low areas can be on the order of about 40% at the local, zip code level. In this paper we develop a model that can be used to investigate the effect of this heterogeneity. We purposely keep the model as simple as possible so that we can isolate the effect of spatial heterogeneity and also to be able to pinpoint the externalities that result between high and low vaccine regions.

### 3 The Model

Our model consists of the spread of an infectious disease across a population divided into a set of non-overlapping regions. Each agent interacts with a subset of agents from her own region; we call these local contacts. The agents also interact with a smaller number of agents from other regions; we call these global interactions. Each interaction creates the possibility of disease transmission if one of the agents is infectious and the other is susceptible to the disease.

#### 3.1 Agents and the Hubs of Social Interaction

There is a fixed population of agents,  $Z \equiv \{1, \dots, z\}$ . These agents are distributed in a physical space. There exists a fixed number,  $H$ , of hubs of activity (social interaction) that the agents belong to. We refer to these hubs as regions interchangeably throughout the paper. In our model “hubs” are locations where person to person interactions take place. We assume these hubs are geographically based such that a hub may be thought of as a neighborhood or school district. We will assume that the hubs are of only one type and that every agent in the population belongs to one and only one hub. Let  $h(i)$  denote the hub for agent  $i$ . The group of agents belonging to a particular hub is denoted  $\Gamma_k \equiv \{\text{all } i \in Z | h(i) = k\}$ , where  $k \in \{1, \dots, H\}$ . Since every agent is assigned only one hub,  $\bigcup_{k=1}^H \Gamma_k = Z$  and  $\sum_{k=1}^H |\Gamma_k| = z$ .

In our computational experiments, we assign agents to hubs by creating a fixed number of equal-sized regions, where each region is served by a single hub—i.e., all agents in a given region belong to the same hub.

Even though each agent ( $i$ ) belongs to an activity hub,  $h(i)$ , and shares that hub with all others in  $\Gamma_{h(i)}$ , that does not mean that  $i$  actually interacts with everyone in  $\Gamma_{h(i)}$ —e.g., one does not necessarily interact with everyone in one's own school district; regular interactions are often limited to only those within one's own school or perhaps classroom. We assume that each agent has a set of agent-specific contacts that are comprised of only a small subset of individuals from  $\Gamma_{h(i)}$ . In addition, some contacts for a given agent may also contain agents from outside an agent's hub. The structure of contacts is described in detail next.

### 3.2 Agent Contacts

Note that the disease we model spreads strictly through interpersonal interactions among the individuals. As such, it is the global pattern of contacts along with vaccination patterns that determine the eventual paths for the disease to spread. Each agent  $i$  has a set of contacts,  $C(i)$ , which is the group of individuals who interact, in close physical proximity, with agent  $i$  such that transmission of the disease can occur. These contacts are, on average, of a size  $\bar{c}$ . In our computational model, the contacts are created via random matching in a sequential manner until the average number of contacts per agent in the population reaches  $\bar{c}$ . In our model a large proportion of one's contacts come from  $\Gamma_{h(i)}$ , the agent's own hub—think of neighbors, co-workers, or classmates one interacts with on a regular basis. A smaller percentage of the contacts include agents outside of the hub (e.g., people he regularly rides with on the bus to the workplace or school). This mix of local and global contacts is similar in spirit, though different in detail, to the canonical models of Watts and Strogatz (1998). As such, we assume that a  $p^C$  share of agent  $i$ 's contacts come from  $\Gamma_{h(i)} - \{i\}$ , while the remaining share  $(1 - p^C)$  of  $i$ 's contacts come from  $Z - \Gamma_{h(i)}$ . The contacts for an agent, once created, stay fixed over time for all agents. We shall call  $p^C$  the “local specificity” of contacts. The higher the value of  $p^C$ , the more local, or hub based, are the individual agents' contacts. As the value of  $p^C$  decreases, the contacts of the agents in different regions are increasingly inter-connected and the individual contacts of agents become more global in nature. It should be noted that our contact network model is a version of the well-known stochastic block model—see Abbe (2018) for a recent and comprehensive survey of stochastic block models.

In Fig. 1, we visualize the contacts within our model of a sample population of 1000 agents who are distributed among 10 exclusive regions (hubs)—100 agents per region—when  $p^C$  takes on the values from  $\{1, 0.995, 0.99, 0.98, 0.97, 0.95\}$ . When  $p^C = 1$ , the regions are entirely self-contained—the agents interact with only a subset of those who are in the same region: An outbreak occurring in one region cannot spread to another region. As  $p^C$  is lowered, inter-regional interactions replace intra-regional interactions at an increasing rate: The likelihood of a disease spreading to other regions starts to go up. As  $p^C$  decreases one expects the possibility of global epidemics to emerge because of increasing inter-regional connections.

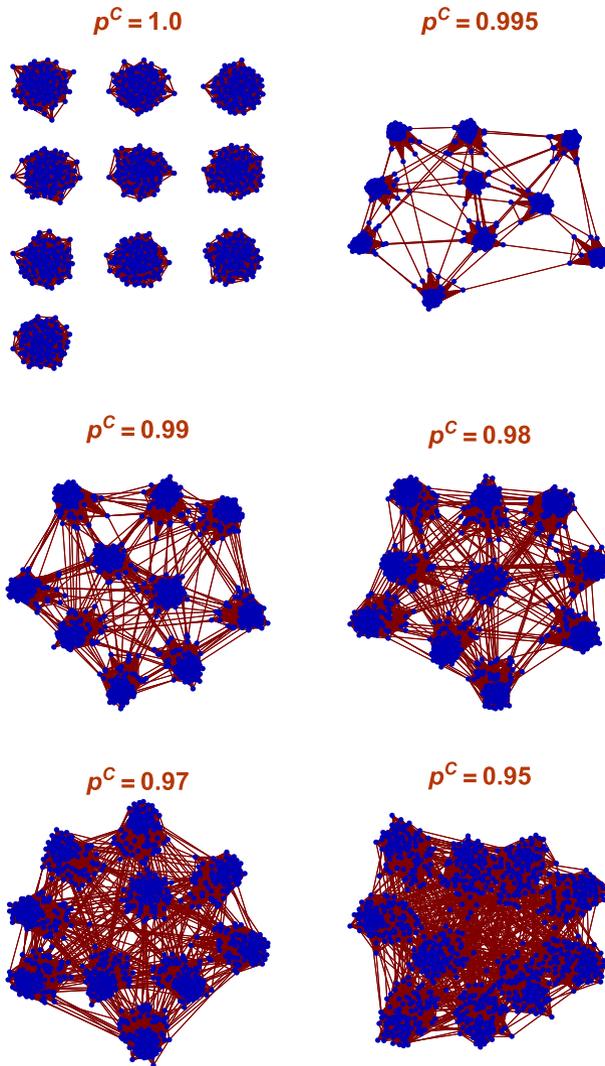


Fig. 1 Visualization of contact links for a population of 1000 agents in 10 regions

### 3.3 Spatial Infection Dynamics

We model a standard Susceptible Exposed Infectious Recovered (SEIR) epidemic process of which seasonal influenza is an example. When a susceptible agent contacts an infectious agent, the susceptible agent can transition to the exposed state with a given probability (specified below). Once in the exposed state the agent then transitions to the infectious state and then to the recovered state in a fixed number of periods. Once recovered the agent is immune to future infection for the current season. We assume that some agents are effectively vaccinated at the beginning of a season. Once

vaccinated the agent immediately moves to the recovered (immune) state. The details of this process are discussed below.

The disease spreads in this population in seasonal epidemics, each season being indexed by  $s$ . A season begins with a new strain of virus that is exogenously introduced into the population. Within each season, the agents interact with one another each period through their contacts, where a period is equivalent to a day. We assume that all agents are exogenously assigned a vaccination status from  $\{vaccinated, unvaccinated\}$  at the start of each season,  $t = 0$ . Specifically, each region has a fixed rate of vaccination,  $\pi_k^V$ , such that  $\pi_k^V * |\Gamma_k|$  agents in that region are vaccinated in the given season on average. For the experiments performed in this paper, we hold  $\pi_k^V$  fixed for all seasons, within a given experiment. How  $\pi_k^V$  is specified across regions is central to our investigation.

**(A.1)** Each agent in region  $k$  is vaccinated with a fixed probability of  $\pi_k^V$ .  $\pi_k^V$  is specific to each region and can be heterogeneous across regions.

Given their vaccination status, the agents engage in daily interactions with their contacts from  $t = 1$  to the end of the season. The interactions are such that each agent  $i$  is in direct contact with all  $j \in C(i)$  each period.

As mentioned above, the timeline of infection in each season follows the basic *SEIR* model given the vaccination status of all agents at the beginning of the season. More specifically, in each time period of a given season, the state of an agent is captured by his vaccination status,  $v_i^t(s)$ , and his infection status,  $f_i^t(s)$ . The vaccination status of an agent can be either *vaccinated* ( $V$ ) or *unvaccinated* ( $NV$ ). The infection status can take one of the following four states: (1) *susceptible* ( $S$ ); (2) *exposed* ( $E$ )—*infected but not yet infectious*; (3) *infected and infectious* ( $I$ ); (4) *recovered and immune* ( $R$ ). An agent can be in the exposed state for  $\tau^E$  periods before becoming infectious. Once an agent becomes infectious, he remains in that state for  $\tau^I$  periods before recovering and becoming immune to the current epidemic. Both  $\tau^E$  and  $\tau^I$  are common for all agents and remain constant over time and across all seasons.

There are five mutually exclusive states that an agent can be in at any time  $t$  of a given season  $s$ :

$$(v_i^t(s), f_i^t(s)) = \begin{cases} (V, R) & : \text{Vaccinated and immune;} \\ (NV, S) & : \text{Unvaccinated and susceptible;} \\ (NV, E) & : \text{Unvaccinated and exposed;} \\ (NV, I) & : \text{Unvaccinated and infectious;} \\ (NV, R) & : \text{Unvaccinated and recovered;} \end{cases} \tag{1}$$

Note that the infection state of a vaccinated agent is denoted  $R$ , which is the same as the agent who is recovered from an infection. We are assuming that vaccination is fully effective so that a vaccinated individual is immune to the virus for the entire season.

The state of an agent changes from one period to the next based on his interactions with the agent's contacts. As stated earlier, an agent  $i$  comes into close contact with *all* agents in  $C(i)$  each period. If the agent is susceptible, he can get infected through his contact with an infectious agent. However, transmission of the virus is probabilistic

so that a susceptible agent, when he comes into contact with an infectious agent, gets infected with a probability  $\beta$ , which is fixed and identical for all agents.

- (A.2) A susceptible agent has a fixed probability  $\beta$  of transitioning to the exposed state when he interacts with an infectious agent.

When there are  $n$  infectious agents in  $C(i)$ , the likelihood of a susceptible agent  $i$  getting infected through interaction with them is:

$$q(n) = 1 - (1 - \beta)^n. \tag{2}$$

Note that  $q'(n) > 0$ . Hence, the larger the number of infectious individuals among one's contacts, the higher is the probability that a susceptible individual will be infected.

The epidemic unfolds as each agent in the population updates her infection state over time. We assume that the interactions with the close contacts occur simultaneously for all agents and the infection states are updated simultaneously at the end of each period for all agents. Given that at any time  $t$  the state of an agent can be in any one of the five states as described in (1), the transition dynamic for each individual from period  $t$  to  $t + 1$  can now be described. Denote by  $n_i^t(s)$  the number of 'infectious' agents among  $i$ 's contacts,  $C(i)$ , in period  $t$  of a given season,  $s$ . If a susceptible agent gets infected through his contact with  $n_i^t(s)$  infectious agents in  $t$ , which happens with a probability of  $q(n_i^t(s))$  as specified in (2), he becomes exposed ( $E$ ) in  $t + 1$ . Denote by  $\widehat{t}_i(s)$  the first period his infection state became  $E$ . The state transition from  $t$  to  $t + 1$  takes the following form:

- For all  $i$  with  $(v_i^t(s), f_i^t(s)) = (V, R)$ ,

$$(v_i^{t+1}(s), f_i^{t+1}(s)) = (V, R) \forall t \text{ in season } s. \tag{3}$$

Those agents who were vaccinated in  $t = 0$  become immune to the virus and remain protected for the entire season.

- For all  $i$  with  $(v_i^t(s), f_i^t(s)) = (NV, S)$ ,

$$(v_i^{t+1}(s), f_i^{t+1}(s)) = \begin{cases} (NV, E) & \text{with probability } q(n_i^t(s)) \\ (NV, S) & \text{with probability } 1 - q(n_i^t(s)) \end{cases} \tag{4}$$

An unvaccinated agent who is susceptible gets infected and enters the *exposed* state with a probability,  $q(n_i^t(s))$ , where  $n_i^t(s)$  is the number of agents in  $C(i)$  who are infectious in  $t$ . With probability  $1 - q(n_i^t(s))$ , the agent avoids infection and stays susceptible.

- For all  $i$  with  $(v_i^t(s), f_i^t(s)) = (NV, E)$ ,

$$(v_i^{t+1}(s), f_i^{t+1}(s)) = \begin{cases} (NV, E) & \text{if } t \leq \widehat{t}_i(s) + \tau^E; \\ (NV, I) & \text{if } t > \widehat{t}_i(s) + \tau^E. \end{cases} \tag{5}$$

**Table 3** State transition rule

$(v_i^t(s), f_i^t(s))$	$(v_i^{t+1}(s), f_i^{t+1}(s))$
$(V, R)$	$(V, R)$
$(NV, S)$	$\begin{cases} (NV, E), \text{ with probability } q(n_i^t(s)), \\ (NV, S), \text{ with probability } 1 - q(n_i^t(s)). \end{cases}$
$(NV, E)$	$\begin{cases} (NV, E), \text{ if } t \leq \widehat{t}_i(s) + \tau^E, \\ (NV, I), \text{ if } t > \widehat{t}_i(s) + \tau^E. \end{cases}$
$(NV, I)$	$\begin{cases} (NV, I), \text{ if } \widehat{t}_i(s) + \tau^E < t \leq \widehat{t}_i(s) + \tau^E + \tau^I \\ (NV, R), \text{ if } t > \widehat{t}_i(s) + \tau^E + \tau^I. \end{cases}$
$(NV, R)$	$(NV, R)$

An agent who is infected but not yet infectious ( $E$ ) remains in that state for  $\tau^E$  periods after the date of infection,  $\widehat{t}_i(s)$ . Once past the  $\tau^E$  periods, he becomes infectious ( $I$ ) himself.

- For all  $i$  with  $(v_i^t(s), f_i^t(s)) = (NV, I)$ ,

$$(v_i^{t+1}(s), f_i^{t+1}(s)) = \begin{cases} (NV, I) & \text{if } \widehat{t}_i(s) + \tau^E < t \leq \widehat{t}_i(s) + \tau^E + \tau^I \\ (NV, R) & \text{if } t > \widehat{t}_i(s) + \tau^E + \tau^I \end{cases} \quad (6)$$

An agent who was infected at  $\widehat{t}_i(s)$  gets infectious starting  $\widehat{t}_i(s) + \tau^E + 1$ . The state of  $I$  lasts for  $\tau^I$  periods—i.e., from  $\widehat{t}_i(s) + \tau^E + 1$  to  $\widehat{t}_i(s) + \tau^E + \tau^I$ —at the end of which the infected agent recovers and becomes immune to the virus for the remainder of the season.

- For all  $i$  with  $(v_i^t(s), f_i^t(s)) = (NV, R)$ ,

$$(v_i^{t+1}(s), f_i^{t+1}(s)) = (NV, R). \quad (7)$$

A recovered agent remains in that state for the remainder of the season.

The state transition dynamic is summarized in Table 3. Note that both  $(V, R)$  and  $(NV, R)$  are absorbing states, since neither a vaccinated agent nor an infected agent who recovers is ever infected again during that season.

To complete the description of the dynamic, we need to specify the condition under which the epidemic can start and the condition under which it terminates. First, to initiate the contagion dynamic, we seed the population with a small number of infected agents at the beginning of each season. This is done by randomly selecting a fixed number,  $m$ , of non-vaccinated/susceptible agents and artificially infecting them at the start of each season. These *seed agents* then have their state set to  $(NV, E)$  at  $t = 1$ . Note that we are careful to set this number to a very low level so that the epidemic dynamics are driven by the interactions of agents in the model. This is meant to mimic the manner in which new influenza epidemics in the U.S. enter from outside at infrequent rates and then run an epidemic course during the given season based on the prevailing vaccination states of the agents and the mutual interactions among them.

Finally, the seasonal epidemic ends when there is no *infected* agent in the population—i.e., no one is in the state of  $(NV, E)$  or  $(NV, I)$ . This is an absorbing state for the population, since, once it reaches this state, there is no one who can pass on the virus to others either in the current period or in the future periods: The particular epidemic has come to an end until the next season starts with a new strain of virus.

## 4 Design of Computational Experiments

### 4.1 Overview

We simulate a sequence of repeated ‘epidemic seasons,’ with the total number of seasons being  $\tilde{s}$ . For each season, we start with the vaccination status exogenously assigned in  $t = 0$  for all agents in the population for the current strain of virus. This gives us  $v_i^1(s)$  for all  $i \in Z$  in the given season  $s$ . All those agents who are vaccinated,  $v_i^1(s) = V$ , gain immunity to the current virus such that  $f_i^1(s) = R$ .<sup>5</sup> Those agents who are unvaccinated,  $v_i^1(s) = NV$ , become susceptible such that  $f_i^1(s) = S$ . We randomly pick a fixed number of these unvaccinated agents as seed agents and artificially infect them with the virus—i.e., for each seed agent,  $k$ , we specify  $(v_k^1(s), f_k^1(s)) = (NV, E)$ , setting  $\hat{t}_k(s) = 1$  (the date of infection for the seed agent  $k$ ).

Given  $(v_i^1(s), f_i^1(s))$  for all  $i \in Z$ , we follow the previously specified state transition dynamic from  $t = 1$  and on, keeping track of  $(v_i^t(s), f_i^t(s))$  for all agents each period. The epidemic continues as long as there are individuals who are infected—i.e., those unvaccinated agents who are either exposed or infectious. The epidemic ends when no one in the population is in the state of exposed ( $E$ ) or infectious ( $I$ ) such that the population contains only susceptible ( $S$ ) and/or immune ( $R$ ) individuals— $f_i^t(s) \in \{S, R\}$  for all  $i \in Z$ . Denote by  $\hat{T}(s)$  the earliest period that no infected individual is observed. Once the population reaches  $\hat{T}(s)$ , there is no possibility of further contagion; the epidemic season has officially come to a close. A new season,  $s + 1$ , starts with a new strain of virus, where the population goes through the same sequence of events until it reaches  $\hat{T}(s + 1)$ .

For each season, the epidemic is the chain of infections by the individuals in the population. The ‘size’ of an epidemic is indicated by the total number of people who were infected over the course of the epidemic. The simulation outputs on the epidemic sizes are collected for all  $s \in \{1, \dots, \tilde{s}\}$ . It is the distribution of these outputs over the entire  $\tilde{s}$  seasons that we analyze in this paper.

<sup>5</sup> Note that most actual vaccines are not 100% effective as we model them here. As an alternative representation of our model, without loss of generality, one can envision our vaccination coverage parameter as the vaccine effectiveness that results from the number of vaccinations actually performed.

**Table 4** Parameter specification

Parameter	Description	Values
$z$	Population size	10,000
$H$	Number of activity hubs	100
$p^C$	Share of contacts from inside the agent's activity hub	{1, 0.975, 0.95, 0.9, 0.8}
$1 - p^C$	Share of contacts from outside the agent's activity hub	{0, 0.025, 0.05, 0.1, 0.2}
$\bar{c}$	Mean size of a contact network	20
$\tau^E$	Length of the latent periods	2
$\tau^I$	Length of the infectious periods	7
$\beta$	Infection probability	0.02
$m$	Number of the susceptible agents infected in $t = 0$	10
$\tilde{s}$	Number of seasons (replications)	100
$\pi_k^V$	Region-specific vaccination rate	–

## 4.2 Parameter Specification

The relevant parameters and their values used in our baseline computational experiments are listed in Table 4. Our computational experiments assume a population of 10,000 agents, distributed over 100 activity hubs of equal size so that each hub has exactly 100 agents.

The contacts of each agent consist of 20 other agents on average, where the contacts are chosen through the sequential process described. The matching process terminates when the average contacts across all agents in the population reaches 20.<sup>6</sup> The contacts are constructed in a structured way such that each agent has  $p^C$  fraction of his contacts composed of individuals from his own activity hub and  $(1 - p^C)$  fraction from outside the hub. We carry out a comparative dynamics analysis with respect to the fraction of cross-regional connections by considering multiple values for  $p^C$  as specified in Table 4, ranging from 1.0 to 0.8.

For the infection dynamic, an infected person remains in an exposed (latent) state for two periods (days) before becoming infectious –  $\tau^E = 2$ . Once an agent becomes infectious, he remains in that state for seven periods (days) before recovering –  $\tau^I = 7$ . A susceptible person has 0.02 ( $\beta$ ) probability of becoming infected when he comes into contact with an infectious individual.<sup>7</sup> Finally, we consider 100 independent seasons and assume that each epidemic is initiated by ten randomly chosen seed agents for each parameter setting.

The region-specific vaccination rate,  $\pi_k^V$ , takes on values according to the assumptions made for each set of experiments carried out in this study. The exact assignment

<sup>6</sup> A typical set of contacts for the population has sizes that range between 7 and 36 with its mean at 20 and a standard deviation of about 4.

<sup>7</sup> Parameters for periods of latency and infectiousness are chosen to be reasonable approximations for seasonal influenza. See: CDC (2017d): <https://www.cdc.gov/flu/professionals/acip/clinical.htm>. Infection probability is chosen to yield reasonable sized epidemics at levels of influenza vaccinations observed in the U.S. and to yield levels of herd immunity consistent with data, see Plans-Rubio (2012).

of their values will be discussed in detail as we describe these experiments in the sections that follow.

### 4.3 Vaccination Rates

As mentioned previously, the vaccination status of each agent is determined exogenously in this paper. The main focus of our study is to explore the impact that the spatial pattern of vaccination has on the size of the epidemics, both at the global and local levels, as well as how this relationship is affected by the fraction of cross-regional contacts.

There are three sets of experiments we perform to achieve these objectives:

*Experiment 1 (Homogeneous vaccination coverage):* The entire population has the same rate of vaccination. Hence, there is no inter-regional heterogeneity in vaccination. We explore how the uniform vaccination rate affects the size of the epidemics. This special case will provide the first benchmark against which the outputs from our main two experiments can be compared.

*Experiment 2 (Heterogeneous vaccination coverage with two levels of vaccination coverage):* The population regions are divided into two equal-sized groups: one set of regions with a high vaccination rate and another with a low vaccination rate.

*Experiment 3 (Heterogeneous vaccination coverage with and without the high to low cross-regional contacts):* We explore the effect that the presence of high to low region contacts has on the externalities that arise between regions.

In the following sections, we describe these experiments in detail and report the results obtained. In all experiments, we run the epidemic process for 100 independent seasons and report the distributional characteristics of the observed epidemics over those 100 realizations of the model.

### 4.4 Herd Immunity

Central to an understanding of these epidemic processes is the concept of “herd immunity.” Herd immunity occurs when a sufficient percentage of the population is vaccinated (or naturally immune) such that the epidemic does not spread because there are not enough susceptible individuals to sustain the epidemic (Anderson and May 1985). Theoretically this occurs when an average person in the population is unable, on average, to infect more than one additional person during the time she spends in the infectious state. If the population has randomly mixed contacts, one can find the level of vaccinations where this occurs as the value  $V^h$  such that the product of the number of contacts, transmission rate, the time spent infectious, and the fraction of susceptible individuals in the population is equal to one. In terms of the notation of this paper this is,  $\bar{c}\beta\tau^I(1 - V^h) = 1$ . Thus the level of vaccinations that produces herd immunity is:  $V^h = 1 - \frac{1}{\bar{c}\beta\tau^I}$  or greater. For the parameters of this paper, it is about 64% of the population.<sup>8</sup> Of course our model will deviate somewhat from this value

<sup>8</sup> The specific level of herd immunity depends on interaction patterns, vaccine effectiveness, spatial patterns of vaccination and other variables. Additionally, as pointed out by an anonymous reviewer, because the

because the contact structure in our model is not randomly mixed, but instead contains an explicit non-random structure due to the hubs and the parameter,  $p^C$ . However this will give us some understanding of the range of values of  $\pi_k^V$  which will produce reasonable sized epidemics.

## 5 Results

### 5.1 Experiment 1: Perfectly Uniform Vaccination Coverage

As a starting point, we first consider a simple experiment in which the entire population has the uniform rate of vaccination such that  $\pi_k^V = \pi^V$  for all  $k \in \{1, 2, \dots, H\}$ . Hence, there are no regional differences in the proportion of individuals who are vaccinated. A comprehensive set of values are considered for  $\pi^V$  in this experiment, where  $\pi^V \in \{0, 0.1, 0.2, \dots, 0.9\}$ . In  $t = 0$  of each season,  $(\pi^V * z)$  agents are randomly selected from the population and vaccinated. Of the remaining population of unvaccinated agents, ten are randomly selected as the seed agents and artificially infected in  $t = 1$  to start the epidemic process. This first experiment is designed in order to create a baseline expectation for epidemic size as a function of vaccine coverage and to show that the model generates standard epidemic features such as the “reverse S” relationship between vaccine coverage and epidemic size.

Figure 2 presents three typical epidemics for three different values of  $\pi^V$ . The  $\pi^V = 0.2$  case, shown in the top panel, demonstrates the canonical shape for an SEIR epidemic. In this case there is a peak period around time 80 after which the epidemic begins to decline before eventually disappearing. Note the following: (1) the peak of the epidemic has far fewer people than are infected overall. About 1000 people are infected in the peak period of the epidemic but over 6700 people are infected over the entire course of the epidemic from beginning to end. (2) The “S” shape of the recovered population and reverse “S” shape of the susceptible population are typical of this type of epidemic process. (3) The 2nd and 3rd panels of the figure display the same general relationships but the epidemics are much smaller. Thus the typical shape is somewhat masked.

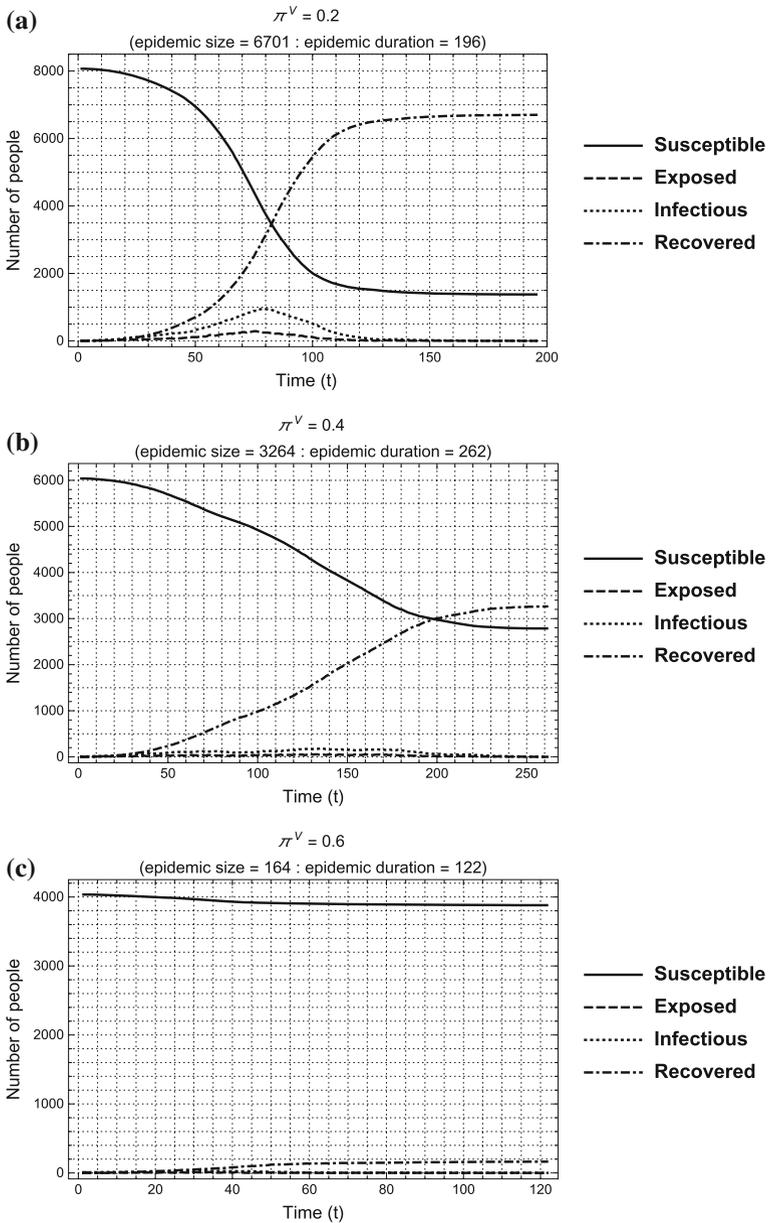
Table 5 reports the size of the epidemics averaged over the 100 seasons for all values of  $\pi^V$  for each  $p^C \in \{1, 0.975, 0.95, 0.9, 0.8\}$ . Additionally, we report at the bottom of the table three measures that characterize the contact structure for each case: (1) mean path length ( $PL$ ); (2) mean global clustering coefficient ( $GCC$ ); and (3) mean local clustering coefficient ( $LCC$ ). The numbers reported for the three measures are the averages over the 100 independent seasons (replications).

The “mean path length” is the average length of the shortest paths for all possible pairs of agents in the population, where the “length of a path” is defined as the number of edges that the path contains (Jackson 2008). If agents are closer to each other, on average, one should expect an epidemic to be larger because it will more easily

---

Footnote 8 continued

methods we use to create our network result in an increase in clustering but not an increase in degree distribution variance, herd immunity in our model is likely less than 64%. However, 64% is in the plausible range of estimates of influenza herd immunity in the U.S. (Plans-Rubio 2012).



**Fig. 2** Time paths of a typical epidemic when  $p^C = 0.95$

spread to others in the population. The contact structure may also be characterized by the global and local clustering coefficients, both of which measure the extent to which one's contacts are also in direct contact with one another. More specifically, the global clustering coefficient ( $GCC$ ) is obtained by dividing the number of closed

**Table 5** Epidemics with uniform vaccination coverage

$\pi^V$	Mean (SD) of epidemic sizes (100 seasons)				
	$p^C = 1$	$p^C = 0.975$	$p^C = 0.95$	$p^C = 0.9$	$p^C = 0.8$
0	810.02 (96.95)	9095.91 (54.21)	9109.54 (39.44)	9112.95 (37.02)	9110.76 (39.15)
0.1	668.06 (93.60)	7781.78 (98.48)	7876.35 (50.92)	7880.58 (53.69)	7876.47 (51.00)
0.2	526.92 (81.11)	6262.17 (186.49)	6559.77 (106.61)	6581.15 (77.99)	6594.25 (82.93)
0.3	379.29 (80.23)	4142.72 (401.08)	5001.79 (151.84)	5172.66 (106.79)	5206.75 (107.49)
0.4	241.73 (63.51)	1168.02 (519.67)	2796.18 (311.52)	3485.69 (166.50)	3690.60 (114.67)
0.5	127.01 (44.39)	208.09 (95.64)	428.21 (297.44)	1220.76 (426.64)	1948.52 (224.81)
0.6	63.45 (20.91)	75.79 (34.29)	79.69 (41.09)	113.65 (67.42)	190.74 (140.86)
0.7	35.93 (12.96)	36.25 (15.00)	36.73 (15.53)	38.28 (16.29)	41.82 (19.25)
0.8	19.94 (6.35)	19.79 (6.51)	20.03 (5.44)	20.25 (5.81)	20.83 (7.64)
0.9	13.33 (2.55)	13.25 (2.24)	13.49 (2.12)	13.62 (2.64)	13.14 (2.70)
$PL^a$	–	5.06	4.52	4.10	3.76
$GCC^b$	0.2019	0.1854	0.1702	0.1426	0.0979
$LCC^c$	0.2019	0.1856	0.1705	0.1432	0.0986

<sup>a</sup>Mean path length

<sup>b</sup>Mean global clustering coefficient

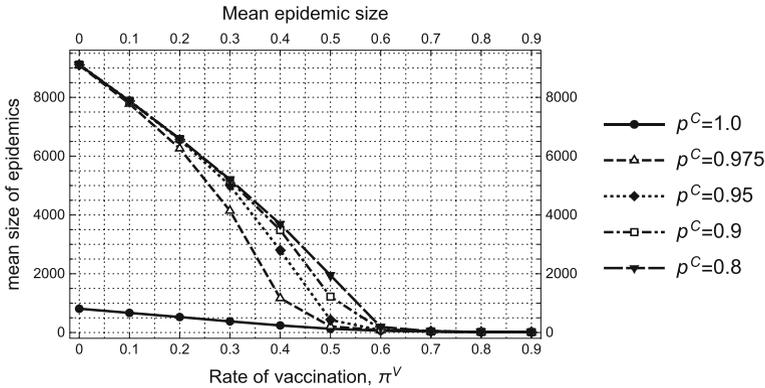
<sup>c</sup>Mean local clustering coefficient

triplets—e.g.,  $i$  is connected to  $j$ ,  $j$  is connected to  $k$ , and  $k$  is connected to  $i$  in a closed loop—by the total number of open and closed triplets in the contact structure. Alternatively, one can measure the clustering coefficient for each individual agent and then take the grand average over the entire population. Denote by  $LCC(i)$  the local clustering coefficient for agent  $i$ , where

$$LCC(i) = \frac{\text{number of pairs of } i\text{'s contacts, } (j, k), \text{ such that } j \in C(k) \text{ and } k \in C(j)}{\text{total number of pairs of } i\text{'s contacts}} \tag{8}$$

Averaging over the population, we obtain the *mean local clustering coefficient*:  $LCC = \sum_{\forall i \in Z} LCC(i)$ . For both clustering coefficients we would expect that the size of the epidemic is likely to be smaller when there is more clustering in the contact structure, since the infectious disease is more likely to be trapped in isolated clusters. As can be seen from Table 5, the effects of  $p^C$  on the three measures of the contact structure are as expected: (1) The mean path length of the global network (consisting of all individual contacts) is shorter when  $p^C$  is lower, ranging from 5.06 for  $p^C = 0.975$  to 3.76 for  $p^C = 0.8$ ;<sup>9</sup> and (2) both measures of the clustering coefficients are smaller when  $p^C$  is lower such that the contacts are more global.

<sup>9</sup> Note that the mean path length is undefined for  $p^C = 1$  as the regions are completely separated with no links between them.



**Fig. 3** Impact of  $\pi^V$  and  $p^C$  on the mean epidemic size

The main results as reported in Table 5 may now be summarized: (1) The mean size of the epidemics decreases with  $\pi^V$ ; and (2) a decline in the local specificity of the contacts ( $p^C$ )—i.e., an increase in the share of inter-regional contacts—raises the mean size of the epidemics, when the rate of vaccination is low enough to induce epidemics of significant size. These results are intuitive and as expected: A higher rate of vaccination reduces the size of the epidemics, while more global contacts (which result in shorter mean path length and less clustered interaction structure) tend to further increase the size of the epidemic when the rate of vaccination is relatively low – when the rate of vaccination is high, the sizes of the epidemics are too small for the resulting comparisons to have any statistical significance. It should be noted that the epidemics are very small for values of  $\pi^V > 0.7$  as expected from herd immunity. It is not until  $\pi^V$  is sufficiently low that the typical epidemic takes off and is governed by the complexities of the contact structure. Furthermore, there are not significant differences in epidemic size, as a function of the contact structure, until vaccination levels drop below 0.7.

For a visual representation, we provide Fig. 3, in which the mean sizes of the epidemics are plotted for all values of  $\pi^V$  and for  $p^C \in \{1, 0.975, 0.95, 0.9, 0.8\}$ . Note that once the vaccination coverage percentage falls below 0.5 the average epidemic can be large. In this figure one also can see that the effect of global contacts is most important for intermediate vaccination rates. If the vaccination rate is sufficiently large (0.7 and above) infections are very close to zero regardless of how large the vaccination rate grows and regardless of how many global interactions take place. At the other extreme, if vaccination rates are very low (0.2 and below), the fraction of global contacts is also relatively unimportant, except for the case where no global interactions are allowed. If even a small number of global interactions are allowed, the epidemic spreads very widely and the average size of the epidemic is determined primarily by the number of vaccinations performed;  $p^C$  has a negligible effect for these very low vaccination levels (provided it is less than 1). Finally, for intermediate levels of vaccinations,  $0.2 < \pi^V < 0.6$ , both the number of global interactions and the vaccination rate have a significant effect on average epidemic size. It is this range of values that we concentrate our efforts on for the remainder of the paper.

We observe the following property from Fig. 3:

**Property 1:** For intermediate levels of uniform vaccination coverage, the average size of an epidemic increases with the share of the inter-regional contacts.

The obvious implication of this property is the increasing importance of vaccinating the population as the contact structure becomes more inter-regional. If globalization in social and economic activities leads to a general increase in the interactions between people of distant regions, an outbreak in one region is likely to have severe and far-reaching consequences; a significant increase in the vaccination effort will be required to keep the epidemic under control. Again, this is an expected feature of an epidemic process.

One important and general feature of epidemiology is the “reverse S” shape in the relationship between vaccination coverage and epidemic size. Note that in Fig. 3, for sufficiently low vaccination levels, an increase in the vaccination rate yields a nearly linear decrease in epidemic size. As the number of vaccinations increases, additional vaccinations yield larger effects - the curve becomes steeper. But, as the population approaches the level of vaccinations that yield herd immunity the curve flattens out again. This general pattern is most apparent visually in the  $p^C = 0.975$  graph, but is also present in the other graphs as well. Note that the inflection point for the  $p^C = 0.975$  graph occurs at about 0.4. This non-linear relationship between the number of vaccinations performed and the marginal effect of additional vaccinations will play a key role in the results that we report later in the paper. We also want to emphasize that the shape of these curves (in terms of the amount of curvature and the position of the inflection point) depend on the interaction structure of the model. Without running the agent-based model we would not be able to accurately identify with any precision these elements of the model.

With our baseline results established, we now consider the effect of the main features of our model. Before we move forward, we stress that the results of Experiment 1 match with expectations for a simple SIR epidemiological model. The results of Experiment 1 are used as a base result for comparison with the following sections of the paper.

## 5.2 Experiment 2: Heterogeneous Vaccination Coverage

As the second experiment, we consider a simple case of spatial heterogeneity where the vaccination rate in each region is set to one of two levels; some regions have a high vaccination rate ( $\pi_H^V$ ) and others have a low vaccination rate ( $\pi_L^V$ ). Because we have 100 regions, each with 100 agents, we assume that the agents in fifty of those regions (5000 agents) are vaccinated at the rate of  $\pi_H^V$  and the remaining agents at the rate of  $\pi_L^V$ .

For a systematic investigation of the impact the spatial variation in vaccination rates has on the infection dynamics, we specify the two rates to be characterized by a mean-preserving spread with a fixed mean at  $\bar{V}$ :

$$\begin{aligned}\pi_H^V &= \bar{V} + \Delta V, \\ \pi_L^V &= \bar{V} - \Delta V.\end{aligned}\tag{9}$$

In order to best understand the effect of spatial heterogeneity we first consider  $\bar{V} = 0.4$  and  $\bar{V} = 0.5$  which correspond to the range of U.S. influenza vaccination levels in recent years according to CDC: “Among all people  $\geq 6$  months, flu vaccination coverage during the 2015-16 flu season was 45.6%, which was a 1.5% decrease compared with the 2014-20-15 season (47.1%).”<sup>10</sup> We also consider two higher levels that are just below and just above levels that provide herd immunity,  $\bar{V} = 0.6$  and  $\bar{V} = 0.7$ . After establishing the property for sufficiently high levels of vaccination,  $\bar{V} \geq 0.4$ , we then consider  $\bar{V} = 0.3$ , which is a level significantly below the level providing herd immunity, to show that the property can be reversed when the rates of vaccination are sufficiently low.

We consider four values for the differential:  $\Delta V \in \{0, 0.05, 0.1, 0.2\}$ . For instance, when the global mean rate is 0.4 and  $\Delta V = 0.2$ , one region has 0.6 vaccination rate and the other has 0.2 vaccination rate. Of course, the case of  $\Delta V = 0$  corresponds to the previously considered case of uniform fixed rate of vaccination where  $\pi^V = 0.4$  or 0.5 for the entire population. Again, we have chosen values for  $\Delta V$  that are in line with observed data. Recall, that the differences between the maximum and minimum zip code level flu claims data were a little over 40% (two-times our maximum value for  $\Delta V$ ) for the cities listed.

The sizes of the epidemics generated from 100 independent seasons were generated for various levels of cross-regional contacts,  $p^C \in \{1, 0.975, 0.95, 0.9, 0.8\}$ . The mean and the standard deviation of the corresponding outputs are reported in Table 6 for each level of  $\bar{V} \in \{0.4, 0.5, 0.6, 0.7\}$ . Most strikingly, the mean epidemic size uniformly increases in  $\Delta V$  for all values of  $p^C$  for  $\bar{V} \geq 0.4$ . In addition, we find that Property 1 is robust to allowing for spatial variation in vaccination rates: For all values of  $\Delta V$  a decrease in  $p^C$  (which implies more cross-regional linkages in contacts) raises the mean size of the epidemics. In sum, the epidemics can be quite large when there are cross-regional variations in the vaccinations rates, and increasing the number of cross-regional linkages further amplifies this effect.

**Property 2:** For sufficiently high rates of vaccination (below the level providing herd immunity), the size of an epidemic, on average, increases as the variation in the rates of vaccination increases between the two regions.

To best understand this result, again return to the earlier discussion of Fig. 3. Specifically, recall the shape of the relationship between vaccination coverage and epidemic size. For all values of  $\bar{V} \geq 0.4$ , an increase in vaccination rate is pushing the population into an area of the curve where there is decreasing marginal effect of additional vaccinations. In other words, in this area of the curve, additional vaccination coverage above 0.4 produces a smaller change in epidemic size than do decreases in vaccinations below 0.4. Thus the increase in vaccination coverage in the high vaccination regions, does less good than the harm done by lowering vaccination coverage in the

<sup>10</sup> <https://www.cdc.gov/flu/pdf/fluavaxview/2015-16/nfid-coverage-2015-16-final.pdf>. Also see CDC (2017b).

**Table 6** Epidemics with spatial heterogeneity in vaccination coverage case of two regions with  $\pi_H^V (= \bar{V} + \Delta V)$  and  $\pi_L^V (= \bar{V} - \Delta V)$

$p^C$	Mean (SD) of epidemic sizes (100 seasons)			
	$\Delta V = 0$	$\Delta V = 0.05$	$\Delta V = 0.1$	$\Delta V = 0.2$
(a) $\bar{V} = 0.4$				
1.0	241.73 (63.51)	251.39 (68.65)	270.10 (84.51)	369.67 (91.90)
0.975	1168.02 (519.67)	1220.23 (582.28)	1562.48 (618.82)	2562.67 (547.49)
0.95	2796.18 (311.52)	2930.31 (245.89)	3074.23 (253.67)	3588.12 (139.36)
0.9	3485.69 (166.50)	3534.20 (158.55)	3589.99 (131.49)	3872.58 (94.13)
0.8	3690.60 (114.67)	3706.94 (129.73)	3772.19 (109.99)	4003.77 (87.17)
(b) $\bar{V} = 0.5$				
1.0	127.01 (44.39)	139.64 (51.70)	168.35 (62.34)	277.81 (67.21)
0.975	208.09 (95.64)	251.23 (141.34)	326.97 (206.38)	942.89 (400.95)
0.95	428.21 (297.44)	470.94 (308.64)	772.94 (379.83)	2089.94 (371.09)
0.9	1220.76 (426.64)	1361.79 (480.59)	1740.53 (412.10)	2627.97 (144.20)
0.8	1948.52 (224.81)	1987.92 (272.17)	2206.17 (159.77)	2746.94 (109.47)
(c) $\bar{V} = 0.6$				
1.0	63.45 (20.91)	76.07 (28.52)	93.96 (41.27)	180.85 (68.64)
0.975	75.79 (34.29)	89.37 (41.19)	126.74 (65.44)	297.94 (157.18)
0.95	79.69 (41.09)	105.35 (54.31)	148.62 (82.51)	528.81 (302.86)
0.9	113.65 (67.42)	141.80 (88.37)	237.85 (161.56)	1144.63 (397.51)
0.8	190.74 (140.86)	235.22 (176.69)	451.11 (281.17)	1502.12 (168.83)
(d) $\bar{V} = 0.7$				
1.0	35.93 (12.96)	39.36 (16.05)	50.96 (20.83)	112.68 (41.82)
0.975	36.25 (15.00)	39.61 (18.59)	54.63 (26.40)	132.81 (60.09)
0.95	36.73 (15.53)	41.39 (18.41)	55.36 (24.43)	154.80 (81.56)
0.9	38.28 (16.29)	42.30 (20.20)	57.07 (33.90)	209.05 (119.49)
0.8	41.82 (19.25)	49.33 (29.41)	63.91 (40.06)	307.86 (191.63)

low vaccination regions where there is a larger marginal effect. Thus, once the mean vaccination coverage level is past the inflection point on Fig. 3, an increase in heterogeneity in regional vaccination coverage should produce an increase in mean epidemic size. This is what we observe in the computational experiment.

This effect, however, is not uniform across all vaccination levels. Note that the inflection point of the curve and the rates of marginal effect of vaccination (the slope of the curves in Fig. 3) are determined by the interaction structures of the population. Thus we see different magnitudes for the effects depending on values of  $p^C$  and different base vaccination rates. That the property identified above may be reversed for sufficiently low values of  $\bar{V}$  is then shown in Table 7, where the epidemic sizes for  $\bar{V} = 0.3$  are reported for the same set of parameter configurations as in Table 6. Take the case of  $\bar{V} = 0.3$ ,  $p^C = 0.975$ , and  $\Delta V = 0.05$ . For these parameters, the average epidemic size is smaller than when the vaccination coverage is uniformly

**Table 7** Epidemics with spatial heterogeneity in vaccination coverage case of two regions with  $\pi_H^V (= \bar{V} + \Delta V)$  and  $\pi_L^V (= \bar{V} - \Delta V)$  for  $\bar{V} = 0.3$

$p^C$	Mean (SD) of epidemic sizes (100 seasons)			
	$\Delta V = 0$	$\Delta V = 0.05$	$\Delta V = 0.1$	$\Delta V = 0.2$
1.0	379.29 (80.23)	395.43 (81.29)	413.74 (84.91)	474.60 (116.27)
0.975	4142.72 (401.08)	4014.39 (391.35)	4120.63 (365.04)	4332.80 (227.67)
0.95	5001.79 (151.84)	5002.40 (152.05)	4953.34 (171.96)	4938.12 (138.53)
0.9	5172.66 (106.79)	5167.57 (102.33)	5160.25 (102.56)	5180.00 (85.40)
0.8	5206.75 (107.49)	5216.15 (86.96)	5232.27 (100.28)	5316.26 (73.87)

0.3 across all regions. In other words, for these parameters, spatial heterogeneity in vaccination coverage *decreases* epidemic size. One can understand this result if we again return to Fig. 3 for  $p^C = 0.975$ . Notice that at vaccination coverage of 0.3 the curve is still becoming steeper—i.e., the curve is concave from below for this range of vaccination rates. As vaccinations are increased, there is a larger marginal effect of the additional vaccinations. Thus, when vaccination heterogeneity is introduced, the effect of increasing vaccinations in some regions out-weighs the effect of decreasing the vaccinations in other regions, all else equal. As seen in this example, it is possible that *increasing the spatial heterogeneity in vaccination coverage can decrease expected epidemic size when vaccine coverage is sufficiently below herd immunity*. For the other parameter values in this range, note that there are several other combinations that result in smaller epidemic size and also some that result in increases in epidemic size. Also, for some of the combinations the effect is relatively small compared to other changes observed in Table 7. The smallest changes occur because the relationship between epidemic size and vaccine coverage is relatively linear for these parameter combinations. Again, the various different effects are due to the different interaction structures creating different relationships between vaccine coverage and epidemic size.

For the remainder of the paper, we will focus on the range of sufficiently high vaccinations rates,  $\bar{V} \geq 0.4$ , which contains the parameter values that are consistent with the actual U.S. influenza vaccination levels. Going back to the results in Table 6, one can compare the effect of increasing spatial heterogeneity in vaccination coverage and changes in overall vaccination rates. As an example, consider results for  $p^C = 0.95$ . Recall that a value of  $\bar{V} = 0.7$  is above the level which provides herd immunity when vaccination coverage is uniform; no significant epidemic occurs. With uniform coverage, the average epidemic for these parameters yields only 36.73 infections. But with  $\Delta V = 0.2$ , the average epidemic size jumps to 154.80, over a four-fold increase. Also note that this epidemic is nearly two-times as large as the corresponding epidemic with  $\bar{V} = 0.6$  and uniform vaccination. The effect of this increase in spatial variation is much larger than a 10% decrease in vaccination rates. Further, this demonstrates that the spatial heterogeneity of vaccination coverage can greatly increase the overall level of vaccinations needed to produce herd immunity. Negligible epidemics with spatially uniform coverage can become large epidemics with spatially heterogeneous coverage.

Taking these results together, one can consider when spatial heterogeneity in vaccination coverage can be most important from a policy perspective. The changes in expected epidemic size and the effect on herd immunity are largest when the vaccine coverage in the overall population is close to herd immunity. Thus if one considers the effects described above, it is most important to think about heterogeneity in coverage for an infectious disease like measles where we tend to be close to herd immunity. The effects of heterogeneity here, could be very large.

### 5.2.1 Experiment 3: Externalities Arising from Spatial Heterogeneity

Note that the population in experiment 2 consists of two groups of equal size. Given the discussion above, it may seem that the mean epidemic size of the overall population can simply be inferred from the convex combination of the two mean values of the epidemic sizes corresponding to the respective rates of vaccination. This perspective would be conceptually appropriate if individuals in our model population limited their contacts to only those others who come from the regions having the same rate of vaccination; hence, the convex combination of the two uniform rate cases. However, the relationship is more complex. The previously described perspective neglects the fact that in our model the two groups with differential vaccination rates can also directly interact with one another across the regional border, where the cross-regional contacts exist at the rate of  $1 - p^C$ . These cross-regional contacts create deviations from a convex combination prediction and also create difficulties in finding a closed form solution for expected epidemic size. These cross-regional interactions between the high and low vaccination areas can have a differential impact on the risks of infection for the two groups of individuals; hence, ultimately influencing the size of the global epidemic.

In this section, we attempt to isolate the effect of the cross-regional interactions in our model and evaluate the benefits and costs to the two types of agents—i.e., those living in high-vaccination areas versus those living in low-vaccination areas—from such interactions. To address this issue, we perform the following computational experiment. Suppose the agents in a high-vaccination area interact only with other agents who also live in a high-vaccination area and the same for those in the low-vaccination area.

This what-if scenario can be implemented in our model by specifying the contacts of the agents in such a way that the cross-regional contacts for each agent are restricted to only those agents in the regions that belong to the set of regions with the same vaccination rate. For example, an agent who belongs to a high-vaccination region will have  $p^C$  share of his contacts from the same region/hub (which, of course, will have a high rate of vaccination) and  $(1 - p^C)$  share of contacts from those regions outside of his own but still having the same high rate of vaccination. Let us call this the case of “segregated contacts” and the baseline case we examined previously the case of “mixed contacts.”<sup>11</sup>

<sup>11</sup> We note that this type of social engineering of interactions is unlikely to be executed in practice by policy makers. It is possible that engineering like this could occur by schools choosing to place children in classrooms according to vaccine status or cities could vary bus routes dependent on knowledge of vaccine

Our question can now be answered by looking at the mean epidemic size (mean number of infections) in the high vaccination areas compared to that in the low vaccination areas under the two different types of contact mixing. The results from the experiment are presented in Table 8 for  $\bar{V} \in \{0.4, 0.5, 0.6, 0.7\}$ , when  $p^C = 0.95$ . For each value of  $\bar{V}$ , we consider  $\Delta V \in \{0.05, 0.1, 0.2\}$ . Note that these values of  $\bar{V}$  represent those cases where increased spatial heterogeneity in vaccine coverage leads to increased expected epidemic size. The results are organized as follows. Given a value of  $\bar{V}$ , we consider the three values of  $\Delta V$  in the column specific to the given  $\bar{V}$ . For each case, we report the within-group epidemic size—i.e., the number of infected individuals within each group (low-vaccination or high-vaccination group)—under the two scenarios, segregated vs. mixed population. To clarify the results in Table 8, for each parameter set, we display the changes in infections for each group and for the total population in Table 9. For example, when moving from segregated to mixed contacts at  $\bar{V} = 0.4$ ,  $\Delta V = 0.05$ , infections in the low vaccination regions drop from 1963.33 to 1816.41 (reported on Table 8) a change of  $-146.92$  (reported on Table 9).

We separate our discussion of these results into two parts. Initially we discuss the  $\bar{V} = 0.4$  and  $\bar{V} = 0.5$  cases. Then we discuss the cases with larger overall vaccination rates. We make two observations from the  $\bar{V} = 0.4$  and  $\bar{V} = 0.5$  experiments. First, when the contacts are “mixed,” the mean number of infections is lower for the agents in the low-vaccination group (low-vaxx) and higher for those in the high-vaccination group (high-vaxx), compared to when the contacts are “segregated.” For instance, for  $\bar{V} = 0.4$  and  $\Delta V = 0.1$  the epidemic size for the low-vaccination group goes down from 2508.76 to 2256.71, while the epidemic size for the high-vaccination group goes up from 177.80 to 817.52. Hence, the mixing of the two groups raises the risk of infection for those in the high-vaccination group, while it reduces the risk of infection for those in the low-vaccination group. This corresponds with intuition and expectations. When the agent contacts are mixed across regions with different levels of vaccine coverage, the agents in high vaccine regions come in contact with a higher percentage of unvaccinated agents (in low vaccine regions) and encounter increased risk of infection. The agents in low vaccine regions come in contact with more vaccinated agents when contacts are mixed, and are thus more protected from infection.

Second, the epidemic size for the whole population is higher in the “mixed” population than in the “segregated” population when  $\bar{V} = 0.4$ , while the relationship is reversed when  $\bar{V} = 0.5$ . For instance, when  $\bar{V} = 0.4$  and  $\Delta V = 0.1$ , the epidemic size rises from 2686.56 to 3074.23, a change of 387.67, as we go from “segregated” to “mixed.” Conversely, when  $\bar{V} = 0.5$  (and  $\Delta V = 0.1$ ), the epidemic size falls from 1339.51 to 772.94, a change of  $-566.57$ .

One sees a similar, though slightly more nuanced pattern in the  $\bar{V} = 0.6$  and  $\bar{V} = 0.7$  examples. When the vaccination rates are this large, some regions of the population are significantly above the level of herd immunity. Thus, looking at Table 9, one sees essentially no change for  $\Delta V = 0.05$  for both the high and low vaccination regions. This also is true for all levels of  $\Delta V$  for the high vaccine regions. All of these

---

coverage and thus engineer interactions. But, we view this as primarily a comparative statics exercise to better understand effects as opposed to a suggestion of realistic social engineering policy.

**Table 8** Region-specific epidemic sizes under two different contact structures  $p^C = 0.95$

$\Delta V$	Contacts <sup>1</sup>	Region	$\bar{V} = 0.4$	$\bar{V} = 0.5$	$\bar{V} = 0.6$	$\bar{V} = 0.7$
0.05	Segregated	Low-vaxx:	1963.33 (262.62)	565.77 (368.28)	76.46 (63.59)	26.79 (17.59)
		High-vaxx:	565.77 (368.28)	76.46 (63.59)	26.79 (17.59)	12.08 (5.68)
		Total:	2529.10 (508.19)	642.23 (379.46)	103.25 (66.29)	38.87 (19.07)
	Mixed	Low-vaxx:	1816.41 (154.55)	325.84 (227.55)	77.29 (47.39)	29.74 (16.25)
		High-vaxx:	1113.90 (138.44)	145.10 (95.43)	28.06 (18.05)	11.65 (7.86)
		Total:	2930.31 (245.89)	470.94 (308.64)	105.35 (54.31)	41.39 (18.41)
0.1	Segregated	Low-vaxx:	2508.76 (117.29)	1292.52 (439.73)	177.80 (116.35)	46.99 (35.86)
		High-vaxx:	177.80 (116.35)	46.99 (35.86)	17.98 (9.55)	10.52 (5.11)
		Total:	2686.56 (159.94)	1339.51 (440.83)	195.78 (117.13)	57.51 (37.11)
	Mixed	Low-vaxx:	2256.71 (179.22)	621.26 (317.65)	128.02 (76.22)	46.88 (24.59)
		High-vaxx:	817.52 (109.19)	151.68 (73.27)	20.60 (13.41)	8.48 (4.85)
		Total:	3074.23 (253.67)	772.94 (379.83)	148.62 (82.51)	55.36 (24.43)
0.2	Segregated	Low-vaxx:	3272.84 (55.19)	2508.76 (117.29)	1292.52 (439.73)	177.80 (116.35)
		High-vaxx:	46.99 (35.86)	17.98 (9.55)	10.52 (5.11)	6.90 (1.73)
		Total:	3319.83 (66.25)	2526.74 (118.17)	1303.04 (439.06)	184.70 (116.07)
	Mixed	Low-vaxx:	3125.92 (122.01)	1951.39 (348.84)	510.04 (296.38)	151.45 (81.17)
		High-vaxx:	462.20 (69.80)	138.55 (36.80)	18.77 (9.79)	3.35 (1.93)
		Total:	3588.12 (139.36)	2089.94 (371.09)	528.81 (302.86)	154.80 (81.56)

For the mixed contact case, the ten initial seed agents (infected at  $t = 0$ ) are randomly distributed across the population of 10,000 agents. For the segregated contact case, we randomly select 5 seed agents for each region (of 5000 agents), which then add up to 10 seed agents for the entire population

**Table 9** Change in the number of infections by region  $p^C = 0.95$ 

	$\bar{V}$	Low-vaxx regions	High-vaxx regions	Total population
$\Delta V = 0.05$	0.4	-146.92	548.13	401.21
	0.5	-239.93	68.64	-171.29
	0.6	0.83	1.27	2.10
	0.7	2.95	-0.43	2.52
$\Delta V = 0.1$	0.4	-252.05	639.72	387.67
	0.5	-671.26	104.69	-566.57
	0.6	-49.78	2.62	-47.16
	0.7	-0.11	-2.04	-2.15
$\Delta V = 0.2$	0.4	-146.92	415.21	268.29
	0.5	-557.37	120.57	-436.80
	0.6	-782.48	8.25	-774.23
	0.7	-26.35	-3.55	-29.90

regions are above the level of vaccine coverage that produces herd immunity; thus, when contacts are made with the low vaccine regions there is a negligible effect on the number of infections. However, when looking at the low vaccine regions one does see an effect for large enough values of  $\Delta V$ . Here the low vaccine regions gain noticeable protection from shifting contacts from other low vaccine regions to regions with much higher levels of vaccine coverage. Thus, on net, the overall epidemic size decreases significantly. Note that these decreases can be very large for some parameter sets. For instance, shifting from segregated to mixed contacts decreases average epidemic size by 774 agents for  $\bar{V} = 0.6$ ,  $\Delta V = 0.2$ .

More generally, one can understand these results through the concept of herd immunity. When the vaccination rate for a specific region is above or close to the level of herd immunity, adding a few additional connections from a low vaccination region does little to increase the infection risk for agents in the high vaccine area. But, for agents in the low vaccine area, there is significant benefit to moving links from a low vaccine area to a high vaccine area that is relatively infection free. Thus when the high vaccination regions have vaccine coverage near or above 60% the effect of additional protection for the low vaccine regions dominates the additional risk for the high vaccine regions and average epidemic size falls. In the case where the overall vaccination rate is lower, average vaccine coverage of 40%, the results go in the opposite direction. Here, even agents in the high vaccine region are below levels of vaccine coverage that grant herd immunity. Thus when links are added that connect to a low vaccine region they face a significant increase in infection risk. And, again because the high vaccine regions are below the level of herd immunity, the agents from the low vaccine region get less protection from moving links from the low vaccine regions to the high vaccine regions. The overall effect is an increase in average epidemic size when connections between low and high vaccine regions are created.

These results contain a fair amount of nuance. Note that we are holding the level of global interactions,  $p^C$ , constant as we move from a segregated to a mixed contact

structure. The only difference between the segregated and mixed contact structures is that the global connections under the mixed structure allow connections to other areas that have the different vaccination level. But, as mentioned above, this comes at the expense of increasing the risk of infection for unvaccinated agents in the high vaccination regions. So, what we have observed here are two externalities being generated when we connect high vaccine and low vaccine regions. These inter-regional linkages create a positive externality for agents living in low vaccine regions because these agents are now more likely to be connected to vaccinated agents in the high vaccine regions. In the other direction, a negative externality has been created because unvaccinated agents from high vaccine regions are more likely to be connected to unvaccinated agents in the low vaccine regions. These agents now have a higher risk of infection. In our results the positive externality dominates when  $\bar{V}$  is high (near herd immunity) such that the epidemic size decreases when high to low vaccine regions are connected. Alternatively, when  $\bar{V}$  is low (significantly below herd immunity), the negative externality dominates and the epidemic size increases when the two regions are connected.

It is worthwhile to think about the implications that the presence of these externalities may have for individual behavior with respect to their decision to vaccinate or not. Recall that vaccination choices are specified exogenously in our model. If, however, the individual agents are to make these choices independently based on their assessment of the benefits vs costs of vaccination, there is potential that these externalities could create perverse incentives on the part of those in the low-vaccination group. For these agents, additional contacts with the members of the high-vaccination group can reduce the risk of infection for them without their having to be vaccinated themselves. Specifically, the members of the low-vaccination group could end up *free-riding* on the members of the high-vaccination group. The consequence of such behavior cannot be examined in the current paper with its exogenous specification of the vaccination rates. However, an extension that allows endogenization of the individuals' vaccination decisions in the presence of these externalities would be a worthwhile endeavor that can directly address the issue in question.<sup>12</sup>

## 6 Conclusion

In this paper we develop a specific model of spatial vaccine coverage heterogeneity. The model isolates the effects of heterogeneity by holding overall vaccination coverage rates constant. We find that spatial heterogeneity by itself has significant implications for understanding epidemic dynamics. Specifically, an increase in spatial heterogeneity leads to larger epidemics on average as long as the population has a sufficiently high level of vaccine coverage. At these levels, it also increases the level of vaccinations that are required to produce herd immunity. Both of these effects are magnified when more interregional connections exist in the contact structure of society. Conversely, if the coverage rate is significantly below the level needed to create herd immunity, then increasing spatial heterogeneity may decrease epidemic size. One policy relevant

<sup>12</sup> The authors currently are developing such a model.

highlight of these results concerns the closeness of different infectious diseases to reaching herd immunity. The results above suggest that increased spatial heterogeneity in vaccine coverage is most dangerous when the population is near herd immunity. This suggests that this research may be very important for examining MMR vaccinations and other instances of vaccine coverage that are at or near levels that provide herd immunity.

We also highlight the externalities present between regions with different levels of vaccination coverage. Regions with low levels of vaccine coverage gain significant protection from having a small number of their interregional connections with other low vaccine regions shifted to regions with higher levels of vaccine coverage. Conversely, unvaccinated individuals in the high vaccine regions are harmed through these high to low interregional connections. The net effect of these two externalities depends on the overall level of vaccinations in the population. If vaccination coverage is sufficiently close to herd immunity then the change in connections leads to a smaller mean epidemic size. However, if the population is sufficiently far from herd immunity, then the shifted connections can lead to increased mean epidemic size.

## References

- Abbe, E. (2018). Community detection and stochastic block models: Recent Developments. *Journal of Machine Learning Research*, 18, 1–86.
- Anderson, R. M., & May, R. M. (1985). Vaccination and herd immunity to infectious diseases. *Nature*, 318, 323–329.
- Atwell, J. E., Van Otterloo, J., Zipprich, J., Winter, K., Harriman, K., et al. (2013). Nonmedical vaccine exemptions and pertussis in California, 2010. *Pediatrics*, 132(4), 624–630.
- 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.
- Birnbaum, M. S., Jacobs, E. T., Ralston-King, J., & Ernst, K. C. (2013). Correlates of high vaccination exemption rates among kindergartens. *Vaccine*, 31(5), 750–756.
- Buttenheim, A., Jones, M., & Baras, Y. (2012). Exposure of California kindergarteners to students with personal belief exemptions from mandatory school entry vaccinations. *American Journal of Public Health*, 102(8), e59–e67.
- Carrel, M., & Bitterman, P. (2015). Personal belief exemptions to vaccination in California: A spatial analysis. *Pediatrics*, 135(1), 80–88.
- CDC (2017a). <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>. Accessed 1 Oct 2019.
- CDC (2017b). <https://www.cdc.gov/flu/fluavaxview/reportshtml/trends/index.html>. Accessed 1 Oct 2019.
- CDC (2017c). <https://www.cdc.gov/nchs/fastats/immunize.htm>. Accessed 1 Oct 2019.
- CDC (2017d). <https://www.cdc.gov/flu/professionals/acip/clinical.htm>. Accessed 1 Oct 2019.
- Chen, F., & Toxvaerd, F. (2014). The economics of vaccination. *Journal of Theoretical Biology*, 363, 105–117.
- 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.
- Feikin, D. R., Lezotte, D. C., Hamman, R. F., Salmon, D. A., Chen, R. T., & Hoffman, R. E. (2000). Individual and community risks of measles and pertussis associated with personal exemptions to immunization. *Journal of the American Medical Association*, 284(24), 3145–3150.

- Galeotti, A., & Rogers, B. W. (2013). Strategic immunization and group structure. *American Economic Journal: Microeconomics*, 5(2), 1–32.
- Galeotti, A., & Rogers, B. W. (2015). Diffusion and protection across a random graph. *Network Science*, 3(3), 361–376.
- Goeffard, P.-Y., & Philipson, T. (1997). Disease eradication: Public vs. private 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.
- Gersovitz, M. (2013). Mathematical epidemiology and welfare economics. In Manfredi & d’Onofrio (Eds.), *Modeling the interplay between human behavior and the spread of infectious diseases*. New York: Springer.
- Gust, D. A., Strine, T. W., Maurice, E., Smith, P., Yusuf, J., Wilkinson, M., et al. (2004). Underimmunization among children: Effects of vaccine safety concerns on immunization status. *Pediatrics*, 114(1), e16–e22.
- Ingraham, C. (2015). The devastating impact of vaccine deniers, in one measles chart. January 22, 2015. *Washington Post*, <https://www.washingtonpost.com/news/wonk/wp/2015/01/22/the-devastating-impact-of-vaccine-deniers-in-one-measles-chart/>.
- Jackson, M. O. (2008). *Social and economic networks*. Princeton, NJ: Princeton University Press.
- Lieu, T. A., Ray, G. T., Klein, N. P., Chung, C., & Kulldorff, M. (2015). Geographic clusters in underimmunization and vaccine refusal. *Pediatrics*, 135(2), 280–289.
- McNutt, L.-A., Desemone, C., DeNicola, E., El Chebib, H., Nadeau, J. A., et al. (2016). Affluence as a predictor of vaccine refusal and underimmunization in California private kindergartens. *Vaccine*, 34(14), 1733–1738.
- May, T., & Silverman, R. D. (2003). Clustering of exemptions as a collective action threat to herd immunity. *Vaccine*, 21(11–12), 1048–1051.
- Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ: Princeton University Press.
- Olpinski, M. (2012). Anti-vaccination movement and parental refusals of immunization of children in USA. *Pediatrica Polska*, 87, 381–385.
- Oster, E., & Kocks, G. (2018). January 16, 2018. After a debacle, how California became a role model on measles. *New York Times*, <https://www.nytimes.com/2018/01/16/upshot/measles-vaccination-california-students.html>.
- Philipson, T. (2000). Economic epidemiology and infectious diseases. In A. J. Cuyler & J. P. Newhouse (Eds.), *Handbook of health economics*. Amsterdam: North Holland.
- Plans-Rubio, P. (2012). The vaccination coverage required to establish herd immunity against influenza viruses. *Preventive Medicine*, 55(1), 72–77.
- Poland, G. A., & Jacobson, R. M. (2001). Understanding those who do not understand: A brief review of the anti-vaccine movement. *Vaccine*, 19(17–19), 2440–2445.
- 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., Sydorak, 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.
- Sugarman, D. E., et al. (2010). Measles outbreak in a highly vaccinated population, San Diego, 2008: Role of the intentionally undervaccinated. *Pediatrics*, 125(4), 747–755.
- Tizzoni, M., et al. (2012). Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine*. <https://doi.org/10.1186/1741-7015-10-165>.
- 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. J., & Strogatz, S. H. (1998). Collectives dynamics of small-world networks. *Nature*, 393, 440–442.