

A framework for discovering health disparities among cohorts in an influenza epidemic

Lijing Wang¹ · Jiangzhuo Chen² · Achla Marathe³

Received: 11 December 2017 / Revised: 30 March 2018 / Accepted: 31 May 2018 /
Published online: 14 June 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Infectious diseases such as Influenza and Ebola pose a serious threat to everyone but certain demographics and cohorts face a higher risk of infection than others. This research provides a computational framework for studying health disparities among cohorts based on individual level features, such as age, gender, income, etc. We apply this framework to find health disparities among subpopulations in an influenza epidemic and evaluate vaccination prioritization strategies to achieve specific objectives. We explore the heterogeneities in individuals' demographic and socioeconomic attributes as the potential cause of health disparities. An agent-based model is used to simulate an influenza epidemic over a synthetic social contact network of the Montgomery County in Southwest Virginia to identify infected cases which are then labeled with a specific clinical outcome by using a predefined probability distribution based on age and risk level. We divide the population into age and income based cohorts and measure the direct and indirect economic impact of

This article belongs to the Topical Collection: *Special Issue on Social Computing and Big Data Applications*

Guest Editors: Xiaoming Fu, Hong Huang, Gareth Tyson, Lu Zheng, and Gang Wang

✉ Jiangzhuo Chen
chenj@vt.edu

Lijing Wang
lijingw9@vt.edu

Achla Marathe
amarathe@vt.edu

¹ Department of Computer Science, Virginia Tech, Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24061, USA

² Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24061, USA

³ Department of Agricultural and Applied Economics, Virginia Tech Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24061, USA

vaccination for each cohort. Simulation-based results find strong health disparities across age and income groups. Various vaccine distribution strategies are considered and outcomes are measured through metrics such as death count, total number of infections, net return per capita, net return per dollar spent and net return per vaccinated person. The results, framework, and methodology developed here can assist public health policy makers in efficiently allocating limited pharmaceutical resources.

Keywords Agent based · Computational framework · Health disparity · Simulation · Vaccination

1 Introduction

Infectious diseases such as influenza and Ebola pose a serious threat to global public health. In the United States, it is estimated that each year seasonal influenza causes 31.4 million outpatient visits, over 200,000 hospitalizations, 3,300–49,000 deaths, and is responsible for 44.0 million days of lost productivity [20, 34, 39, 41]. Worldwide influenza causes about three to five million cases of severe illness and about 250,000 to 500,000 deaths [45].

To control the spread of influenza, the Advisory Committee on Immunization Practices (ACIP) recommends seasonal influenza vaccination annually for individuals aged 6 months and older without contraindications [15]. However, this recommendation assumes an unlimited supply of vaccines to be available which is not always the case, especially during an influenza pandemic [16, 20, 32].

This work focuses on the study of health disparities and vaccination priorities among different subpopulations. Health disparities can arise due to differences in age, gender, geographic location, income level, and other socioeconomic variables [23] since personal attributes influence behavior, activities, and social interactions, which in turn affect an individual's vulnerability and infectivity. To effectively control the spread of a disease it is important to understand these differences so scarce public health resources can be distributed efficiently [1, 4, 21, 24, 36, 43].

Health disparity, as measured by case outcome and economic return, varies by the way a population is partitioned. For instance, the disparity may be high among age groups but not so much among income groups, implying that age is a bigger determinant of disparity than income. It is important to understand health disparity with respect to demographic features and identify the most significant ones. This would help identify cohorts that are at a disadvantage in terms of encountering more infections and may require more resources to achieve herd immunity not only within their own cohorts but among the entire population. Ultimately it will guide the distribution of health care resources so that overall infection rates can be curtailed and economic loss can be minimized. The literature lacks this kind of detailed, individual-based study of health disparities and their consequences on the control of epidemics.

To study cohort level disparities, we use an individual-based model where a synthetic representation of a real population is created at an individual level. The synthetic individuals are statistically equivalent to the Census population when aggregated to a block group level. Each synthetic individual is assigned an activity sequence which is derived from time use surveys. It includes the types of activities performed and the time of day they are performed at [5]. A location for each activity is also assigned and when individuals are co-located at these locations, they are assumed to be in contact with each other, resulting in the creation of a synthetic social contact network. This approach supports fully heterogeneous

individuals and does not require assumptions about large-scale regularity of interactions, unlike compartmental models [3, 8].

There have been many articles that study the cost of influenza infections [6, 9, 26]. For example, there are studies focusing on the cost effectiveness of vaccination for children [19, 25, 29, 33, 35, 40], healthy working adults [10, 28], and people of age 65 years and above [37]. Research in [16] compares the cost effectiveness of vaccination using a dynamic model and a static model and prioritizes individuals by age and risk groups. These works focused on the measurement of direct and indirect impacts that take medical cost and productivity loss into account. Meltzer et al. [32] used a Monte Carlo simulation model to analyze the direct economic impact of vaccine-based interventions in pandemic influenza and proposed multiple criteria to generate vaccination priorities for subpopulations. Work by [17] examined herd immunity to further learn the direct and indirect effects of influenza vaccination. However, prior work has not examined health disparities that originate from differences in individual level attributes.

In this work, we build a computational framework to study the impact of heterogeneities in individuals' demographic and socioeconomic attributes, on health disparities during an influenza epidemic. We also design vaccination prioritization strategies by simulating epidemics where vaccine-based interventions are applied to targeted groups of people in the population to fulfill specific policy objectives. Our contributions include:

- A framework to study health disparities in an influenza epidemic using a synthetic social contact network and an agent-based simulation model. This modular framework is highly generic and can be used to study other disease epidemics and intervention strategies.
- Studied the impact of heterogeneities in individual level attributes on health and economic outcomes. Identified attributes that cause significant health disparities among cohorts.
- Calculated direct and indirect impacts of vaccination-based interventions.
- Designed cohort-based vaccination prioritization strategies using multiple objectives and measured their performance.

The rest of this paper is organized as follows. Section 2 describes the framework and the models used in this research. Section 3 shows simulation results, and discusses health disparities with respect to age and income attributes. Various vaccination strategies are designed and discussed with respect to their case and economic outcomes. Section 4 concludes the paper. [Appendix](#) provides details of the data used in the paper.

2 Framework and methods

In this section, we present the details of the framework along with the models and the methods used. Although this work focuses on influenza and vaccine based intervention, the framework is generic and can be applied to diseases and be used for other interventions such as antivirals and social distancing.

2.1 Framework

The framework consists of three major components: (1) real/synthesized data, (2) modeling and simulation, and (3) analysis (see [Figure 1](#)). The first component collects data from a variety of sources and synthesizes it [38]. This includes building the social contact network,

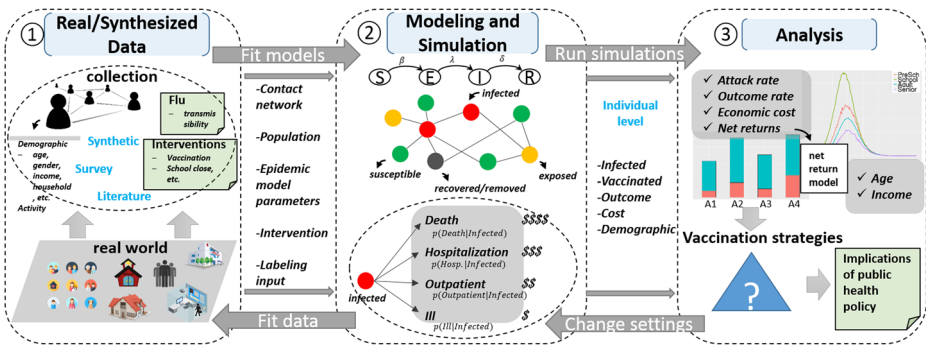


Figure 1 Framework for studying health disparities based on individuals, cohorts, and specific objectives. *Real/synthesized data* module collects data on costs, disease parameters, interventions etc. and synthesizes data from multiple sources to create the synthetic social contact network. *Modeling and simulation* module consists of the disease model, net return model which labels health outcomes and associated costs. The inputs are collected by the first module, the outputs generated by the second module are individual level infected cases with a specific outcome (i.e., death, hospitalization, outpatient, ill but not seek medical care), and vaccinated cases. *Analysis* module performs health disparity analysis and intervention analysis. Health disparity is analyzed with respect to age and household income. Vaccination prioritization strategies are designed to minimize death rate/total death count, or to maximize net returns per dollar spent/total net returns

collecting data on disease model parameters, data on case outcome distributions conditional on demographics, and data on costs conditional on clinical outcomes. The second component consists of a disease spread model which determines the health state (susceptible, exposed, infected or removed) of each synthetic individual at the end of the simulation; and a labeling algorithm for classifying each infected individual into four health outcomes i.e., death, hospitalization, outpatient, and ill but not seek medical care. The third component is the analytical engine of the framework that analyzes health disparities based on individual level demographic and health attributes, calculates the impact of interventions, and builds intervention priorities targeting towards specific economic and social objectives.

The following steps illustrate how the framework processes the data, runs the simulations, and analyzes the results.

- (a) *collect and synthesize data on the population and social contact network* → *build influenza disease model* → *run epidemic simulations* → *label infected cases with clinical outcomes* → *analyze health disparity with respect to individual level attributes*.

To simulate an epidemic, we need individual level data of the population, its contact network on which the transmissions occur, disease model parameters such as transmission rate, i.e., the probability of transmission from an infectious person to a susceptible person per unit of contact time, incubation duration, infectious duration, etc. If vaccination intervention is applied, then vaccine efficacy (reduction on disease transmission probability) and compliance rate (probability that a person will take the vaccine) are also required. The simulation output provides health label (infected or not) and vaccine label (vaccinated or not) for each person. This is joined with the person’s individual level attributes such as age, gender, income etc.

Further, each synthetic individual is assigned with a risk level based on age regardless of health and vaccine label, i.e., whether the person is high-risk or non-high-risk [32]. Then each infected case is labeled with a specific clinical outcome among:

death, hospitalization, outpatient, and ill but not seeking medical care, based on age and the risk level of the individual. For each clinical outcome, there is an associated cost of treatment. The clinical outcomes of the infected are decided by age and the risk level as given in [32]; and the data on the treatment costs for each outcome is given in [12]. We have included these data in Appendix for completeness.

We divide the whole population into subgroups based on a select attribute, e.g. age, and compute the difference between age-groups with respect to cumulative infection rate, health outcome, and economic cost. The statistical significance tests are used to determine whether health disparities exist in terms of the select attribute.

- (b) *design vaccination strategies* → *change intervention parameter settings* → *run simulations* → *analyze the effects of revised vaccination strategies* → *find public health policy implications*.

We consider different vaccine distribution schemes assuming vaccines are in limited supply. A no-priority strategy is defined for comparative analysis where all vaccines are distributed randomly in the population. In other cases, a specific demographic is selected to divide the population into subgroups on which a vaccine prioritization is determined. In this study, vaccination priorities are designed based on a subgroup's ability to achieve specific objectives such as minimum number of total infections or maximum net return per capita. Different vaccination priorities are considered and compared using simulation results.

Our goal is to answer the following kinds of questions through this framework: (1) Can health disparities be explained by the demographics? If yes, what are the causal reasons behind them? (2) Is it possible to design vaccination prioritization strategies assuming vaccines are in limited supply based on this knowledge? (3) What is the level of improvement provided by these vaccination priorities with respect to various measures of performance?

2.2 Method

This section provides details of the data and methods used in the study.

2.2.1 Synthetic social contact network

A synthetic population and its social contact network are used to simulate the spread of the disease. In this study we use the synthetic social contact network of the Montgomery county in Virginia [38]; its basic properties are summarized in Table 1. A methodology for constructing the synthetic population and contact network for any area in the US can be found in [3, 5, 8, 18, 38]. In what follows we briefly describe the procedures; interested readers can find details in the aforementioned references.

Table 1 Social contact network of Montgomery County, VA

Nodes	Edges	Avg. degree	Network density
77820	3936774	50.7519	0.0013

The attributes of the synthetic social contact network of Montgomery County, VA. The contact network is represented as an edge-weighted network. Nodes correspond to individuals, edges represent contacts, and edge weights represent contact durations

First, a statistical representation of the each individual in the population is built using the US Census data. This synthetic population is statistically equivalent to the real population as given in the US census, when aggregated up to a census block group level. Individuals in the synthetic population are endowed with a complete range of demographic attributes as available in the Census [5, 7], including variables such as age and income level. The population synthesis process preserves the confidentiality of the individuals in the original data sets, yet produces realistic attributes and demographics for the synthetic individuals. Joint demographic distributions are reconstructed from the marginal distributions available in typical census data together with joint distributions in Public Use Microdata Samples (PUMS) using an iterative proportional fitting technique. This technique guarantees that a census of our synthetic population is statistically indistinguishable from the input census when aggregated to block groups.

Next, a set of activity templates for households are determined, based on American time-use surveys [11] and National Household Travel Survey. These activity templates provide a daily set of activities for individuals and the time of day they are performed. Each synthetic household is then matched with one of the survey households, using a decision tree based on demographics such as the size of the household, number of workers in the household, number of children, etc. The synthetic household members are then assigned the activity templates of its matching survey household members, giving each synthetic member a daily sequence of activities. The activities can be of type home, work, shop, school and “other”. For each activity, and for each individual, a geographic location is identified based on land-use patterns, transportation network and data from commercially available databases such as Dun and BradStreet.

A social network is formed when individuals are simultaneously present at a location. The co-location based social network is dynamic and changes as people visit different locations and come in contact with individuals at these locations. The contact network $G(V, E, w)$ is an edge-weighted network. Nodes correspond to individuals, edges represent the contact between two end nodes, and edge weights represent contact durations. Edge (u, v) with weight $w(u, v)$ represents that node u is in contact with node v for duration $w(u, v)$, during which the disease may transmit from node u to node v with probability $p(w(u, v))$ if u is infected and v is susceptible.

2.2.2 Disease model

In this study, a standard SEIR model is used for modeling the spread of influenza. SEIR models are widely used in mathematical epidemiology literatures [2, 27] and have been extensively developed and applied to model contagious disease outbreaks [30, 31]. Each person is in one of the following four health states at any time: susceptible, exposed, infectious, and removed. A person v is in the susceptible state until he becomes exposed. If v becomes exposed, he remains so for $\Delta t_E(v)$ days, which is called incubation period, during which he is not infectious. Then he becomes infectious and remains so for $\Delta t_I(v)$ days, which is called infectious period. Finally he becomes removed (or recovered) and remains so permanently.

The health states of person v during the epidemic period, denoted by (v) , can be equivalently represented by $\tau(v) = (t_{S \rightarrow E}(v), t_{E \rightarrow I}(v), t_{I \rightarrow R}(v))$, where $t_{S \rightarrow E}(v)$ is the day when v becomes exposed, $t_{E \rightarrow I}(v) = t_{S \rightarrow E}(v) + \Delta t_E(v)$ is the day when v becomes infectious, and $t_{I \rightarrow R}(v) = t_{E \rightarrow I}(v) + \Delta t_I(v)$ is the day when v becomes removed. The model of our framework assumes that $\Delta t_E(v)$ and $\Delta t_I(v)$ for each person v are known at the

beginning of the simulation. It also assumes the health state transitions are not reversible and they are the only possible state transitions [14, 30, 31].

With the SEIR model, the disease spreads in a population in the following way. It can only be transmitted from an infectious node to a susceptible node. On any day, if node u is infectious and v is susceptible, disease transmission from u to v occurs with probability $p(w(u, v)) = 1 - (1 - r)^{w(u, v)}$ where r is the probability of disease transmission for a contact of one unit time. So the disease propagates probabilistically along the edges of the contact network.

To simulate the spread of influenza in the synthetic contact network, we use EpiFast [8], a high-performance agent-based simulation model. It follows the standard SEIR disease model to capture the within-host and between-host disease progressions, and uses a distributed algorithm to compute stochastic disease propagation in the synthetic network. EpiFast provides individual level details while maintaining heterogeneity among individuals.

2.2.3 Label infected people with clinical outcomes

Each infected person can have one of four clinical outcomes attributed to influenza infection: death (D), hospitalization(H), outpatient(O), ill but not seeking medical care(I). These outcomes depend on a person's underlying risk condition and age. One's risk condition (high risk/non-high risk) arising from a pre-existing medical condition often depends on age too. Economic cost is assigned to each infected person according to the outcome. The economic cost includes medical costs related to the treatment of infected cases (direct costs) and productivity loss of infected people (indirect costs). Once individual level outcomes and costs are known, we can calculate the total costs for any subgroup in the population.

For an infected individual whose age is known, the risk level and clinical outcomes are determined by the distributions given in [32]; and the treatment costs for each outcome under each age and risk level are determined by the data given in [12]. More details about these distributions are shown in Appendix Tables 19, 20, and 21.

2.2.4 Interventions

Two scenarios regarding interventions are considered in this work: base case, where no intervention is applied to contain the epidemic; and intervention case, where an intervention is applied to a subset of the population. The intervention may be targeted towards a subpopulation chosen by individuals' specific attributes or completely random where each person is equally likely to be chosen. In our simulation, we focus on mass vaccination, which is applied at the beginning of the epidemic. Each person chosen to be vaccinated complies with a specified probability, called compliance rate. We assume that each person within a cohort follows a same compliance rate. Compliance rate varies across different cohorts.

2.2.5 Net return model

The total economic costs under base case include direct and indirect costs. The total costs under intervention case include direct, indirect, and intervention costs. Direct costs are the medical costs associated with the clinical outcome. Indirect costs are measured by the loss of productivity due to illness. Net return (NR) is defined as the reduction in total cost due to intervention. It provides a measure of economic efficiency.

By aggregating individual costs up to a cohort level or population level, and comparing the base case and the intervention case, we can compute the net return of an intervention as:

$$NR = \sum_{i \in S_{base}} C_i - \sum_j C_j - \sum_k C_k^{vax}, \quad (1)$$

where S_{base} , S_{int} are sets of infected individuals under base case and intervention case, respectively. S_{vax} is the set of vaccinated individuals. C_i represents the economic cost of individual i who gets infected. C_k^{vax} is the cost of vaccination for individual k who gets vaccinated. We assume that for any k , C_k^{vax} is US \$30.53 [32]. C_i is a predefined variable that varies based on age, risk factors, and clinical outcomes. The details on cost are provided in the Appendix, Table 21. We can compute total net return, direct net return, and indirect net return by using corresponding C_i of total cost, direct cost, indirect cost, respectively in Table 21. All costs are adjusted to 2016 USD.

To account for difference in group sizes we calculate net return per capita (NRPC), net return per vaccinated person (NRPV), and net return per dollar spent (NRPD) as:

$$NRPC = \frac{NR}{N_c}, NRPV = \frac{NR}{N_v}, NRPD = \frac{NR}{N_d}, \quad (2)$$

where N_c , N_v , N_d represent the group size, number of vaccinated people within the group, and total dollar spent on vaccinating the group respectively.

Note that our individual based framework allows us to calculate net return for any subpopulation and for any health outcome. More details are shown in Section 3.

2.2.6 Subpopulations

To discover health disparities with respect to various individual level attributes, we split the population into cohorts by the selected attributes. In this work, we use age and household income to build the cohorts. We create: (i) four age groups that are distinct with respect to economic activity and health care related costs: 0-4 year olds (preschool); 5-19 year olds (school); 20-64 year olds (adult); and people who are 65 and above years old (senior); (ii) four quartiles by annual household income: \$0-18400 (Q1); \$18400-41620 (Q2); \$41620-75000 (Q3); and above \$75000 (Q4).

2.3 Applications

This framework can be applied to undertake various studies using alternative interventions, disease models, and datasets. Three major tasks become easier within this framework include: (a) study attributes based health disparity; (b) determine which attributes better explain health disparity; (c) design effective intervention strategies and build prioritization for distributing limited health care resources.

We list all the notations defined in Sections 2 and 3 in Table 2.

3 Simulation results and analysis

This section describes simulation settings and results. Vaccination is the only intervention considered here. Firstly, health disparities and economic disparities are discussed separately with respect to age and household income. We assume that vaccines can fully cover the whole population, and the compliance rates of each cohort are the same. Sensitivity of the

Table 2 Notations and their meanings

Notation	Meaning
$G(V, E, w)$	A contact network
V, E	The sets of nodes, edges
$w(u, v)$	The contact duration of node u with node v on the day
$p(w(u, v))$	The probability of disease transmits from node u to node v during $w(u, v)$
$\tau(v)$	The health states of node v during the epidemic period
$t_{S \rightarrow E}(v)$	The day when v becomes exposed
$t_{E \rightarrow I}(v)$	The day when v becomes infectious
$t_{I \rightarrow R}(v)$	The day when v becomes recovered
$\Delta t_E(v)$	Incubation period of node v
$\Delta t_I(v)$	Infectious period of node v
NR	Net return
C_i	The cost of an infected individual i
C_k^{vax}	The cost of vaccinating individual k
$NRPC(\overline{NRPC})$	Net return per capita (average net return per capita)
$NRPD(\overline{NRPD})$	Net return per dollar spent (average net return per dollar spent)
$NRPV(\overline{NRPV})$	Net return per vaccinated person (average net return per vaccinated person)
$TNR(\overline{TNR})$	Total net return (average total net return)
$AR(\overline{AR})$	Attack rate or cumulative infection rate (average attack rate)
$DR(\overline{DR})$	Death rate (average death rate)
$HR(\overline{HR})$	Hospitalization rate (average hospitalization rate)
$OR(\overline{OR})$	Outpatient rate (average outpatient rate)
$IR(\overline{IR})$	Ill but not seeking medical care rate (average ill rate)
$CMP L$	Compliance rate for vaccination
$DC(\overline{DC})$	Death count (average death count)

disparity results are considered with respect to disease transmission rate and compliance rate for vaccination. Secondly, vaccination prioritization strategies are designed assuming vaccines are in limited supply. The compliance to vaccination can be uniform or vary by age.

3.1 Simulation settings

Table 3 shows list of parameter values and their sources. We seed the epidemic by randomly infecting 10 individuals at the start of the simulation. A base attack rate of 40% is assumed when no interventions are applied. The corresponding disease transmission rate for this network is 0.00007 per minute of contact time. The transmission rate keep the same across base case and intervention case. In intervention case, vaccination is applied randomly with a compliance rate of 50% and the vaccine efficacy is assumed to be 90%. A 90% vaccine efficacy means that the vaccine reduces the probability of disease transmission by 90%. Note that these parameter values do not correspond to any specific epidemic, as they vary from season to season and differ between populations. For each scenario, we run 30 replicates for 300 days and report the average results.

Table 3 List of parameter values and their sources

Parameter	Value	Source
Attack rate (AR)	40-60%	[30, 32]
Transmission rate	calibrated to attack rate	not applicable
Proportion of symptomatic	67%	[13, 22, 42]
Avg. incubation period	1.9 days	[31]
Avg. infectious period	4.1 days	[31]
Diagnosis rate	60%	[22]
Compliance to vaccination	25-70%	[16]
Efficacy of vaccine	50-90%	varies by season

3.2 Health disparity

Here we study disparities in attack rate between different age groups (or income groups) when no intervention is applied (base case). The disease model determines the health state of each individual at the end of simulation. We identify all infected cases and split them into age (or income) groups to calculate the attack rates for each group.

Age-based groups The epidemic curves, which show the number of new infections each day, for each age group are shown in Figure 2. We find that the epidemic in the school age group starts earlier and peaks higher. To check for significant differences in attack rates (cumulative infection rate) across age groups, the t-tests are performed. Results in Table 4 show p-values of 2-sample t-tests. The average attack rate (\overline{AR}) for each group is reported in their respective rows. The p-values show that there is a significant difference in \overline{AR} between any two age groups, and school age group has the greatest attack rate at $\overline{AR} = 70\%$, making it the most vulnerable group to influenza infection.

Table 5 shows that there is also a significant difference in average death rate (\overline{DR}) between any two age groups, and senior age group has the greatest death rate at $\overline{DR} = 0.296\%$.

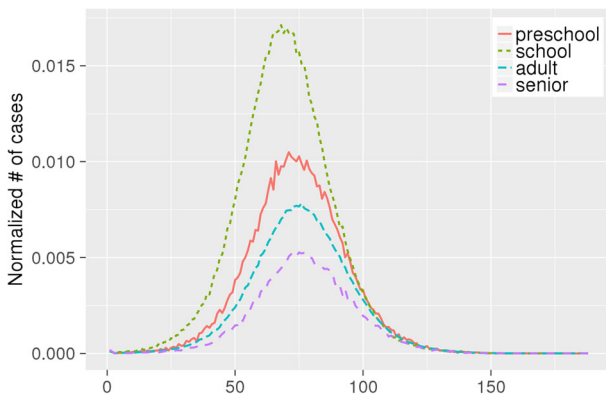


Figure 2 The epidemic curves of age groups along days. X-axis denotes days, Y-axis denotes normalized number of cases (as a fraction of the size of the group)

Table 4 Disparity in attack rate among age groups: p-value (base-case)

Attack rate by age group	Preschool	School	Adult	Senior
Preschool ($\overline{AR} = 44\%$)	–	0.0001***	0.0001***	0.0001***
School ($\overline{AR} = 70\%$)	–	–	0.0001***	0.0001***
Adult ($\overline{AR} = 34\%$)	–	–	–	0.0001***
Senior ($\overline{AR} = 22\%$)	–	–	–	–

¹ \overline{AR} is the average attack rate computed from 30 simulation replicates

²* $p < .05$, ** $p < .01$, *** $p < .001$

We also observe significant difference in average hospitalization rate (\overline{HR}), average outpatient rate (\overline{OR}), average ill rate (\overline{IR}) with respect to age. Results are omitted for the sake of brevity.

Income-based groups Figure 3 shows the epidemic curves of income groups. We observe that households with high income tend to have an earlier and higher epidemic peak. Table 6 shows p-values of 2-sample t-tests where attack rates between any two income groups are tested. Here too, we find a significant difference in \overline{AR} between any two income groups, and \overline{AR} increases as household income increases. The higher quartile income groups encounter higher infection rates. This is likely due to the fact that higher income households have larger families, as shown in Table 7. In larger families, once one person is infected, the risk of secondary infection can go up by 38% [44].

Unlike age groups, we observe no significant difference in \overline{DR} among income groups in Table 8. This is because the elderly population which is at a higher risk of death is more evenly distributed among income groups as compared to age groups.

3.3 Economic disparity

Next we examine the economic impact of intervention by age and income groups. For each group we compute the direct, indirect, and total net returns. To do comparative analysis across groups we measure net return per capita (NRPC), net return per vaccinated person (NRPV), and net return per dollar spent (NRPD) using (1) and (2). All returns are measured in 2016 USD.

Age-based groups We show direct, indirect, and total net return for each age group in Figure 4. For all age groups the indirect return is the dominant component of total net

Table 5 Disparity analysis of death rate among age groups: p-value (base-case)

Death rate by age group	Preschool	School	Adult	Senior
Preschool ($\overline{DR} = 0.035\%$)	–	0.0181*	0.0001***	0.0001***
School ($\overline{DR} = 0.046\%$)	–	–	0.0001***	0.0001***
Adult ($\overline{DR} = 0.069\%$)	–	–	–	0.0001***
Senior ($\overline{DR} = 0.296\%$)	–	–	–	–

¹ \overline{DR} is the average death rate computed from 30 simulation replicates

²* $p < .05$, ** $p < .01$, *** $p < .001$

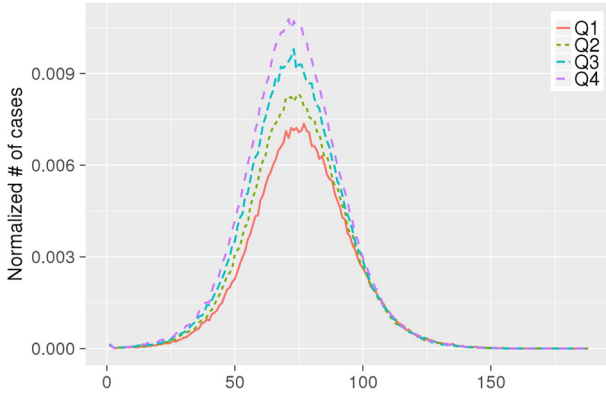


Figure 3 The epidemic curves of income groups along days. X-axis denotes days, Y-axis denotes normalized number of cases

Table 6 Disparity analysis of attack rate among income groups: p-value (base-case)

Income quartiles	Q1	Q2	Q3	Q4
Q1 ($\overline{AR} = 32\%$)	–	0.0001***	0.0001***	0.0001***
Q2 ($\overline{AR} = 36\%$)	–	–	0.0001***	0.0001***
Q3 ($\overline{AR} = 41\%$)	–	–	–	0.0001***
Q4 ($\overline{AR} = 46\%$)	–	–	–	–

¹ \overline{AR} is the average attack rate computed from 30 simulation replicates

²* $p < .05$, ** $p < .01$, *** $p < .001$

Table 7 Average household size of income groups

Income quartiles	Average household size
Q1	2.00
Q2	2.19
Q3	2.44
Q4	2.85

Table 8 Disparity analysis of death rate among income groups: p-value (base-case)

Income quartiles	Q1	Q2	Q3	Q4
Q1 ($\overline{DR} = 0.079\%$)	–	0.1	0.02*	0.4
Q2 ($\overline{DR} = 0.087\%$)	–	–	0.5	0.3
Q3 ($\overline{DR} = 0.090\%$)	–	–	–	0.1
Q4 ($\overline{DR} = 0.082\%$)	–	–	–	–

¹ \overline{DR} is the average death rate computed from 30 simulation replicates

²* $p < .05$, ** $p < .01$, *** $p < .001$

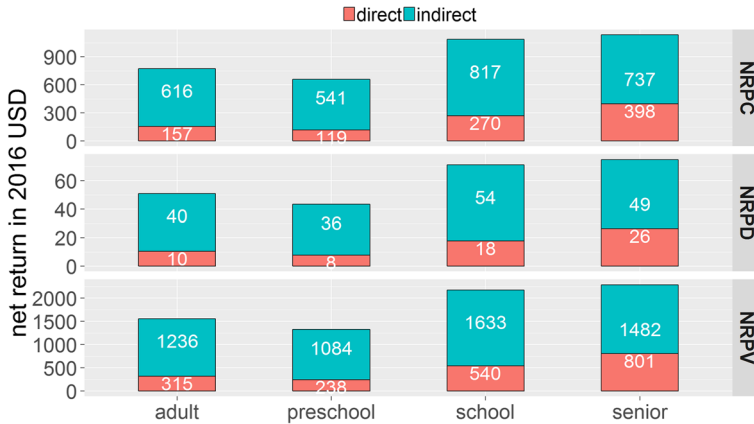


Figure 4 Net return per capita (*NRPC*), per dollar spent (*NRPD*), and per vaccinated person (*NRPV*). The bar height represents value of net return within each one. Each total net return (blue + red) is the sum of indirect net return (upper blue bar) and direct net return (lower red bar). Indirect net returns are significantly larger than direct net returns across all age groups

return. This is because the indirect cost, i.e. loss in wages, from death is much greater than the direct (medical) cost, for all individuals regardless of age and risk level, as shown in Table 21. School-aged and senior groups have significantly higher net return compared to the other two age groups because vaccine intervention reduces the attack rate in school-aged group and the death rate in senior group more. This observation is reflected in all per capita calculations as shown in Table 9. Vaccinations to school-aged and senior groups are economically more beneficial than to preschoolers and adults. However, this does not mean that vaccination based intervention is always more effective in school-aged and senior groups in reducing the attack rate. We will explain this later in the sensitivity analysis section.

Figure 5 shows distributions of death, hospitalization, outpatient, and ill not seeking medical care among the infected cases, in each age group, and Figure 6 shows the corresponding cost distributions. Note that for each age group, although the percentage of death count (Figure 5) is the smallest, the cost of death (in Figure 6) is the largest. This is because the economic cost of death includes lifetime productivity and hence is higher than the cost for any other clinical outcome (see Table 21).

Table 9 Disparity analysis of the net return per dollar spent among age groups: p-value (intervention case)

NRPD by age group	Preschool	School	Adult	Senior
Preschool ($\overline{NRPD} = 44.31$)	–	0.0001***	0.14	0.0001***
School ($\overline{NRPD} = 72.23$)	–	–	0.0001***	0.44
Adult ($\overline{NRPD} = 51.80$)	–	–	–	0.0001***
Senior ($\overline{NRPD} = 75.79$)	–	–	–	–

¹* $p < .05$, ** $p < .01$, *** $p < .001$

² \overline{NRPD} is computed over 30 simulation replicates

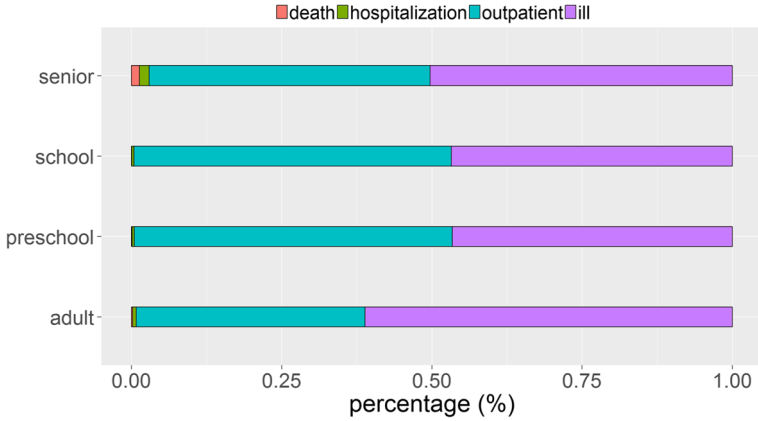


Figure 5 Count distribution of death (first from left in red), hospitalization (second from left in green), outpatient (third from left in blue), and ill but not seeking medical care (forth from left in purple), among the infected cases in each age group. For each age group, the percentage of death count is the smallest

Income-based groups Unlike age groups, the disparity in net return per dollar spent, among income groups is not as significant, see Table 10. The only exceptions are first income quartile versus third and fourth income quartiles.

3.4 Sensitivity analysis

Next we perform a sensitivity analysis on compliance rate and attack rate (by adjusting the disease transmission rate) to see whether our findings are robust to these changes.

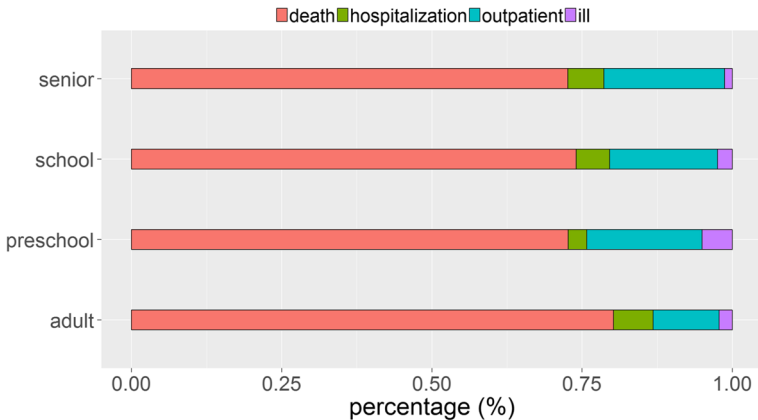


Figure 6 Cost distribution of death (first from left in red), hospitalization (second from left in green), outpatient (third from left in blue), and ill but not seeking medical care (forth from left in purple) in each age group. For each age group, the cost of death is the largest

Table 10 Disparity analysis of net return per dollar among income groups: p-value (intervention case)

Income quartiles	Q1	Q2	Q3	Q4
Q1 ($\overline{NRPD} = 51.57$)	–	0.3	0.0051**	0.0069**
Q2 ($\overline{NRPD} = 54.93$)	–	–	0.0999	0.1278
Q3 ($\overline{NRPD} = 60.40$)	–	–	–	0.8689
Q4 ($\overline{NRPD} = 59.88$)	–	–	–	–

¹* $p < .05$, ** $p < .01$, *** $p < .001$

² \overline{NRPD} is computed over 30 simulation replicates

3.4.1 Compliance rate

The simulation settings remain the same as described in Section 3.1 except that compliance to vaccination is reduced from 50% to 25% now. The economic returns of both school-aged group and senior group were significantly higher than for preschoolers and adults at 50% compliance rate, but at lower levels of compliance, the senior group provides the highest benefits, see Table 11. The reason is that the attack rate drops by 66% in senior group and only by 51% among the school-aged children as shown in Table 12. The lower impact on attack rate among school-aged group is probably due to the fact that lower compliance to vaccination prevents this group from gaining herd immunity and the indirect gains from vaccination are disproportionately reduced. The economic returns are not found to be significantly different across preschoolers, schoolers, and adults under 25% compliance.

3.4.2 Attack rate

Next we change the attack rate from 40% to 60%, corresponding to disease transmission rate of 0.00011. All other simulation parameters remain the same as before. We consider both 50% and 25% compliance rate to vaccination. We find the results are qualitatively the same as they are for 40% attack rate and are omitted here to avoid duplication.

The sensitivity analysis results show that the following findings are robust to simulation settings: 1) disparities in health and clinical outcomes exist among age and income groups, but disparities in economic return are only significant with respect to age; 2) school age group is much more vulnerable to influenza infection than other age groups due to their higher contact rates; 3) in random allocation, the vaccination of school-aged group is the most effective in containing the disease except when compliance to intervention is low; 4) deaths have the biggest impact on net return.

Table 11 Disparity analysis of the net return per dollar among age groups: p-value (CMPL=25%)

NRPD by age group	Preschool	School	Adult	Senior
Preschool ($\overline{NRPD} = 59.38$)	–	0.058	0.13	0.0001***
School ($\overline{NRPD} = 79.65$)	–	–	0.28	0.012*
Adult ($\overline{NRPD} = 66.97$)	–	–	–	0.0001***
Senior ($\overline{NRPD} = 106.65$)	–	–	–	–

¹* $p < .05$, ** $p < .01$, *** $p < .001$

² \overline{NRPD} is computed over 30 simulation replicates

Table 12 Reduction in attack rate (%)

	AR (base)	CMPL=50%		CMPL=25%	
		AR	Reduction	AR	Reduction
Preschool	44	0.46	99	17	61
School	70	1.09	98	34	51
Adult	34	0.29	99	12	64
Senior	22	0.18	99	7	66

3.5 Vaccination prioritization strategy

If vaccines are limited it is important to prioritize their use based on specific objectives. Here we consider a simple heuristic for determining priorities based on various objectives, which include: minimizing the total number of deaths, minimizing death rate, maximizing net return per dollar spent, maximizing total net returns, and minimizing attack rate. Our heuristic assigns higher priority to a subpopulation that has a higher value of the measure we want to minimize (maximize). For example, if the objective is to minimize death rate, then we prioritize age groups according to their death rates, i.e. the age group with a higher death rate is given a higher priority. Given that there is no significant difference in death rate and net return among income-based groups, we only consider age-based groups for deciding vaccination priorities. We use the simulation results based on the setting given in Section 3.1 to determine the vaccination priorities for age-based groups. The results are shown in Table 13.

3.5.1 Fixed compliance rate across age groups

We first assume the vaccination coverage is 50% and the compliance rate of each cohort is 1. The number of vaccines used is limited to 50% of the population size which is $77,820 * 50\% = 38,910$ for the Montgomery county. Table 14 shows the various prioritization strategies with the actual vaccinated fractions under different objectives. The strategies S_{4321} , S_{3421} , S_{2134} , S_{4231} , and S_{3241} correspond to minimizing \overline{DR} , \overline{DC} , \overline{AR} , and maximizing \overline{NRPD} , \overline{TNR} , where subscript numbers of each strategy describe vaccination prioritization of the four age groups. The column values in Table 14 under each strategy

Table 13 Priorities for age-based groups

Priority	Criteria for prioritization				
	\overline{DR}	\overline{DC}	\overline{NRPD}	\overline{TNR}	\overline{AR}
1(top)	senior	adult	senior	adult	school
2	adult	senior	school	school	preschool
3	school	school	adult	senior	adult
4(bottom)	preschool	preschool	preschool	preschool	senior

¹ \overline{DR} - average death rate; \overline{DC} - average death count; \overline{NRPD} - average net return per dollar spent; \overline{TNR} - average total net return; \overline{AR} - average attack rate. \overline{DR} , \overline{NRPD} and \overline{AR} are normalized values that are comparable across subgroups

Table 14 Vaccinated fraction of each age group under different vaccination strategies

Subgroup (size)	S_{np}	S_{4321}	S_{3421}	S_{2134}	S_{4231}	S_{3241}
Preschool (4617)	0.5	0	0	1	0	0
School (13310)	0.5	0	0	1	1	0
Adult (52558)	0.5	0.6	0.74	0.4	0.35	0.74
Senior (7335)	0.5	1	0	0	1	0

1-preschool, 2-school, 3-adult, 4-senior; S_{np} denotes strategy with no priority; S_{4321} denotes vaccinating seniors first, then adults, then school-aged and then preschoolers. Similarly strategies S_{3421} , S_{2134} , S_{4231} , S_{3241} follow priorities given by their subscripts. Note that S_{3421} and S_{3241} result in the same allocation because adult group has the first priority and it has more people than the size of the vaccine stockpile so no other age groups could be vaccinated

represent the actual vaccinated fractions of the corresponding age groups under that strategy. For example S_{4321} column, which corresponds to minimizing death rate, shows that vaccines are prioritized according to the following ordering: seniors, adults, school-aged children, and preschoolers. Seniors are vaccinated first which uses 7,335 vaccines to cover 100% of the senior group. The remaining 31,575 vaccines are distributed to adults, which cover only 60% of the adults. None of the children receive any vaccines in this case. Strategy with no priority (S_{np}) is introduced as a baseline in which 50% of each age group is selected uniformly at random for vaccination.

To compare the performance of strategies, we compute death rate, death count, net return per dollar spent, total net return, and attack rate over the whole population. Note that death rate, net return per dollar spent, and attack rate are normalized values and hence are directly comparable. The results are shown in Table 15. S_{2134} turns out to be the optimal strategy across all objectives. Followed by S_{4231} and S_{np} , whose results are also competitive. These three strategies are significantly better than the other ones. Comparison of S_{2134} , S_{4231} , and S_{np} with other strategies shows that they have a higher vaccinated fraction of school-aged group.

We conduct the same analysis under other settings to determine optimal vaccination priorities; these include attack rate=40%, compliance=0.25; attack rate=60%, compliance=0.5; attack rate=60%, compliance=0.25. We observe that the optimal strategy always has the highest number of school age people vaccinated.

S_{np} in Table 15 performs very close to the optimal strategy. Similar observations were made in other settings, except when attack rate=60% and compliance=0.25. S_{np} performs

Table 15 Performance of different vaccination strategies

Strategy	$\overline{DR}(\%)$	\overline{DC}	$\overline{TNR}(\text{M}\$)$	$\overline{NRPD}(\text{\$})$	$\overline{AR}(\%)$
S_{np}	0.000900	0.7	67.64	57.12	0.43
S_{4321}	0.015806	12	47.11	38.46	14.64
S_{3421}	0.017305	13	48.44	39.13	13.20
S_{2134}	0.000600	0.47	68.15	59.69	0.29
S_{4231}	0.000643	0.50	67.88	59.54	0.42
S_{3241}	0.017305	13	48.44	39.13	13.20

Low values of $\overline{DR}(\%)$, \overline{DC} , $\overline{AR}(\%)$ are desirable and high values of $\overline{NRPD}(\text{\$})$, \overline{TNR} (million $\text{\$}$) are desirable. Values in bold are optimal

Table 16 Compliance rates based on age

Subgroup	CMPL
Preschool	0.65
School	0.53
Adult	0.54
Senior	0.66

poorly when the disease transmission rate is high and the compliance rate is low. This indicates that vaccination priority is particularly important when vaccines are in short supply and the virus strain is highly infective.

3.5.2 Age based compliance rates

In previous section, we assume the compliance rate is the same across all age groups. However, this is unlikely to be the case so we estimate compliance rates based on age (showed in Table 16). The compliance rates are extracted from a survey conducted by Gfk.com, under the National Institute of Health grant no. 1R01GM109718. This survey collects data on demographics of the respondents and their preventive health behaviors during a hypothetical influenza outbreak. We categorize survey respondents into four age groups corresponding to the age groups used in this paper and estimate their compliance to vaccination.

The new average compliance rate is 52%. If we still assume the vaccination coverage is 50%, the vaccine supply will be enough to meet the needs of almost everyone, so in order to observe the effect of different vaccine prioritization, we assume the vaccination coverage is 30%, i.e. only 43288 vaccines are available. The vaccinated fraction table (similar to Table 14) is shown in Table 17. It considers both vaccination coverage and compliance rate across each age group. S_{3421} is exactly the same with S_{3241} .

The results are listed in Table 18. The death rate, net return per dollar spent, and attack rate are directly comparable. We can still observe that the optimal strategy is S_{2134} in terms of all measures and following by S_{4231} and S_{np} . The observation is consistent with the former results where the compliance rate is assumed to be uniform.

Table 17 Vaccinated fraction of each age group under different vaccination strategies (based on age based compliance)

Subgroup (size)	S_{np}	S_{4321}	S_{3421}	S_{2134}	S_{4231}	S_{3241}
Preschool (4617)	0.3	0	0	0.65	0	0
School (13310)	0.3	0	0	0.53	0.53	0
Adult (52558)	0.3	0.35	0.44	0.25	0.22	0.44
Senior (7335)	0.3	0.66	0	0	0.66	0

1-preschool, 2-school, 3-adult, 4-senior; S_{np} denotes strategy with no priority; S_{4321} denotes sequentially vaccinating seniors first, then adults, then school-aged and then preschoolers, according to their priorities. If after vaccines are given to an age group based on its compliance rate, there are vaccines remaining, then the next age group on the priority list is considered. Similarly strategies S_{3421} , S_{2134} , S_{4231} , S_{3241} follow priorities given by their subscripts. Note that S_{3421} and S_{3241} result in the same allocation because adult group has the first priority and it has more people than the size of the vaccine stockpile so no other age group could be vaccinated

Table 18 Performance of different vaccination strategies

Strategy	$\overline{DR}(\%)$	\overline{DC}	$\overline{TNR}(\text{M}\$)$	$\overline{NRPD}(\$)$	$\overline{AR}(\%)$
S_{np}	0.020946	16.30	89.85	126.57	11.18
S_{4321}	0.029427	22.90	76.47	104.50	21.02
S_{3421}	0.032254	25.10	79.35	107.95	18.40
S_{2134}	0.009766	7.60	102.07	148.69	3.61
S_{4231}	0.013279	10.33	97.38	141.12	7.79
S_{3241}	0.032254	25.10	79.35	107.95	18.40

Low values of $\overline{DR}(\%)$, \overline{DC} , $\overline{AR}(\%)$ are desirable and high values of $\overline{NRPD}(\$)$, \overline{TNR} (million \$) are desirable. Values in bold are optimal

3.6 Public health policy implications

The public health authorities often publish recommendations of interventions to the population. In case of limited resources, e.g. when vaccines are not enough to satisfy the needs of the whole population, some subpopulations are given higher priorities, to optimize the efficiency of interventions towards specific objectives of containing the epidemic.

Through the simulation results, we observe that school age group is much more critical than other age groups, and a vaccination strategy targeting school age group first can maximize net returns and improve herd immunity. School-aged children tend to have longer contact times with each other and form dense subnetworks at school locations; thus the epidemic spreads faster and more easily within the school age group. If the school age group receives no vaccine or few vaccines, the highly connected subnetworks of students continue to be able to spread the disease among students, impeding the intervention strategy to effectively reduce overall infections and costs. On the other hand, if the school-aged group is well vaccinated, the epidemic is effectively contained not only in this group but across all groups.

3.7 Limitations

This computational framework makes several assumptions about parameter settings. For example, in health and economic disparity study, it assumes that individuals follow the same compliance rate within a cohort and it is forced to be the same across cohorts. In vaccination strategy study, we relax the assumption and make the compliance rate follow a realistic distribution based on age. However, these parameter settings may not correspond to any specific scenario in the real world. Our goal is to provide a framework that can be used to study hypothetical scenarios to support planning, response and decision-making during epidemic outbreaks.

4 Conclusions

In this paper, we build a computational framework that provides a generic methodology for finding health and economic disparities among cohorts based on individual level features. We apply this framework to study disparities with respect to age and income during an influenza epidemic in the Montgomery county of Southwest Virginia. In one scenario,

no interventions are applied to control the epidemic and in other scenarios vaccine-based intervention is applied to cohorts that are split by age and income.

We use an agent based disease propagation model to study the transmission dynamics of influenza on the social network. The infected cases are then assigned different clinical outcomes and cost of treatment depending on the severity of clinical outcome. Various vaccine allocation strategies are designed and simulated in this study.

Our results show significant health disparities across age groups and income groups, and economic disparities among age groups. The metrics for measuring the outcome disparities include attack rate, death rate, and death count. The metrics for measuring economic disparities include net return per capita, net return per vaccinated, and net return per dollar spent. We also find that in a severe flu season with a limited vaccination coverage, if vaccines are assigned randomly without priorities, then the intervention is least effective for the school aged cohort. Given the high connectivity of school aged children in the social contact network, they will be at a disadvantage if the vaccine assignment is completely random. After measuring performance of vaccination prioritization strategies under both simulated and realistic scenarios, the strategy that prioritizes the school aged children maximizes the normalized net returns and minimizes attack rate and death rate.

Acknowledgments This work has been partially supported by the National Institutes of Health (NIH) (grant number 1R01GM109718), NSF Research Traineeship (grant number NRT-DESE-154362), Defense Threat Reduction Agency (DTRA) (grant number HDTRA1-11-1-0016), and DTRA Comprehensive National Incident Management System contracts (grant number HDTRA1-11-D-0016-0001, grant number HDTRA1-17-D-0023).

Appendix: Supplementary data

Distribution of risk level

The high risk individuals encounter a different distribution of clinical outcomes than non-high risk individuals. For instance, high-risk individuals have a higher probability of hospitalization or death compared to non-high risk individuals. High-risk can be caused by many factors; it could be due to preexisting conditions; age; occupation e.g. health care worker or emergency worker; immuno-compromised health condition; pregnancy etc. We determine high risk or non-high risk based on the data from Meltzer et al. [32].

Distribution of health outcomes

The clinical outcomes resulting from influenza infection are: death; hospitalization; outpatient; and ill but not seeking medical care. Work by Meltzer et al. [32] provides the distribution of health states per 1,000 people and clinical outcome rates per 1,000 cases. We

Table 19 Percentage of population at high/non-high risk level(%)

Age	High risk	Non-high risk
0-19	6.4	93.6
20-64	14.4	85.6
65+	40	60

Table 20 Clinical outcome rates for high-risk and non-high risk (%)

Variable	Death	Hospitalization	Outpatient	Ill
High risk				
0-19	0.76	1.37	89.15	8.72
20-64	1.29	1.46	61.50	35.76
65+	2.63	5.09	66.90	25.39
Non-high risk				
0-19	0.01	0.37	50.95	48.66
20-64	0.03	0.68	35.15	64.14
65+	0.34	1.42	38.20	60.04

use our simulation model to determine the health states and case counts but use the clinical outcome rates given by [32] for age groups (0 to 19 years, 20 to 64 years, and 65 years and older) and risk conditions (high risk and non-high risk). Table 20 shows the outcome rates on age groups and risk conditions.

Cost of health outcomes

Data on costs are collected from Carias et al. [12] for different clinical outcomes for cohorts characterized by age and underlying risk condition (Table 21). The population was divided into age categories that are distinct with respect to economic activity and health care-related costs: 6 months-4 year olds; 5-19 year olds; 20-64 year olds; and above 65 years. For each age cohort, costs were calculated for high-risk and non-high risk individuals.

Table 21 Economic cost (\$)

Variable	Direct (medical cost)	Indirect (productivity losses)	Total
Death High risk			
0-4	51078.87	1391426.07	1442504.94
5-19	208758.81	1651810.98	1860569.79
20-64	79726.86	862321.26	942048.12
65+	49734.66	240579.18	290313.84
Non-high risk			
0-4	54766.29	1391426.07	1446192.36
5-19	185465.46	1651810.98	1837276.44
20-64	73453.14	862321.26	935774.4
65+	36642.21	240579.18	277221.39
Hospitalization High risk			
0-4	23230.08	1165.5	24395.58
5-19	44744.1	1665	46409.1
20-64	32776.08	1998	34774.08
65+	21199.89	1298.7	22498.59

Table 21 (continued)

Variable	Direct (medical cost)	Indirect (productivity losses)	Total
Non-high risk			
0-4	10422.9	832.5	11255.4
5-19	21376.38	1165.5	22541.88
20-64	25037.16	1332	26369.16
65+	14084.79	909.09	14993.88
Outpatient High risk			
0-4	669.33	166.5	835.83
5-19	935.73	333	1268.73
20-64	738.15	166.5	904.65
65+	2877.12	259.74	3136.86
Non-high risk			
0-4	376.29	166.5	542.79
5-19	307.47	166.5	473.97
20-64	468.42	166.5	634.92
65+	1023.42	259.74	1283.16
Ill (no medical care) High risk			
0-4	4.44	166.5	170.94
5-19	4.44	83.25	87.69
20-64	4.44	83.25	87.69
65+	4.44	129.87	134.31
Non-high risk			
0-4	4.44	166.5	170.94
5-19	4.44	83.25	87.69
20-64	4.44	83.25	87.69
65+	4.44	129.87	134.31

In Table 21, direct costs are the medical costs associated with the clinical outcome as well as the cost of vaccination. Indirect costs are the productivity loss due to illness. The total costs are the sum of direct and indirect costs. All costs are in 2016 USD.

References

1. Badham, J., Stocker, R.: The impact of network clustering and assortativity on epidemic behaviour. *Theor. Popul. Biol.* **77**(1), 71–75 (2010)
2. Bailey, N.T.J.: *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd edn. Griffin (1975)
3. Barrett, C.L., Beckman, R.J., Khan, M., Kumar, V.S.A., Marathe, M.V., Stretz, P.E., Dutta, T., Lewis, B.: Generation and analysis of large synthetic social contact networks. In: *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pp. 1003–1014 (2009)
4. Barrett, C., Bisset, K., Leidig, J., Marathe, A., Marathe, M.: Economic and social impact of influenza mitigation strategies by demographic class. *Epidemics* **3**(1), 19–31 (2011)
5. Beckman, R.J., Baggerly, K.A., McKay, M.D.: Creating synthetic baseline populations. *Transp. Res. A Policy Pract.* **30**(6), 415–429 (1996)

6. Biggerstaff, M., Reed, C., Swerdlow, D.L., Gambhir, M., Graitcer, S., Finelli, L., Borse, R.H., Rasmussen, S.A., Meltzer, M.I., Bridges, C.B.: Estimating the potential effects of a vaccine program against an emerging influenza pandemic, United States. *Clin. Infect. Diseases*. **60**, S20–S29 (2015)
7. Bisset, K., Marathe, M.: A cyber-environment to support pandemic planning and response. *DOE SciDAC Mag.*, 36–47 (2009)
8. Bisset, K.R., Chen, J., Feng, X., Kumar, V.A., Marathe, M.V.: EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In: *Proceedings of the 23rd International Conference on Supercomputing, ICS '09*, pp. 430–439. ACM, New York (2009)
9. Borse, R.H., Shrestha, S.S., Fiore, A.E., Atkins, C.Y., Singleton, J.A., Furlow, C., Meltzer, M.I.: Effects of vaccine program against pandemic influenza A(H1N1) virus. United States, 2009–2010 **19**(3), 439 (2013)
10. Bridges, C.B., Thompson, W.W., Meltzer, M.I., Reeve, G.R., Talamonti, W.J., Cox, N.J., Lilac, H.A., Hall, H., Klimov, A., Fukuda, K.: Effectiveness and cost-benefit of influenza vaccination of healthy working adults: A randomized controlled trial. *JAMA* **284**(13), 1655–1663 (2000)
11. Bureau of Labor Statistics: American time use survey. <https://www.bls.gov/tus/> (2017)
12. Carias, C., Reed, C., Kim, I.K., Foppa, I.M., Biggerstaff, M., Meltzer, M.I., Finelli, L., Swerdlow, D.L.: Net costs due to seasonal influenza vaccination—united states, 2005–2009. *PLoS one* **10**(7), e0132,922 (2015)
13. Carrat, F., Vergu, E., Ferguson, N.M., Lemaître, M., Cauchemez, S., Leach, S., Valleron, A.J.: Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *Am. J. Epidemiol.* **167**(7), 775–85 (2008)
14. Chen, J., Chu, S., Chungbaek, Y., Khan, M., Kuhlman, C., Marathe, A., Mortveit, H., Vullikanti, A., Xie, D.: Effect of modelling slum populations on influenza spread in Delhi. *BMJ Open* **6**, 9 (2016). <https://doi.org/10.1136/bmjopen-2016-011699>
15. Committee NVA: Strategies to achieve the healthy people 2020 annual influenza vaccine coverage goal for health-care personnel: Recommendations from the national vaccine advisory committee. *Public Health Reports* **128**(1), 7–25 (2013). pMCID: PMC3514716
16. Dorratoltaj, N., Marathe, A., Lewis, B.L., Swarup, S., Eubank, S.G., Abbas, K.M.: Epidemiological and economic impact of pandemic influenza in chicago: Priorities for vaccine interventions. *PLoS Comput. Biol.* **13**(6), e1005,521 (2017)
17. Eichner, M., Schwehm, M., Eichner, L., Gerlier, L.: Direct and indirect effects of influenza vaccination. *BMC Infect. Dis.* **17**(1), 308 (2017)
18. Eubank, S., Guclu, H., Kumar, V.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. *Nature* **429**(6988), 180–184 (2004)
19. Glezen, W.P., Gaglani, M.J., Kozinetz, C.A., Piedra, P.A.: Direct and indirect effectiveness of influenza vaccination delivered to children at school preceding an epidemic caused by 3 new influenza virus variants. *J. Infect. Dis.* **202**(11), 1626–1633 (2010)
20. Grohskopf, L.A., Sokolow, L.Z., Broder, K.R., Olsen, S.J., Karron, R.A., Jernigan, D.B., Bresee, J.S.: Prevention and control of seasonal influenza with vaccines. *MMWR Recommend. Rep.* **65**(5), 1–54 (2016)
21. Halloran, M.E., Longini, I.M., Cowart, D.M., Nizam, A.: Community interventions and the epidemic prevention potential. *Vaccine* **20**(27–28), 3254–3262 (2002)
22. Halloran, E.M., Ferguson, N.M., Eubank, S., Longini, I.M., Cummings, D.A., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T.C., et al.: Modeling targeted layered containment of an influenza pandemic in the united states. *Proc. Natl. Acad. Sci.* **105**(12), 4639–4644 (2008)
23. Health People: Disparities. Available at <https://www.healthypeople.gov/2020/about/foundation-health-measures/disparities>, Accessed 01 Sept 2016 (2016)
24. Kang, G.J., Ewing-Nelson, S.R., Mackey, L., Schlitt, J.T., Marathe, A., Abbas, K.M., Swarup, S.: Semantic network analysis of vaccine sentiment in online social media. *Vaccine* (2017)
25. King, J.C.J., Beckett, D., Snyder, J., Cummings, G.E., King, B.S., Magder, L.S.: Direct and indirect impact of influenza vaccination of young children on school absenteeism. *Vaccine* **30**(2), 289–293 (2012)
26. Kostova, D., Reed, C., Finelli, L., Cheng, P.Y., Gargiullo, P.M., Shay, D.K., Singleton, J.A., Meltzer, M.I., Lu, P.J., Bresee, J.S.: Influenza illness and hospitalizations averted by influenza vaccination in the united states, 2005–2011. *PLoS ONE* **8**(6), 1–8 (2013)
27. Kuznetsov, Y.A., Piccardi, C.: Bifurcation analysis of periodic seir and sir epidemic models. *J. Math. Biol.* **32**(2), 109–121 (1994). <https://doi.org/10.1007/BF00163027>
28. Lee, P.Y., Matchar, D.B., Clements, D.A., Huber, J., Hamilton, J.D., Peterson, E.D.: Economic analysis of influenza vaccination and antiviral treatment for healthy working adults. *Ann. Intern. Med.* **137**(4), 225–231 (2002)

29. Longini, I.M., Halloran, M.E., Nizam, A., Wolff, M., Mendelman, P.M., Fast, P.E., Belshe, R.B.: Estimation of the efficacy of live, attenuated influenza vaccine from a two-year, multi-center vaccine trial: Implications for influenza epidemic control. *Vaccine* **18**(18), 1902–1909 (2000)
30. Marathe, A., Lewis, B., Barrett, C., Chen, J., Marathe, M., Eubank, S., Ma, Y.: Comparing effectiveness of top-down and bottom-up strategies in containing influenza. *PLoS ONE* **6**, e25,149 (2011)
31. Marathe, A., Lewis, B., Chen, J., Eubank, S.: Sensitivity of household transmission to household contact structure and size. *PLoS ONE* **6**, pMCID: PMC3148222 (2011)
32. Meltzer, M.I., Cox, N.J., Fukuda, K.: The economic impact of pandemic influenza in the united states: priorities for intervention. *Emerg. Infect. Dis.* **5**(5), 659–671 (1999)
33. Meltzer, M.I., Neuzil, K.M., Griffin, M.R., Fukuda, K.: An economic analysis of annual influenza vaccination of children. *Vaccine* **23**(8), 1004–1014 (2005)
34. Molinari, N.A.M., Ortega-Sanchez, I.R., Messonnier, M.L., Thompson, W.W., Wortley, P.M., Weintraub, E., Bridges, C.B.: The annual impact of seasonal influenza in the us: Measuring disease burden and costs. *Vaccine* **25**(27), 5086–5096 (2007)
35. Monto, A.S., Davenport, F.M., Napier, J.A., Francis, T.J.: Effect of vaccination of a school-age population upon the course of an A2-Hong Kong influenza epidemic. *Bull World Health Organ* **41**(3), 537–542 (1969)
36. Moreno, Y., Pastor-Satorras, R., Vespignani, A.: Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B - Cond. Matter Complex Syst.* **26**(4), 521–529 (2002)
37. Mullooly, J.P., Bennett, M.D., Hornbrook, M.C., Barker, W.H., Williams, W.W., Patriarca, P.A., Rhodes, P.H.: Influenza vaccination programs for elderly persons: Cost-effectiveness in a health maintenance organization. *Ann. Intern. Med.* **121**(12), 947–952 (1994)
38. NDSSL: Synthetic data of Montgomery county, Virginia. <http://ndssl.vbi.vt.edu/synthetic-data/> (2014)
39. Presanis, A.M., De Angelis, D., Hagy, A., Reed, C., Riley, S., Cooper, B.S., Finelli, L., Biedrzycki, P., Lipsitch, M.: The New York City Swine Flu Investigation Team (2009) The severity of pandemic H1N1 influenza in the United States, from April to July 2009. A Bayesian analysis. *PLoS Med* **6**(12), 1–12 (2009)
40. Prosser, L.A., Bridges, C.B., Uyeki, T.M., Hinrichsen, V.L., Meltzer, M.I., Molinari, N.A.M., Schwartz, B., Thompson, W.W., Fukuda, K., Lieu, T.A.: Health benefits, risks, and cost-effectiveness of influenza vaccination of children **12**(10), 1548 (2006)
41. Reed, C., Chaves, S.S., Daily Kirley, P., Emerson, R., Aragon, D., Hancock, E.B., Butler, L., Baumbach, J., Hollick, G., Bennett, N.M., Laidler, M.R., Thomas, A., Meltzer, M.I., Finelli, L.: Estimating influenza disease burden from population-based surveillance data in the united states. *PLoS ONE* **10**(3), 1–13 (2015)
42. Sander, B., Nizam, A., Garrison, L.P., Postma, M.J., Halloran, M.E., Longini, I.M.: Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation transmission model. *Value Health* **12**(2), 226–233 (2009)
43. Sydenstricker, E.: The incidence of influenza among persons of different economic status during the epidemic of 1918. 1931. *Public Health Rep* **121** Suppl **1**, 191–204 (2006)
44. Tsang, T.K., Lau, L.L., Cauchemez, S., Cowling, B.J.: Household transmission of influenza virus. *Trends Microbiol.* **24**(2), 123–133 (2016). <https://doi.org/10.1016/j.tim.2015.10.012>
45. WHO: Influenza (season) fact sheet. <http://www.who.int/mediacentre/factsheets/fs211/en/>, Accessed 01 Sept 2016 (2016)