

# Distribution of vaccine/antivirals and the 'least spread line' in a stratified population

E. Goldstein, A. Apolloni, B. Lewis, J. C. Miller, M. Macauley, S. Eubank, M. Lipsitch and J. Wallinga

*J. R. Soc. Interface* 2010 **7**, 755-764 first published online 14 October 2009 doi: 10.1098/rsif.2009.0393

Supplementary data	"Data Supplement" http://rsif.royalsocietypublishing.org/content/suppl/2009/10/12/rsif.2009.0393.DC1.htm I
References	This article cites 14 articles, 3 of which can be accessed free http://rsif.royalsocietypublishing.org/content/7/46/755.full.html#ref-list-1
	Article cited in: http://rsif.royalsocietypublishing.org/content/7/46/755.full.html#related-urls
Rapid response	Respond to this article http://rsif.royalsocietypublishing.org/letters/submit/royinterface;7/46/755
Subject collections	Articles on similar topics can be found in the following collections
	biomathematics (112 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here

To subscribe to J. R. Soc. Interface go to: http://rsif.royalsocietypublishing.org/subscriptions



## Distribution of vaccine/antivirals and the 'least spread line' in a stratified population

E. Goldstein<sup>1,\*</sup>, A. Apolloni<sup>2</sup>, B. Lewis<sup>2</sup>, J. C. Miller<sup>1</sup>, M. Macauley<sup>5</sup>, S. Eubank<sup>2</sup>, M. Lipsitch<sup>1,†</sup> and J. Wallinga<sup>3,4,†</sup>

<sup>1</sup>Harvard School of Public Health, Boston, MA 02115, USA
 <sup>2</sup>Virginia Bioinformatics Institute, Blacksburg, VA 24061, USA
 <sup>3</sup>RIVM, 3720 BA, Bilthoven, The Netherlands
 <sup>4</sup>University Medical Center, 3508 GA, Utrecht, The Netherlands
 <sup>5</sup>Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA

We describe a prioritization scheme for an allocation of a sizeable quantity of vaccine or antivirals in a stratified population. The scheme builds on an optimal strategy for reducing the epidemic's initial growth rate in a stratified mass-action model. The strategy is tested on the EpiSims network describing interactions and influenza dynamics in the population of Utah, where the stratification we have chosen is by age (0-6, 7-13, 14-18, adults). No prior immunity information is available, thus everyone is assumed to be susceptible—this may be relevant, possibly with the exception of persons over 50, to the 2009 H1N1 influenza outbreak. We have found that the top priority in an allocation of a sizeable quantity of seasonal influenza vaccinations goes to young children (0-6), followed by teens (14-18), then children (7-13), with the adult share being quite low. These results, which rely on the structure of the EpiSims network, are compared with the current influenza vaccination coverage levels in the US population.

Keywords: vaccine; stratified population; optimal allocation; antiviral

### 1. INTRODUCTION

Prioritization for influenza vaccinations and improvement in coverage levels continue to be the subjects of investigation (Longini & Halloran 2005; Patel et al. 2005; Halloran & Longini 2006; Dushoff et al. 2007; Hadeler & Mueller 2007). Until 2008, the priority groups for influenza vaccination allocation according to the US Center for Disease Control (CDC) guidelines were children 0.5-4 years old, adults over 50 as well as people with certain medical conditions. Data up to 2007 suggest that adults aged 18-49 had coverage levels on par with children aged 5-17 (CDC 2008). In 2008, all children aged 0.5-17 were included in the priority group, with no data available yet on the effect of the recent prioritization policy on coverage levels. For the 2009 H1N1 vaccination plan, individuals aged 0.5-24 make up one of the priority cohorts for vaccine allocation (CDC 2009).

Motivated in part by the need for evidence-based guidance on those issues, we have devised a method to prioritize the allocation of a sizeable quantity of a vaccine or antivirals in a stratified population. As our strategy is identical for vaccine and antivirals (see the

\*Author for correspondence (egoldste@hsph.harvard.edu).  $^{\dagger}\text{Co-senior}$  author.

Electronic supplementary material is available at http://dx.doi.org/ 10.1098/rsif.2009.0393 or via http://rsif.royalsocietypublishing.org.

electronic supplementary material), we concentrate on vaccine for the rest of this paper. The setting we consider in this paper is a large population, where interactions between people and influenza dynamics are represented by a network; the explicit example we use is the EpiSims network describing the population of Utah (Barrett et al. 2008). The population is stratified (usually by age), and vaccine is distributed between the strata and then allocated at random within each stratum—thus, we cannot relate individuals who receive the vaccine to particular nodes on a network describing the population. We use simulations on the network to estimate the next generation matrix, whose entries represent the number of persons in one stratum infected by an average infectious person in another stratum during the exponential growth period of the epidemic. Using this matrix, we present an explicit algorithm for an allocation of a sufficiently large quantity of a vaccine between the population strata, namely we specify the (generally unequal) coverage levels in each stratum as a function of the total vaccine quantity available. In the process we also determine prioritization (ranking) for vaccine allocation among the strata—the available vaccine is distributed simultaneously for all strata, but the strata with the higher ranking get higher coverage levels. The latter prioritization may be the most important practical outcome of our approach given a number of uncertainties related

to the exact allocation proportions, such as the total number of people who will get vaccinated, and the precision with which a network describes a real population. We then apply our algorithm to the EpiSims network for the population of Utah, and test its optimality by simulations.

The theoretical framework for our strategy is the optimal vaccination policy for reducing the epidemic's reproductive number in a stratified mass-action model. Equivalently, the initial growth rate of the infection is minimized when the infection is introduced into a population vaccinated (or treated) according to this allocation—all this is explained in detail in the electronic supplementary material. We note that the emphasis is not on the smallest quantity of a vaccine needed to get the reproductive number below 1 (such a quantity may be unavailable, and the precise reproductive number of the epidemic may be unknown before the epidemic starts); rather, the emphasis is on the policy for distributing some quantity of a vaccine between the strata to minimize the initial growth rate of the epidemic. Formulating such a policy with this approach requires knowledge (up to a scaling factor) of the next generation matrix, whose entries are defined as the number of new infections in one stratum caused by an average infected individual in another stratum during the exponential growth period of the epidemic; it also requires knowledge of the initial distribution of susceptibles among the strata. Our method is appropriate for a 'sizeable' (sufficiently large) quantity of vaccine; the optimal allocation of a 'small' quantity may be different.

A related allocation policy in a stratified massaction model appeared before in Cairns (1989); we have derived our policy following a more general model introduced in Wallinga et al. (submitted), which is more flexible and allows for the treatment of a wider class of next generation matrices and varying vaccine efficacies among the strata. A key concept that emerges from the strategy is the least spread line. This is a line in an N-dimensional space, where N is the number of strata in the population; points (vectors) on that line represent the distributions of susceptibles to be left in the strata after an optimal allocation of a sufficiently large quantity of a vaccine; movement along the line corresponds to the varying total amounts of vaccine—see §2.2 for more details. The least spread line can be found explicitly using the next generation matrix and the initial distribution of susceptibles between the strata, by solving several systems of linear equations—see the electronic supplementary material. One can then compare the least spread line with the initial distribution of susceptibles, and devise a prioritization scheme for vaccine allocation between the strata.

This paper shows that these theoretical results can be proven for a stratified population that interacts via stratified mass-action-type kinetics, in which individuals within each stratum are exchangeable. Disease transmission in real populations departs from such kinetics in several ways. Specifically, the local depletion of susceptibles among the contacts of infectious persons (owing to the clustering of transmission

within infected households or school classrooms) violates exchangeability; moreover, an average infected person differs from an average person, in terms of their contacts. To test the robustness of the method to departures from its assumptions, we tested it on an explicit, individual-based simulation on the EpiSims platform, in which influenza dynamics were simulated in the population of Utah, and in which we have focused on the stratification of the population by age (0-6), 7-13, 14-18, adults). In the absence of information on prior immunity, we simulate a situation in which everyone is susceptible—this may be relevant, possibly with the exception of persons over 50, to the 2009 H1N1 influenza outbreak. To find the least spread line, one needs an estimate of the next generation matrix  $K = (k_{ij})$ , where  $k_{ij}$  equals the number of persons in stratum i infected by an average infectious person in stratum j during the exponential growth period of the epidemic. To assess K, we have run 100 epidemic simulations with five randomly chosen persons initially infected in Utah. There was a large degree of initial stochasticity, but in the early exponential growth period, the next generation matrices were similar in different simulations.

We note that the 'dynamical' next generation matrices estimated in this fashion using an exponential growth period on a 'large' network are not rigorously defined in this paper, in part because there is no explicit analytical structure describing the network (such as a community of households)—we consider this issue further in the discussion. The dynamical next generation matrices estimated by simulations can be compared to the more classical 'static' stratified mass-action next generation matrices (Diekmann & Heesterbeek 2000; Wallinga et al. 2006; Mossong et al. 2008), which are obtained by measuring the number of contacts in the model between a typical member of each stratum and the members of all strata, without any simulation of infection dynamics. The dynamical matrix differs from the static matrix because as the epidemic runs, local structure in the network becomes involved (infected individuals will cluster, e.g. in households, so that even at the early phase of the epidemic many contacts will be with immune persons) and because within a stratum, an 'average' person has different characteristics from an 'average infected person' (e.g. the average infected person in a stratum is likely to live in a bigger household than the average person in the same stratum). To illustrate that point, we have computed both the dynamical and the static next generation matrices and formulated 'optimal' allocation strategies for both. The matrices were quite different, giving different prioritization for vaccine allocation for the strata we have chosen. We have shown by simulations that the dynamical one is very nearly optimal, and in particular it gives lower initial growth rate for the epidemic than for the static one.

The dynamic estimate of the least spread line in our simulations consists of all multiples of the vector (0.029, 0.115, 0.041, 0.814). This means that given a number T of susceptibles left after a vaccine allocation, it is optimal for minimizing the initial growth rate of an

epidemic to have 0.029T susceptibles in age group 0-6, 0.115T susceptibles in the age group 7–13, etc. This constitutes a simple, explicit guideline for optimal vaccine allocation. The initial distribution of individuals according to those four age groups in Utah is (0.13,0.125, 0.094, 0.651). This renders an explicit formula for optimal coverage levels in the four strata as a function of the total vaccine quantity to be distributed. In particular, top priority (highest coverage levels) in an allocation of a sizeable quantity of seasonal influenza vaccinations (enough to reach the least spread line, which in our case means coverage of 20.1 per cent or more of Utah's population) goes to young children (0-6), followed by teens (14-18), then children 7-13, with the adult share being quite low. We also note that for a 'small' quantity of influenza vaccinations, with no prior immunity, top priority goes to teens, who have a disproportionate number of infections in the early stages of simulated epidemics, with no vaccine used—see the electronic supplementary material for more details.

We want to point out that minimizing the initial growth rate of an epidemic is not equivalent to minimizing the epidemic's final size (see Ball & Clancy 1993; Britton 2001 for relations between the latter and the reproductive number), though the two are related. We have tested our 'optimal initial growth rate' strategy by simulations against other vaccination strategies with the same total quantity of a vaccine used. In one set of simulations, we have allowed for seasonality in transmission parameters. One hundred epidemics were simulated for each vaccination policy, and the functions C(t) representing the average cumulative number of persons infected by day t were plotted. We have found that the 'optimal' strategy indeed had the smallest initial growth rate. Moreover compared to policies which were 'far enough' from it, it did better in terms of C(t) at all times, and correspondingly in terms of the final size. For some of the policies which were 'close' enough to the optimal, the function C(t) eventually descended below the one for the optimal policy, with a modest improvement in the final size. Previous work (e.g. Dushoff et al. 2007) has shown that optimizing the final size is extremely difficult and sensitive to small differences in parameter values. Moreover, the final size calculations (whether or not weighted by severity) rely on the strong assumptions of fixed conditions (no effects of seasonality, behaviour change, control measures, etc. and no further availability of vaccines), which will be violated in practice. Because these are uncertain, a strategy focused on the present (minimizing the initial growth rate) may be justified as more reliable than the one designed to optimize long-term outcomes.

Finally, our findings were compared to the current seasonal influenza vaccination coverage levels in the USA (CDC 2008). Data up to 2007 suggest that adults aged 18–49 had coverage levels on par with children aged 5–17. Our simulations/analysis, which rely on the structure of the EpiSims network, in particular suggest that if we vaccinated the same number of people as currently receiving the vaccine, but reduced the adult share considerably in favour of

children, transmission could be significantly reduced; moreover, incorporating the existence of prior immunity should only strengthen this conclusion, assuming that prior immunity on the average increases with the age for children, and is lower than the one for adults. For the same reason, our prioritization of teens as the second most important compared to 7-13 year olds is questionable, though both groups should receive attention, and the top priority should remain with preschool children. In that context, we are supportive of the CDC decision (CDC 2009) to include young individuals in the priority cohort for H1N1 influenza vaccination.

#### 2. METHODS AND RESULTS

### 2.1. Minimizing the initial growth rate of an epidemic in a stratified mass-action model

Suppose we have a population, divided into N strata  $1, \ldots, N$ , with deterministic transmission according to mass-action kinetics between each pair of strata. A key quantity describing the epidemic spread in a stratified mass-action model is the next generation matrix  $K = (k_{ij})$ . Here  $k_{ij}$  equals the number of persons in stratum *i* infected by an average infected person in stratum *j* in the early exponential growth stages of an epidemic. The quantity  $k_{ij}$  depends on a number of factors (see Wallinga *et al.* submitted and the electronic supplementary material):

- (i) the number  $s_i$  of susceptibles in stratum i,
- (ii) the contact rate  $b_{ij}$  (per unit of time) between an average person in stratum *i* and the one in stratum *j*. We assume reciprocity of contacts, thus  $b_{ij} = b_{ji}$  (Wallinga *et al.* 2006; Mossong *et al.* 2008),
- (iii) the total infectivity (the quantity of pathogen shedding)  $c_i$  of an average person in stratum j and
- (iv) the susceptibility (instanteneous risk of getting infected per infectious contact) parameter  $a_i$  for an average person in stratum *i*.

Explicitly (see the electronic supplementary material for more details),

$$k_{ij} = s_i a_i b_{ij} c_j.$$

The largest eigenvalue of K is the reproductive number  $R_0$  of the epidemic, in the absence of control measures (Diekmann & Heesterbeek 2000). Suppose we have a fixed quantity Q of a vaccine, and we give  $q_i$  doses to persons in stratum i, with  $\sum q_i = Q$ . To simplify notation, for the rest of this paper we deal with a perfect vaccine, though the argument would also work for a 'leaky' vaccine which reduces susceptibility by a certain factor (which may also depend on the stratum), or for an antiviral, which reduces infectivity—see the electronic supplementary material. Let

$$T = \Sigma s_i - Q$$

be the total number of susceptibles left after vaccination. We will have  $s_i^{\mathrm{T}} = s_i - q_i$  susceptibles in stratum *i*. No other parameters affecting the next generation matrix have changed. We can ask the following question:

**Question.** Given a positive number T, how do we choose the distribution of susceptibles  $s^{\mathrm{T}} = (s_1^{\mathrm{T}}, \ldots, s_N^{\mathrm{T}})$ , with constraint that the total number of susceptibles  $\sum_i s_i^{\mathrm{T}} = T$ , to minimize the largest eigenvalue of the new next generation matrix  $K_T = (s_i^{\mathrm{T}} a_i b_{ij} c_j)$ ?

It turns out that the question above has an explicit answer  $s^{T}$ —see theorem 2.2.1 and the remark following it in the electronic supplementary material, as well as Cairns (1989) for a slightly less general case. In addition,  $s^{T}$  depends linearly on *T*. Thus, we have the least spread line

$$L = (s^{\mathrm{T}} | T > 0),$$

which is ideally the target of a vaccination or antiviral distribution strategy. Finding this line amounts to solving several systems of linear equations with the coefficients depending on the next generation matrix K, and the initial numbers  $s_i$  of susceptibles in each stratum.

Remark. Here, we concentrate on a vaccine which is perfect. Data from placebo-controlled trials (Belshe et al. 1998, 2000; Longini et al. 2000) suggest that this is a good approximation. Vaccine efficacy for susceptibility is estimated in the papers above to be between 83 and 92 per cent depending on the vaccine efficacy against infectivity strain: of vaccinated individuals is estimated to be about 80 cent. Thus, a vaccinated individual would per cause verv few infections compared  $\operatorname{to}$ an unvaccinated one.

#### 2.2. Prioritization using the least spread line

In this section, we describe how to use the least spread line to prescribe the optimal allocation strategy and prioritization for vaccine distribution. We illustrate this by numerical examples in subsequent sections.

Suppose we have found the least spread line  $L = \text{span}(s_L)$  (all positive multiples of the vector  $s_L$ ), where the vector  $s_L = (s_1^L, \ldots, s_N^L)$  is scaled so that the sum of its components equals 1. Let  $s_{\text{init}} = (s_1^{\text{init}}, \ldots, s_N^{\text{init}})$  be the initial distribution of susceptibles, as a proportion of the total, so that the sum of the components of  $s_{\text{init}}$  is also 1.

Suppose we distribute quantity q of the vaccine among the strata so that stratum i gets  $q_i$  doses, with  $\sum q_i = q$ —here and for the rest of this section each number is viewed as a proportion of the total initial population. Thus, the resulting distribution of susceptibles left after vaccine allocation is according to the vector  $s^q = (s_1^{\text{init}} - q_1, \ldots, s_N^{\text{init}} - q_N)$ . The optimal allocation aims to get vector  $s^q$  on the least spread line. This means that  $s^q$  is proportional to the spanning vector  $s_L$ , namely

$$s^q = ps_L,$$

for some p > 0. Summing the components of the vectors in the equality above, we get that 1 - q = p. Also, stratum *i* received

$$q_i = s_i^{\text{init}} - p s_i^L,$$

vaccine doses. Thus, the proportion of individuals vaccinated in stratum i is

$$\frac{s_i^{\text{init}} - ps_i^L}{s_i^{\text{init}}} = 1 - p \frac{s_i^L}{s_i^{\text{init}}}.$$
(2.1)

The above gives an explicit strategy for distributing the vaccine among the strata for the overall coverage level of the fraction q = (1 - p) of the population. Our 'optimal' vaccination strategy allows for the simultaneous coverage of all strata (no time ordering for coverage). However, the higher is the ratio  $s_i^{\text{init}}/s_i^L$ , the higher is the prescribed proportion of persons vaccinated (coverage level) in stratum *i*. Hence, the notion of prioritization—we order the strata according to the ratios  $s_i^{\text{init}}/s_i^L$  for prioritization (higher coverage levels) for vaccine distribution. Let the order of priority for the strata be

$$i_1,\ldots,i_N.$$

Here, stratum  $i_1$  gets the highest priority (coverage level) for vaccine distribution (the ratio  $s_{i_1}^{\text{init}}/s_{i_1}^L$  is the highest), and the stratum  $i_N$  has the lowest priority (coverage level).

Having described the optimal allocation scheme and prioritization among the strata, let us calculate the smallest vaccine quantity needed to reach the least spread line from the initial population of susceptibles. Suppose some vector  $p_0 s_L$  on the least spread line requires the smallest quantity of a vaccine to reach the least spread line from the initial distribution of susceptibles. This means that one of the strata got no vaccine—otherwise, we can increase  $p_0$  and decrease the vaccine quantity needed to reach the least spread line. As the stratum  $i_N$  gets the lowest coverage upon reaching the least spread line, it must be the stratum which received no vaccine. Hence, the  $i_N^{\text{th}}$  component of  $p_0 s_L$  must equal the corresponding component of  $s_{\text{init}}$ . In other words,  $p_0 = s_{i_N}^{\text{init}}/s_{i_N}^L$ , and the smallest vaccine quantity needed to reach the least spread line is

$$q_0 = 1 - p_0 = 1 - \frac{s_{i_N}^{\text{init}}}{s_{i_N}^L}.$$
 (2.2)

### 2.3. Simulation to study the robustness of optimal allocations

In the sections that follow, we describe simulation experiments to test whether our approach to defining optimal vaccination allocations is effective in synthetic populations which violate some of its assumptions. Before doing so, we briefly describe the simulation approach.

Numerical simulations for an influenza epidemic were performed using two agent-based simulation platforms developed by the Network Dynamics and Simulation Science Laboratory (NDSSL) at the Virginia Bioinformatics Institute: EpiSimdemics and EpiFast. These two tools fuse data from the US census, activity surveys, Dun and Bradstreet business information, and NavTeq map information to build up an 'in silico' society among whose individuals epidemics spread. These tools are the improved versions of the original EpiSims. The overall process consists of four main steps (Barrett *et al.* 2008).

- (i) Creation of a synthetic population. A synthetic population of households and persons in these households is generated from the census data. Household locations are assigned to each household according to its census block and geolocated using the NavTeq street data. Using the data from the activity survey, a sequence of activities (i.e. a daily schedule of activities ordered in time) is associated to a household according to household demographics. Activities in the sequence are then associated with each household member, depending on demographics. Each activity in the sequence has a start time, an end time and an activity type (work, shop, etc.).
- (ii) Generation of an interacting network. Activities are located in areas where they can occur and an attractor for each possible kind of activity is associated to each activity location. For example, the attractor for 'work' activities is related to the number of employees at that location; the attractor for 'shop' activities is related to the area of retail space. A two-step gravity model employs these attractors to assign locations probabilistically for each activity in the activity sequences. Once all activities are assigned locations, sublocations are assigned to represent individual mixing groups at a single location (i.e. office, single store at a mall, etc.). The maximum number of persons simultaneously at the same sublocation is constrained by the sublocation size. Different sublocation sizes can lead to very different social networks.
- (iii) Simulating the epidemic process. Each person in the simulation is given a state of health, in this case drawn from the commonly used SEIR (susceptible, exposed, infectious, removed) disease model. An epidemic is then simulated as a dynamic process that traverses this social network. Nodes (i.e. people) begin the simulation as susceptible to this infectious process. Nodes can be selected at random or by their characteristics as candidates for initial infections. Infected nodes are assigned an incubation period and infectious period from a specifiable distribution. After the incubation period has expired, the node becomes infectious. During the infectious period, all connected nodes experience a daily probability of infection that depends on the duration of the contact that the edge represents. Infections occur stochastically based on these probabilities. After the infectious period expires,

the node becomes removed and is no longer a member of network.

interventions. (iv) Representing Interventions designed to disrupt the flow of infections across the network can be represented by altering the edges between different classes of nodes. As the edges convey the risk of infection, the absence of the edge prevents infection between two formerly connected nodes, and the scaling of the weight along this edge can represent the decreased probability of infection owing to a particular intervention. For instance, the simple case of 100 per cent effective vaccinations can be represented by removing all the edges of the vaccinated node, while the use of a leaky vaccination will result in a scaling of the weight of all the links. A social distancing measure can be represented by eliminating all the edges between individuals in the same activity location, or just some portion if one presumes that some out-of-activity contacts will persist despite the closure. In the case of school closure, all the links among children attending the same school will be removed except those arising from out-of school activities (i.e. recreational activities, shopping).

EpiSimdemics can process this data structure to calculate efficiently the second-by-second interactions of all the agents in the population and output the social network. Although EpiSimdemics can be used to simulate epidemic outbreaks in the social network, we used a much more efficient EpiFast simulation. EpiFast gains efficiencies through novel graph analytic techniques (for more details, Bisset et al. 2009) allowing largescale individual simulations in a fraction of the time required by the more flexible EpiSimdemics modelling platform. EpiFast supports dynamic interventions that can be triggered based on the time or level of infecthe simulation and can be applied tion in probabilistically to the members of specified populations. Simulations were conducted on a large cluster (112 four-core 2.2 GHz cpus for HPL performance of 1.471 TFlops). Running on 20 blades, a set of 100 simulations for the Utah population would take 130 min of wall-clock time, and produce 40 MB to 1.2 GB of raw output.

### 2.4. Finding the next generation matrix

Entries of the next generation matrix are defined as the number of new infections in one stratum caused by an average infected individual in another stratum during the early, exponential growth period of the epidemic. One way of estimating those entries is through simulations on a network describing interactions and disease dynamics in a population. Another, more classical method, which ignores the network's local structures and assumes stratified mass-action mixing, is to estimate the 'static' next generation matrix using the contact data in a population (Wallinga *et al.* 2006; Mossong *et al.* 2008). In this section, we define the two next generation matrices and describe how their entries may be estimated for real populations (in the static case) and for synthetic populations (in both cases).

2.4.1. Static next generation matrix. We define the 'static' next generation matrix as the matrix of potentially infectious contacts from persons in stratum j to persons in stratum i, which is defined up to a multiplicative constant that reflects the transmissibility of the agent and the duration of infectiousness. This is exactly the matrix measured by survey and diary studies, in which the individuals of various strata (age groups) were asked to estimate the number of contacts they had over a fixed period of time with others in each of the age groups (Wallinga et al. 2006; Mossong et al. 2008). Entries of this matrix measure the expected number of persons in one stratum to be infected by an average individual in another stratum, in a completely susceptible population (Diekmann & Hoesterbeek 2000). This matrix is often used as the next generation matrix for a mass-action model.

The 'dynamical' next generation matrices, counting the number of infections in one stratum caused by an average infected person in the other stratum during exponential growth periods of epidemics, are a bit different. For one thing, as the epidemic runs, local structure in the network becomes involved (infections cluster in households and other transmission groups). Also, the average infected adult in the early stage of the epidemic will differ from the average adult; for example, she s/he will tend to have more children in the household (hence be at higher risk of infection). We will estimate the 'dynamical' next generation matrices for the population of Utah using simulations in the next section. In  $\S2.5.2$ , we will also compare the results of optimal vaccination strategies for the 'static' and the 'dynamical' next generation matrices.

The static next generation matrix is of the form

$$K^{\mathrm{st}} = (k_{ij}^{\mathrm{st}}) \quad \text{with } k_{ij}^{\mathrm{st}} = q \cdot D \cdot \frac{t_{ij}}{s_j}.$$

Here,  $t_{ij}$  is the total amount of time spent in a common location between all persons in strata *i* and *j* during a time unit (say one day),  $s_j$  is the number of persons in stratum *j*, *D* is the mean duration of an individual infectiousness period, and *q* is the transmissibility parameter. We can also write

$$k_{ij}^{\rm st} = q \cdot D \cdot s_i \cdot b_{ij},$$

where the symmetric matrix  $B = (b_{ij}) = (t_{ij}/s_is_j)$ measures the rate of contact between average persons in strata *i* and *j*. Such matrices were already considered in Cairns (1989). With the Utah synthetic population and the transmissibility parameter chosen, one gets the static matrix

$$K^{\rm st} = \begin{bmatrix} 0.351 & 0.328 & 0.069 & 0.089 \\ 0.316 & 0.896 & 0.287 & 0.06 \\ 0.05 & 0.216 & 1.051 & 0.056 \\ 0.444 & 0.313 & 0.388 & 0.597 \end{bmatrix}.$$

J. R. Soc. Interface (2010)

The initial distribution of susceptibles in Utah by age groups (as proportions of the total population) is

$$s_{\text{init}} = \begin{bmatrix} 0.130\\ 0.125\\ 0.094\\ 0.651 \end{bmatrix}.$$

The least spread line for the matrix  $K^{\text{st}}$  and the initial distribution of susceptibles can be found from equation (2.2.5) in the electronic supplementary material, with  $E_i = 1$ . It is spanned by the vector

$s_L^{\rm st} =$	0.070
	0.032
	0.022
	0.876

Note that we chose a spanning vector whose sum of coordinates equals 1. Thus, given a quantity T of susceptibles left after vaccine allocation, it is optimal for minimizing the initial growth rate of an epidemic to have  $0.07 \cdot T$  susceptibles in age group 0-6,  $0.032 \cdot T$  susceptibles in age group 7-13, etc.

To formulate the prioritization scheme as in §2.2, we look at the ratios  $s_{\text{init}}(i)/s_L^{\text{st}}(i)$  of the entries of the vector  $s_{\text{init}}$  and the corresponding entries of the vector  $s_L^{\text{st}}$ :

$$s_{
m ratio}^{
m st} = egin{pmatrix} 1.865 \\ 3.894 \\ 4.332 \\ 0.743 \end{bmatrix}.$$

Higher values for those ratios correspond to higher priority strata, in terms of receiving a disproportionate quantity of a vaccine—see equation (2.1). The ratios suggest that top priority in an allocation of a sizeable quantity of seasonal influenza vaccinations (enough to reach the least spread line) goes to teens (14-18), followed by children (7-13), then young children (0-6), then adults. Finally, the minimal quantity of a vaccine needed to reach the least spread line is 1-0.743 = 0.257(see equation (2.2)), or the coverage of 25.7 per cent of the Utah population. Any additional 'optimal' coverage beyond the 25.7 per cent of the population should go proportionally to the vector  $s_L^{\text{st}}$ —thus, we descend along the least spread line after reaching it. For instance, vaccine doses at the total coverage level of 30 per cent of the population should be distributed according to the vector

$$s_{\text{init}} - 0.7 \cdot s_L^{\text{st}} = \begin{bmatrix} 0.081\\ 0.103\\ 0.079\\ 0.037 \end{bmatrix}$$

2.4.2. Dynamical next generation matrix. In this section, we describe how to estimate the 'dynamical' next generation matrix for the synthetic population, and use this matrix to find the least spread line. The dynamical next generation matrix counts the number of infections in one stratum caused by an average infected person in another stratum during the exponential growth period of an epidemic. We have used two methods to assess the next generation matrix—one more classical, and one using infectivity ratios.

In the first case, for each day t and for each simulated epidemic, we have found the  $X_j(t)$  persons in stratum jwho got infected on day t. We have found the  $Y_{ij}(t)$ persons in stratum i whom those  $X_j(t)$  persons have subsequently infected. We have estimated the next generation matrix  $K(t) = (k_{ij}(t))$  with

$$k_{ij}(t) = \frac{Y_{ij}(t)}{X_j(t)}$$

The practical disadvantage of such a definition is that it requires epidemic conditions (exponential growth) to persist for the subsequent infectious period of individuals who got infected on day t. As a result, the stability of such matrices does not persist for too long in simulations; we have found that  $k_{ij}(t)$ begins to decline in the end of the exponential growth period of the epidemic, representing future infections which happen under the saturation of susceptibles.

A more computationally convincing method involves infectivity ratios. Conveniently, individual infectiousness profiles on EpiSims take a whole number of days, with constant infectivity during the infectious period. For each day t, we have found the  $X'_{j}(t)$  persons in stratum j who were infectious on day t. We then found the  $Y'_{ij}(t)$  persons in stratum i who got infected by those  $X'_{j}(t)$  persons on day t. We have estimated the next generation matrix to be  $K^{2}(t) = (k_{ij}^{2}(t))$  with

$$k_{ij}^2(t) = \frac{Y_{ij}'(t)}{X_i'(t)} \cdot D.$$

Here, D is the mean duration of an individual infectious period (which is 4.2 days on EpiSims). Thus,  $k_{ij}^2(t)$  would equal the mean number of infections in stratum i caused by the infected individuals in stratum *i* during one day (namely day t), times the mean duration of the infectious period. If the conditions related to epidemic's progression stay constant over the infectious period of individuals in stratum j who got infected on some day t, the mean number of infections in stratum *i* subsequently caused by them  $(k_{ij}^1(t))$ would equal  $k_{ii}^2(t)$ ; thus, the next generation matrix  $K^{2}(t)$  equals the next generation matrix K(t). In practice, the period when the two matrices coincide in a simulation is rather brief; we found the infectivity ratio definition to be more computationally stable, though in principle, for an 'infinite' network, the two definitions are the same for the 'infinite' exponential growth period.

There was a large degree of initial stochasticity in the entries of K(t), but in the exponential growth period, the next generation matrices K(t) were similar in different simulations. We have estimated the next generation matrix during the exponential growth period to be

$$K^{\rm dyn} = \begin{bmatrix} 0.362 & 0.30 & 0.077 & 0.187 \\ 0.430 & 0.824 & 0.273 & 0.113 \\ 0.086 & 0.274 & 1.313 & 0.104 \\ 0.830 & 0.382 & 0.375 & 1.104 \end{bmatrix}.$$

Using this matrix, the initial distribution of susceptibles and equation (2.2.5) in the electronic supplementary material, we have found the least spread line to be  $L = \operatorname{span}(s_L^{\text{dyn}})$  with

$$s_L^{\rm dyn} = \begin{bmatrix} 0.029\\ 0.115\\ 0.041\\ 0.814 \end{bmatrix}$$

To formulate the prioritization scheme using the dynamical next generation matrix, we look, as in §2.2, at the ratios  $s_{\text{init}}(i)/s_L^{\text{dyn}}(i)$  of the entries of the vector  $s_{\text{init}}$  and the corresponding entries of the vector  $s_L^{\text{dyn}}$ 

	4.507	
	1.083	
$s_{\rm ratio} =$	2.280	
	0.799	

The ratios suggest that top priority in an allocation of a sizeable quantity of seasonal influenza vaccinations goes to young children (0-6), followed by teens (14-18), then children 7–13, then adults. The minimal quantity of a vaccine needed to reach the least spread line is 1-0.799 = 0.201, or the coverage of 20.1 per cent of the Utah population.

### 2.5. Testing prioritization schemes by simulations

2.5.1. Prioritization strategy using the dynamical next generation matrix. In this section, we use simulations to compare the 'optimal' allocation strategy obtained from the dynamical next generation matrix against several alternative strategies at the 20.1 per cent coverage level. In the next section, we compare the 'optimal' strategies obtained from the dynamical and the static next generation matrices, and show that the 'dynamical' one gives a lower initial growth rate.

Figure 1 plots several allocation strategies with 20.1 per cent total coverage, in addition to the no coverage scenario (top graph). A hundred simulations were run with each vaccination strategy and five randomly chosen individuals initially infected. Outbreaks that did not become epidemics (under 3000 infected in total) were discarded. Each graph C(t) represents the average cumulative number of people infected up to day t for a given vaccination strategy. We see that there is a decreasing order for the graphs as we approach the 'optimal' strategy. The strategy with 55 per cent going to adults roughly corresponds to the current proportions for seasonal influenza vaccinations. Our main emphasis is on moving in the direction of the least spread line, rather than whether the least spread line is indeed optimal in terms of minimizing the final size—the latter question is addressed in the electronic supplementary material.

In the electronic supplementary material, we also consider the strategies in figure 1 with seasonality added to the transmission parameters. There, all transmission parameters are multiplied by a factor of

$$s(t) = 1 + 0.5 \sin(\pi t / 180)$$



Figure 1. Each of the five graphs plots the average cumulative number of people infected up to each day for a given vaccination strategy. Black, no vaccination. The other four graphs represent different allocations with the coverage level of 20.1 per cent of the population. Blue, random allocation. Purple, 55 per cent going to adults, 45 per cent to children (distributed at random). Green, 40 per cent going to adults, 60 per cent to children. Red, optimal (dynamical) allocation.

Here, t is time (in days) since the seeding on the five initially infected. The optimal dynamical strategy will still have the smallest initial growth rate; however, the strategy of 55 per cent going to adults will catch up in the final size.

2.5.2. Comparing the strategies obtained from the dynamical and the static next generation matrices. In this section, we have considered allocation strategies with 26 per cent coverage for the Utah population. Optimal 'dynamical' allocation at this coverage level corresponds to the vector  $s_{\text{init}} - 0.74 s_L^{\text{dyn}}$ , and the optimal 'static' allocation goes according to the vector  $s_{\text{init}} - 0.74 s_L^{\text{st}}$ . We have compared these allocation strategies by simulations, using the same protocol as in  $\S2.5.1$ . The results are plotted in figure 2, where random allocation at the 26 per cent coverage level as well as no vaccination is also considered. We see that both 'optimal' strategies are better than random allocation in all regards. Comparing the 'dynamical' and 'static' optimal allocations, we see that the dynamical one has an initial advantage, yet the static one catches up and eventually has a slightly smaller final size—such a phenomenon is seen further in some simulations in the electronic supplementary material. We want to point out that the actual influenza transmission coefficients are highly seasonal, with real epidemics waning earlier than the simulated ones (which have no seasonality in transmission coefficients), and the modest gain in the final size should be further mitigated and perhaps even reversed.

### 3. DISCUSSION

We have described a prioritization scheme for an allocation of a sizeable quantity of a vaccine or antivirals in a stratified population. The scheme builds on an optimal strategy for reducing the epidemic's initial growth



Figure 2. Each of the four graphs plots the average cumulative number of people infected up to each day for a given vaccination strategy. Black, no vaccination. The other three graphs represent different allocations with the coverage level of 26 per cent of the population. Blue, random allocation. Red, optimal static allocation. Green, optimal dynamical allocation.

rate in a stratified mass-action model. The strategy is tested on the EpiSims network describing interactions and influenza dynamics in the population of Utah, where the stratification we have chosen is by age (0-6, 7-13, 14-18, adults). Under the assumption that all individuals are susceptible, we have found that top priority in an allocation of a sizeable quantity of seasonal influenza vaccinations goes to young children (0-6), followed by teens (14-18), then children 7-13, with the adult share being quite low. The latter assumption of a fully susceptible population may be relevant, possibly with the exception of persons over 50, to the 2009 H1N1 influenza outbreak.

Some of our findings overlap with the papers mentioned in the first paragraph of the introduction, in terms of setting up an optimization procedure to minimize certain quantities related to epidemic's spread, and emphasizing the role of children in the spread of communicable diseases, particularly influenza. At the same time, our emphasis is different from much of the previous work, which concentrated on the epidemic's final size, and the computational methods (such as a genetic algorithm or hill climbing) for minimizing it. As it was noticed before (Dushoff et al. 2007), such optimal strategies are very sensitive to the parameters in the models. In particular they are not scalable, namely they depend on the infectivity parameter, which may be a priori unknown and may change owing to seasonality. Moreover, other conditions related to an epidemic's progression (such as people's behaviour, availability of additional vaccine and other control measures) affect the final size a great deal and make a strategy based on the final size minimization questionable. Our approach, which strives to minimize the epidemic's initial growth rate, may provide a more practical alternative. We believe that our strategy (which is scalable) is novel in the usage of the network to estimate dynamically the next generation matrix, the explicit analytical formula for the optimal coverage levels and the unified framework it provides for addressing practical questions such as prioritization for vaccine allocation in real populations.

We want to point out that the actual optimal proportions for vaccine allocation among the strata depend on the network used for simulations. Even for a specific network, these proportions vary with changing coverage levels, as seen in §2.2, while in reality, we do not know in advance how many people will get vaccinated. Moreover, the next generation matrix, assessed during the exponential growth period of a simulation, varies somewhat with simulations; this variability is amplified in some simulations owing to the spatial aspects in the case of Utah. Thus, the main result reported here is the development of the prioritization scheme for vaccine allocation, rather than the precise numbers reported for the simulated Utah population. These qualitative results may create awareness about the vulnerable strata, potentially increasing the total number of vaccinated and, at the same time, attaining a more efficient distribution of a vaccine.

We did not assess the next generation matrix statistically as we did not use averaging over simulations, a small portion of which do not have a pronounced exponential growth period when an epidemic in Utah does not pick up soon enough in Salt Lake City. Instead, we took an exponential growth period of one arbitrarily chosen simulation and selected the next generation matrix from it. We have checked that the incidence vector during the exponential growth period was very close to a leading eigenvector of the selected next generation matrix. Moreover, the next generation matrices in other simulations which exhibited clear exponential growth were similar to the one chosen. Once the next generation matrix for the network is assessed, the 'optimal' strategies prescribed by it can be tested for optimality in terms of minimizing the initial growth rate by simulations (as they were in our paper). If they perform well (as they do in our simulations), ultimately the original source of their derivation (the next generation matrix) becomes unimportant.

The dynamical next generation matrices used here are not rigorously defined. In fact, it is unclear how to proceed with such a definition for a large but finite network of significant complexity, such as a real population. It may be possible to define a next generation matrix for an infinite, self-similar network. Selfsimilarity means that a finite number of distributions (such as household/classroom/workplace sizes, school/ work/commuting times) are used to generate an infinite, synthetic population. We have not pursued this theory as developing a mathematical theory of selfsimilar networks probably requires a separate paper, and the proof of the existence of a next generation matrix for such a network sheds no light on how to estimate it for a given network.

We also want to point out that prioritization orders for the strata resulting from dynamical and static next generation matrices are different. Simulations have shown that the vaccination strategy derived using the dynamical next generation matrix gives a lower initial growth rate for the epidemic than the vaccination strategy at the same coverage level obtained from the static

Table 1. Influenza vaccination coverage levels in 2006 and 2007, for different age groups.

age group/year	2006 (%)	2007 (%)
6-23 months	32.2	NA
2-4 years	26.4	37.9
5-17 years	12.4	17.5
18-49 years	13.4	15.3

next generation matrix. Moreover, in our case, the dynamical next generation matrix puts more emphasis on young children (0-6); if prior immunity were incorporated (and assumed to be increasing with age), this conclusion should only be strengthened. All this further suggests that while contact matrices (Wallinga et al. 2006; Mossong et al. 2008) can suggest good schemes for allocating interventions (the static matrix allocation is far better than random allocation), network modelsif realistic—can provide better estimates than those based on the contact matrix alone, because they incorporate the effects of local depletion of susceptibles owing to the clustering of transmission. We have found in other recent work that incorporating clustering leads to changes in the estimates of reproductive numbers and control efforts required, relative to the estimates from pure mass-action models. Once again, the differences are subtle, but suggest that the use of network information can improve the estimates of control measures (Goldstein *et al.* 2009).

Finally, we can compare our findings with the current influenza vaccination coverage situation. Our main reference for this is CDC (2008). In 2008, CDC adopted new guidelines which include all children aged 0.5-17 in a priority group for influenza vaccinations (previously only children 0.5-4 were in that group, in addition to adults over 50 and persons with certain medical conditions). No information is available yet about the effect of this measure on coverage levels. The 2006–2007 data on influenza vaccination coverage (before the new guidelines were enacted) are given in table 1.

Our model/simulations, which rely on the EpiSims network, suggest a very modest role for adult vaccination in reducing influenza spread, even under the conservative assumption that all persons are equally susceptible. If immunity on average increases with age, as one would expect for seasonal influenza, then the role of adults should be even smaller. For the same reason, our prioritization of teens over 7-13 year olds is questionable, though both groups should receive attention, and the top priority should remain with preschool children. In that context, we are supportive of the CDC decision (CDC 2009) to include young individuals in the priority cohort for H1N1 influenza vaccination.

This work is supported by the US National Institutes of Health cooperative agreement 5U01GM076497 'Models of Infectious Disease Agent Study' (M.L. and E.G.); the RAPIDD Program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health (J.C.M.). We thank our external collaborators and the members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their suggestions and comments. This work has also been partially supported by NSF Nets grant CNS-0626964, NSF HSD grant SES-0729441, CDC Center of Excellence in Public Health Informatics grant 2506055-01, NIH-NIGMS MIDAS project 5 U01 GM070694-05, DTRA CNIMS grant HDTRA1-07-C-0113 and NSF NETS CNS-0831633 (A.A., B.L., S.E.).

#### REFERENCES

- Ball, F. & Clancy, D. 1993 The final size and severity of a generalized stochastic multitype epidemic model. Adv. Appl. Prob. 25, 721–736. (doi:10.2307/1427788)
- Barrett, C., Eubank, S. & Marathe, M. 2008 An interactionbased approach to computational epidemiology. *Proc. XXIII AAAI Conf. on Artificial Intelligence*, *Chicago, IL*, pp. 1590–1593
- Belshe, R. G. et al. 2000 Correlates of immune protection induced by live, attenuated, cold-adapted, trivalent, intranasal influenza virus vaccine. J. Infect. Dis. 181, 1133–1137. (doi:10.1086/315323)
- Belshe, R. G. et al. 1998 The efficacy of live attenuated, coldadapted, trivalent, intranasal influenzavirus vaccine in children. New Eng. J. Med. 338, 1405–1412. (doi:10. 1056/NEJM199805143382002)
- Bisset, K., Cheng, J., Feng, X., Vullikanti, V. S. A. & Marathe, M. 2009 Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. *Proc. XXIII Int. Conf. on Supercomputing*, 8–12 *June 2009*, pp. 430–439.
- Britton, T. 2001 Epidemics in heterogeneous communities: estimation of r0 and secure vaccination coverage. J. R. Stat. Soc. B 63, 705–715. (doi:10.1111/1467-9868. 00307)
- Cairns, A. 1989 Epidemics in heterogeneous populations: aspects of optimal vaccination policies. *Math. Med. Biol.* 6, 137–159. (doi:10.1093/imammb/6.3.137)
- CDC 2008. See http://www.cdc.gov/flu/professionals/acip/ coveragelevels.htm.
- CDC 2009. See http://www.cdc.gov/h1n1flu/vaccination/acip.htm.

- Diekmann, O. & Heesterbeek, J. A. P. 2000 Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. New York, NY: John Wiley & Sons.
- Dushoff, J., Plotkin, J., Viboud, C., Simonsen, L., Miller, M., Loeb, M. & Earn, D. 2007 Vaccinating to protect a vulnerable subpopulation. *PLoS Med.* 4, e174. (doi:10.1371/ journal.pmed.0040174)
- Goldstein, E., Paur, K., Fraser, C., Kenah, E., Wallinga, J. & Lipsitch, M. 2009 Reproductive numbers, epidemic spread and control in a community of households. *Math. BioSci.* 221, 11–25. (doi:10.1016/j.mbs.2009.06.002)
- Hadeler, K. P. & Mueller, J. 2007 Optimal harvesting and optimal vaccination. *Math. BioSci.* 206, 249–272. (doi:10.1016/j.mbs.2005.09.001)
- Halloran, M. & Longini, I. J. 2006 Community studies for vaccinating schoolchildren against influenza. *Science* **311**, 615616 (doi:10.1126/science.1122143)
- Longini, I. J. & Halloran, M. 2005 Strategy for distribution of influenza vaccine to high-risk groups and children. Am. J. Epidemiol. 161, 303–306. (doi:10.1093/aje/ kwi053)
- Longini, I. J., Halloran, M., Nizam, A., Wolff, M., Mendelman, P., Fast, P. & Belshe, R., 2000 Estimation of the efficacy of live, attenuated influenza vaccine from a two-year, multi-center vaccine trial: implications for influenza epidemic control. *Vaccine* 18, 1902–1909. (doi:10. 1016/S0264-410X(99)00419-3)
- Mossong, J. et al. 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, 381–390 (doi:10.1371/journal.pmed. 0050074)
- Patel, R., Longini, I. & Halloran, M. 2005 Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. J. Theor. Biol. 234, 201–212. (doi:10.1016/j. jtbi.2004.11.032)
- Wallinga, J., Teunis, P. & Kretzschmar, M. 2006 Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Amer. J. Epidemiol.* 1644, 936–944 (doi:10.1093/aje/ kwj317).
- Wallinga, J., Van Boven, M. & Lipsitch, M. Submitted. Real-time estimation of impact of targeted intervention strategies during an epidemic: mathematical analysis of impact of targeted interventions.