Spatial considerations for the allocation of pre-pandemic influenza vaccination in the United States

Joseph T. Wu*, Steven Riley and Gabriel M. Leung

Department of Community Medicine and School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

The impact of the next influenza pandemic may be mitigated by inducing immunity in individuals prior to the start of national epidemics using a pre-pandemic vaccine targeted against current avian influenza strains. The US Department of Health and Human Services (HHS) intends that pre-pandemic vaccines will be allocated to states in proportion to the size of their population in predefined priority groups, i.e. approximately pro-rata. We show that such an equitable policy is likely to be the least efficient in terms of the number of infections averted. We demonstrate that the potential benefits could be substantial if a fully discretionary policy is allowed, i.e. if some regions are allocated sufficient vaccines to achieve herd immunity while other regions are allocated no vaccine. Since such an inequitable policy may be impractical, we consider the sensitivity of an intermediate policy (in which 50% of the stockpile is allocated on a pro-rata basis) to key transmission uncertainties. The benefits of the 50% discretionary policy are sensitive to parameter values which cannot be known in advance. Therefore, despite substantial potential benefits of non-pro-rata policies, our results suggest that the current HHS policy of pro-rata allocation by state is a good compromise in terms of simplicity, robustness, equity and efficiency.

Keywords: influenza vaccines; pandemic; herd immunity; resource allocation; efficiency; equity

1. INTRODUCTION

The use of a pre-pandemic vaccine against the dominant avian influenza strain circulating is now being considered as an option to mitigate the next influenza pandemic (Bresson et al. 2006; Lin et al. 2006; Treanor et al. 2006). However, it is unlikely that supply will be sufficient to provide high coverage for most populations, including the United States (Monto 2006; Nichol & Treanor 2006). According to the current US Health and Human Services (HHS) Pandemic Influenza Plan (www.hhs.gov/pandemicflu/plan/sup6.html), each state will receive pandemic vaccines in proportion to the size of its population in predefined priority groups. These groups include frontline healthcare workers, emergency service personnel and the elderly. Because the size of priority groups scales approximately linearly with the total population size, this policy approximates to pro-rata allocation under which vaccines are allocated according to the population size of each state.

A pro-rata policy is intuitively appealing because it is equitable and simple, i.e. states are allocated vaccines according to population size. Thus, every individual in principle has an equal chance of vaccination (equitable) and the policy requires no additional epidemiological information for implementation (simple). However, its public health consequences have not been examined in terms of efficiency and robustness which are also desirable properties of any vaccine distribution policy, i.e. would such an allocation policy avert as many infections as possible (efficiency) and would it perform well under a range of as yet uncertain epidemiological scenarios (robustness). Here we show that pro-rata or equitable allocation may be the least efficient policy in terms of minimizing the population infection attack rate (IAR). Motivated by this observation, we then quantify the tradeoff among the other three goals of equity, simplicity and robustness for more efficient spatial allocation policies in the distribution of pre-pandemic vaccines. Because tradeoff among these objectives is common in control of epidemics, the insights from this study are not only restricted to pre-pandemic vaccinations but also applicable to general problems of epidemic control.

We analyse a range of policies under which some portion of the vaccine stockpile is distributed pro-rata while the remaining is allocated to minimize IAR. Following the terminology introduced in Kaplan & Merson (2002), we call the latter portion of the stockpile the discretionary stockpile and the associated policies discretionary policies. Specifically, a d% discretionary policy reserves d% of the vaccines for the discretionary stockpile with the remaining (100−d)% distributed pro-rata. The associated IAR is denoted by IAR(d). In this study, an allocation policy is defined as a map, in the mathematical sense, that takes the basic reproductive number $R_0$ and coverage $c$ as inputs and generates a geographical vaccine allocation $\alpha=(\alpha_1, \ldots, \alpha_K)$, where $\alpha_i$ is the proportion of stockpile allocated to geographical area $i$. We use a deterministic meta-population model parameterized with detailed travel data (see §2) to predict the epidemiological consequences of alternative allocation.
infectious contacts are made with other individuals present in region \( j \) at rate \( \beta_j \) and the proportion of infectious individuals in population \( j \) at time \( t \) is \( \sum_{i=1}^{K} m_{ij} I_i / \sum_{i=1}^{K} m_{ij} N_i \). Since we are interested only in final attack rates, we can omit the latent stage. In the presence of vaccination, the model further classifies individuals by their immune states but otherwise follows the same dynamics (see electronic supplementary material). We develop an algorithm to compute the final number of infections in each region, denoted by \( U_1, \ldots, U_K \), by solving a system of \( K \) nonlinear equations regardless of the number of immune states (see electronic supplementary material).

(b) Vaccine efficacy

We assume that the pre-pandemic vaccine has efficacy for both susceptibility and infectiousness (Longini et al. 2004; Patel et al. 2005). We use a multi-state leaky vaccine response model that we developed in a previous study (Riley et al. 2007) to describe the distribution of reduction in susceptibility given by the most potent pre-pandemic vaccine that has been tested in clinical trials (Lin et al. 2006). We do not assume a distribution of reduction in infectiousness because it is not important as long as the mean reduction in infectiousness is the same (see electronic supplementary material for details on the vaccine response model). We assume that vaccine efficacy for both susceptibility and infectiousness are multiplied by \( \mu \in [0, 1] \), which reflects the degree of antigenic match between vaccine and pandemic strains (\( \mu = 1 \) for perfect match).

(c) Optimization of vaccine allocation

Let \( (\alpha_1, \ldots, \alpha_K) \) denote an arbitrary vaccine allocation. Given a set of transmission parameters and the size of vaccine stockpile \( V \), the allocation for the \( \alpha_k \) discretionary policy is obtained by solving the following nonlinear programme in which the total number of infections is minimized:

\[
\text{minimize } \sum_{i=1}^{K} U_i
\]

subject to

\[
1 - \frac{d}{100} \frac{N_i}{N} \leq \alpha_i \leq \min \left\{ \frac{N_i}{V}, 1 \right\}
\]

\[1 \leq i \leq K, \quad \sum_{i=1}^{K} \alpha_i = 1.\]

The lower-bound on \( (\alpha_1, \ldots, \alpha_K) \) is the discretionary stockpile size constraint. The upper bound says that vaccines allocated to a region cannot exceed its population size or the stockpile size.

3. RESULTS

(a) Pro-rata may be the least efficient policy

We first give a simple example to illustrate why pro-rata may be inefficient by failing to minimize the population IAR. When we present results from the US 10-region model later in this section, we will use this example again to explain the structure of the discretionary policies. We define the function \( r(\cdot) \) to be the reduction in IAR achieved by vaccine coverage level \( c \), which is defined as the proportion of population vaccinated. Owing to herd immunity, the marginal number of infections averted per additional individual vaccinated increases as coverage rises until control is achieved (see electronic supplementary material for justification of this claim). That is, as \( c \) increases, the
The allocation of pre-pandemic vaccines  J. T. Wu et al.  2813

Table 1. Parameter values and assumptions. Local transmissibility of region i is given by \( R_0(1 + \epsilon)hD_i \), where \( R_0 \) is the basic reproductive number assumed when constructing the discretionary policies.

<table>
<thead>
<tr>
<th>parameter</th>
<th>base case assumptions</th>
<th>sensitivity</th>
<th>sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z^i ) (% of ( V ))</td>
<td>vaccine efficacy for susceptibility ( V_{El} = 0.69 ) when antigenic match is perfect</td>
<td>driven by ( \mu: z^i \leftrightarrow \mu: z^i )</td>
<td>Riley et al. (2007)</td>
</tr>
<tr>
<td>( V_{El} ) (vaccine efficacy for ( I ))</td>
<td>( V_{El} = 0.8 ) when antigenic match is perfect</td>
<td>driven by ( \mu: V_{El} \leftrightarrow \mu: V_{El} )</td>
<td>Longini et al. (2004); Patel et al. (2005)</td>
</tr>
<tr>
<td>( \epsilon ) (error in estimate of transmissibility)</td>
<td>no error, i.e. ( \epsilon = 0 )</td>
<td>([-0.5, 0.5])</td>
<td>Mills et al. (2004); Viboud et al. (2006)</td>
</tr>
<tr>
<td>( h_i ) (relative hazard for transmissibility in region i)</td>
<td>transmissibility is the same for all regions, i.e. ( h_i = 1 ) for all ( i )</td>
<td>([0.5, 1.5])</td>
<td>Mills et al. (2004); Viboud et al. (2006)</td>
</tr>
<tr>
<td>( \delta_i ) (relative level of inter-regional mixing)</td>
<td>inter-regional mixing is at normal level; ( \delta = 1 ) by definition</td>
<td>([0, 2])</td>
<td></td>
</tr>
<tr>
<td>( \mu ) (antigenic match between vaccine strain and pandemic strain)</td>
<td>imperfect match with ( \mu = 0.7 )</td>
<td>([0, 1])</td>
<td></td>
</tr>
</tbody>
</table>

gradient of \( r(\xi) \) increases (i.e. \( r(\xi) \) is convex) and \( r(\xi) \) becomes flat only after the critical coverage level \( c^* \) has been achieved (figure 1b). (A function is convex if its gradient is increasing.) Suppose \( K \) populations of sizes \( N_1, \ldots, N_K \) are isolated from each other but have the same disease dynamics (i.e. \( r(\xi) \) is the same for all populations). Given \( V \) doses of vaccines and an allocation \( \alpha = (\alpha_1, \ldots, \alpha_K) \), where \( \alpha_i \) is the proportion of the overall stockpile allocated to population \( i \), the total number of infections averted is \( T(\alpha) = N_1r(\alpha V/N_1) + \cdots + N_Kr(\alpha_K V/N_K) \). Throughout this paper, we consider only ‘sensible’ allocations that do not prescribe coverage higher than the critical level \( c^* \) in any population; coverage beyond \( c^* \) has no benefit. We denote the pro-rata allocation by \( \alpha = (N_1/\sum_{i=1}^{K} N_i, \ldots, N_K/\sum_{i=1}^{K} N_i) \) and note that \( \alpha \) minimizes \( T \) because \( T \) is convex and \( \alpha \) is the unique stationary point of \( T \) (see electronic supplementary material). Therefore, in terms of minimizing IAR, pro-rata is the least efficient policy (among all sensible allocations) in this hypothetical example.

In the remainder of this section, we present results from the US 10-region model.

(b) The 100% discretionary policy
We first compare pro-rata with the 100% discretionary policy. Figure 2 shows that as \( R_0 \) varies between 1 and 3, \( \Delta IAR(100) \) can be as large as 17.8% but remains below 6% when \( R_0 \leq 2 \), which is suggested as the most probable range in recent studies of influenza pandemic mitigation (Mills et al. 2004; Ferguson et al. 2006; Germann et al. 2006). At low coverage, \( \Delta IAR(100) \) is close to zero because allocation policy makes no difference when only a small amount of vaccines is available. At high enough coverage, \( \Delta IAR(100) \) also approaches zero because control is achieved under both policies. \( \Delta IAR(100) \) becomes non-negligible with higher \( R_0 \) values. If there is no inter-regional mixing, \( \Delta IAR(100) \) shows similar patterns but can be as large as 26.6% (see figure 2 in the electronic supplementary material). In general, \( \Delta IAR(100) \) is inversely related to the degree of inter-regional mixing. Since the level of inter-regional mixing will probably be lower than normal during a pandemic, our base case estimates of gain in efficiency will tend to be conservative.

Figure 3 shows that the 100% discretionary policy typically achieves a lower IAR by sacrificing some regions to which no or relatively few vaccines are allocated, i.e. concentration of coverage and absence of equity. Furthermore, the policy is ‘choppy’ with many discontinuities as \( R_0 \) and \( c \) change, i.e. the associated allocation may exhibit sudden swings with small changes in \( R_0 \) and \( c \). For example, with an \( R_0 \) of 2.4, the allocation to Region 8 falls from 100 to 0% as \( c \) increases from 2 to 2.1%. All discretionary policies exhibit concentration of coverage and choppiness, the degree of which scale with the discretionary stockpile size. These properties are the consequence of the convex relation between coverage and reduction in IAR (figure 1b). To see this, consider the simple example in figure 4, which is a special case of the example given in the beginning of this section where two non-interacting populations with sizes \( N_1 > N_2 \) have the same transmission dynamics (see electronic supplementary material for mathematical details). Because the number of infections averted is a convex function of allocation, it is never optimal (in terms of minimizing IAR) to split the stockpile between the two populations unless control has been achieved in at least one of them. The optimal policy can therefore be identified by comparing
only two policies: (i) vaccinate Population 1 then Population 2 and (ii) vaccinate Population 2 then Population 1. It is this immediate consequence of local herd-immunity effects that accounts for concentration of coverage. Figure 4 shows that there is a switching point $s^*$ such that it is optimal to follow policy (i) if the stockpile size exceeds $s^*$ and follow policy (ii) otherwise. As a result, at the switching point $s^*$, the optimal allocation to Population 2 drops discontinuously from 100 to 0%. This accounts for choppiness. In our meta-population model, inter-regional mixing is low because individuals spend more than 97% of their time in their home regions on average. This level of mixing is low enough such that concentration of coverage and choppiness are preserved in the 100% discretionary policy as shown in figure 3. Concentration of coverage has also been observed in previous studies that optimize vaccine allocations to different age groups (Longini et al. 1978; Patel et al. 2005). In those studies, however, mixing among age groups is high and concentration of coverage is a result of targeting high-prevalence groups (e.g. school children) or high-mortality groups (e.g. elderly). Here in our spatial model, inter-population mixing is low and concentration of coverage arises owing to different population sizes and the convex relation between reduction in IAR and coverage.

**Gain in efficiency as a function of discretionary stockpile size**

Consider the quantity $\rho(d) = \Delta IAR(d)/\Delta IAR(100)$ which represents the proportion of maximal gain in efficiency (given by the 100% discretionary policy) that can be attained by the $d\%$ discretionary policy. The relation between discretionary stockpile size $d$ and $\rho(d)$ is a critical component of the equity–efficiency tradeoff. If the gain in efficiency is close to maximal (i.e. $\Delta IAR(100)$) with a relatively small discretionary stockpile size (say 30%), an inequitable allocation, which is a necessary condition under discretionary policies, may be justifiable (Kaplan & Merson 2002). This corresponds to the $d-\rho(d)$ relation being concave (i.e. the gradient of $\rho(d)$ decreases as $d$ increases). We consider only cases where $\Delta IAR(100) \geq 2\%$. Figure 5a shows that the relation is roughly linear for large values of $\Delta IAR(100)$, which means that if $d\%$ of the vaccines is reserved for the discretionary stockpile, then $\Delta IAR(d)$ is approximately $d\%$ of that shown in figure 2. As $\Delta IAR(100)$ drops below 6% (which corresponds to $R_0 \leq 2$), the $d-\rho(d)$ relation becomes mostly convex. Thus,
the efficiency gain of discretionary policies described here is minimal unless equity is substantially compromised.

(d) The risk associated with efficiency maximization

We have shown that regardless of the transmission scenario, the 100% discretionary policy prescribes highly inequitable allocations and is thus likely to be impractical. As a result, we switch our focus to the 50% discretionary policy, which represents a more equitable allocation policy, to test the robustness of reduction in IAR provided by discretionary policies (constructed using the base case parameter values in table 1). Qualitative results obtained from the sensitivity analyses below hold for all discretionary policies.

First, we conduct univariate sensitivity analyses on the 50% discretionary policy for each of the following parameters: (i) relative level of inter-regional mixing $\delta_i$, (ii) antigenic match $\mu_i$, and (iii) error in estimate of transmissibility $\epsilon$. Graphical illustrations of these sensitivity analyses can be found in figure 3 in the electronic supplementary material. $\Delta$IAR(50) increases as $\delta$ decreases because concentration of coverage is more effective when regions are isolated from each other (figure 3a in the electronic supplementary material). As antigenic match $\mu$ deviates from the base case value of 70%, $\Delta$IAR(50) drops (figure 3b in the electronic supplementary material). $\Delta$IAR(50) can become negative when $\mu$ is near 1 (i.e. antigenic match is underestimated) or when $\epsilon$ is below zero (i.e. transmissibility is overestimated; figure 3b,c in the electronic supplementary material). The effect of the former condition is much weaker because $\mu$ affects all discretionary policies while $\epsilon$ has no effect on pro-rata. Both conditions translate into underestimating vaccine effectiveness in reducing transmission, thus overestimating critical coverage. In these cases, the strategy of concentrating coverage into a few regions, which is the cornerstone of discretionary policies, performs poorly when the prescribed coverage is higher than the critical levels in these regions. Coverage beyond the critical level is misallocated because it generates very little or no reduction in IAR. We term this situation overshoot. Pro-rata has a smaller risk of overshoot because it has no dependence on transmissibility and it diversifies coverage among regions. Underestimating transmissibility ($\epsilon>0$) has little effect on $\Delta$IAR(50) in the range of $\epsilon$ considered ($\pm 50\%$). This suggests that when constructing the discretionary policies, assuming a lower transmissibility (within the range of equally plausible estimates) may alleviate the risk of overshoot and thereby enhance robustness. Finally, we perform a multivariate sensitivity analysis on $\delta_i$, $\mu_i$, $\epsilon$, as well as the relative hazards for local transmissibility $h_i$, $i=1,\ldots,K$ (table 1). Together, these parametric variations represent model uncertainties that cannot be accurately predicted until the pandemic strikes. These include heterogeneities in local transmissibility and efficacies of mitigation interventions besides pre-pandemic vaccination. In the presence of these parametric variations, the resultant $\Delta$IAR(50) is mostly smaller (up to 28%) than the base case $\Delta$IAR(50) (figure 3d in the electronic supplementary material), which is the reduction in IAR expected from the 50% discretionary policy if all base case assumptions are satisfied. Figure 4b shows that the 50% discretionary policy can perform significantly worse than pro-rata (by more than 5% in IAR) when (i) there is overshoot and (ii) the pro-rata IAR is low ($\Delta$IAR(0)$\leq 50\%$). The latter condition corresponds to an intrinsically weak pandemic or a very effective mitigation
programme. Such a scenario is unlikely when pre-pandemic vaccination is the only intervention but becomes more plausible in the presence of other interventions. In general, the larger the size of the discretionary stockpile, the higher the risk of overshoot and the associated magnitude of underperformance (relative to pro-rata).

4. DISCUSSION

Other studies have considered targeted allocation of influenza vaccine based on non-spatial population heterogeneities (Longini et al. 1978; Patel et al. 2005). Generally, our results are consistent with these other studies in which we find that policies other than pro-rata are potentially more efficient. However, for the spatial sub-populations of the United States, although pro-rata may be the least efficient policy, the work presented here suggests that it does strike an attractive balance between equity, efficiency, robustness and ease of implementation. Because the basic reproductive number, for pandemic influenza is likely to be less than 2, our estimated gain in efficiency from discretionary policies is not large enough (reduction in IAR ≤ 6%) to justify the associated inequity. Moreover, this gain in efficiency can be severely attenuated when the assumptions used to construct the discretionary policies are violated (figure 3d in the electronic supplementary material). In some cases, significant overshoot occurs and renders discretionary policies less efficient than pro-rata (figure 5b). Such cases become more likely when other mitigation interventions are present (see below). Finally, while discretionary policies require detailed knowledge of the epidemiology of pandemic influenza (a transmission model and parameter estimates), pro-rata has no such requirement and is therefore much more straightforward to implement. These conclusions are robust against a range of model assumptions and structures. Therefore, despite substantial potential benefits of non-pro-rata policies, our results suggest that the current HHS policy of pro-rata allocation by state is a good compromise in terms of equity, efficiency, simplicity and robustness.

The validity of our results, as with all similar studies, is somewhat limited by the underlying assumptions of our model. We have used a deterministic model to evaluate vaccine allocation policies. Although stochastic effects may be important when control is nearly achieved in one or more regions, we expect that a stochastic model will lead to similar conclusions because the population sizes are large and parameter uncertainties have a much larger effect. We have assumed in our base case that transmissibility is the same in all regions. If transmissibility differs among regions, the relation between reduction in IAR and coverage (figure 1b) will also differ (e.g. different c∗) but the general shape (convexity) will remain unchanged for all regions. Consequently, discretionary policies will still exhibit concentration of coverage and choppiness, and their efficiency remains susceptible to overshoot. The gain in efficiency over pro-rata will require further analysis. However, it is not yet clear how a systematic difference in local–regional transmissibility, if it exists, can be estimated to support such an analysis: estimation of R0 for the 1918 pandemic influenza strain in 45 US cities showed no correlation between R0 and various city characteristics (including population size, latitude and longitude; Mills et al. 2004).

Here, we have considered the minimization of attack rate to be the main objective of pre-pandemic vaccination policies. There are a number of alternative approaches that could be taken. One would be to minimize the post-vaccination reproductive number (Ball & Lyne 2002). In the electronic supplementary material, we show that the reproductive number under pro-rata is near optimal (within 15%) under this objective. We also show that allocations which minimize the reproductive number do not necessarily minimize the attack rate, and vice versa, because the reproductive number describes the early dynamics of the epidemic while the attack rate is a summary outcome of the entire epidemic. Minimization of the reproductive number may be suitable if the objective is to minimize the initial epidemic growth rate (instead of the eventual total number of infections averted) which may help maintain healthcare services during the early phase of the pandemic. A second alternative policy goal might be to minimize the number of deaths. Because the case fatality ratio will probably be age-specific, this objective would require detailed knowledge of the age-specific mortality rate for the novel influenza strain, which was very different during the different pandemics of the twentieth century. The resulting formulation is similar to cases where geographical vaccine allocation is further optimized by targeting high-risk groups or high-prevalence groups (Patel et al. 2005; Bansal et al. 2006). In these cases, the relation between reduction in IAR (c) and coverage c will take a more complicated form than that shown in figure 1b. For instance, if there are two risk groups in each population and vaccination priority is given to the group that has higher infectiousness (e.g. children), the c–r(c) relation comprises two convex regions (see figure 4 in the electronic supplementary material). Suppose all high-risk groups must be vaccinated (Phase 1) before any of the low-risk groups can be vaccinated (Phase 2) within each of these two phases, the discretionary policies will have the same structure shown here because convexity is preserved. However, if some low-risk groups can be vaccinated before all high-risk groups have been fully vaccinated, the structure of the discretionary policies will be more complex and would require further investigation. Recent studies of pandemic influenza mitigation have come to the consensus that an effective programme will comprise multiple targeted, layered interventions including quarantine, isolation, vaccination and antiviral prophylaxis (Ferguson et al. 2006; Germann et al. 2006; Wu et al. 2006). Our conclusions will probably remain valid in the context of multi-component mitigation as long as the relation between reduction in IAR and coverage takes the form as shown in figure 1b. For instance, the presence of other interventions is roughly equivalent to lower transmissibility (lower R0), which in turn means a smaller ΔIAR(100) (figure 2). In this case, the relationship between gain in efficiency and discretionary stockpile size that we describe remains applicable because we have characterized how this relation depends on ΔIAR(100) (figure 5a): ΔIAR(d) will probably be lower when other interventions are in place and the tradeoff between discretionary stockpile size and gain in efficiency is worsened. This further strengthens our conclusion that
pro-rata is the most appropriate policy for allocating pre-pandemic vaccines to different geographical areas.

Although we have focused on evaluating pre-pandemic influenza vaccine allocation policies throughout this paper, our major results are applicable in a wider epidemiological context. Specifically, we have shown that for a meta-population with different population sizes and low inter-population mixing: (i) the convex relation between IAR reduction and vaccination coverage result in concentration of coverage and choppiness for policies that minimizes attack rate, (ii) the performance of these policies are susceptible to overshoot and therefore very sensitive to estimates of key epidemiological parameters, and (iii) minimization of attack rate is different from minimization of reproductive number both in terms of the underlying policies and the associated outcomes.

The authors thank the following institutes for research funding: The Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong SAR Government (J.T.W., S.R., G.M.L.); Modeling of Infectious Disease Agents Study (MIDAS), National Institute of General Medical Sciences (J.T.W., S.R., G.M.L.) and The University of Hong Kong SARS Research Fund (S.R., G.M.L.). We thank the School of Industrial and Systems Engineering at Georgia Tech for access to their computation equipment. We thank the three reviewers for their helpful comments.

REFERENCES


