Ensemble model of human immune response to influenza A virus infection and its application to evaluation of treatment strategies

Baris Hancioglu *, Gilles Clermont †‡, and David Swigon *

*Department of Mathematics, 301 Thackeray Hall, University of Pittsburgh, Pittsburgh, PA 15260, USA, †Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA, ‡CIRM (Center for Inflammation and Regenerative Modeling), 100 Technology Drive Suite 200, Pittsburgh, PA 15219-3110 and CRISMA Laboratory, University of Pittsburgh, Pittsburgh, PA 15261, USA, and †Department of Critical Care Medicine, 3550 Terrace St, University of Pittsburgh Medical Center, Pittsburgh, PA 15261, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Ensemble models of biological systems provide probabilistic predictions of the dynamics that approximate the variability of response among individuals. We present an ensemble model of the human immune response to influenza A virus infection, consisting of a set of ordinary differential equations with parameters characterized by a probability distribution reflecting the goodness of fit of the equations to empirical data. This ensemble model is used to compute probabilistic estimates on the trajectories of the immune response, duration of disease, maximum tissue damage, likelihood of rebound of disease and the probability of occurrence of superspreaders, i.e., individuals with significant viral load but without apparent signs of infection. Antiviral drug effectiveness is evaluated and it is found that the strength, duration and time of initiation of treatment have significant influence on treatment benefits and on possible adverse effects.

Abbreviations: IAV, influenza A virus, ODE, ordinary differential equation, APC, antigen presenting cells, ROC, receiver operating characteristic

The ideal treatment of a communicable infectious disease would improve the individual outcome of those infected and reduce the risk to society by decreasing the propensity of disease transmission to uninfected hosts; that is, reduce the reproduction number of the disease. How closely the goals of individual health restoration and infection containment in a population are related, and whether there are conditions under which they might actually diverge remains to be explored. We address this issue in the case of influenza A virus infection, using ensemble modeling.

Mathematical models have proven to be valuable tools in the understanding of host immune response to infectious diseases such as IAV, HIV, hepatitis B, and hepatitis C [1, 2, 3, 4, 5, 6, 7, 8]. The effect of antiviral treatment on the in-host course of influenza infection has been modeled [9] for in vitro culture, in the absence of the immune system, and hypothesized in a mouse model [5]. Recently, ensemble methods have gained attention in large scale simulations of weather systems as well as in systems biology [10, 11, 12, 13].

We previously developed an ODE model of host immune response to IAV infection [14] that includes a simplified representation of key interactions and effects between host tissue and IAV. Parameterized from literature data, this model reproduces well the existing observations on the generally described clinical course of influenza, but is insufficient for shedding additional insight on population heterogeneity in disease duration and severity variability, and in response to interventions such as vaccines or antiviral drugs.

In this work, we constructed an ensemble of models designed to reflect uncertainty about parameter values, data sparsity and likely variation across a population of hosts exposed to IAV. This ensemble preserves the mathematical structure of the reference model, but assumes that parameters take values from a probability distribution. Every set of parameters in the distribution leads to a disease trajectory (realization) of the ensemble and the probability of this parameter set is determined by the likelihood that the trajectory captures existing data [15, 16]. Once a full probability distribution over parameter sets is computed, the ensemble model provides probabilistic predictions of the system’s evolution and response to perturbations. Applied to a biological system, the ensemble approach has the potential to quantify the variety of possible outcomes associated with a perturbation of this system, such as a drug intervention, and estimate the risk associated with various intervention scenarios.

Traditionally, the efficacy of an antiviral drug is judged by its effect on the duration of disease, measured as the duration of elevated temperature levels and symptoms [17]. However, a more detailed description of the system, such as the one provided by our model, immediately presents other criteria that may be taken into account, such as the peak value of the virus, peak value of the damage to the system, duration of elevated virus levels, and the probability of disease rebound. Some of these criteria may be important from the point of view of the individual (controlling symptoms, risk of secondary infection, or death), others from the point of view of society (disease containment). We have employed our ensemble approach to analyze the effect of the initiation, duration and strength of these treatments on the course of the disease, specifically, the level of the virus in the host (and hence its infectiousness), the level of tissue damage to the host and the level of interferons in the host (which are correlated with symptoms).

Results

The ODE model used here for the dynamics of influenza virus is essentially identical to the one in [14] and is based on the following simplified set of interactions (Fig. 1): Influenza virus (V) attacks the host respiratory tract mucosa, interacts with healthy epithelial cells (H) and infects them by binding to cell surface receptors via one of the major surface glycopro-
The virus replicates in infected cells (I), and several hours after cellular infection the infected cell dies and newly synthesized virus particles are released by the action of another major glycoprotein, NA [19]. Antigen presenting cells (M) are stimulated by the remnants of dead cells (D) [20]. APC and infected cells stimulate the innate immunity by secreting α and β interferons (F) [21] which interact with healthy cells and convert them to infection resistant cells (R) [22]. Additionally, APC stimulate the proliferation of effector cells (cytotoxic T cells, CTL, or natural killer cells NK, both denoted by E) that destroy infected cells before they can release mature viruses. Finally, APC activate the proliferation of virus-specific plasma cells (P) which produce antibodies (A) that bind with IAV and render it ineffective. Additional variable (S) quantifies the affinity between antibodies and virus and is assumed to be increasing in proportion to the number of plasma cells, to mimic the clonal selection of antibodies.

At baseline, a healthy host has no dead, infected or resistant cells, no detectable interferon levels, and no activated APCs. Initial levels of effectors, plasma cells, and antibodies are assumed to be at homeostatic values. In the naive host, we assume a relatively low compatibility with the infecting strain (S(0) = 0.07). Such compatibility may have resulted from previous exposure to a different strain of IAV or sub-optimally matched vaccine. In a typical IAV infection, the initial aliquot of aerosolized virus particles that the host receives is about 10^3 particles per ml, which with the present scale for V corresponds to V(0) = 10^{-3}. Initial conditions are fixed over the ensemble. There are two distinct classes of trajectories produced by parameter sets in our sample - (i) standard trajectories for which virus levels show a single peak followed by a monotone decrease, and (ii) trajectories with a rebound of the disease showing a second peak in viral titers and immune markers. The occurrence of a second peak is surprising but not completely unexpected because our data put no restrictions on viral titers beyond day 7. No trajectories showed more than two peaks or oscillatory behavior. As there is no clinical evidence currently available that would point to the existence of patients with rebounding disease, we focus here on standard trajectories.

**Standard behavior.** The computed full probability distribution over parameter sets producing standard course of the disease is fairly broad, reflecting the sparsity of the data, but nonetheless far from uniform. Marginal parameter distributions can be approximately divided into three classes: localized, biased, and uniform. Localized distributions, with either a well-defined peak, or a negligible value in some region of the parameter range, include the bursting rate a_I of infected epithelial cells, the deactivation rate for APC’s, a_M, the rates of stimulation of effector and plasma cells by APC, b_EM and b_PM, and the degradation rate a_P of interferon (Fig. 2). Biased distributions are either increasing or decreasing over the parameter range, and are found for the rate of IAV secretion by infected cells γ_V, the antibody-virus interaction rate γ_A, the nonspecific virus removal rate a_V, the rate of infectivity of the virus γ_HV, the homeostatic maintenance rate for plasma cells a_P, and the antibody degradation rate a_A. All other parameters are approximately uniform over the range explored (Fig. S1). Covariance analysis of the log parameter values shows negative correlation of γ_V and γ_HV (r = −0.63), a_V and a_I (r = −0.51), and positive correlation of b_EM with a_M (r = 0.55), a_E (r = 0.71) and b_PM (r = 0.64) and of b_PM and a_M (r = 0.84).

The distribution allows for probabilistic predictions of the course of infection in a naive host, which are shown in Fig. 3 not as individual trajectories, but as percentile levels of the distribution of the variable versus time. (For example, if the 50% median curve of V has the value V_1 at time t_1 then one half of all trajectories in the ensemble have V(t_1) ≤ V_1.) In accord with the data used to create the ensemble model, virus level peaks at day 1.5, with maximal loads of approximately 10^5 times the inoculum size, and declines slowly for approximately 4 days [23]. Virus level declines sharply after day 6 and falls below the initial inoculum after 8-10 days in 95% of the cases. We assume that the host is considered infectious when virus level exceeds V = 1, which happens 1 day before symptoms begin [24].

Ensemble trajectories for healthy cells are essentially identical up to day 7 with a sharp drop in H at day 1. The proportion of healthy cells at day 15 varies widely over the ensemble, which reflects an uncertainty between the apportioning of cells to healthy and resistant classes. Infected cells reach a maximum proportion of 15-55% of all epithelial cells by day 1.5, and disappear by day 7 in 95% of cases. Resistant cells increase sharply at day 1 and reach a maximum level of 80-99% on day 5-10. APCs are activated after day 1, peaking anywhere between days 1.5 and 2.5 and returning to homeostatic levels within 25 days. The interferon response peaks between days 2 and 3 and then very slowly decreases. Effectors and plasma cells start to depart from a homeostatic value after day 1. Effector cells peak around day 7-9. Plasma cells are produced 2-3 days before virus-specific antibodies are detectable and their population peaks between days 10 and 20, in accord with empirical observations [25]. Antibody levels drop from the homeostatic levels to a minimum on day 1.5, replenishment to homeostatic level occurs on day 7 followed by approximately 10^2 - 10^3 fold maximum increase. Antigenic compatibility grows monotonically starting once adaptive immunity is activated (after day 8). Antibodies are capable of inhibiting viral particles with 80% effectiveness in about 15 days after infection. Once the infection is over, subtype-specific antibodies constitute 60 to 100% of active antibodies.

The resulting loss of respiratory epithelial cells is a major reason for several of the symptoms that accompany infection, such as cough, depressed tracheobronchial clearance, and altered pulmonary function in severe cases [26]. We consider the host "symptomatic" if the maximal respiratory damage load, quantified by the proportion of dead respiratory epithelial cells, exceeds 10% [27]. Maximal damage varies across the ensemble from insignificant levels (less than 1% damage) to about 40% damage, mostly attained between days 1.5 and 2. The median peak damage level is 13%. About 25% of the population would be considered asymptomatic whereas 5% of them suffer from severe symptoms. The symptomatic period lasts from day 1 to day 4 after which time most of the cells become resistant to the infection.

**Superspreaders.** It is suspected that many cases of influenza infections are asymptomatic, yet with ability to shed virus and infect others [24]. Adults shed virus for 3 to 5 days whereas young children and some immuno-compromised individuals have been reported to shed influenza virus for 3 weeks and even more [28, 29, 30]. We found that in our ensemble of models about 1.8% of trajectories result in little damage to the system (max(D) < 5%) and little activation of interferon (max(F) < 5000), although virus levels achieve peak values comparable with the standard disease. Such characteristics correspond to a patient with little or no symptoms that is capable of transmitting significant amounts of the virus. The marginal distributions of parameters are similar to the distri-
butions of the combined data, except for $\gamma_{VA}$, $\gamma_{HV}$, $a_V$, $a_R$ and $b_F$ being shifted toward lower values, resulting in lower infectivity and production of interferon, and the distributions of $b_D$, $b_H$, $b_{3UV}$ and $a_P$ shifted toward higher values, resulting in faster healthy cell recovery, conversion to resistant state, activation of macrophages and degradation of interferon (see Fig. S2).

**Predicting phenotypic behavior.** Identifying a parsimonious set of model parameters highly predictive of clinically relevant phenotypes would be valuable because it could be used to design optimal treatment strategies tailored to individuals. Predicting quartile of maximal damage is more difficult than predicting superspreaders and rebounders, as it requires at least seven parameters to achieve 60% accuracy (Fig. 4), compared with three (interferon-related) parameters ($b_F$, $b_{MF}$, $a_F$) for superspreaders and three (antibody and CTL-related) parameters ($a_M$, $b_{EM}$, $a_E$) for rebounders to achieve discrimination $>0.9$ as measured by receiver-operator characteristic (ROC) value (Fig. S3). Tree-based classifiers performed consistently better than other methods.

**Drug therapy.** The systemic response to antiviral drug treatment depends on multiple factors including the type of antiviral drug, the strength/potency of the drug, and the times of initiation and duration of treatment.

We here describe the effect of M2 inhibitor treatment, although essentially identical results are obtained for neuraminidase inhibitor treatment as well. The positive effects of the treatment depend greatly on the initiation time and treatment duration. The main effects are summarized in Fig. 5 for treatment with high potency and low potency drug. (Detailed plots of effects on individual variables can be found in Figs. S4 and S5). A single treatment strategy is characterized by the time of treatment start (days after the infection) and treatment duration in days, and it corresponds to a point in the graph. A benefit is characterized by a region of treatment strategies for which the benefit is achieved in at least 90% of the ensemble.

The potential benefits of high efficacy treatment (Fig. 5A) are: (i) shortening of the duration of disease below 4 days (blue region), (ii) simultaneous lowering of maximum damage to below 0.1 and maximum interferon level to below 1000 (magenta), (iii) simultaneous lowering of maximum damage to below 0.1 and maximum virus level to below 10 (green), or (iv) lowering the probability of disease to below 50% (gray). Major possible adverse effect of a high efficacy treatment is more than 50% chance of a serious rebound with damage above 0.25 and virus levels above 30 (red). The potential benefits of low efficacy treatment (Fig. 5B) are: (i) shortening of the duration of disease in 90% of cases below 5 days (blue region), or (ii) simultaneous lowering of maximum damage to below 0.15 and maximum virus level to below 30 (green). The major adverse effect of low efficacy treatment is extension of the duration of the disease above 8 days.

In summary, for the model presented here we have not found any treatment strategy that would be universally beneficial in that it would achieve simultaneous decrease in the virus level, damage, interferon level, and rebound probability. The treatment strategy utilized in practice is to initiate antiviral drugs once symptoms have appeared. This typically happens after the peak of the disease, roughly 2 days after infection. The only positive effect of such a treatment is a shortening of the duration of the disease by about 1-2 days, depending on the potency of the drug, which is accomplished within 4 days and further treatment has no additional benefits. The most beneficial time to begin treatment is close to day 1 after infection. In that case both low and high potency drugs lower peak viral load (by 22%-83% on log scale), interferon level, and damage (by 75%-95%), with higher potency drug having a stronger effect. The optimum length of the treatment to accomplish these results is 2-3 days for high efficacy drug; treatment longer than 10 days has no further benefits. The effect of an early treatment, i.e., is beneficial for the patient, in that it lowers the damage and increases the probability of no disease proportionally with the length of the treatment, up to 20% (low potency drug) or 60% (high potency drug) for a 16 day treatment. It is not beneficial for the control of the epidemic because it does not significantly lowers virus levels.

Adverse effects antiviral drugs are chiefly caused by an insufficient stimulation of the immune system and either a rebound of the disease upon discontinuation of the treatment, or a prolonged disease is characterized by high infectivity of the individual, but low damage and interferon levels. Both rebound and prolonged disease may be problematic from epidemiological point of view, as the infected host becomes an asymptomatic carrier of the disease.

**Discussion**

In this paper, we capture data sparsity and uncertainty by adapting parallel tempering, a variant of Markov Chain Monte Carlo sampling to create an ensemble of models of the human immune response to IAV infection and study the influence of antiviral treatment on the course of the disease. This time course, although calibrated from a limited human sample of longitudinal viral titers, closely follows recently reported experimental results for human and mouse models [3, 7]. Simulations illustrate significant variations in IAV-host response across the ensemble of models representing a heterogeneous population, manifested in varying propensity to communicate disease, duration of symptoms, and overall severity of illness. This model is most appropriate for seasonal influenza, as it does not allow for significant early innate immune activation. A model appropriate for highly pathogenic IAV would require an adequate description of such an activation and the systemic repercussions of the ensuing cytokine storm [31].

The model presented displays some surprising results and unique predictions that have not yet been verified experimentally, but warrant further investigation as they may influence how antiviral therapies should be administered. A proportion of models in the ensemble produce disease trajectories that show a rebound of the disease and second peak in the viral levels starting at days 10-13 post infection (Figs. S6 and S7). Such a relapse was indeed described clinically in a population of younger individuals with non-intact immune systems [32], as suggested by the subset of parameters more predictive of this behavior (Fig. S3). Our model does not imply the likelihood of multiple relapses or cyclic behavior. A smaller proportion of models correspond to trajectories displaying a normal time course of viral titers, yet low interferon and tissue damage, therefore apparently healthy or minimally symptomatic individuals with preserved ability to communicate disease. These parameter sets would correspond to individuals that would be particularly efficient at spreading disease, in a manner similar to superspreaders as described for other communicable diseases [33, 34].

Ensemble analysis of an intervention with a highly effective antiviral drug preventing viral budding includes a surprising prediction that if the drug is started around 12 hrs after initial infection, it may not prevent, rather delay disease and...
cause a rebound. Furthermore, if antivirals are given early (as in prophylactic treatment), they tend to benefit the individual by lowering damage and symptoms, but do not help in control of the epidemic because they do not lower the viral count. Such treated individuals can therefore spread the disease equally well as untreated. This problem is yet another aspect to consider in the difficult decision on whether influenza treatment should primarily benefit individuals or the society as a whole [35, 36, 37].

The posterior distribution on parameter sets provides a solution to the inverse problem presented by the dynamical system describing the IAV-host interaction given the available empirical data. This approach is similar in spirit that of combining predictions of trajectories of complex dynamical systems employed to model weather and climate dynamics [11]. In the present paper we have limited the ensemble to models that are identical in initial conditions and equation structure, but differ in model parameters. The solution presented herein could be modified to include variations in initial conditions such as pre-existing immunity to IAV and baseline variations in homeostatic levels of immune effector cells and IAV type-specific antibodies generated by prior infection of immunization.

Materials and Methods

Ensemble model. The influenza ODE model described in [14] and used in this paper has 10 variables and 27 parameters:

\[
\begin{align*}
\frac{dV}{dt} &= \gamma_V I - \gamma_V A S V - \gamma_V H V - a_V V - \frac{a_V V}{1 + a_V V} \\
\frac{dH}{dt} &= b_H D (H + R) + a_R R - \gamma_H V H - b_H F H F \\
\frac{dI}{dt} &= \gamma_H V H - b_E E I - a_I I \\
\frac{dM}{dt} &= (b_M D + b_M V) (1 - M) - a_M M \\
\frac{dF}{dt} &= b_F M + c_F I - b_F H F H - a_F F \\
\frac{dR}{dt} &= b_H F H - a_R R \\
\frac{dE}{dt} &= b_E M E - b_E I E + a_E (1 - E) \\
\frac{dP}{dt} &= b_P M P + a_P (1 - P) \\
\frac{dA}{dt} &= a_A (P - A) - \gamma_V S A V \\
\frac{dS}{dt} &= r P (1 - S)
\end{align*}
\]

The variable \(D\) serves as a marker for tissue damage [20] and an indicator of the severity of disease. It also represents the available space for tissue cells, and therefore \(D\) is given by a conservation law \(D = 1 - H - R - I\). All variables have been rescaled by their constant homeostatic values as in [14] and hence the system (1)-(10) is dimensionless.

Our ensemble model consists of the ODE model (1)-(10) (written, for short, as \(X(t; \alpha, x_0)\)) with (related) independent variables \(x = (x_1, \ldots, x_n)\) (here \(n = 10\)), a parameter set \(\alpha = (\alpha_1, \ldots, \alpha_p)\) (here \(p = 27\)), and with an associated probability distribution \(p(\alpha)\) that represents, for each \(\alpha\), the likelihood that the model captures available data \(d = (d_1, \ldots, d_q)\) with standard deviations \(\sigma = (\sigma_1, \ldots, \sigma_p)\) (here \(q = 15\)). A trajectory \(x(t; \alpha, x_0)\) of the ODE model with initial condition \(X_0\) is considered to be a realization of a random process with a distribution given by \(p(\alpha)\). Therefore, the value of any variable at any given time point is also a random variable with distribution induced by \(p(\alpha)\).

The distribution \(p(\alpha)\) is determined by comparing the data \(d = (d_1, \ldots, d_q)\) with the corresponding computed quantities \(c = (c_1, \ldots, c_q)\) which are functions of the trajectory \(X(t; \alpha, x_0)\) of the system and hence functions of \(\alpha\). The residuals \(\Delta r_j(\alpha) = c_j(\alpha) - d_j, j = 1, \ldots, q\) determine the likelihood function \(L(\alpha)\), and consequently the distribution \(p(\alpha)\) as

\[
L(\alpha) = \prod_{j=1}^q \left(\frac{1}{\sqrt{2\pi\sigma_j^2}}\right) \exp\left(-\frac{(\Delta r_j(\alpha))^2}{2\sigma_j^2}\right)
\]

\[
p(\alpha) = Q^{-1} L(\alpha) \theta(\alpha)
\]

where \(\theta(\alpha)\) is the prior distribution and \(Q = \int L(\alpha) \theta(\alpha)\) is the normalizing constant.

We here choose \(\alpha_i\) to represent the logarithm of the value of the corresponding parameter (i.e., \(\alpha_1 = \log \gamma_V\), etc.) because trajectories of the system (1)-(10) are sensitive to relative changes in parameters, and we wish to avoid biasing the sample toward large parameter values. This choice affects the sampling metric on the parameter space. The prior distributions for all parameters were taken to be a uniform distribution between 1/4 and 4 times a biologically reasonable baseline value [4, 14] (see Table 1).

The likelihood function was computed from the available clinical data for human subjects and a set of observations considering the response of a naive host to the standard initial conditions [4, 14] (see Table 2). The standard deviations for the virus data are taken from [26] while others are chosen according to earlier observations [4]. More data points can be added when they become available. Initial conditions were \(V(0) = 0.001\), \(H(0) = 1\), \(I(0) = M(0) = F(0) = R(0) = 0\), \(E(0) = P(0) = A(0) = 1\) (rescaled to homeostatic level units) and \(S(0) = 0.07\).

Ensemble sampling.

The distribution \(p(\alpha)\) is represented by a sample of parameter sets \(\alpha^1, \ldots, \alpha^M\) (each of which is a vector with \(p\) components) found using the Metropolis-Hastings Monte Carlo method (MMC) enhanced by parallel tempering [36, 39]. We use a formulation in which we assume that \(p(\alpha)\) is a Gibbs-Boltzmann distribution derived from an energy function \(E(\alpha)\) as \(p(\alpha) = \exp(-B E(\alpha))\). The original MMC method samples the distribution at a fixed value of \(B\) by proposing, at every step \(k\), a random perturbation \(\Delta\) of the current parameter set \(\alpha^k\) and accepting this perturbed set as the next set in the sample, \(\alpha^{k+1}\), with probability \(P = \min\{1, \exp((-B E(\alpha)) - E(\alpha^{k+1}))\}\). If \(\alpha^k\) is rejected then \(\alpha^{k+1}\) is made equal to \(\alpha^k\). Parallel tempering algorithm utilizes several MMC algorithms running simultaneously at several values of \(B\) lower than 1 in order to improve the coverage of the parameter space and lower autocorrelation of the sample. The parameter sets \(\alpha^1, \ldots, \alpha^M\) in two chains with \(\beta_1 \leq \beta_2 \leq \ldots \leq \beta_M\) were chosen to be 1, 0.4, 0.12, 0.033, 0.0079, respectively. Optimal convergence of MMC algorithm is achieved if the acceptance ratio (i.e., the ratio of accepted to proposed sets) is in the range 0.3-0.4 [40]. The points \(\alpha^{1, 1}, \ldots, \alpha^{M, 1}\) sample the distribution \(p(\alpha)\).

We used MATLAB 7.9 to run computer simulations. The ensemble models are obtained by numerical integrations of ODEs using the MATLAB built in solver, ode23s. We have performed 6 distinct runs of the algorithm starting from the randomly chosen initial points. The quality of the sample was tested using Geweke test [41], which suggests that the chains have converged. We also applied Heidelberger & Welch test [42] to the chains and all of them passed the stationary test which further gives evidence for the convergence of the algorithm for each chain. In addition, Gelman & Rubin convergence diagnostic [43] showed that the 6 chains, each with different starting values which are overdispersed with respect to the target distribution, have converged to the same posterior distribution. We performed the tests in R using Bayesian Output Analysis (BOA) package.

The sample was used to estimate the ensemble averages of any trajectory-dependent quantity \(W\) as \(\bar{W} = \sum_{k=1}^M W(x(t; \alpha^k, x_0))\). The percentile value \(P_X(W)\) of any trajectory-dependent quantity \(W\) was found as the smallest number that is larger than \(X\%\) of values of \(W(x(t; \alpha^k, x_0))\).

Phenotype prediction models.
Non-parametric testing of full marginal distributions, subsampled at a frequency lower than the first lag with negative partial autocorrelation function (typically 15-18), demonstrates that full marginals are significantly different as measured by Mann-Whitney U and Kolmogorov-Smirnov Z for the vast majority of parameters between standard and rebounding, superspreaders and non-superspreader, and across quartile of maximal tissue damage groups. Using information gain as ranking criteria, we build a variety of classifiers on 2/3 of the samples, adding predictors to the models sequentially according to their associated information gain. Model predictions were validated on the remaining 1/3 of samples and assessed by receiver-operating characteristic area under the curve and correct classification rate.

Antiviral Treatment Modeling.

Since M2 inhibitor (such as amantadine or rimantadine) blocks the M2 channel protein of the virus, stopping protons from entering the virus and blocking its uncoating inside of the infected cell [44], its effect was simulated by lowering \( \gamma_H \), the rate of infectivity of the virus, to either 1/3 (for low efficacy drug) or 1/10 (for high efficacy drug) of its original value for the duration of the treatment, i.e., for \( t_1 < t < t_1 + t_2 \) where \( t_1 \) is the time of initiation of treatment and \( t_2 \) is the duration of treatment. None of the other parameters were changed. The trajectory \( x_k'(t), x_k(t) \) for each so modified time-dependent parameter set \( \alpha_k(t), k = 1, \ldots, M \) was then computed and statistics were obtained for various quantities of interest including the mean time of the peak and peak value of \( V, F, D \) the probability of disease (proportion of trajectories with any increase in \( V \)), the probability of occurrence of a rebound (second peak), and the duration of the disease (measured as the length of time for which \( V > 1 \) during a peak). Graphs of these quantities as functions of \( t_1 \) and \( t_2 \) can be found in Figs. 54 and 55.

Neuraminidase inhibitor (such as oseltamivir or zanamivir) blocks the function of neuraminidase which prevents the virus from budding from the host cell [45], and hence its action was simulated by lowering \( \gamma_V \) in a manner similar to the above.

ACKNOWLEDGMENTS. This work was supported by Alfred P Sloan fellowship (DS) and under grant R01-GM36802. B.H. gratefully acknowledges support from Virginia Tech. This research utilized in part computational resources of PittGrid. www.pittgrid.pitt.edu.
Fig. 1. Schematic representation of interactions included in the model. Each black connector represents an interaction that affects the variables attached to the connector, proportionally to the indicated rate constant: downregulation is indicated by a bar, upregulation by an arrow, no bar or arrow indicates no change in the variable (it participates in the interaction as a catalyst or a source). Degradation terms and homeostatic maintenance loops have been omitted.

Fig. 2. Marginal probability distributions of logarithms of selected parameters of the ensemble model, normalized to the baseline values in Table 1.

Fig. 3. Statistical description of trajectories of the ensemble model for standard case showing percentile levels of the distribution of each variable (not trajectories) across the ensemble as a function of time: 5th percentile (bottom dotted green), 25th (bottom dashed blue), 50th (solid red), 75th (top dashed blue) and 95th (top dotted green).

Fig. 4. Prediction of disease burden, as a quartile of dead epithelial cells, from parameter values. Percentage of correct predictions is shown on the left axis for various classifiers: linear SVM (magenta), regularized logistic (green), neural network (blue), naive Bayes (brown), random forest (orange) and CART (red). Bar graph (right axis) shows the information gained by adding parameters.

Fig. 5. Diagrams of the response of the system to treatments with an antiviral drug of high efficacy (A) or low efficacy (B). The treatment benefits and adverse effects are indicated by shaded areas (see text for explanation).

Table 1. Baseline parameter values

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_V$</td>
<td>582</td>
<td>$b_{HF}$</td>
<td>0.0190</td>
<td>$a_F$</td>
<td>4.41</td>
</tr>
<tr>
<td>$\gamma_{VA}$</td>
<td>343</td>
<td>$b_{IE}$</td>
<td>0.0088</td>
<td>$b_{EM}$</td>
<td>2.65</td>
</tr>
<tr>
<td>$\gamma_{vH}$</td>
<td>1.23</td>
<td>$a_I$</td>
<td>1.63</td>
<td>$b_{EI}$</td>
<td>1.16</td>
</tr>
<tr>
<td>$a_V$</td>
<td>3.90</td>
<td>$b_{MD}$</td>
<td>4.18</td>
<td>$a_E$</td>
<td>0.750</td>
</tr>
<tr>
<td>$a_{V1}$</td>
<td>78.9</td>
<td>$b_{MV}$</td>
<td>0.0867</td>
<td>$b_{EM}$</td>
<td>2.00</td>
</tr>
<tr>
<td>$a_{V2}$</td>
<td>17764</td>
<td>$a_M$</td>
<td>0.658</td>
<td>$a_F$</td>
<td>0.0736</td>
</tr>
<tr>
<td>$b_{HD}$</td>
<td>4.75</td>
<td>$b_F$</td>
<td>85203</td>
<td>$a_A$</td>
<td>0.00484</td>
</tr>
<tr>
<td>$a_R$</td>
<td>2.00</td>
<td>$c_F$</td>
<td>1024</td>
<td>$\gamma_{AV}$</td>
<td>277</td>
</tr>
<tr>
<td>$\gamma_{HV}$</td>
<td>0.688</td>
<td>$r$</td>
<td>$1.1 \times 10^{-5}$</td>
<td>$b_{FH}$</td>
<td>19.8</td>
</tr>
</tbody>
</table>

Table 2. Data used in the generation of ensemble.

<table>
<thead>
<tr>
<th>Variable Value</th>
<th>Value</th>
<th>St. Dev.</th>
<th>Variable Value</th>
<th>Value</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $V(0)$</td>
<td>1.88*</td>
<td>0.41*</td>
<td>$S(15)$</td>
<td>0.8†</td>
<td>0.1</td>
</tr>
<tr>
<td>log $V(1)$</td>
<td>3.82*</td>
<td>0.53*</td>
<td>$D_{max}$‡</td>
<td>0.35†</td>
<td>0.2</td>
</tr>
<tr>
<td>log $V(2)$</td>
<td>3.41*</td>
<td>0.35*</td>
<td>$t(F_{max})$</td>
<td>2‡</td>
<td>0.5</td>
</tr>
<tr>
<td>log $V(3)$</td>
<td>2.65*</td>
<td>0.41*</td>
<td>log $F_{max}$</td>
<td>4‡</td>
<td>0.5</td>
</tr>
<tr>
<td>log $V(4)$</td>
<td>2.53*</td>
<td>0.35*</td>
<td>log $E_{max}$</td>
<td>2‡</td>
<td>0.5</td>
</tr>
<tr>
<td>log $V(5)$</td>
<td>1.32*</td>
<td>0.35*</td>
<td>log $P_{max}$</td>
<td>4‡</td>
<td>0.5</td>
</tr>
<tr>
<td>log $V(6)$</td>
<td>0.82*</td>
<td>0.35*</td>
<td>log $A_{max}$</td>
<td>3‡</td>
<td>0.5</td>
</tr>
<tr>
<td>log $V(7)$</td>
<td>0.18*</td>
<td>0.12*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*from [26]  
†from [4]  
‡from [14]