

LA-UR-05-1915

Approved for public release;  
distribution is unlimited.

*Title:* Sensitivity Analysis of an Infectious Disease Model

*Author(s):* Dennis R. Powell,  
Jeanne Fair,  
Rene J. LeClaire,  
Leslie M. Moore,  
David Thompson

*Submitted to:* 2005 International System Dynamics Conference  
July 17-21, 2005  
Boston, MA



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Form 836 (8/00)

# Sensitivity Analysis of an Infectious Disease Model

**Dennis Powell, Jeanne Fair, Rene LeClaire, Leslie Moore**

Los Alamos National Laboratory

Los Alamos, New Mexico 87545

(505) 665-3839

[drpowell@lanl.gov](mailto:drpowell@lanl.gov), [jmfair@lanl.gov](mailto:jmfair@lanl.gov), [rjl@lanl.gov](mailto:rjl@lanl.gov), [lmoore@lanl.gov](mailto:lmoore@lanl.gov)

**David Thompson**

Sun Microsystems Inc.

4150 Network Circle

Santa Clara, CA 95054

(509) 624-1018

[dt.home@comcast.net](mailto:dt.home@comcast.net)

*Abstract: A model of infectious diseases has been developed for integration within a larger simulation structure to assess the interdependencies of critical infrastructures. The model has been parameterized to model a disease outbreak in a large metropolitan area. The model calculates the spread of the infection and the influence of vaccination policies, quarantine and isolation procedures. Consequences are deaths, illnesses, and estimates of economic costs. Sensitivity analysis is a statistical technique to investigate how uncertainty in the input variables affects the model outputs and which input variables tend to drive variation in the outputs. Such analysis can provide critical information for decision makers and public health officials who may have to deal with the realities of a virulent infectious disease. This paper presents the results of preliminary analyses of the effects of inputs to the infectious disease model on the calculated consequences.*

## Introduction

The whole world lives under the threat of infectious diseases. The effects of virulent infections with high lethality rate, whether from natural causes or from explicit acts of war, are quite similar. Many people will die; medical and support systems will be stressed. People will demand protection from the threat of disease, either through vaccination, new drugs, or self-isolation. The desire to avoid contact, and thus infection, may result in people staying home from work, resulting in disruption of many services and substantial economic losses. Interventions to mitigate the effects of an infectious disease epidemic include preventative treatments such as vaccination, the use of effective treatment regimes, application of strategies to reduce the rate of spread of the disease, and the use of policies and technologies to provide early detection and rapid response to the presence of the disease.

With the greatly expanded trade and travel, infectious diseases can spread at a fast pace within and across country borders, resulting in potentially significant loss of life, major economic crises, and political instability. With an increasingly mobile society, infectious diseases can spread rapidly across state and national borders. Health officials must be

prepared to recognize and respond to these new and emerging threats to public health in order to reduce illness and death. Recent examples of the rapid transmission and consequences of newly emerging diseases include the severe acute respiratory syndrome (SARS) and avian influenza (H5N1 strain) in Asia. As witnessed in Asia, infectious disease outbreaks could impact national critical infrastructure, cause large fatalities, and substantial economic damage. While many studies have sought to estimate the impacts of infectious disease on total illnesses and deaths, few studies and modeling efforts have sought to determine the intangible costs that often surround disease spread and containment, such as social stigmatization, loss of worker productivity, apprehension, impacts to transportation, and the economic sector. However, it can be the intangibles that can quickly add up and cause extreme economic and emotional harm to a region or country.

Natural biological variation in pathogen impacts and responses such as transmission rate, in addition to uncertainty in disease parameters can greatly impact the outcomes of modeling efforts of the impacts on our Nation's critical infrastructure. It is important to then understand if the model's pattern of behavior is strongly influenced by changes in the uncertain parameters. For many diseases these uncertainties can be based on centuries-old data. Some examples of these uncertain parameters would include transmission, infectivity rate, incubation time, mortality, recovery rate, and the stage of the greatest infectivity.

Some simple characteristics of the epidemiology of the disease may remain ambiguous at best, or a large range of values exist for disease parameters. At worst, a complete disagreement of the disease characteristics may appear in the literature. Clearly, there will be uncertainties in epidemiological modeling that must be incorporated into the modeling effort. Completing sensitivity analyses with a dynamics system model will elucidate the role of uncertainties and enable the extrapolation of the results into decision support systems.

## **Infectious Disease Model**

In general, infectious disease models are classed [Koopman] as either compartmental or as network models. The former are characterized by differential equations under the assumption of continuous mixing of the population. Network models represent individuals with unique connections and habits (typically stochastically assigned). Compartmental models have considerable history, while network models are relatively recent, e.g., see Eubank, et al. The infectious disease model described here is a modified SEIR (Susceptible-Exposed-Infected-Recovered) model [Murray] implemented in Vensim Version 5.2. The model described here uses an expanded set of disease stages, demographic groupings, an integrated model for vaccination, quarantine and isolation, and demographic and stage-dependant contact rates. As a variant on the SEIR model paradigm this implementation represents the populations as homogeneous and well mixed with exponentially distributed residence times in each stage (characterized with a nominal residence time) [Hethcote]. However, the use of additional stages and demographic groupings is designed to add additional heterogeneity where it can be useful in capturing key differences between sub-populations for disease spread and response.

## **Disease Stages**

The disease stage representation within a stock and flow structure is shown in Figure 1 (the use of prefixes and color-coding for variable names in the diagram is due to compliance with coding standards [Thompson, et al.] defined for the parent critical infrastructure simulation project [Bush, et al.]). The stages are represented in a generic manner so that the model can be used for a large number of infectious agents by selecting the input parameters appropriately. Each stage is designed to carry some combination of four key characteristics, namely whether at this stage, the disease shows clinical symptoms (patient exhibits symptoms characteristic of the disease), is infectious (the disease can spread to another person due to contact or proximity) to other persons, is preventable by vaccine, and the characteristic residence time in the stage. This is accomplished by indexing infectivity, contact rates, residence time and vaccine effectiveness by disease stage. In this manner, disease parameters such as infectivity (the probability that a given contact will result in an infection) are a characteristic of the persons in a given stage of the disease. Each stage with the right set of characteristics, in combination with the remaining stages, can be made to simulate a variety of diseases. Two stages of recovery are used to accommodate diseases whose immunity extends for a relatively short period.

The population tracked throughout the model is divided into demographic groups that allow the differing behavior between age groups and other demographics to be represented. In the current model six groups are used. The first five groups represent infants, youth, younger adults, older adults and the elderly. The sixth group is used to treat the class of people who respond to a disease outbreak. This class includes health care workers, emergency responders and law enforcement personnel who behave differently from the bulk of the population in such a crisis. These groups are represented in the model using subscripted variables for the stocks representing the disease stages and associated auxiliary variables. The model assumes that the duration of the outbreaks to be modeled is short enough that the effect of movement between the demographic groups is small.

## **Vaccination**

The process of targeted or ring vaccination is modeled starting with the identification of an index case with disease-specific symptoms who identifies some fraction of the people with which they have had recent contact. Each contact is queued for quarantine and/or vaccination. Of the contacts identified by each index, a portion will have been infected and be in some stage of the disease depending on how long the contact took place prior to the identification of the index case. Therefore contacts are queued out of every stage up to index case identification, as well as the susceptible population (for those not infected), in proportion to their contact rates and the time since the contact (as estimated by the residence time in each stage). This is illustrated in Figure 2, which isolates a portion of the model for stages zero and one. Note that the queued and vaccinated populations are represented with a parallel set of stages to model the fact that people continue to get sick while they are waiting in a queue or have been unsuccessfully vaccinated. Since it is generally unknown which stage a contact is in when they are identified, many will be queued for vaccination and vaccinated after the point where it is useful to do so.

The initial design of the model represented the competing flows from each stage – for moving to the next stage (based on average residence time) versus being queued for vaccination or being vaccinated – as multiple flows out of each stock. However this created some difficulties because of the very different characteristic times associated with moving between stages versus moving out to be queued or vaccinated. The results tended to be unpredictable and result in an artificial slowing of progress of an outbreak because of the competing stage flows across and vaccination flows down. A more reliable method of splitting the movement of people was implemented by dividing the flows as they enter each stock according to the probability that people will flow out of the stock for queuing or vaccination before their residence time in the stage expires.

Mass vaccination is treated separately because of the differing dynamics with mass vaccination. For example, people identified for mass vaccinated are generally not quarantined unless they become symptomatic in contrast to targeted vaccination where there is greater reason to believe that they are at risk. The stock and auxiliary variables for this mass vaccinated population begin with the word ‘Mass’ (such as ‘Mass Susceptible’) as illustrated at the bottom of Figure 2.

## **Contact Rates**

Contact rates are one of the most important drivers in modeling the spread of an infectious disease, analogous to the use of a reproduction number in many infection models. The use of contact rate was chosen because it can be more directly related to the differing behavior amongst age groups and the choices made with regard to quarantine and isolation.

The contact rate model is shown in Figure 3. It is divided into two sections, one concerned with the behavior of infected individuals and the other with the rest of the population. The quarantine behavior of the bulk of the population is estimated based on the population targeted for vaccination typically quarantined as a precaution and an estimate of the tendency of persons to self-quarantine based on government alerts and the severity of the disease. Quarantine behavior can have an important effect on the spread of the disease but also affects business operations and the demand for business and infrastructure services.

The contact behavior for infected persons is treated separately for individuals not queued for vaccination, those queued for targeted vaccination, undergoing a targeted vaccination and those undergoing mass vaccination. Contact behavior is also stage dependent for these individuals since they may be symptomatic and reduce their interactions with others. Contact rate is also allowed to vary with demographic group using an age group index.

Transmission of infectious diseases depends on the two factors that may differ in each stage of the disease, the contact rate within each age group and the relative infectiousness of the disease. Differences in infectivity between stages may also be disagreed upon. Halloran et al. (2002) and Kaplan et al. (2002) made the assumption that prodromal

period in a disease such as smallpox is highly infectious whereas Legrand et al. (2003) do not consider individuals in the prodromal stage to be infectious. This infectivity rate difference in stage could have potentially large consequences in model output. This disease model was designed to capture each aspect of transmission within each stage using the contact rate model.

Depending on when detection of a specific disease outbreak occurs, treatment or vaccination may be too late for substantial control of the disease and resulting mortality. Therefore, each stage is modeled to include the entire progression of the disease which would be inevitable for many of the initial cases. Recovery can occur at all stages of the disease if treatment or vaccination is initiated.

### **Other Features**

The infection model is designed for integration into a much larger model set that includes a population model and models of vaccination production, detectors, government behavior, economics and key infrastructures including public health, emergency services, energy, water, food and telecommunications. The features of this larger simulation project, called CIP/DSS (Critical Infrastructure Protection Decision Support System), is discussed elsewhere [Bush]. However the model can also be run in isolation although some of the dynamics associated with interdependencies with these other sectors are absent in that case.

Government response in the model in the form of quarantine and vaccination programs is initiated after the discovery of an outbreak either through direct detection of the outbreak or recognition of the first cases in the public health system. Detection can take place relatively early in the outbreak while recognition by the public health system awaits the first infected persons reaching the initial symptomatic stages of the disease with an additional delay for confirmation and government actions. Government action can select from several response options including targeted vaccination, mass vaccination and combinations of these. Vaccination can be biased toward particular sub-populations to model priority vaccinations of children or health care personnel. Allowances for refusing vaccination and separating segments of the population who cannot tolerate vaccination out of the queues can also be made.









The application of this infection model and the larger CIP/DSS simulation package is limited to those where the treatment of the aggregate behavior of the system is adequate. The situation is analogous to the difference between modeling disease spread using an SEIR type model as we do here versus an agent-based model where the behavior of individuals is modeled. Some heterogeneity can be introduced by indexing on demographics and other important characteristics but those sub-groups are still treated in an aggregate fashion.

### **Confidence in Model**

Every simulation model should undergo a process of assessing whether the model provides results in which the users have confidence, at least to the level at which the model is intended for use. Terms such as validation and accreditation are often used for this process, although validation of a simulation model is problematical [Oreskes et al.]. For the purposes of this paper, the term confidence building will be used since models cannot be proved to be valid, only disproved. To build confidence in this model, the results have been assessed against the published results of other models for selected diseases. This assessment is still in process, but early results show that this model closely reproduces the results of other SEIR models for the limited number of diseases examined thus far. More work is needed to better establish confidence in this model.

### **Sensitivity Analysis**

Historical cases of many epidemics display considerable variations that arise from a multitude of sources including the number of initially infected, the rate of spread of the disease, the lethality of the infection, and the total number of fatalities. As a result, depending on the magnitude of these variations, the behavior of an epidemic will vary from one realization to another. While these realizations are individually deterministic, the overall collective effect is to make the characteristics appear in a range of values. Capturing the uncertainty over this data is essentially the specification of probability distributions for the parameters. This parametric uncertainty is a characterization not only of system parameters but also, to some extent, the external environment. This characterization involves the development of methods to model both epistemic (irreducible uncertainty resulting from inherent variability of a quantity) and aleatoric (theoretically reducible uncertainty resulting from lack of knowledge about a quantity) type. The next step is to attribute variance in the outputs to sources of variance in the inputs.

Sensitivity analysis is a statistical technique that can be applied to mathematical or computational models to provide insight on how uncertainty in the input variables affect the model outputs and which input variables tend to drive variation in the outputs. For models such as this infectious disease model, where the output is intended to inform decision-makers, uncertainty in the output can be disconcerting, as a single value is not given. However, the benefit is that a range of output values reveals a suite of possible model outcomes. It is sometimes of value to put in place robust policies that avoid the worst that can happen, although the policies may not achieve optimal results. If mitigating policies and alternatives are included and parameterized within the model,

then sensitivity analysis can provide estimates of how responsive key output measures are with respect to these inputs.

The term “sensitivity analysis” is variously interpreted in different technical communities and problem settings. Until recently, sensitivity analysis was typically considered as a local measure of the effect of a given input on given output, customarily obtained by estimating system derivatives such as:  $S_j = \partial Y / \partial X_j$ , where  $Y$  is the output of interest and  $X_j$  an input factor. While the local approach is valuable for some problems, other problems, such as the analysis of risk and decision support demand a different approach. For these problem domains the degree of variation of the input factors is material, as one of the outputs being sought from the analysis is a quantitative assessment of the uncertainty around some best estimate value for  $Y$ , the output vector (uncertainty analysis). The application of Monte Carlo methods in conjunction with a variety of sampling strategies can achieve this. In this context, sensitivity analysis is aimed at establishing priorities, to determine which factor most needs better determination, and to identify those inputs that propagate most variance in the output. Sensitivity analysis in this context is often performed using techniques related to regression analysis, using the regression coefficient for a given factor as a measure of sensitivity. Methods of this type are termed *global*, to distinguish them from *local* methods, where only one point of the factor’s space is explored, and factors are changed one at a time. A disadvantage of regression based methods is that their performance is poor for non-linear models, and can be misleading for non-monotonic models [Saltelli, et al.].

## Steps in Sensitivity Analysis

Sensitivity analysis is a method that can require iteration and refinement to achieve the desired goals of the analyst. The following steps [Saltelli, et al.], shown in Figure 4, are suggested to provide a working guide to the process. Not all steps must be taken and the analyst may find it worthwhile to embellish one or more steps.

1. Identify simulation outputs and features based on the goals of the analysis.
2. Decide which input factors to include in the analysis.
3. Choose a probability distribution function for each of the input factors.
4. Choose a sensitivity analysis method on the basis of the following:
  - a. The question being asked;
  - b. Number of model evaluations;
  - c. Presence of correlation structure;
  - d. Use of an experimental design.
5. Generate the input sample. This has the form of  $N$  strings of input factor values on which the model is evaluated.
6. Evaluate the model on the generated input sample and produce the output, which contains  $N$  output values in the form specified in Step 1.
7. Analyze the model outputs and draw conclusions, possibly starting a new iteration of the analysis.

## STEPS IN SENSITIVITY ANALYSIS

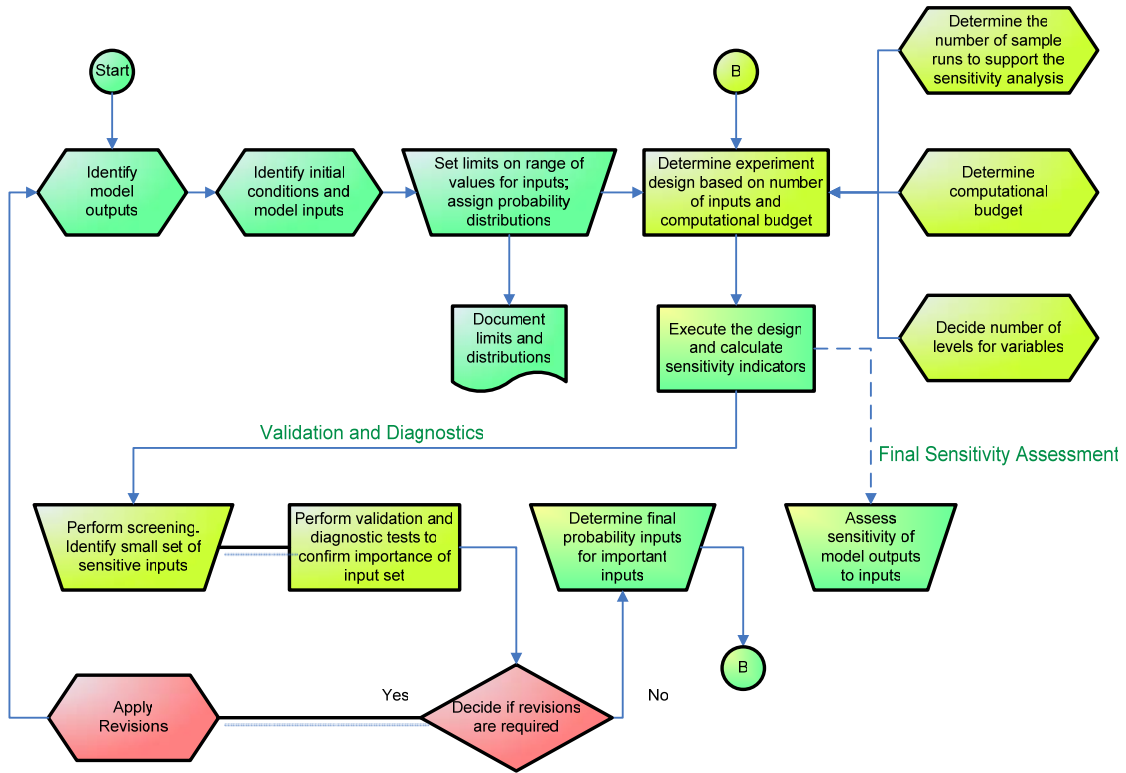


Figure 4: Steps in sensitivity analysis.

If an experimental design is implemented resulting in  $N$  output samples, then the importance measure of  $x_i$  is given by the estimate of the correlation ratio [McKay, 98]:

$$\eta^2 = \text{Var}(\hat{y})/\text{Var}(y).$$

In this case,  $y$  is the set of all outputs from the design and  $\hat{y} = E(y | x_i)$ . The larger the correlation ratio, the more important  $x_i$  is in contributing to the variance of the outputs. For the results presented in this paper, an experimental design of 817 runs varying 40 factors over five levels was used.

### Application

The inputs to the infectious disease model chosen for this study are given in Table I. The selected levels are based on the nominal value for the parameter, as well as a low value and a high value. Thus the value for level 1 is the low value; level 2 is the midpoint between the level 1 and the nominal value; level 3 is the nominal value; level 4 is the midpoint between the level 3 and the high value; and level 5 is the high value. The nominal, low, and high values were determined from the literature and from domain experts.

Table I: Model inputs selected for sensitivity analysis.

Variable Name	Level-1	Level-2	Level-3	Level-4	Level-5	Units
Sdc: Normal Contact Rate[Upo Infant] <sup>1</sup>	3	4	5	6	7	c/day <sup>2</sup>
Sdc: Normal Contact Rate[Upo Youth]	9	12	15	18	21	c/day
Sdc: Normal Contact Rate[Upo Young Adult]	6	7	8	9	10	c/day
Sdc: Normal Contact Rate[Upo Old Adult]	2	4	6	8	10	c/day
Sdc: Normal Contact Rate[Upo Elderly]	1	3	5	7	9	c/day
Sdc: Isolation Contact Rate	0.02	0.06	0.1	0.2	0.3	c/day
Sdc: Quarantine Contact Rate	0.1	0.3	0.5	1	1.5	c/day
Sdc: Quarantine Effectiveness	0.5	0.7	0.9	0.95	1	dmnl <sup>2</sup>
Sdc: Self Quarantine Tendency Maximum[Upo Infant]	0.45	0.7	0.95	0.975	1	dmnl
Sdc: Self Quarantine Tendency Maximum[Upo Youth]	0.4	0.6	0.8	0.9	1	dmnl
Sdc: Self Quarantine Tendency Maximum[Upo Young Adult]	0.15	0.225	0.3	0.65	1	dmnl
Sdc: Self Quarantine Tendency Maximum[Upo Old Adult]	0.2	0.3	0.4	0.7	1	dmnl
Sdc: Self Quarantine Tendency Maximum[Upo Elderly]	0.4	0.6	0.8	0.9	1	dmnl
Sdc: Removal From Quarantine Delay	240	372	504	672	840	hours
Sdc: Fraction Of Symptomatic Not Isolated	0	0.1	0.2	0.35	0.5	dmnl
Sdc: Quarantine Reaction Delay	1	4	7	14	21	days
Sdc: Sickness Contact Factor	0.1	0.175	0.25	0.4	0.55	dmnl
Sdp: Test Initially Infected	200	600	1000	1400	1800	person
Sdp: Unvaccinated Infectivity[Sdp Stage Two]	0.01	0.03	0.05	0.125	0.2	dmnl
Sdp: Unvaccinated Infectivity[Sdp Stage Three]	0.05	0.13	0.21	0.28	0.35	dmnl
Sdp: Vaccinated Infectivity[Sdp Stage Three]	0	0.025	0.05	0.125	0.2	dmnl
Sdp: Sdp Stage Residence Time[Sdp Stage Zero]	2	2.5	3	3.5	4	days
Sdp: Fraction Initially Immune	0.2	0.32	0.44	0.55	0.66	dmnl
Sdp: Disease Fatality[Upo Infant]	0.2	0.27	0.34	0.39	0.44	dmnl
Sdp: Disease Fatality[Upo Youth]	0.2	0.157	0.114	0.207	0.3	dmnl
Sdp: Disease Fatality[Upo Young Adult]	0.2	0.145	0.09	0.195	0.3	dmnl
Sdp: Disease Fatality[Upo Old Adult]	0.1	0.1505	0.201	0.2505	0.3	dmnl
Sdp: Disease Fatality[Upo Elderly]	0.2	0.287	0.374	0.387	0.4	dmnl
Upo: Initial Population	1.0e6	3.0e6	5.0 e6	7.0e6	9.0e6	person
Sdp: Fraction Vaccine Intolerant	0.08	0.14	0.2	0.26	0.32	dmnl
Sdp: Fraction Refusing Vaccination	0	0.025	0.05	0.15	0.25	dmnl
Gpo: Vaccination Policy	1	2	2	3	4	dmnl
Pvr: Maximum Vaccination Rate	1000	2500	4000	5500	7000	p/hour <sup>2</sup>
Sdp: Mass Vaccination Phase In	1	2	3	5	7	days
Sdp: Targeted Vaccination Phase In	2	3	4	6	8	days
Sdp: Name Overlaps	0.5	0.65	0.8	0.9	1	dmnl
Sdp: Fraction Named	0.5	0.75	1	3	5	dmnl
Sdp: Interview Time	3	5	7	9	11	days
Sdp: Index Fraction	0	0.125	0.25	0.375	0.5	dmnl
Sdp: Policy Time	2	8	14	20	26	days

<sup>1</sup> Some variables are indexed by age group. The age groups are 0-4, 5-18, 19-44, 45-64, and 65 and over.

<sup>2</sup> c/day is contacts per day; dmnl is dimensionless; p/hour is person per hour.

These inputs were organized according to an experiment design to support both uncertainty and sensitivity analysis. The design is based on a combination of two orthogonal arrays, one with 3 levels and the other with 2 levels that yield a design with 5 levels for 40 factors [Moore, et al.]. The total number of distinct sets of inputs is 817.

Outputs from the simulation are collected for analysis. For this experiment the outputs are listed in Table II. Most of the output variables are provided for diagnostic purposes. The primary metrics of the simulation are Sdp: Total Fatalities, Sdp: Total Cumulative Afflicted, and Sdp: Outbreak Duration. The first denotes the number of fatalities due to the disease throughout the population. The second is the total number of people who contract the disease, irrespective of the outcome. Third is the duration of the outbreak. The outbreak is defined to be over when the number of patients with the disease in the hospital is less than a specified threshold (say 5) and the infection rate indicates that the disease is dying out, or the simulation ends. The simulation is run for 365 days and the disease outbreak begins at day 3.

Table II. Variables output from the simulation for analysis.

<b>Variable</b>	<b>Unit</b>
Sdc: Alert Time	minute
Sdc: Contact Rate[Upo Age Groups]	c/day
Sdc: Mass Vaccinated Contact Rate[Upo Age Groups]	c/day
Sdc: Queued Contact Rate[Upo Age Groups]	c/day
Sdc: Total Quarantined	person
Sdc: Vaccinated Contact Rate[Upo Age Groups]	c/day
Sdp: Actual Vaccination Rate	p/hour
Sdp: Outbreak Duration	days
Sdp: Sum All Infecteds	person
Sdp: Sum All Mass Vaccinated	person
Sdp: Sum All Queued	person
Sdp: Sum All Susceptibles	person
Sdp: Sum All Vaccinated	person
Sdp: Sum Infecteds	person
Sdp: Total Cumulative Afflicted	person
Sdp: Total Early Recovered	person
Sdp: Total Fatalities	person

## Results

Summary results of the experimental design are shown in Figure 5 as histograms of the observed outputs. The outputs of interest are the total number of people quarantined due to the government imposed vaccination policy (Sdp: Total Quarantined), the duration of the disease outbreak in days (Sdp: Outbreak Duration), the total number of people that are infected with the disease (Sdp: Total Cumulative Afflicted), and the number of fatalities (Sdp: Total Fatalities). Since the disease model generally produces exponential growth, a log transformation is deemed appropriate. Figure 6 depicts the histograms of the logarithm of the data.

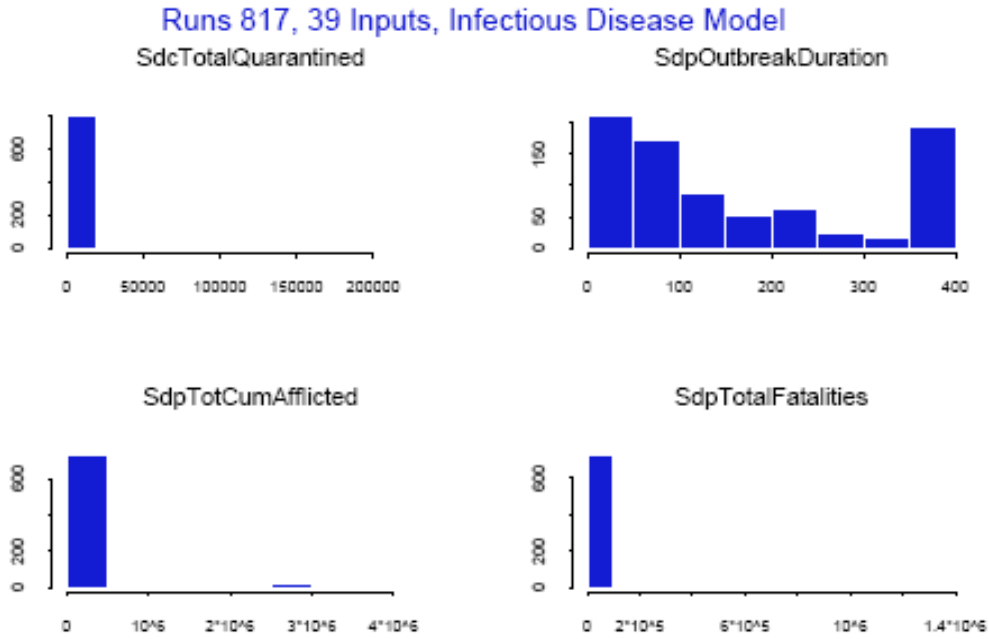


Figure 5: The histograms of key outputs from the infectious disease model.

It should be noted that Sdp: Outbreak Duration expresses the length of time that the size of the exposed population is greater than a specified threshold, in this case 5. There are cases where this condition is not met, i.e., there are more than 5 exposed people at the end of the simulation run. In such cases, the outbreak duration is the number of days from the beginning of the outbreak to the end of the simulation.

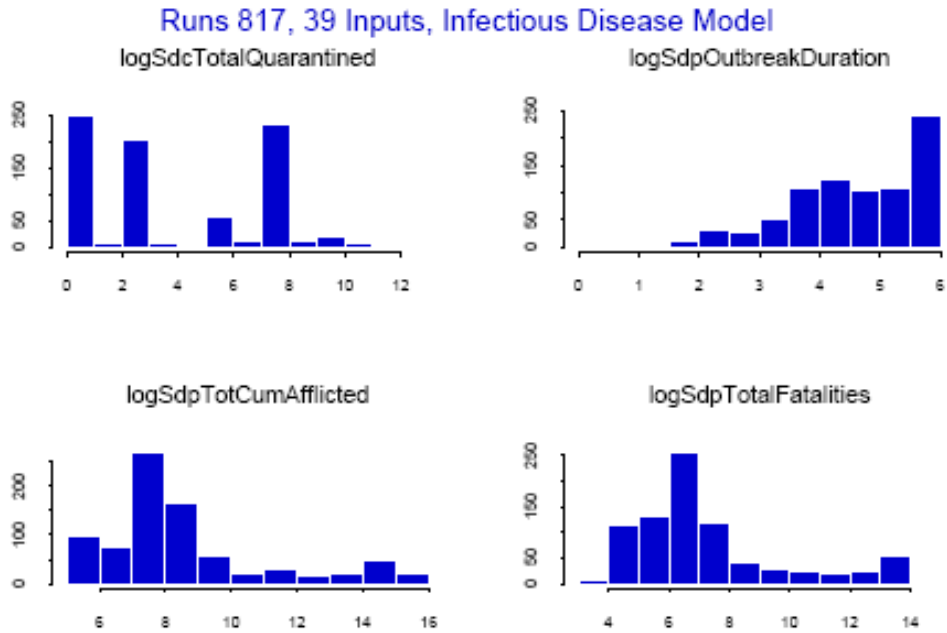


Figure 6: Log transforms of the key output metrics.

Figure 7 shows the sensitivity of the outputs to the input factors, by rank of the correlation value of the input. In each plot the top five variables are listed that have the highest correlation values. Also, each plot has a solid line near the x-axis that depicts the level of significance for the correlation statistic. If a correlation measure falls above the line then it can be considered significant at the 0.05 level. What is notable in these plots is that none of the input tend to be strong predictors of the outputs by themselves. For Sdp: Total Fatalities, the top predictor is Sdc: Quarantine Contact Rate, but its value is only about 0.17. This same input is also the single best predictor of Sdp: Total Cumulative Afflicted and Sdp: Outbreak Duration.

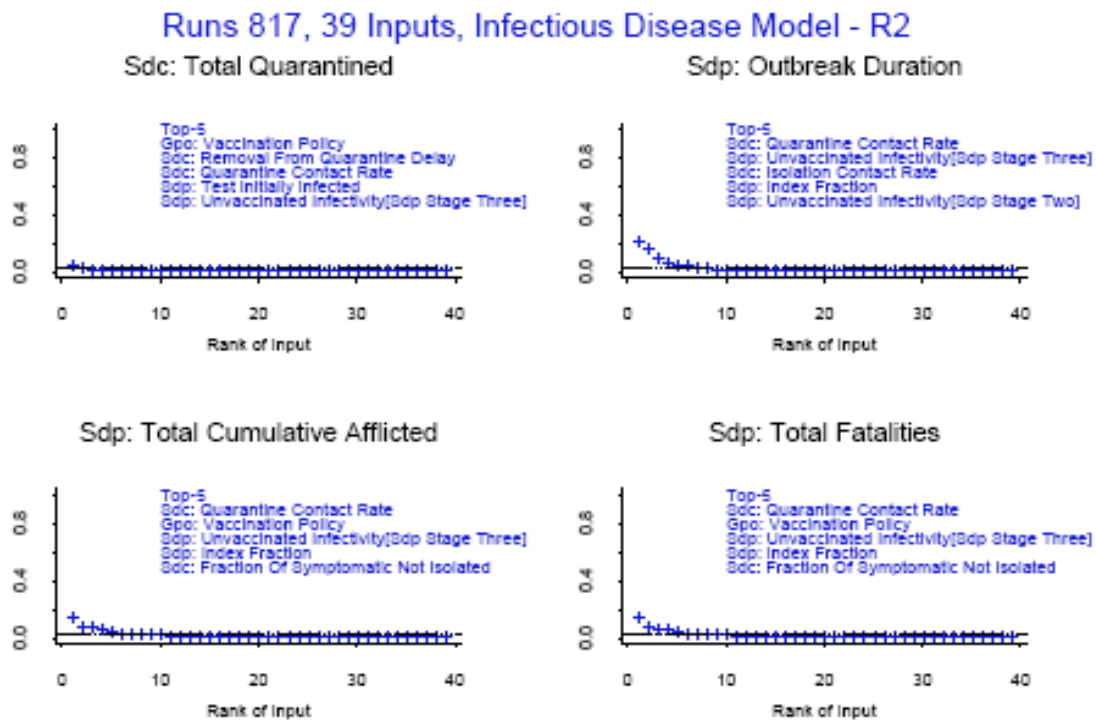


Figure 7: The sensitivity of the major metrics to the model inputs.

Figure 8 shows the sensitivity analysis performed on the (natural) log transformation of the key outputs. The monotonic log transformation has the potential to reveal more structure in the outputs than can be discerned in the raw data.



## Runs 817, 39 Inputs, Infectious Disease Model - R2

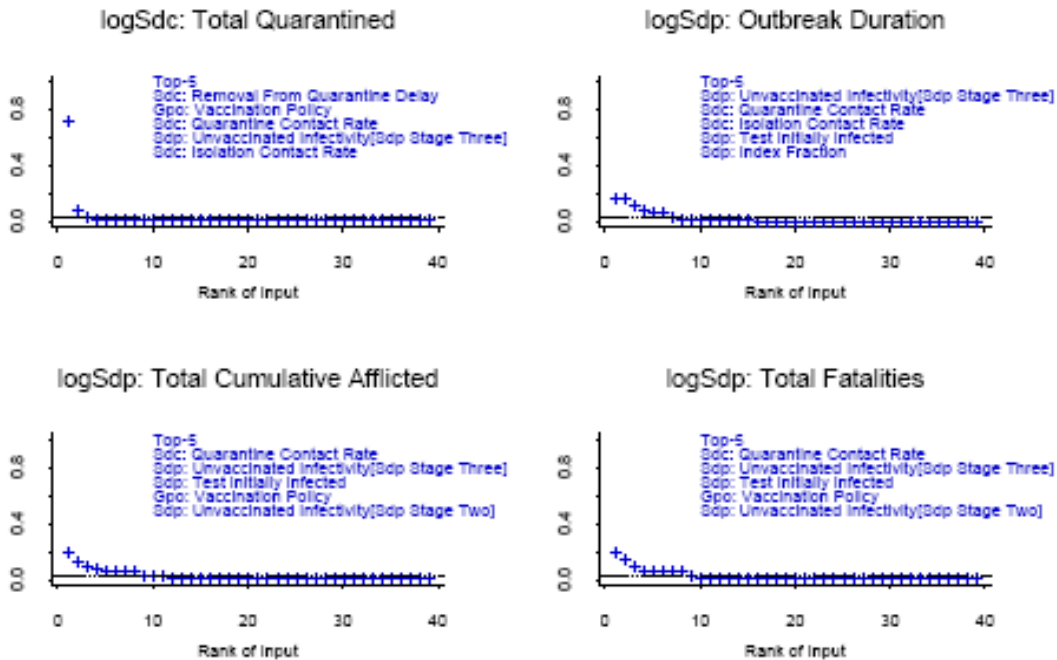


Figure 8: The sensitivity of the log of major metrics to the model inputs.

As can be seen by comparing Figure 7 and Figure 8, the log transform shows an improved relationship between Total Quarantined and Removal From Quarantine Delay. A slight improvement is noted between Total Cumulative Afflicted and Quarantine Contact Rate, and Total Fatalities and Quarantine Contact Rate. The effect on Outbreak Duration is to reverse the order of sensitivity to the top two inputs, thereby slightly raising the measured sensitivity to Unvaccinated Infectivity. Overall, however, the sensitivity of these model outputs is relatively low to the specified inputs. It was expected that contact rate and infectivity would be

If one subsequently looks at how two variables at a time contribute to the variance of the outputs, the logical approach is to use the top correlate and then examine the other variables for their contribution. This is done in Figure 9. The variable  $\log(\text{Sdp:Total Fatalities})$  has four top correlates: Sdc: Quarantine Contact Rate, Sdp: Unvaccinated Infectivity[Stage 3], Sdp: Test Initially Infected, and Gpo: Vaccination Policy. This plot examines the variance that can be accounted by inputs given that the variance associated with one of the top four correlates is already counted. In the plot for “SdpInitInfect” (lower left subplot of Figure 9), shows that Sdp: Initially Infected and Sdc: Quarantine Contact Rate together account for approximately 35% of the variance in the log of total fatalities. Each plot in Figure 9 has a dashed line near the x-axis that depicts the level of significance for the correlation statistic for the specified conditional variable, e.g., for Sdc: Quarantine Contact Rate for the upper left subplot. Each plot also has a heavy dashed line that depicts the fraction of variance accounted by the conditional variable. If the two lines are added, the resulting level is considered the line at which the listed inputs can be considered significant at (approximately) the 0.05 level. For the upper left plot, the

line of significance is 0.21; any of the correlates with a value greater than 0.21 would be considered significant. In this case, the top four correlates are significant in each subplot.

### logSdpTotalFatalities - JR2 / Runs 817, 39 Inputs, Infectious Disease Model

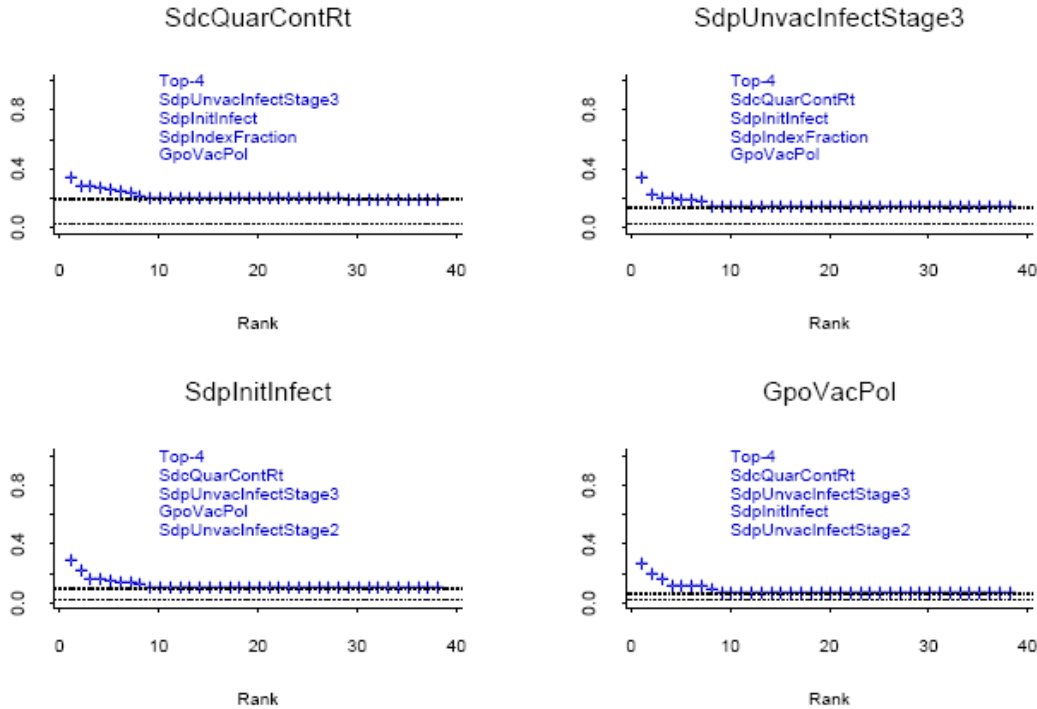


Figure 9. The sensitivity of the log of total fatalities to two inputs.

This type of analysis can compute the fraction of variance due to the top  $k$  correlates. Due to resource limitations only the cases where  $k = 1$  or  $2$  are examined in this study. This analysis reveals that for this model of infectious disease, the most influential inputs are Sdc: Quarantine Contact Rate, Sdp: Unvaccinated Infectivity[Stage 2], Sdp: Unvaccinated Infectivity[Stage 3], Sdp: Test Initially Infected, and Gpo: Vaccination Policy with respect to estimating the total number of fatalities.

## Conclusions

A model of infectious diseases has been developed to assess the interdependencies of critical infrastructures placed under stress by disease outbreaks. The model has been parameterized to represent a scenario where a variety of infectious agents are represented in a large metropolitan area while at the same time working within a much larger model context including simulation of each of the major infrastructures, population and economics. Sensitivity analysis, a statistical technique used to identify input variables driving variation in the outputs, has been used to assess which inputs of this model most affect predicted fatalities. This method provides a quantitative assessment of the relative importance of model inputs. The top correlates for total fatalities include the rate of contact of infected persons, the probability of a contact leading to an infection, and the vaccination policy. Although these correlations are in the range of 0.1 – 0.2, a log transformation on the outputs provides modest improvement. In examining the

contribution of inputs two at a time, up to 35% of the variance can be accounted. The results of this analysis show that the sensitive inputs to the infectious disease model match a priori expectations – contact rates, infectivity, and treatment policy. Although these variables predict the key outputs rather weakly, they can in combination account for larger fractions of the observed output variance.

## Acknowledgements

The authors would like to thank the Science and Technology Office of the Department of Homeland Security for funding this work.

## References

- Bush, B., Dauelsberg, L., LeClaire, R., Powell, D., DeLand, S., and Samsa, M., “Critical Infrastructure Protection Decision Support System (CIP/DSS) Overview,” Los Alamos National Laboratory Report LA-UR-05-1870, Los Alamos, NM 87544, 2005.
- Eubank, S., Guclu, H., Kumar, V., Marathe, M., Srinivasan, A., Toroczkai, Z., and Wang, N. “Modeling Disease Outbreaks in Realistic Urban Social Networks,” *Nature*, Vol. 429, 13 May, 2004, pp. 180-184.
- Fraser, C., Riley, S., Anderson, R., and Ferguson, N., “Factors that Make an Infectious Disease Outbreak Controllable,” *PNAS*, Vol. 101, No. 16, April 20, 2004, pp. 6146-6151.
- Halloran, M.E., I.M. Longini Jr., A. Nizam, and Y. Yang. 2002. Containing bioterrorist smallpox. *Science*. 298: 1428-1432.
- Helton, J.C. and Davis, F.J., Sampling-Based Methods for Uncertainty and Sensitivity Analysis, SAND99-2240, Sandia National Laboratories, Albuquerque, NM, 2000.
- Hethcote, Herbert W., “The Mathematics of Infectious Diseases”, *SIAM Review*, Vol 42, No. 4, pp. 599-653.
- Kaplan, E.H., Craft, D.L., and L.M. Wein. 2002. Emergency Response to a Smallpox Attack: The Case for Mass Vaccination. *Proceedings of the National Academy of Science*. 99 (16): 10935-10940.
- Koopman, J., “Modeling Infection Transmission,” *Annual Review of Public Health*, Vol. 25, 2004, pp. 303-326.
- Legrand, J., C. Viboud, P.Y. Boelle, A.J. Valleron, and A. Flahault. 2003. Modeling Responses to a Smallpox Epidemic Taking into Account Uncertainty. *Epidemiol. Infect.* 132: 19-25.

- McKay, M.D., Beckman, R.J., and Conover, W.J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239-245.
- McKay, M.D., "Evaluating Uncertainty in Stochastic Simulation Models," Second International Symposium on Sensitivity Analysis of Model Output (SAMO '98), Venice, Italy, April 19-22, 1998.
- Moore, L. M., McKay, M. D., and Campbell, K., "Combined Array Experiment Design," Los Alamos National Laboratory Report, LA-UR-04-0374, Los Alamos, NM, 2004.
- Murray, J.D., Mathematical Biology, Springer-Verlag, New York, 1989.
- Oreskes, N., Shrader-Frechette, K., Belitz, K., "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," *Science*, Vol. 263, Issue 5147, pp. 641-646.
- Owen, A., "Orthogonal Arrays for Computer Experiments, Integration, and Visualization," *Statist. Sinica* 2, pp. 439-452, 1992.
- Thompson, D., Powell, D., and Bush, B., "CIP/DSS Metropolitan Team Standardization Guide," Los Alamos National Laboratory Report LA-UR-04-6148, Los Alamos, New Mexico, 2004.
- Saltelli, A., Tarantola, S., Campolongo, F. and M. Ratto, Sensitivity Analysis in Practice, by John Wiley and Sons, Ltd., Chichester, England, 2004, pp. 45-47.
- Walther, B. and P.W. Ewald, 2004, Pathogen survival in the external environment and the evolution of virulence. *Biological Reviews* 79:849-869.
- Wojtkiewicz, S.F., Eldred, M. S., Field, R.V., Jr., Urbina, A., and J.R. Red-Horse, "Uncertainty Quantification in Large Computational Engineering Models," *Proceedings of the 2001 American Institute of Aeronautics and Astronautics*, pg. 1455.