NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ASSESSING THE EFFECTIVENESS OF BIOSURVEILLANCE VIA DISCRETE EVENT SIMULATION

by

Jason H. Dao

March 2011

Thesis Advisor:	Ronald D. Fricker, Jr.
Second Reader:	Rachel T. Silvestrini

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2011	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Assessing the Effectiveness of Biosurveillance Via Discrete Event Simulation		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Jason H. Dao			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number _____N.A._____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited		**12b. DISTRIBUTION CODE** A	

13. ABSTRACT (maximum 200 words)

Bioterrorism is not a new threat, but the potential for disastrous outcomes is greater than it has ever been. In order to confront this threat, biosurveillance systems are utilized to provide early warning of health threats, early detection of health events, and situational awareness of disease activity. To date, there is little known about the performance of such biosurveillance systems in comparison to diagnosis capabilities of medical personnel. In this thesis, a discrete event simulation model of an anthrax outbreak is developed in order to analyze the performance of such biosurveillance systems in comparison to medical personnel. This research found the Early Aberration Reporting System C1 statistical algorithm is useful in early event detection of a bioterror attack. Given an exposed population of 1,000 people, the nominal probability that the algorithm signals first is 31.5% and it is 0.3% for an exposed population of 10,000 people. Given an exposed population of 1,000 people, the nominal time it takes for the algorithm to signal is 3.3 days and 0.38 days for an exposed population of 10,000 people.

14. SUBJECT TERMS Biosurveillance, Syndromic Surveillance, C1, Discrete Event Simulation, EARS, Early Event Detection (EED)			**15. NUMBER OF PAGES** 103
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

ASSESSING THE EFFECTIVENESS OF BIOSURVEILLANCE VIA DISCRETE EVENT SIMULATION

Jason H. Dao
Lieutenant, United States Navy
B.A., University of San Diego, 2004

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 2011

Author: Jason H. Dao

Approved by: Ronald D. Fricker, Jr., PhD
Thesis Advisor

Rachel T. Silvestrini, PhD
Second Reader

Robert F. Dell, PhD
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Bioterrorism is not a new threat, but the potential for disastrous outcomes is greater than it has ever been. In order to confront this threat, biosurveillance systems are utilized to provide early warning of health threats, early detection of health events, and situational awareness of disease activity. To date, there is little known about the performance of such biosurveillance systems in comparison to diagnosis capabilities of medical personnel. In this thesis, a discrete event simulation model of an anthrax outbreak is developed in order to analyze the performance of such biosurveillance systems in comparison to medical personnel. This research found the Early Aberration Reporting System C1 statistical algorithm is useful in early event detection of a bioterror attack. Given an exposed population of 1,000 people, the nominal probability that the algorithm signals first is 31.5% and it is 0.3% for an exposed population of 10,000 people. Given an exposed population of 1,000 people, the nominal time it takes for the algorithm to signal is 3.3 days and 0.38 days for an exposed population of 10,000 people.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ALGO	Algorithm
CDC	Centers for Disease Control and Prevention
CUSUM	Cumulative Sum
DOC	Doctor
DES	Discrete Event Simulation
DNA	Deoxyribose Nucleic Acid
DTRA	Defense Threat Reduction Agency
EARS	Early Aberration Reporting System
EED	Early Event Detection
ESSENCE	Electronic Surveillance System for the Early Notification of Community-Based Epidemics
GI	Gastrointestinal
HPAC	Hazard Prediction and Assessment Capability
HSPD-21	Homeland Security Presidential Directive 21
MSE	Mean Square Error
OLS	Ordinary Least Squares
UN	United Nations

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Bioterrorism is not a new threat, but the potential for disastrous outcomes is greater than it has ever been. The U.S. government recognizes the threat and, via Homeland Security Presidential Directive 21 (HSPD-21), has directed “further improvement in the preparedness of our public health and medical systems to address current and future biological warfare threats and to respond with greater speed and flexibility to multiple or repetitive attacks” (HSPD-21, 2007). In order to confront this threat, biosurveillance systems are utilized to provide early warning of health threats, early detection of health events, and situational awareness of disease activity. To date, there is little known about the performance of such biosurveillance systems in comparison to medical personnel. An open question is under what conditions does biosurveillance tends to detect an outbreak more quickly than medical personnel?

The methodology used to answer this question is discrete event simulation of an anthrax outbreak using the Java programming language. In order to design the simulation in this thesis, a review of Professor Fricker's and Buckeridge's simulations was conducted. The Fricker simulation is too simplistic in its design while the Buckeridge simulation is too detailed. Therefore, the design of the simulation in this thesis seeks to be more realistic than Fricker, but also more generalizable than Buckeridge. The goal is to explore the performance of the EARS' C1 statistical detection algorithm versus medical personnel with the following questions in mind:

(1) Can the C1 statistical algorithm used in the Center for Disease Control and Prevention's Early Aberration Reporting System (EARS) be useful/effective for early event detection (EED) in comparison to medical personnel? If so, under what conditions?

(2) What factors most affect the performance of such an algorithm, in the sense that it results in either C1 algorithm or medical personnel performing significantly better than the other?

To address these questions, two response variables were modeled and analyzed: the probability the C1 algorithm signals first and the number of days it takes for the C1

algorithm to signal. The evaluation was conducted for two scenarios: one for an initial exposed population of 1,000 people and one for 10,000 exposed people. In the worst case scenarios, the probability the algorithm signals first is 13.04% for an exposed population of 1,000 people and it is 0.03% for an exposed population of 10,000 people. In the nominal case scenarios, the probability the algorithm signals first is 31.5% for an exposed population of 1,000 people and it is 0.3% for an exposed population of 10,000 people. In the worst case scenarios, the longest time it takes for the algorithm to signal is 6.63 days for an exposed population of 1,000 people and 4.14 days for an exposed population of 10,000 people. In the nominal case scenarios, the time it takes for the algorithm to signal is 3.3 days for an exposed population of 1,000 people and 0.38 days for an exposed population of 10,000 people.

The parameters with the largest effect on the probability the algorithm signals first are: the probability an individual is infected with Anthrax, the probability a non-infected individual goes to the hospital for non-anthrax related flu, and the daily increase in the probability an infected person will be correctly diagnosed. An increase in the threshold and the transitional probabilities of people getting infected, going to the hospital for non-anthrax related flu and correct diagnosis by doctor all decrease the probability the algorithm signals first, and thus increase the probability the doctor signals first. This finding is consistent with Professor Fricker's simulation results in the sense that as the probability of correct diagnosis by doctor increases, the probability the statistical algorithm detects the outbreak decreases.

The parameters with the largest effect on the number of days to algorithm signal are: the probability an individual is infected, the probability a non-infected individual goes to the hospital for non-anthrax related flu, and the daily increase in the probability an infected person goes to the hospital. An increase in the transitional probabilities of people getting infected, going to the hospital for non-anthrax related flu and an infected person goes to the hospital result in an increase in the time it takes for the algorithm to signal.

This research shows that biosurveillance statistical algorithms, such as the EARS C1, are useful in EED of a bioterror attack. Although the probability the algorithm

signals first may seem low, note that whether the algorithm signaled first was quite situation dependent. And even in the worst case scenario for 1,000 exposed people, the algorithm signaled first more than one time in ten. Thus, at the very least biosurveillance is an effective back-up to clinicians. On the other hand, there were scenarios in which the statistical algorithm almost always signaled first. Follow on research that can build upon this thesis are: evaluating different population sizes, investigating the effects of a wider range for the simulation parameters, comparing the performance among other statistical algorithms, and exploring the parameters that have a significant effect on the number of days to the doctor signaling.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First and foremost, I would like to thank Professor Ron Fricker for affording me the opportunity to explore the field of biosurveillance. Under your tutelage, I have been able to apply my newly mastered Operation Research skill set to address the bioterrorism threat. Your tireless dedication and utmost professionalism inspire me to achieve my educational goal and for that I am grateful.

I would also like to thank my grandfather Captain Tan Nguyen VNN, a prisoner of war for 16 years after the Viet Nam conflict. Your perseverance motivates me to be the best that I can be in my naval career thus far. Lastly, I would like to thank my mother Ms. KieuLan Nguyen for the unconditional support and belief in me.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Bioterrorism is not a new threat, but the potential for disastrous outcomes is greater than it has ever been. The U.S. government recognizes the threat and, via Homeland Security Presidential Directive 21 (HSPD-21), has directed "further improvement in the preparedness of our public health and medical systems to address current and future biological warfare threats and to respond with greater speed and flexibility to multiple or repetitive attacks" (HSPD-21, 2007). In order to confront this threat, biosurveillance systems are utilized to provide early warning of health threats, early detection of health events, and situational awareness of disease activity. To date, little is known about the performance of such biosurveillance systems in comparison to medical personnel. An open question is under what conditions does biosurveillance tend to detect an outbreak more quickly than medical personnel?

This thesis addresses this question via a discrete event simulation of an anthrax-based bioterrorism attack. The goal is to use an idealized model of health-seeking behaviors and medical outcomes of an affected population to assess the relative performance of biosurveillance versus medical personnel in detecting the attack.

A. BACKGROUND

1. Biosurveillance

HSPD-21 defines biosurveillance as "the process of active data-gathering with appropriate analysis and interpretation of biosphere data that might relate to disease activity and threats to human or animal health whether infectious, toxic, metabolic, or otherwise, and regardless of intentional or natural origin" (HSPD-21, 2007). There are three types of biosurveillance: human (epidemiologic) surveillance, animal (zoonotic) surveillance, and agricultural surveillance. Syndromic surveillance is a specific type of epidemiological surveillance that has been defined as "the ongoing, systematic collection, analysis, interpretation, and application of real-time (or near-real-time) indicators of diseases and outbreaks that allow for their detection before public health authorities would otherwise note them." (Sosin, 2003)

Syndromic surveillance differs from the traditional epidemiologic surveillance in a number of ways: it uses health-related data, such as counts of individuals coming into medical facilities, over-the-counter medication sales, and aggregate laboratory test results. The data are prediagnostic or prior to case confirmation. Syndromic surveillance is not supposed to provide a definitive determination that an outbreak is occurring but only to signal that an outbreak maybe occurring (Fricker & Rolka, 2006).

2. Biosurveillance Systems

While there are different types of biosurveillance systems currently in operation, they all share a common goal of improving the chances of detecting an outbreak early. All of them have four main functions: data collection, data management, analysis, and reporting. Three large-scale systems currently in use are BioSense, ESSENCE, and EARS.

BioSense. Launched in 2003 as a result of the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, its purpose is to establish an integrated national public health surveillance system for early detection and rapid assessment of potential bioterrorism-related illness. Developed and operated by the Centers for Disease Control and Prevention (CDC), in 2010 the CDC started redesigning the BioSense program based on input and guidance from local, state, and federal partners. The goal of the redesign effort is to be able to provide nationwide and regional situational awareness for all-hazard health-related threats (beyond bioterrorism) and to support national, state, and local responses to those threats (CDC, 2010a).

ESSENCE. An acronym for Electronic Surveillance System for the Early Notification of Community-based Epidemics, ESSENCE was developed starting in 1999 and is operated by the Department of Defense. It monitors infectious disease outbreaks at more than 300 military treatment facilities worldwide on a daily basis using data from patient visits to the facilities and pharmacy data (Fricker, 2010).

EARS. An acronym for Early Aberration Reporting System, EARS was developed by the CDC. It was pioneered as a method for monitoring bioterrorism during large-scale events where there is little or no "baseline" data. Following the terrorist

attacks of September 11, 2001, various city, county, and state public health officials in the United States and abroad have adopted EARS for routine health surveillance using syndromic and other data from emergency departments, reportable conditions, 911 calls, physician office data, school and business absenteeism, and over-the-counter drug sales (CDC, 2010b).

All of the systems rely on statistical algorithms to trigger an outbreak signal, so that public health official can take appropriate actions. However, little is known about how such a system is likely to perform, particularly in comparison to medical personnel. Furthermore, there are many statistical issues that remain to be resolved. One of the issues is: When do statistical methods add value to the existing medical infrastructure and under what conditions?

As shown in Figure 1, Fricker and Rolka (2006) suggest that if the outbreak is sufficiently large, geographically concentrated, and/or easy to diagnose, then a doctor is likely to be equally fast or faster at detecting an outbreak than a statistical algorithm. In contrast, if the outbreak is very small and/or diffuse, then a statistical algorithm operated in isolation is unlikely to detect the outbreak. In the case of a moderately sized outbreak that is easy to diagnose, a doctor's diagnosis will be faster than a statistical algorithm. The result of these restrictions is that statistical methods are likely to add value only when an outbreak is large and/or concentrated enough to statistically detect, but not so large that the outbreak is obvious, combined with the situation where identification of the type of outbreak is sufficiently hard to diagnose, making the doctor likely to miss it for some time (Fricker & Rolka, 2006). Therefore, biosurveillance can potentially serve as primary detection tool for a rare and hard to diagnose disease or agent and a supplementary tool to medical personnel for a moderately sized outbreak that is moderately hard to diagnose.

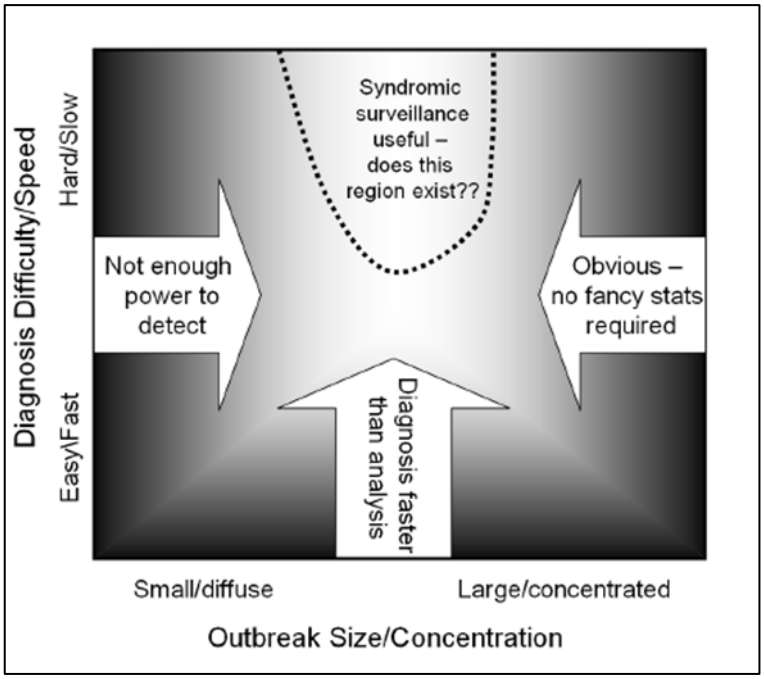

Figure 1. When is syndromic surveillance useful for outbreak detection? From Fricker and Rolka (2006)

3. Anthrax Overview

Anthrax, *Bacillus Anthracis*, has been used as a biological weapon dating back to World War I as a means to cause economic havoc through the loss of livestock. (Grey & Spaeth, 2006). During World War II, the Japanese government formed the research unit 731 at Pingfen to conduct research on anthrax weaponization using prisoners of war as test subjects. It is believed that Japan employed anthrax in its campaign against Manchuria, releasing spores into the atmosphere over the area (Zubay, 2005).

In response to these threats, Britain and United States launched biological weapons initiatives to conduct extensive research on anthrax. In 1942, Britain performed extensive testing at Gruinard Island, off the coast of Scotland by detonating bombs hung on scaffolding structures and examining the extent of contamination of the surrounding area. In 1943, the United States established a pilot plant at Camp Detrick to produce biological weapons and manufactured 5,000 bombs filled with anthrax spores (Christopher et al., 1997).

More recently in 1990, United Nations (UN) inspectors confirmed that Iraq had 100 R400 bombs filled with botulinum toxin, 50 with anthrax, and 16 with aflatoxin. In all, they produced 8500 L of anthrax, 6500 L of which was weaponized into rockets and bombs (Zilinskas, 1997). From 1990 to 1993, the Aum Shinrikyo cult released aerosolized anthrax and botulinum toxin on several occasions at the Diet (the legislature), the Imperial Palace, the U.S. Naval base at Yokosuka, and other places throughout Tokyo (Atlas, 2002). The most recent use of anthrax as a biological weapon occurred in the United States in 2001, when unknown individual or group sent mails containing refined anthrax spores in the form of a highly concentrated dry powder to a variety of media institutions and governmental offices. Of the 22 confirmed cases of anthrax, 11 were due to inhalational and five resulted in casualties. The investigation revealed that the Ames strain of *Bacillus Anthracis* was used in the attack, and this strain was not developed on foreign soil, but rather by scientists associated with the U. S. Army Medical Research Institute of Infectious Diseases (Zubay, 2005).

Following the attacks in 2001, an attempt was made to statistically analyze data regarding symptoms in patients with inhalational anthrax and symptoms from influenza and ambulatory community-acquired pneumonia. The goal was to develop a method to distinguish anthrax from influenza and pneumonia in the early stage of disease progression. Hupert et al. (2003) compared 28 cases of inhalational anthrax, both modern and past occurrences, with more than 2700 cases of influenza and 149 cases of ambulatory community-acquired pneumonia. The study revealed that abnormal lung examination, dyspnea, and nausea or vomiting are statistically greater indicators for anthrax, while sore throat and rhinorrhea[1] are statistically greater indicators for influenza. Cough, chest pain, abnormal temperature, and headache did not demonstrate a statistical difference between anthrax and influenza.

Anthrax is a disease associated mostly with herbivores and has three forms: cutaneous, gastrointestinal, and inhalational. Cutaneous anthrax results from direct contact with infected livestock or livestock products. Mortality for untreated cutaneous

[1] Persistent watery mucus discharge from the nose, commonly referred to as runny nose.

anthrax is about 20%. A pruritic red papular lesion[2] is formed within one week of exposure to the spore. Once the lesion enlarges and ruptures, it forms an ulcer covered by black eschar[3], which then dries up and falls off within two weeks (Grey & Spaeth, 2006). Patients with cutaneous anthrax usually experience headaches and occasional fevers up to 102^{o} F. Unlike the cutaneous form, gastrointestinal (GI) anthrax occurs from the deposition of vegetative bacilli from uncooked meat in the upper or lower portion of the GI track rather than from spore germination. Oral or esophageal ulcers are developed at the initial site of bacterial deposition. Patients usually experience nausea, vomiting, malaise initially and then bloody diarrhea, acute abdominal pain. The actual case numbers for GI anthrax are extremely low, therefore no mortality statistic is available (Zubay, 2005).

In Zubay (2005), inhalational anthrax is described as the most lethal form of the disease, which has a mortality rate of 80%. It is contracted when spores are inhaled and deposited in the alveolar[4]. The spores germinate into active bacilli in the mediastinal lymph nodes[5]. Human to human transmission of the disease is extremely rare, and would occur only through direct transfer of fluids containing the bacteria from one individual to another. The symptoms of inhalational anthrax can be broken down into two stages. In the first stage, which normally last a few days, there are no clinically significant signs. Patients often exhibit only symptoms similar to those of flu and cold, making early diagnosis extremely difficult unless there is prior knowledge of an anthrax outbreak. The second stage develops rapidly with onset of acute dyspnea[6] and subsequent cyanosis[7]. The second stage normally lasts less than 24 hours and leads to death.

Anthrax is considered one of the most dangerous and most likely agents that would be used in a bioterrorist attack due to hardiness of the spores, potency, and

[2] A small, solid, circumscribed elevation characterized by an intense itching sensation.

[3] A piece of dead tissue that is cast off from the surface of the skin.

[4] The tiny air sacs of the lungs.

[5] Region behind the sternum and between the two pleural sacs containing the lungs.

[6] Shortness of breath, a subjective difficulty or distress in breathing.

[7] Bluish discoloration, especially of the skin and mucous membranes, caused by decreases in oxygenated hemoglobin.

availability. The spore is extremely resistant to environmental stresses such as heat, cold, many chemical disinfectants, long dry spells, and low levels of ultraviolet light. It will grow rapidly in a nutrient-rich environment and when the nutrients are exhausted, rather than dying, the bacteria will form dormant spores, which is a method of preserving the Deoxyribose Nucleic Acid (DNA) until conditions return to an optimal state for bacterial growth. The hardiness of the spores requires extensive sterilization efforts and the aerosolized form has no odor, essentially colorless, and virtually undetectable. The first sign that an attack has occurred will probably be the first diagnosis of a patient in a hospital. Besides the hardiness of the spore form, anthrax is extremely potent and deadly bacteria with mortality rates as high as 80% (Zubay, 2005).

In 1993, the U.S. Congressional Office of Technology examined a hypothetical bioterrorist attack utilizing aerosolized spores of *Bacillus Anthracis*. The study concluded an estimated 130,000 to 3 million casualties would result in the event of an aerosolized release of 100 kg of anthrax spores upwind of Washington, DC (Office of Technology Assessment, 1993). Anthrax is readily available throughout the world, will grow relatively easily on most laboratory media, and can also be aerosolized for mass destruction. While anthrax possesses characteristics of an ideal biological weapon, it is more manageable from a biodefense perspective because it is not known to spread from person to person unless there is a direct transmission of bodily fluids, and there is very little risk from secondary aerosolization (Zubay, 2005).

B. LITERATURE REVIEW

In order to develop an idealized discrete event simulation of an anthrax outbreak that is more realistic than Fricker, but also more generalizable than Buckeridge, a literature review of these two simulations is described in the following sections.

1. A Simple Simulation

In his short course, titled “Methodological Issues in Biosurveillance”, at the Twelfth Biennial CDC Symposium on Statistical Methods, Professor Fricker presented the results of a very simple bioterrorism attack simulation study. As illustrated in Figure 2, in Professor Fricker's simulation, on average, 100 people per day (with a standard

deviation of 20 people) go to the hospital with flu-like symptoms. A bioterror attack results in *X* number of people exposed to a bio-agent also going to the hospital with flu-like symptoms, thereby increasing the total number of people at the hospital with flu-like symptoms. A CUSUM (cumulative sum) statistical algorithm monitors the average number of people going to the hospital with flu-like symptoms with a false signal rate fixed at once per 30 days. The CUSUM algorithm will signal an outbreak if there is a statistically unusual increase. Working concurrently with the CUSUM algorithm is a doctor who sees each patient and makes a diagnosis based on his or her expertise. For those exposed to bio-agent, there is some probability *p* that the doctor will correctly diagnose the patient as not having the flu but rather as having been exposed to the bio-agent. The research question for this simple simulation is, what is the probability the clinician diagnoses a case of the bio-agent before the CUSUM algorithm signals? (Fricker, 2009, and Fricker, 2011)

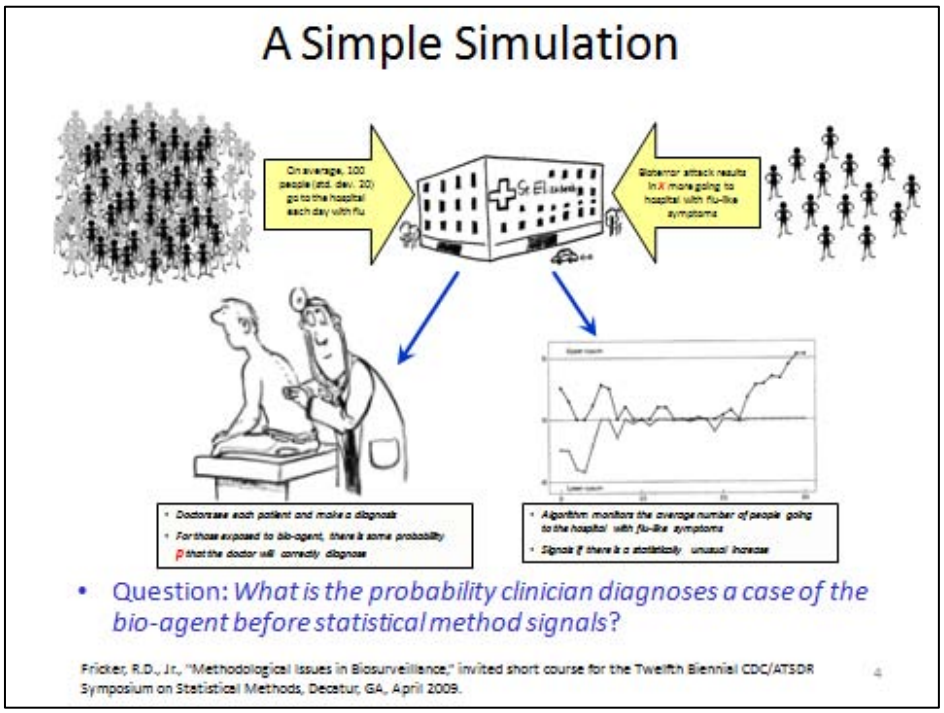

Figure 2. A simple simulation. From Fricker (2009)

The simulation results can be summarized as the higher the probability of correct diagnosis by doctor (*p*), the higher the probability the clinician will detect an outbreak

before the CUSUM signals. As shown in Figure 3, if p is 0.01 and X (the number exposed to the bio-agent) is between 8 and 50 per day, then there is a 50% chance the clinician will detect first. If p is increased to 0.025 and same value for X, then there is a 75% chance the clinician will detect first. If p is 0.05 and X is between 10 and 50 per day, then there is a 90–95% chance the clinician will detect first.

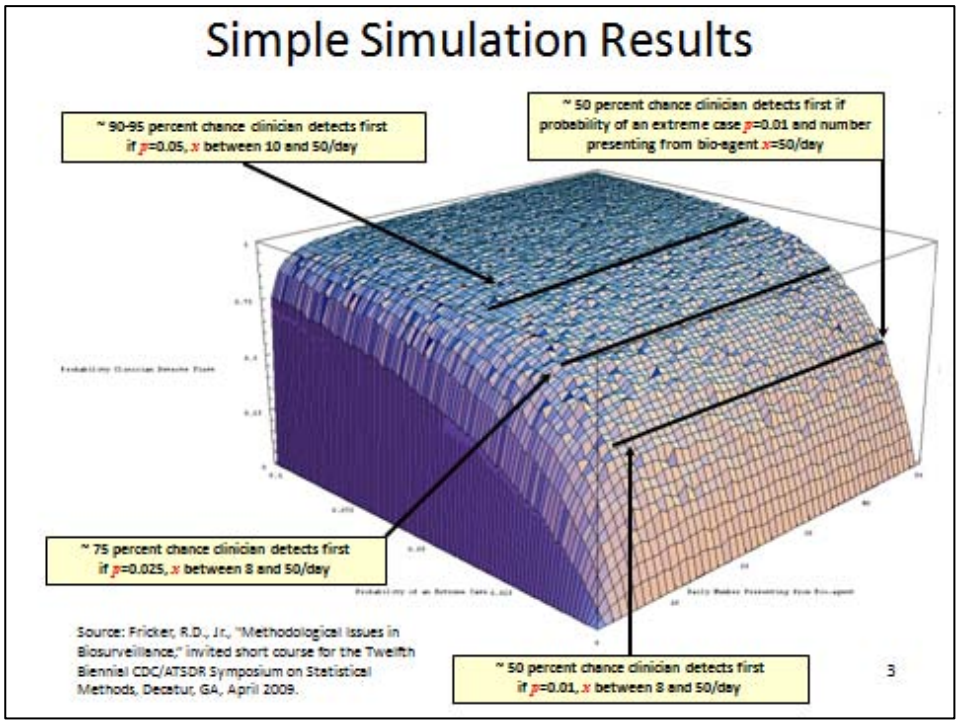

Figure 3. Simple simulation results A. From Fricker (2009)

Consistent with Fricker and Rolka (2006), and as shown in Figure 4, Professor Fricker's simulation results suggest there is a role for statistical algorithms in biosurveillance when the pathogen is hard to diagnose and /or when small numbers of bio-agent are present at the hospital. While this simulation is simplistic with only two parameters p and X, it motivates a more detailed simulation that expands the model, which is the main portion of this thesis.

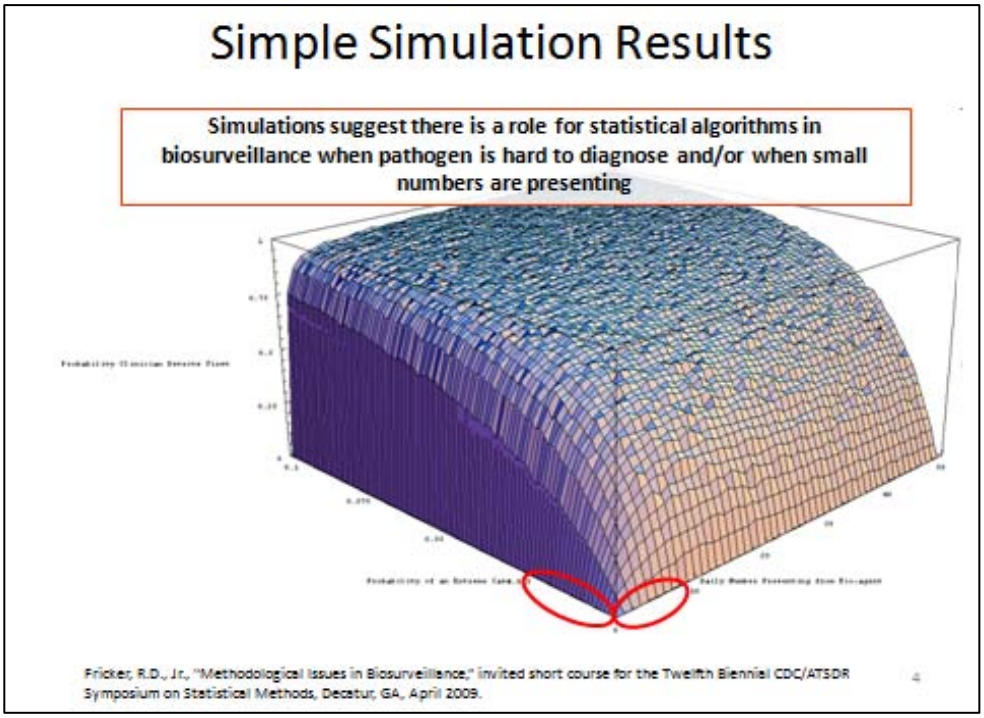

Figure 4. Simple simulation results B. From Fricker (2009)

2. Evaluating Detection of an Inhalational Anthrax Outbreak

In his paper titled "Evaluating Detection of an Inhalational Anthrax Outbreak," Professor Buckeridge conducted a simulation study to compare clinical case finding with syndromic surveillance for detection of an outbreak of inhalational anthrax (the deadliest type with mortality rate of 80%). His aim was to develop a model for simulating the usage of healthcare services after a large-scale exposure to aerosol anthrax spores and then to use this model to estimate the detection benefit of syndromic surveillance when compared with the clinical case finding.

The simulation design consists of four parts: dispersion of released anthrax spores, infection of exposed persons, progression of disease in infected persons, and symptomatic persons' use of the health care system. The dispersion model simulates the number of anthrax spores a person would inhale at locations throughout the region after release of aerosolized spores using the Hazard Prediction and Assessment Capability (HPAC) software developed by the Defense Threat Reduction Agency (DTRA). The infection of exposed person model simulates the number of persons infected using a

semi-Markov process to simulate the progression through three discrete states of disease. Each infected person begins in the incubation[8] state and then progresses through the prodromal[9] state and the fulminant [10] state. The time in each state is sampled from a log normal distribution. The usage of health care system model uses a semi-Markov process to simulate the probability and timing of a symptomatic person seeking care and submission of blood for culture. For patients who are in the prodromal or fulminant state, the probability of seeking care increases linearly over the duration of the state. For patients whose blood samples are cultured, the testing process transitions through two states: growth and isolation. The time spent in these two states is modeled using an exponential distribution.

Three anthrax release scenarios were explored: 1kg, 0.1kg, and 0.01 kg. For each scenario, 1000 simulations were conducted. The evaluation metrics of outbreak detection through syndromic surveillance consists of sensitivity, specificity, and timeliness at a range of decision thresholds. Sensitivity is the probability of correctly detecting an attack, specificity is the probability of not signaling when there is no attack, and timeliness is a measure of the duration between the release of anthrax spores and the first report of an outbreak. The results of the simulation suggest that syndromic surveillance could detect an inhalational anthrax outbreak before clinical case finding. With a simulated 1kg of anthrax spores release, the proportion of outbreaks detected first by syndromic surveillance was 0.59 at a specificity of 0.9 and 0.28 at a specificity of 0.995. When syndromic surveillance was highly sensitive to detect a substantial proportion of outbreaks before clinical case finding, it generated frequent false alarms. The syndromic surveillance system's ability to detect was influenced by both specificity and release size, with specificity being the predominant factor. There was a tradeoff between sensitivity and specificity of syndromic surveillance. In order to reduce the false alarm rate,

[8] The time from the moment of exposure to an infectious agent until signs and symptoms of the disease appear.

[9] Early symptom or set of symptoms that might indicate the start of a disease before specific symptoms occur.

[10] Sudden and severe to the point of lethality.

specificity must be high. However, as specificity is increased, sensitivity is decreased, and the proportion of outbreaks that was detected first by syndromic surveillance decreased more significantly (Buckeridge, 2006).

C. SCOPE OF THESIS

Fricker's simulation is too simplistic in its design while Buckeridge's simulation is too detailed in its design. The Fricker simulation only has two parameters: X (number exposed to the bio-agent) and p (probability the doctor diagnoses correctly). Additionally, the probability of correct diagnosis by the doctor remains the same as time progresses. In contrast, the Buckeridge simulation is too detailed with many parameters in both the dispersion model and the health care usage model. For each parameter value, there are three sets of value intervals due to three anthrax release scenarios of 1kg, 0.1kg, and 0.01 kg, and they are drawn from various probability distributions such as the log-normal, Bernoulli and exponential. If a simulation is too simple or too detailed, then it is difficult to gain some insights into what are the main factors that affect whether an algorithm or clinician is likely to signal an outbreak first. Therefore, the scope of this thesis is to develop an idealized discrete event simulation of an anthrax outbreak that is more realistic than Fricker, but also more generalizable than Buckeridge. In order to explore the performance of the statistical detection algorithm versus medical personnel, this thesis will endeavor to answer these questions:

(1) Can the statistical algorithm be useful/effective for early event detection (EED) in comparison to medical personnel? If so, under what conditions?

(2) What factors most affect the performance of such an algorithm, in the sense that it results in either the algorithm or medical personnel performing significantly better than the other?

II. SIMULATION MODEL

A. DISCRETE EVENT SIMULATION

Discrete event simulation (DES) is a powerful computing technique for understanding the behavior of a system. The operation of such a system is represented as a chronological sequence of events. Each event occurs at a discrete point in time and marks a change of state in the system. The elements of a DES are states, events, and scheduling relationships between events. A state variable in a DES model has a possibility of changing value at least once during any given simulation run. In contrast, a parameter variable does not change during a simulation run. Events are the building blocks in a DES model. Events are responsible for changing a few state variables (possibly none) or many state variables. Once the state transition is done in an event, it will schedule every possible future event. This is the scheduling relationship between events.

The method of time advance in a DES model is called "next event." Simulation time moves in typically unequal increments, jumping from the scheduled time of one event to another. Figure 5 shows that at the start of a simulation, the initial event is scheduled, which is responsible for initializing all state variables as well as scheduling any initial real events of the model. If there are pending events, then simulation time is advanced to the earliest scheduled event, the previous event is removed from the event list, all state transitions associated with the event are executed and the scheduling of any events as specified by the model are performed (Buss, 2010).

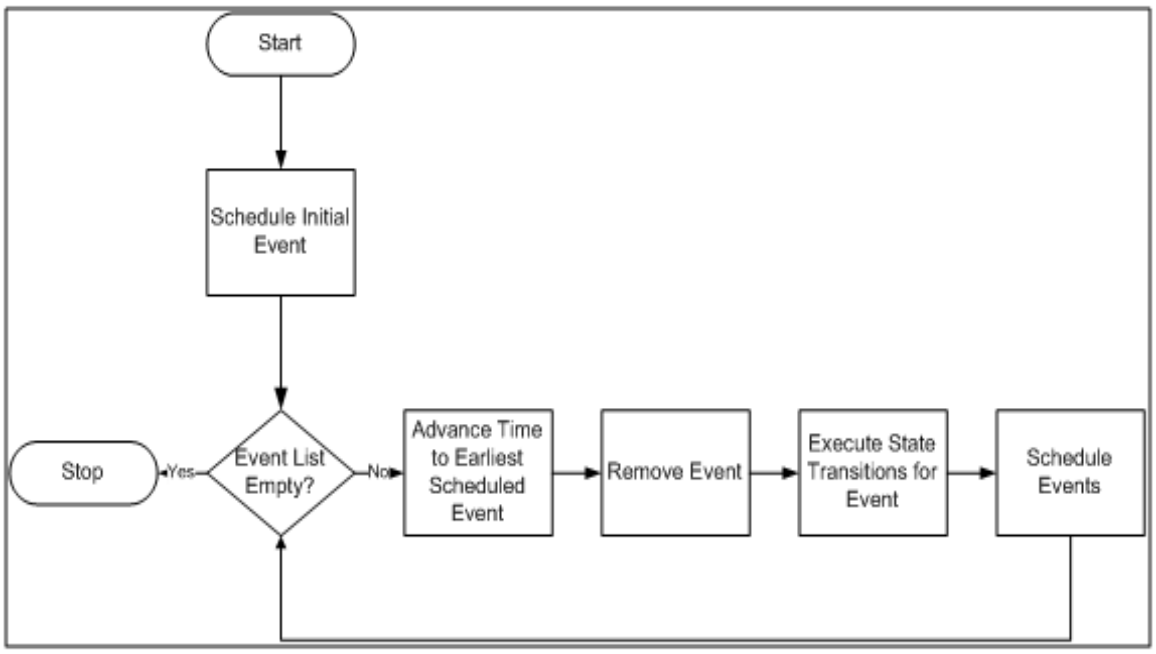

Figure 5. Next event flow chart. From Buss (2010)

An event graph is used to depict the scheduling relationship between events. Each graph consists of nodes and directed edges. Each node corresponds to an event, or state transition, and each edge corresponds to the scheduling of other events. Each edge can optionally have an associated Boolean condition and/or a time delay. Figure 6 shows that the occurrence of Event A causes Event B to be scheduled after a time delay of *t*, providing condition (i) is true (Buss, 2010).

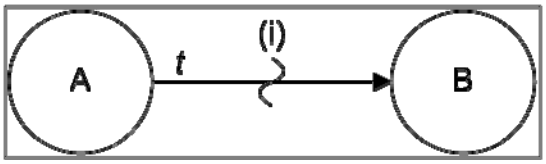

Figure 6. Fundamental event graph construct. From Buss (2010).

B. EARS' C1 ALGORITHM

As described in Fricker et al. (2008), EARS' event detection methods are called "C1-MILD", "C2-MEDIUM", and "C3-ULTRA". The C1 method uses the seven days prior to the current observation to calculate the sample average and sample standard

deviation of a syndrome daily count for day *t*. This thesis only applies the C1 method and uses daily number of people going to the hospital and being classified with flu symptom. The C1 method is defined as

$$C_1(t) = \frac{Y(t) - \overline{Y_1}(t)}{S_1(t)} \tag{1}$$

where

- Y(t) is the observed number of people at hospital for day t
- $\overline{Y_1}(t)$ is the sample mean based on the previous 7 days of data, $\overline{Y_1}(t) = \frac{1}{7}\sum_{j=t-1}^{t-7} Y(j), and$
- $S_1(t)$ is the sample standard deviation based on the previous 7 days of data, $S_1(t) = \frac{1}{6}\sum_{j=t-1}^{t-7}[Y(j) - \overline{Y_1}(j)]^2$

As implemented in EARS, the C1 method signals an outbreak at time *t* when the C1 statistic exceeds a fixed threshold of three sample standard deviations from the sample mean.

C. OUTBREAK SIMULATION MODEL

1. Simulation Design

The goal of the simulation design is to gain insights on which outbreak signal (C1 EARS algorithm or the doctor) occurs first as a function of certain parameters. The approach is to come up with a conceptual design pictorially first, then translates the design into a simplified event graph, and finally into a detailed event graph. The Java programming language with the Simkit library is used to write and execute the outbreak simulation code.

Figure 7 illustrates the design of an outbreak simulation model pictorially. At the start of the simulation, the entire population is susceptible to some disease. Given the

susceptible population, a person can remain susceptible, or go to the hospital with flu-like symptom, or become infected. Given a bioterror attack occurs, an infected person (bio-agent) will go to the hospital seeking care. At the hospital, the doctors see each patient and make a diagnosis. If the doctor correctly diagnoses the patient, then he or she will signal an outbreak. If the doctor misdiagnoses the bio-agent, then that person is still infected and returns to the infected pool of individuals. The C1 algorithm monitors the average number of people going to the hospital with flu-like symptoms (which consists of the sum of those going to the hospital with the flu and those with flu-like symptoms resulting from exposure to the bioterrorism agent) and signals an outbreak, if there is a statistically unusual increase, at which point C1 is greater than the specified threshold.

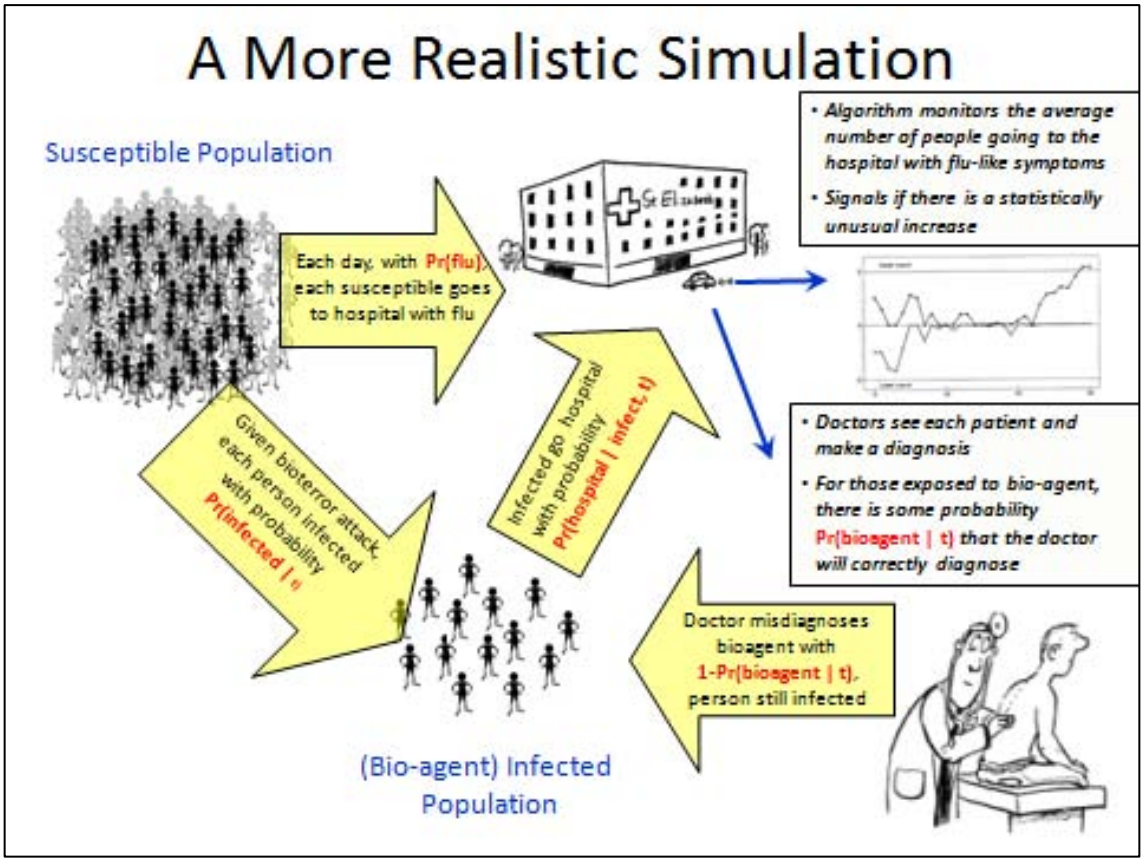

Figure 7. A more realistic simulation

In Figure 8, the conceptual design is translated into a simplified event graph. Each node corresponds to an event such as Susceptible, Stay Susceptible or Go To

Hospital. Each directed edge corresponds to the scheduling of other events. At the beginning of the simulation, the entire population is susceptible to some disease. Given the susceptible population, a person can stay susceptible, go to the hospital with flu-like symptoms, or become infected with the bioterrorism agent. A bioterror attack happens, an infected person may go to the hospital seeking care. Given a person is infected and goes to the hospital, a doctor will perform diagnosis. If the doctor diagnoses the patient correctly, he/she will signal an outbreak. If the doctor misdiagnoses, the patient remains infected and no signal is generated. The C1 algorithm will signal that there is an unusual increase of number of people going to the hospital is the C1 statistic exceeds some prespecified threshold. The number of people going to the hospital used in the C1 statistic calculation represents the people who show up to the hospital from the susceptible population and the infected population.

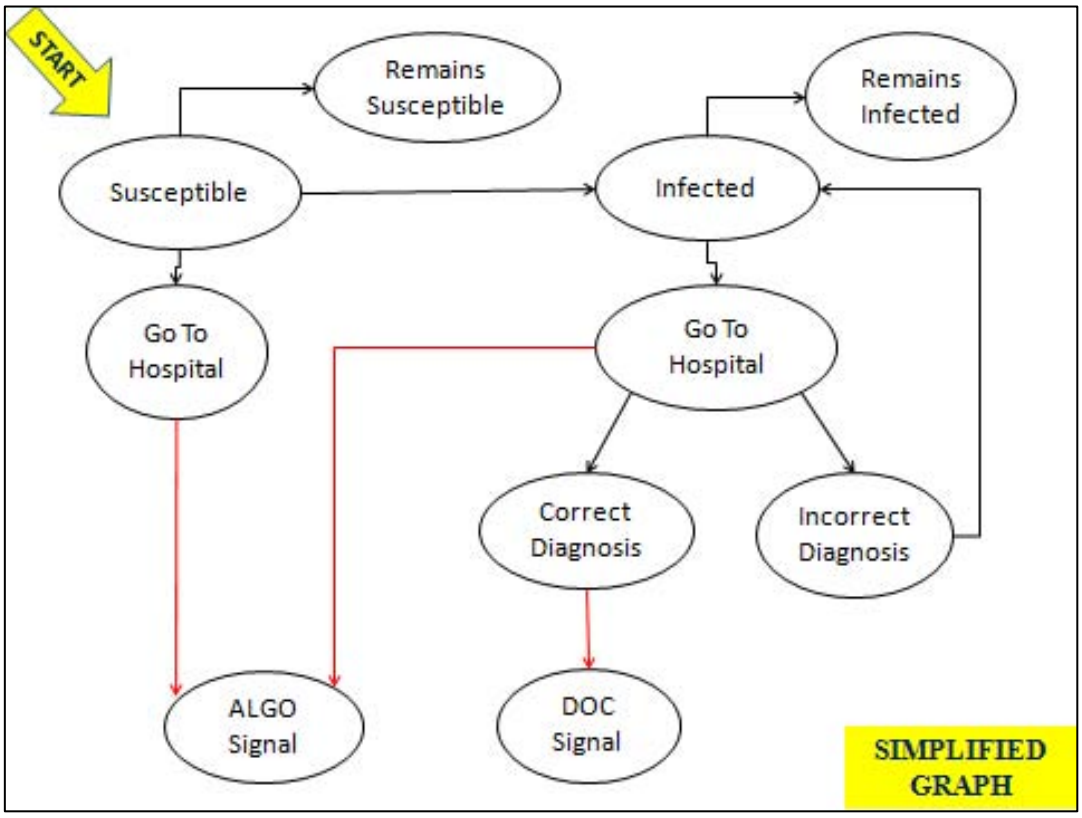

Figure 8. Simplified event graph

The final step before writing the simulation code is drawing a detailed event graph with its corresponding parameters (which will not change during a simulation run), state variables (which will change at least once during a simulation run), state transitions, and the scheduling relationships between events. Four Java classes: *PatientCreator*, *Patient, Outbreak,* and *RunOutbreak* are created to model the bioterrorism attack. Figures 9 and 10 depict the detailed event graph for the *Outbreak* class.

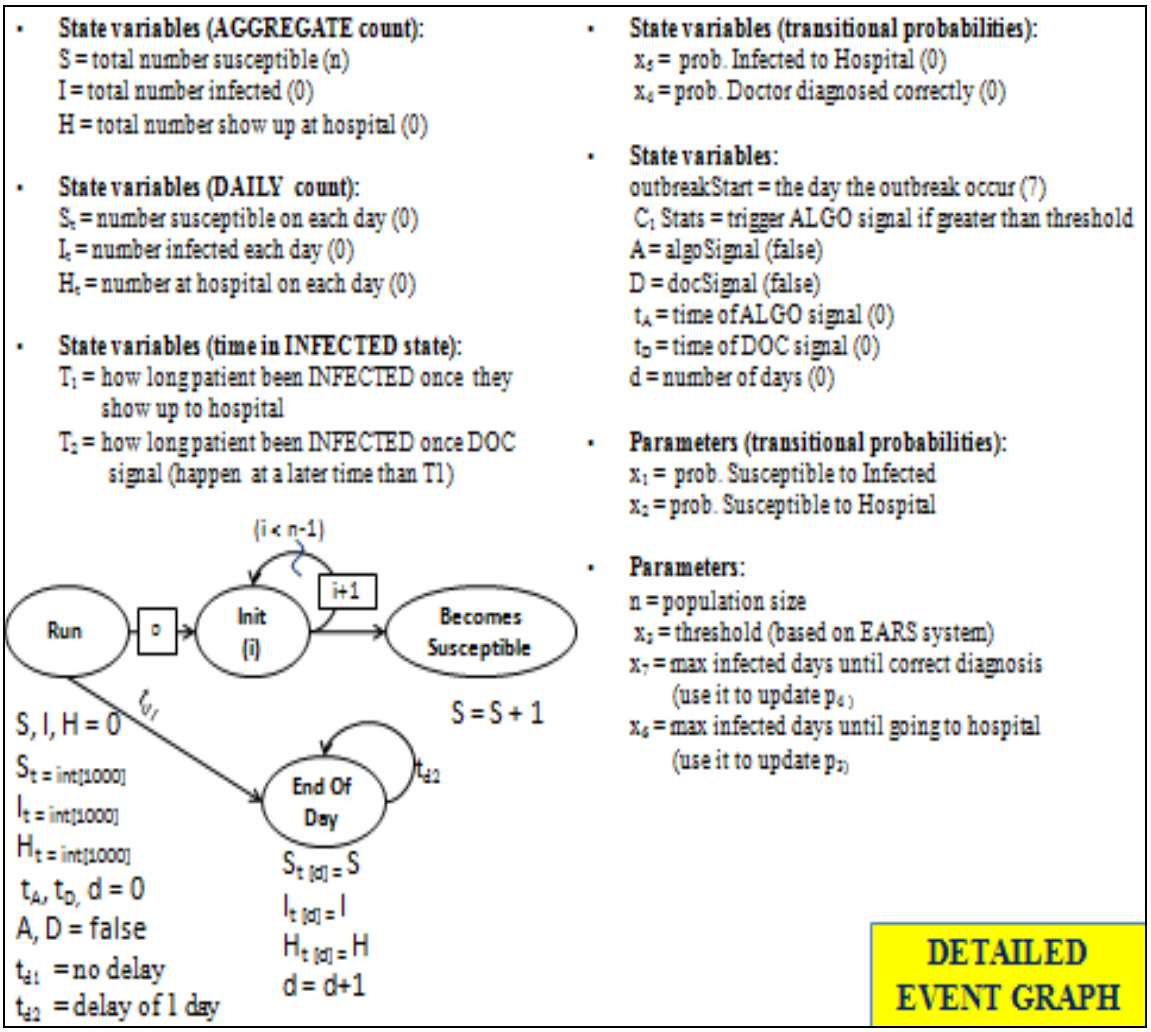

Figure 9. Detailed event graph with parameters and state variables

The *PatientCreator* and *Patient* Java classes are responsible for creating a patient object and keeping track of how long each patient has been infected prior to seeing the doctor at the hospital. How long each patient has been infected will have an impact on

two transitional probabilities: the probability of correct diagnosis by the doctor and the probability of going to the hospital seeking care given a person is infected. This simulation model uses the same approach as in Buckeridge's simulation in the sense that the probability of seeking care increases linearly over the duration of the state. Additionally, the longer a person stays infected, the probability of correct diagnosis by the doctor also increases linearly since the symptoms are becoming more obvious.

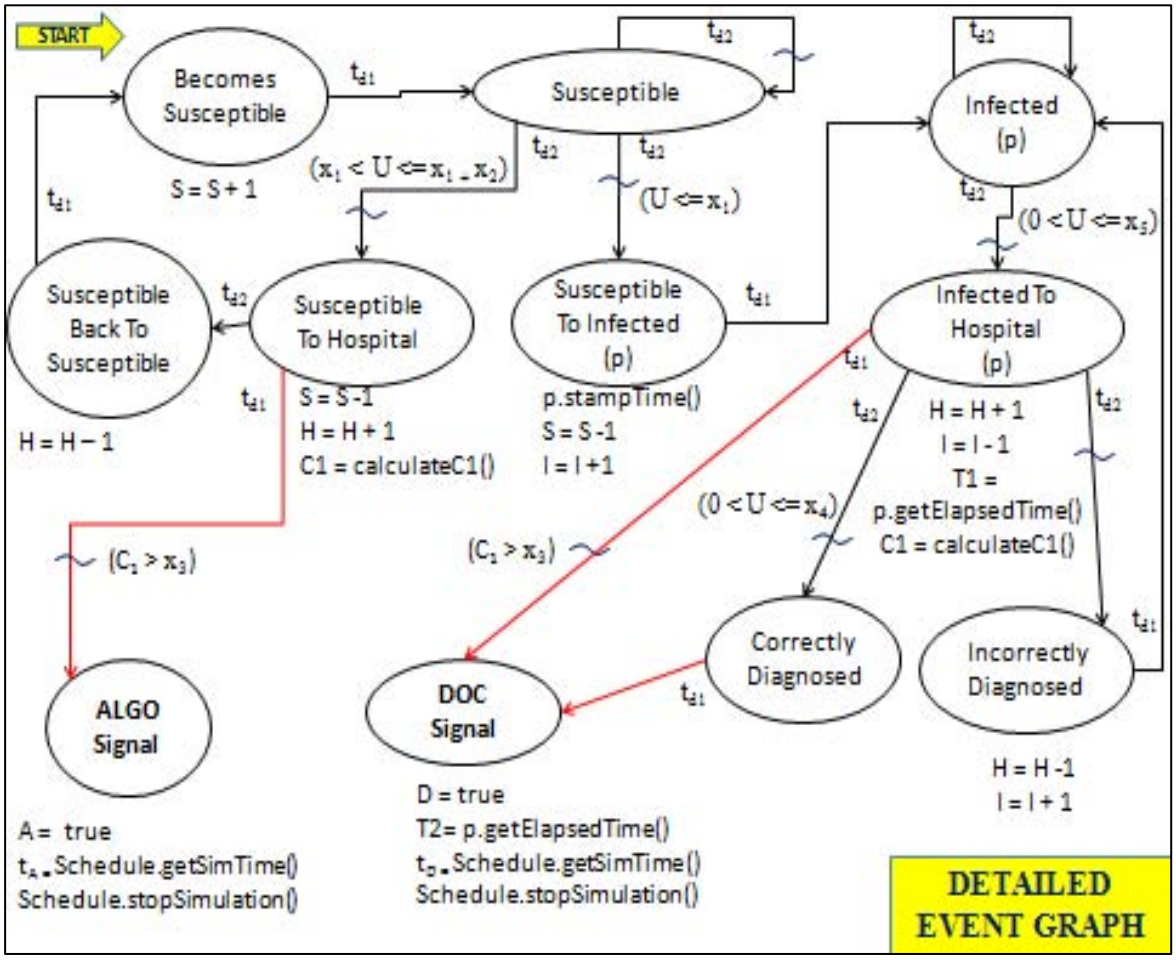

Figure 10. Detailed event graph with state transitions, events and scheduling relationships between events

The *Outbreak* Java class incorporates the detailed event graph from Figures 9 and 10. It contains the simulation's parameters, state variables, state transitions, events and the scheduling relationships between events.

a. Parameters

There are six parameters in the simulation model. Population size (n) is a number representing the population size, which is specified at the beginning of an outbreak simulation run. The population sizes simulated in this thesis are 1,000 or 10,000 people. The transitional probabilities are x_1 and x_2, where x_1 is the probability of transitioning from susceptible to infected and x_2 is the probability a susceptible person goes to the hospital for non-anthrax related flu symptoms. The threshold (x_3) is used as a parameter for an algorithm signal when C1 is greater than the specified threshold. The maximum number of days an infected person is guaranteed to be correctly diagnosed by the doctor is x_7, and the maximum number of days an infected person is guaranteed to go to the hospital seeking care is x_6. Table 1 is the simulation parameters with their name, Java variable type, range and description.

Name	**Java type**	**Range**	**Description**
x_1	double	0.001 to 0.1	probability of transitioning from susceptible to infected
x_2	double	0.001 to 0.1	probability a susceptible person goes to hospital for non-anthrax related flu symptoms
x_3	double	2 to 3	threshold
x_7	double	7 to 21	maximum number of days an infected person is guaranteed to be correctly diagnosed
x_6	double	14 to 28	maximum number of days an infected person is guaranteed to go to hospital seeking care
n	integer	1000 or 10000	population size

Table 1. Simulation parameters

b. State Variables

The state variables can be broken down in five different groups: transitional probabilities, an aggregate count, a daily count, time in the infected state, and other. The probability of an infected person going to the hospital seeking care is x_5 and the probability of correct diagnosis by the doctor is x_4. The initial values for these probabilities start out at 0 and increases linearly over time up to 1. There are three state variables to keep track of an aggregate count: total number susceptible (S) with initial value equal to the population size, total number infected (I) with initial value of 0, and total number show up at hospital (H) with initial value of 0. For the daily count, there are three state variables with initial values of 0: number susceptible each day (S_t), number infected each day (I_t), and number at hospital each day (H_t). The state variable that keeps track of how long each patient is infected before he or she shows up at the hospital is T_1, which has direct impact in updating the probability of correct diagnosis by the doctor (x_4) and the probability of an infected person going to the hospital seeking care (x_5). The last groups of state variables are t_A and t_D, which record the time of the algorithm or doctor signal of an outbreak. The Boolean state variables associated with t_A and t_D are algorithm signal (A) and doctor signal (D), which has an initial value of false. Finally, d represents the current day in the simulation, letting all patient objects know what day it is. The state variable outbreakStart is the day the outbreak occurs with a value of 7. In the simulation, an outbreak does not occur until 7 days has gone by. It is necessary to collect data for 7 days in order to use them in the C1 algorithm. Table 2 is the simulation state variables with their name, Java variable type, and description.

Name	Java type	Description
S	integer	total number susceptible (initial value is population size)
I	integer	total number infected (initial value of 0)
H	integer	total number at hospital (initial value of 0)
S_t	integer array	an array to store number susceptible on each day (size of 1000)
I_t	integer array	an array to store number infected on each day (size of 1000)
H_t	integer array	an array to store number at hospital on each day (size of 1000)
T_1	double	keep track of how long each patient has been infected
x_5	double	probability of transitioning from infected to hospital (initial value of 0), gets updated as the day progresses
x_4	double	probability of correct diagnosis by doctor (initial value of 0), gets updated as the day progresses
t_A	double	record the time of an algorithm signal (initial value of 0)
t_D	double	record the time of a doctor signal (initial value of 0)
A	Boolean	algorithm signal (initial value of false)
D	Boolean	doctor signal (initial value of false)
C1	double	store the value of C1 statistic of the EARS algorithm
d	integer	the current day (initial value of 0)
outbreakStart	integer	the day an outbreak occur (initial value of 7)

Table 2. Simulation state variables

c. Events and State Transitions

Each node in Figure 10 (detailed event graph) represents an event, which corresponds to a public method in the *Outbreak* class. Underneath each event node is the associated state transition or transitions, where certain state variables will be updated during the simulation run. A typical sequence of events can be summarized as: one event occurs (i.e. Susceptible), state transitions are performed for that event, and the next event is scheduled.

(1) The Reset and Run event. The Reset event is responsible for setting the initial values of all state variables at the start of the simulation. The Run event is responsible for scheduling the arrival of each patient into the system. It will stop scheduling the arrival of the patients once it reaches the population size. Additionally, it has an End of Day event, where at the end of each day in the simulation, it is scheduled to record: the number of susceptible (S_t), the number of infected (I_t), and the number of people showing up at the hospital (H_t). Once the daily counts are recorded, End of Day event will increase numDay (d) by 1, which advances the simulation to the next day.

(2) The Becomes Susceptible and Susceptible event. The Becomes Susceptible event is a bookkeeping event, where the occurrence of this event will increment the total number of Susceptible (*S*) by 1. The Susceptible event is responsible for scheduling other events. Given a susceptible person, he or she can either remain susceptible, or go to the hospital, or become infected. The total transitional probabilities for these three events add up to 1. The scheduling of these three events depends on the result of drawing a random uniform variable *U* (0, 1). If *U* is less than or equal to x_1 (the probability of transitioning from susceptible to infected) and *d* (the current day) is greater than or equal to *outbreakStart* (has value of 7), then the person will transition to the Infected event, meaning he or she has gone from being Susceptible to being Infected. The second part of the conditional statement where *d* is greater than or equal to *outbreakStart* ensures that no one can be infected until a bioterror attack happens, which occurs at day 7. If *U* is greater x_1 and U is less than or equal to the sum of x_1 (the probability of transitioning from susceptible to infected) and x_2 (the probability a susceptible person goes to the hospital for non-anthrax related flu symptoms), then the

person will transition to the Susceptible To Hospital event, meaning a susceptible person decides to go to the hospital seeking care. If a susceptible person does not go to the hospital or becomes infected, then he or she remains susceptible to a disease.

(3) The Susceptible To Hospital event. State transitions and the calculation of C1 statistic are performed in this event. If a person shows up to this event, he or she comes from the susceptible population. This event will increment the total number at hospital (*H*) by 1 and decrement the total number susceptible (*S*) by 1. It will call the *calculateC1()* helper method to figure out the value of C1 statistic at that time. If *C1* is greater than x_3 (threshold), it will schedule an ALGO Signal event. This means the algorithm has signaled that there is an outbreak, at which point the simulation will terminate. If there is no outbreak signal from an algorithm, then the person is scheduled to the Susceptible Back To Susceptible event, meaning he or she goes to the hospital and there is nothing wrong with them, therefore they go back to being susceptible.

(4) The Susceptible To Infected event. The Susceptible To Infected event is a bookkeeping event. If a person arrives to this event, that means they were susceptible and then became infected with anthrax due to a bioterror attack. A time is recorded upon an arrival of a person to this event. This is necessary in order to keep track of how long each person has been infected (T_1). After recording the time, the occurrence of this event will decrement the total number of Susceptible (*S*) by 1, and increment the total number of Infected (*I*) by 1. Afterwards, the simulation schedules the person to transition to the Infected event.

(5) The Infected event. Given an infected person, he or she can either remain infected or go to the hospital. The total transitional probabilities for these two events add up to 1. Prior to the scheduling of these two events, the probability of an infected person going to the hospital seeking care (x_5) needs to be updated. This is done due to the fact that the longer a person is infected, the probability of them going to the hospital seeking care increases linearly as the day progresses. Therefore:

$$\text{updated } x_5 = \text{original } x_5 + ((1 - \text{original } x_5) * (T_1 / x_6)) \quad (2)$$

where

- x_5 is the probability of an infected person going to the hospital seeking care

- T_1 is how long each patient has been infected
- x_6 is the maximum number of days an infected person is guaranteed to go to hospital seeking care

Once the update of x_5 is done, the scheduling of other events occurs, which depends on the result of drawing a random uniform variable $U(0,1)$. If U is greater than 0 and less than or equal to the updated x_5, then the person will transition to the Infected To Hospital event, meaning he or she has gone from being Infected to going to the hospital seeking care. If that conditional statement is not true, then a person remains infected.

(6) The Infected To Hospital event. A person who shows up to this event means they were susceptible, became infected with anthrax due to a bioterror attack, and decided to go to the hospital seeking care. The first step is to record how long they have been infected with anthrax prior to showing up to the hospital seeking care (T_1). This is done due to the fact that the longer a person is infected, the probability of correct diagnosis by the doctor (x_4) increases linearly as the day progresses. Therefore:

$$\text{updated } x_4 = \text{original } x_4 + ((1 - \text{original } x_4) * (T_1 / x_7)) \quad (3)$$

where

- x_4 is the probability of correct diagnosis by the doctor
- T_1 is how long each patient has been infected
- x_7 is the maximum number of days an infected person is guaranteed to be correctly diagnosed

Once the update of x_4 is done, the scheduling of other events occurs, which depends on the result of drawing a random uniform variable $U(0,1)$. If U is greater than 0 and less than or equal to the updated x_4, then the person will transition to the Correctly Diagnosed event, meaning an infected person goes to the hospital seeking care and the doctor diagnose them correctly. If that conditional statement is not true, then

the person will transition to the Incorrectly Diagnosed event and ultimately end up at the Infected event, meaning the doctor misdiagnoses the patient and the patient goes back to being infected with the anthrax disease.

This event will also increment the total number at hospital (*H*) by 1 and decrement the total number infected (*I)* by 1. It will call the *calculateC1()* helper method to figure out the value of C1 statistic at that time. If *C1* is greater than x_3 (threshold), it will schedule an ALGO Signal event. This means the algorithm has signaled that there is an outbreak, at which point the simulation will terminate. If there is no outbreak signal from an algorithm, then the person is scheduled to the Correctly Diagnosed or Incorrectly Diagnosed event,

(7) The Incorrectly Diagnosed event. If a doctor misdiagnoses a patient, then he or she will arrive to this event. It will increment the total number of infected (*I*) by 1 and decrement the total number at the hospital (*H*) by 1. After that, a person will transition to the Infected event, meaning an infected person receives an incorrect diagnosis by the doctor will go back to being infected with anthrax.

(8) The ALGO Signal event. The simulation will immediately terminate upon the occurrence of this event. What will trigger the scheduling of this event is when *C1* is greater than the threshold (x_3). There are only two times that C1 is calculated and then compared to the threshold. Once a susceptible person arrives to the hospital seeking care, or an infected person arrives to the hospital seeking care, it will trigger the C1 statistic calculation. In this event, the Boolean state variable *A* is changed from false to true and the time of the algorithm signal (t_A) is recorded. The time of the algorithm signal is recorded to answer the question of how many day(s) does it take for an algorithm to signal an anthrax outbreak. The time it takes for an algorithm to signal will then be compared to the time it takes for a doctor to signal. Prior to ending the simulation, a daily report will be printed out detailing the number of susceptible (S_t), the number of infected (I_t), and the number at the hospital (H_t).

(9) The DOC Signal event. The simulation will immediately terminate upon the occurrence of this event. What will trigger the scheduling of this event is when the doctor correctly diagnoses an infected patient. In this event, the

Boolean state variable *D* is changed from false to true and the time of the doctor signal (t_D) is recorded. The time of the DOC signal is recorded to answer the question of how many day(s) does it take for a doctor to signal an anthrax outbreak. The time it takes for a doctor to signal will then be compared to the time it takes for an algorithm to signal. Prior to ending the simulation, a daily report will be printed out detailing the number of susceptible (S_t), the number of infected (I_t), and the number at the hospital (H_t).

In order to run the simulation, a Java execution class called *RunOutbreak* is required. This is where all the parameters can be changed prior to the start of each simulation run. Various statistical objects are created in order to keep track of the statistics of interest with a 95% confidence interval. The statistics of interest are: average number of algorithm signals, average number of doctor signals, average number of days it takes for an algorithm signal, and average number of days it takes for a doctor signal. Each simulation run consists of 10,000 replications. Figure 11 illustrates a typical output print out as a result of the *RunOutbreak* class.

```
RUN #1:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  874.6673 +/-  0.3439
      Avg no. Infected:  43.8027 +/-  0.3289
      Avg no. At The Hospital:  81.5301 +/-  0.0493

      AVG NO. OF ALGORITHM SIGNALS:  0.1304 +/-  0.0066
      AVG NO. OF DOCTOR SIGNALS:  0.8696 +/-  0.0066

      AVG No. of Days from Susceptible to Algo Signal:  1.5613 +/-  0.0496
      AVG No. of Days from Susceptible to Doc Signal:  4.1611 +/-  0.0077
```

Figure 11. Simulation outputs example

2. Experimental Design

In order to determine the settings of the parameters of the simulation at the start of each run, a *D*-optimal custom designed experiment with five factors resulting in 25 runs is chosen. JMP statistical software is utilized to generate the design matrix using the parameters in Table 1. The *D*-optimal design is presented in Table 3 where:

- x_1 is the probability of transitioning from susceptible to infected
- x_2 is the probability a susceptible person goes to hospital for non-anthrax related flu symptoms
- x_3 is the threshold
- x_7 is the maximum number of days an infected person is guaranteed to be correctly diagnosed
- x_6 is the maximum number of days an infected person is guaranteed to go to hospital seeking care

There is a restriction placed on the values of x_7 in relation to x_6 in the sense that x_6 must be greater than x_7 by 7.

Run #	x1	x2	x3	x7	x6
1	0.1000	0.1000	2.00	7.10	14.00
2	0.0010	0.0010	2.00	21.00	28.00
3	0.1000	0.1000	2.50	7.00	21.84
4	0.0010	0.0505	3.00	7.00	22.01
5	0.1000	0.0505	2.00	7.00	28.00
6	0.0505	0.0010	3.00	21.00	28.00
7	0.0505	0.1000	3.00	7.00	28.00
8	0.1000	0.1000	3.00	14.93	21.83
9	0.1000	0.0010	3.00	7.00	28.00
10	0.0010	0.1000	2.50	14.05	28.00
11	0.0505	0.0505	2.50	14.00	21.00
12	0.0010	0.1000	3.00	7.10	14.00
13	0.0010	0.1000	2.00	13.49	20.39
14	0.1000	0.0505	3.00	14.62	28.00
15	0.1000	0.1000	2.00	21.00	28.00
16	0.0010	0.0010	2.50	7.00	28.00
17	0.0010	0.0010	3.00	13.24	20.14
18	0.0010	0.1000	3.00	21.00	27.90
19	0.1000	0.0010	2.50	21.00	27.90
20	0.1000	0.0010	2.00	13.06	19.96
21	0.0505	0.0010	2.00	11.93	28.00
22	0.0010	0.1000	2.00	7.00	28.00
23	0.1000	0.0010	3.00	7.00	14.00
24	0.0010	0.0010	2.00	7.00	14.00
25	0.0505	0.0505	2.50	11.35	18.25

Table 3. Simulation parameter values generated by JMP D-optimal design matrix

THIS PAGE INTENTIONALLY LEFT BLANK

III. ANALYSIS OF THE SIMULATION RESULTS

There are two response variables of interest in the analysis of the simulation results: the probability of an algorithm signaling first and the number of days it takes for the algorithm to signal. In the probability of algorithm signaling first case, there are two models: one for an initial exposed population of 1,000 people (Model 1) and one for 10,000 exposed people (Model 2). In the number of days it takes for the algorithm to signal case, there are two models: one for an initial exposed population of 1,000 (Model 3) and one for 10,000 exposed people (Model 4). Prior to developing and analyzing the main effects of the four models, the general logistic regression model is explained.

A. LOGISTIC REGRESSION MODEL

Logistic regression models the probability of an event or outcome (p) as

$$\text{logit}(p) \equiv \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \quad . \tag{4}$$

In this model, the log odds of p, often called the logit, is a linear function of the independent variables $x_1,\ldots,x_k$. Note that the odds of p, which is $p/(1-p)$, can range from 0 (when p=0) to infinity (when p=1), while the log odds has domain $(-\infty,+\infty)$. This relationship allows the independent variables to range over the whole real line while p is constrained to the unit interval (as a probability should be constrained).

In Equation 4, we see that for positive coefficients ($\beta_0, \beta_1,\ldots, \beta_k$) increases in the associated independent variable (holding all others constant) results in an increase in the log odds. Similarly, for negative coefficients, decreases in the associated independent variable (holding all others constant) results in a decrease in the log odds. Increasing log odds corresponds to increasing p.

Solving Equation 4 for p and substituting the estimated coefficients (denoted as $\hat{\beta}_0, \hat{\beta}_1,\ldots, \hat{\beta}_k$) resulting from fitting the logistic regression model to data gives in the following equation for estimating p:

$$\hat{p} = \frac{1}{1 + e^{-\left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k\right)}} \qquad (5)$$

Using Equation 5, for a simple logistic regression model with one independent variable, we can plot *x* versus *p* and show that *p* is appropriately constrained to the unit interval. For example, Figure 12 shows the resulting logistic curve for $\hat{\beta}_0 = 1$ and $\hat{\beta}_1 = 2$.

.

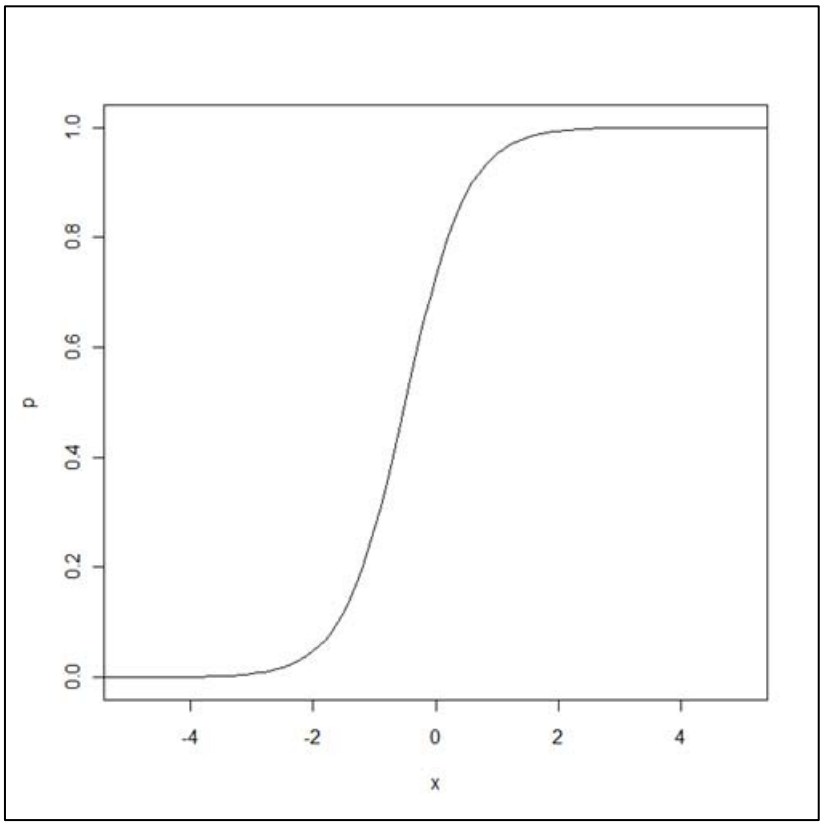

Figure 12. Plot of p = 1/(1+exp(-1-2x))

When estimating the probability *p* per Equation 5, increases in independent variables with positive coefficients correspond to increases in $\hat{p}$; the larger the coefficient (holding all else constant), the more dramatically the probability changes from small (near 0) to large (near 1). Figure 13 illustrates this for four different β_1 values.

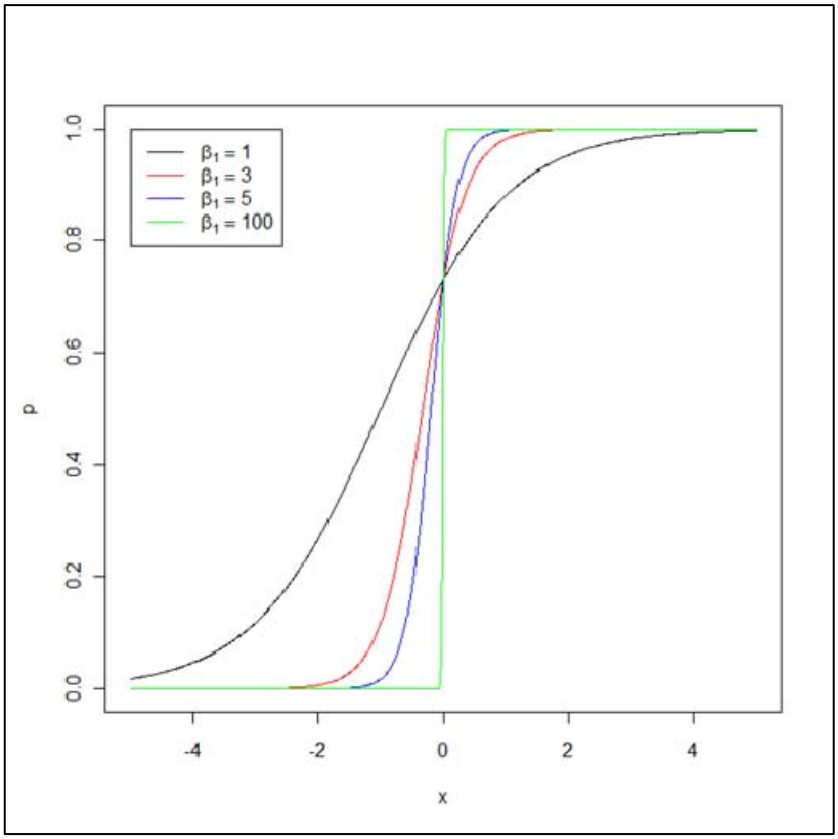

Figure 13. Plot of $p = 1/\left(1+\exp\left(-1-\beta_1 x\right)\right)$ for various values of β_1

The models resulting from the biosurveillance algorithm are not as simple as Equation 4, since they have quadratic and interaction terms in them. Also, the models are not fit in the usual way, where one usually has observed some sort of binary outcome and the logistic regression model is fit as a generalized linear model. Rather, in this case, we have estimated probabilities from the simulation and we fit the estimated log odds as a linear function of the various covariates using ordinary least squares (OLS).

B. PROBABILITY OF ALGORITHM SIGNALING FIRST RESULTS

In analyzing the probability of algorithm signaling first results, there are two versions: one for an initial exposed population of 1,000 people (Model 1) and one for 10,000 exposed people (Model 2). Main effects, interaction, and quadratic terms are included in both models. JMP stepwise function is utilized to determine which terms are significant. After each simulation run, the probability of algorithm signaling first is

estimated in the simulation. Then, it is transformed into the logit in order to fit and analyze the models. Table 4 is a modified version of Table 1, with the variables used in the analysis.

$\hat{p}$	estimated probability (from the simulation) that the algorithm signals first
x_1	probability of transitioning from susceptible to infected, $0.001 \le x_1 \le 0.1$
x_2	probability a susceptible person goes to the hospital for non-anthrax related flu symptoms, $0.001 \le x_2 \le 0.1$
x_3	threshold, $2 \le x_3 \le 3$
x_4	daily increase in the probability an infected person will be correctly diagnosed, beginning at zero on the day of infection and increasing linearly up to a probability of one (when the infected person will have such obvious symptoms he or she is guaranteed to be correctly diagnosed), $1/21 \le x_4 \le 1/7$
x_5	daily increase in the probability that an infected person goes to the hospital, where the probability increases linearly from zero to one (at which time the person is so sick he or she will definitely go to the hospital), $1/28 \le x_5 \le 1/14$

Table 4. Analysis model variables

1. Population Size of 1,000 (Model 1)

Table 5 shows the results of 25 simulation runs, where 10,000 replications are executed within each run. The time it takes to complete the simulation run is 36 hours using a personal computer laptop and a desktop. The probability of algorithm signaling first is estimated via the simulation, translated into the logit, and then entered into JMP (along with the parameters used in the simulation) for model fitting and analysis.

Run #	Probability of algorithm signaling first	Logit of propability of algorithm signaling first
1	0.1304	-1.8974
2	0.9390	2.7339
3	0.0524	-2.8950
4	0.2810	-0.9395
5	0.2490	-1.1040
6	0.9320	2.6178
7	0.0373	-3.2507
8	0.0241	-3.7011
9	0.9178	2.4128
10	0.3903	-0.4461
11	0.2263	-1.2293
12	0.1694	-1.5899
13	0.6307	0.5352
14	0.0739	-2.5283
15	0.9353	2.6711
16	0.7409	1.0507
17	0.6810	0.7584
18	0.2356	-1.1770
19	0.9845	4.1513
20	0.9957	5.4448
21	0.9694	3.4557
22	0.6041	0.4226
23	0.9866	4.2990
24	0.8637	1.8464
25	0.2227	-1.2500

Table 5. Probability of algorithm signaling first results (population size of 1,000)

The probability of algorithm signaling first ranges from 0.1304 (lowest value in run number 1) to 0.9957 (highest value in run number 20). Run numbers 1 and 20 have the same probability of transitioning from susceptible to infected state (x_1 = 0.1) and the same threshold (x_3 = 2). They differ in the probability a susceptible person goes to the hospital for non-anthrax related flu (x_2). In run number 1, the probability is higher at 0.1 while it is at 0.001 for run number 20. The daily increase in the probability an infected person will be correctly diagnosed (x_4) and an infected person goes to the hospital (x_5) in

run number 1 are both lower than in run number 20. In run number 1, the daily increase in the probability an infected person will be correctly diagnosed is 1/7and the daily increase in the probability that an infected person goes to the hospital is 1/14. In run number 20, the daily increase in the probability an infected person will be correctly diagnosed is 1/13 and the daily increase in the probability that an infected person goes to the hospital is 1/20 days.

The model is fit in JMP using stepwise regression, regressing the estimated logit on the various simulation parameters. The results using OLS to fit the logit of the estimated probabilities to the covariates are seen in Equation 6.

Model 1: 1,000 people exposed with quadratic and interaction terms (R^2=0.92)

$$\begin{aligned}\text{logit}(\hat{p}) = {} & 12.0385 + 21.4731x_1 - 48.3301x_2 - 3.6478x_3 \\ & - 78.8509x_4 - 371.0377x_1x_2 - 20.4053x_2x_3 + 26.9634x_3x_4 \\ & + 776.8660(x_2)^2\end{aligned} \qquad (6)$$

In order to graphically depict the effects of the variables with the largest effect (x_2, x_3, and x_4) on the probability of algorithm signaling first, Figure 14 through 16 shows the results for Model 1 where the other variables are set to their nominal values (x_1 = x_2 = 0.05, x_3 = 2.5, x_4 = 1/14, x_5 = 1/21) and then plot the estimated probability of algorithm signaling first as a function of the variables with the largest effect (x_2, x_3, and x_4).

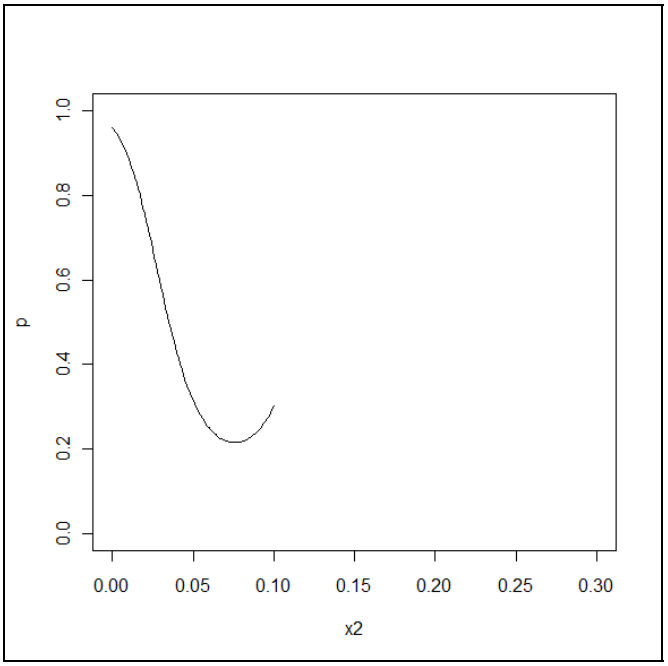

Figure 14. Plot of Model 1 made by varying x_2 over its range while setting all other variables to their nominal values

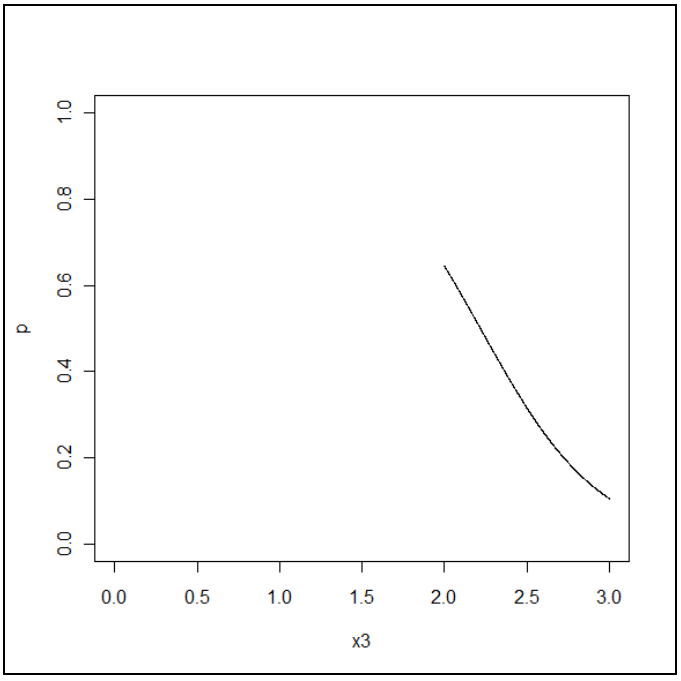

Figure 15. Plot of Model 1 made by varying x_3 over its range while setting all other variables to their nominal values

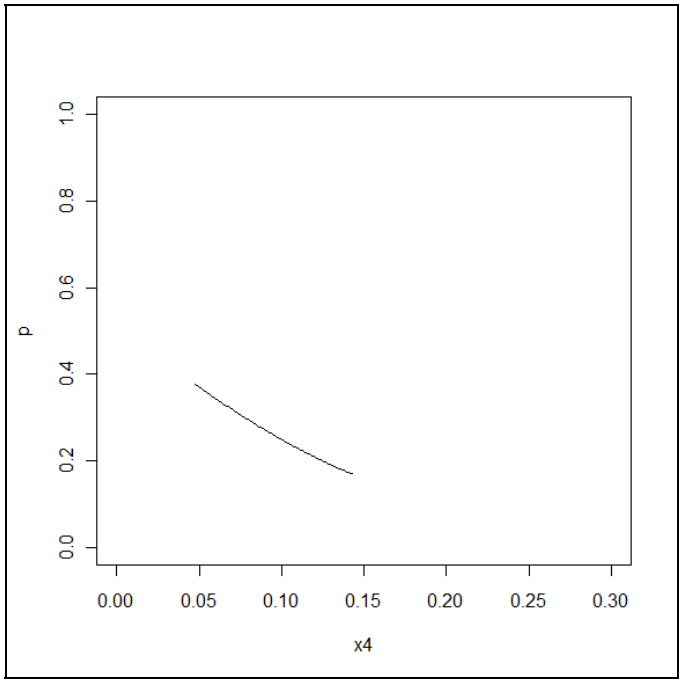

Figure 16. Plot of Model 1 made by varying x_4 over its range while setting all other variables to their nominal values

The variables with the largest effect on the probability the algorithm signals first are x_2 (probability going to the hospital for non-anthrax related flu), x_3 (threshold), and x_4 (daily increase in the probability an infected person will be correctly diagnosed). The results for x_2 (probability going to the hospital for non-anthrax related flu), x_3 (threshold), and x_4 (daily increase in the probability an infected person will be correctly diagnosed) are in the expected direction:

- As the probability of going to the hospital for non-anthrax related flu (x_2) increases, the probability the algorithm signals first decreases,
- As the threshold (x_3) increases, the probability the algorithm signals first decreases, and
- As the daily increase in the probability an infected person will be correctly diagnosed (x_4) increases, the probability the algorithm signals first decreases.

Interestingly, the probability of people getting infected (x_1) only modestly affects the probability the algorithm signals first, at least over the range of that variable. This is a surprising result, as we expected that:

- As the probability of people getting infected (x_1) increases, we then expected that there would be more infected people going to the hospital that could be correctly diagnosed and thus the probability the algorithm signals first decreases.

However, variable x_1 is very modestly associated with a positive increase in the probability the algorithm signals first (though the increase is very small over the range of probabilities considered: $0.001 \le x_1 \le 0.1$). And, since x_5 is not in Model 1, the probability the algorithm signals first is not even associated with the daily increase in the probability infected persons go to the hospital (over the range considered: $1/28 \le x_5 \le 1/14$).

A natural question is which levels of the variables maximize and minimize the probability that the algorithm signals first. The probability the algorithm signals first is maximized ($\hat{p}$ = 0.996) at the boundaries for each of the variables: x_1 = 0.1, x_2 = 0.001, x_3 = 2, and x_4 = 1/21. On the other hand, the probability the algorithm signals first is minimized ($\hat{p}$ = 0.027) at x_1 = 0.1, x_2 = 0.094, x_3 = 3, and x_4 = 1/21. For both the maximization and minimization, since x_5 is not in this model, it can take on any value between $1/28 \le x_5 \le 1/14$.

Model adequacy checks examining the residuals are seen in Figure 17 and 18. There is no pattern in the residuals, therefore the constant variance and independent assumptions are met.

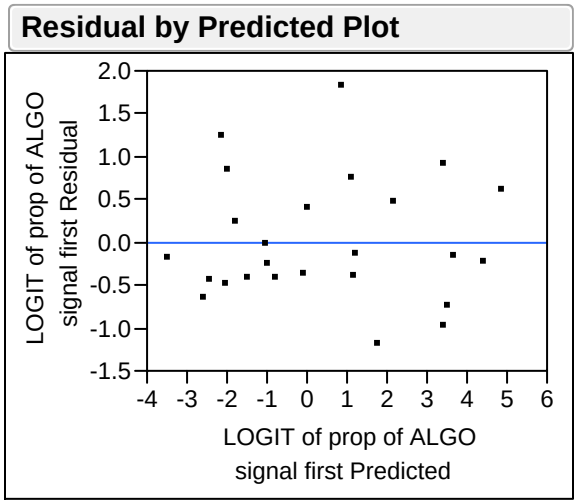

Figure 17. Model 1 residual by predicted plot

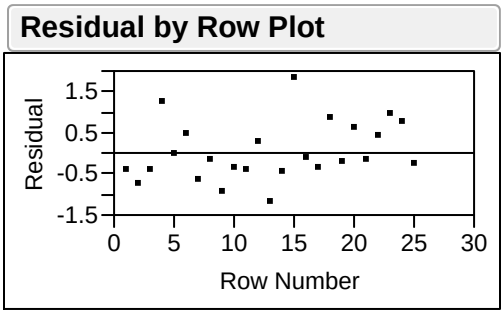

Figure 18. Model 1 residual by row plot

2. Population Size of 10,000 (Model 2)

Table 6 shows the results of 25 simulation runs, where 10,000 replications are executed within each run. The time it takes to complete the simulation run is 96 hours using a personal computer laptop and a desktop. The probability of algorithm signaling first is estimated via the simulation, translated into the logit, and then entered into JMP (along with the parameters used in the simulation) for model fitting and analysis.

Run #	Probability of algorithm signaling first	Logit of probability of algorithm signaling first
1	0.0187	-3.9604
2	0.8390	1.6508
3	0.0023	-6.0725
4	0.1086	-2.1051
5	0.0844	-2.3840
6	0.9788	3.8323
7	0.0014	-6.5699
8	0.0003	-8.1114
9	0.9998	8.5170
10	0.1383	-1.8295
11	0.0488	-2.9700
12	0.0538	-2.8672
13	0.2771	-0.9589
14	0.0139	-4.2619
15	0.0181	-3.9936
16	0.5377	0.1511
17	0.4293	-0.2847
18	0.0770	-2.4838
19	1.0000	13.8155
20	1.0000	13.8155
21	0.9987	6.6441
22	0.2440	-1.1309
23	1.0000	13.8155
24	0.7421	1.0569
25	0.0444	-3.0691

Table 6. Probability of algorithm signaling first results (population size of 10,000)

The probability of algorithm signaling first ranges from 0.0003 (lowest value in run number 8) to 1 (highest value in run number 19, 20, and 23). According to the simulation results, the algorithm will always signal an outbreak first in run number 19, 20, and 23. Table 7 consists of the parameter values in the simulation run number 8, 19, 20, and 23 for comparisons.

Parameter	Run #8	Run #19	Run #20	Run #23
x_1	0.1	0.1	0.1	0.1
x_2	0.1	0.001	0.001	0.001
x_3	3	2.5	2	3
x_4	1/14	1/21	1/13	1/7
x_5	1/21	1/27	1/19	1/14

Table 7. Model 2 parameters for simulation run number 8, 19, 20 and 23

The four simulation runs (in Table 7) all have the same probability of transitioning from susceptible to infected state (x_1 = 0.1). Run number 19, 20, and 23 (where the probability of algorithm signaling is 1) have the same probability a susceptible person goes to the hospital for non-anthrax related flu (x_2 = 0.001). However in run number 8 (where the probability of algorithm signaling is 0.0003), the probability a susceptible person goes to the hospital for non-anthrax related flu is much higher (x_2 = 0.1). Run number 8 and run number 23 have the same threshold (x_3 = 2), while run number 19 has a threshold of 2.5 and run number 20 has a threshold of 2. All four simulation runs differ in the daily increase in the probability an infected person will be correctly diagnosed (x_4) and an infected person goes to the hospital (x_5). The daily increase in the probability an infected person will be correctly diagnosed for runs number 8, 19, 20, and 23 are 1/14, 1/21, 1/13, and 1/7 days respectively. The daily increase in the probability an infected person goes to the hospital for runs number 8, 19, 20, and 23 are 1/21, 1/27, 1/19, and 1/14 days respectively

The model is fit in JMP using stepwise regression, regressing the estimated logit on the various simulation parameters. The results using OLS to fit the logit of the estimated probabilities to the covariates are seen in Equation 7.

Model 2: 10,000 people exposed with quadratic and interaction terms (R^2=0.95)

$$\text{logit}(\hat{p}) = 5.879 + 33.419x_1 - 102.4512x_2 - 2.1464x_3 - 1660.58x_1x_2 + 1716.2717(x_2)^2 \quad (7)$$

In order to graphically depict the effects of the variables with the largest effect (x_1, x_2, and x_3) on the probability of algorithm signaling first, Figure 19 through 21 shows the results for Model 2 where the other variables are set to their nominal values (x_1 = x_2 = 0.05, x_3 = 2.5, x_4 = 1/14, x_5 = 1/21) and then plot the estimated probability of algorithm signaling first as a function of the variables with the largest effect (x_1, x_2, and x_3).

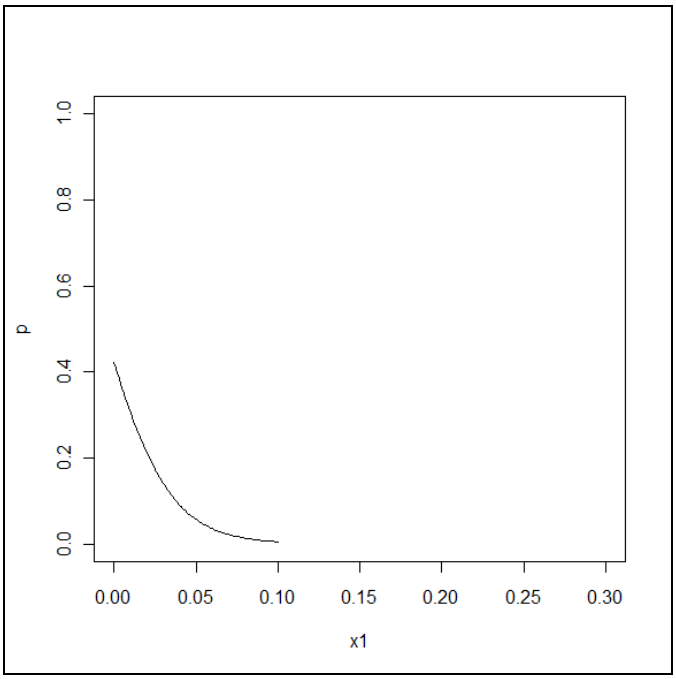

Figure 19. Plot of Model 2 made by varying x_1 while setting all other variables to their nominal values

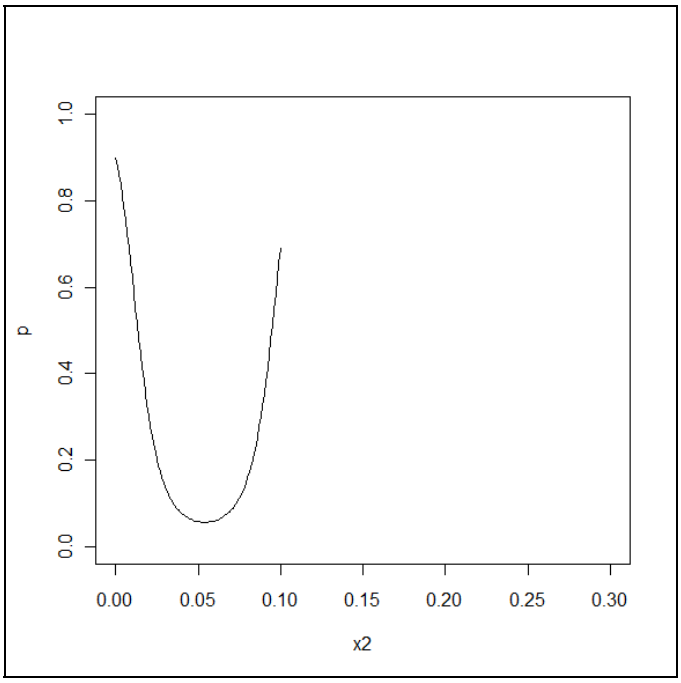

Figure 20. Plot of Model 2 made by varying x_2 over its range while setting all other variables to their nominal values

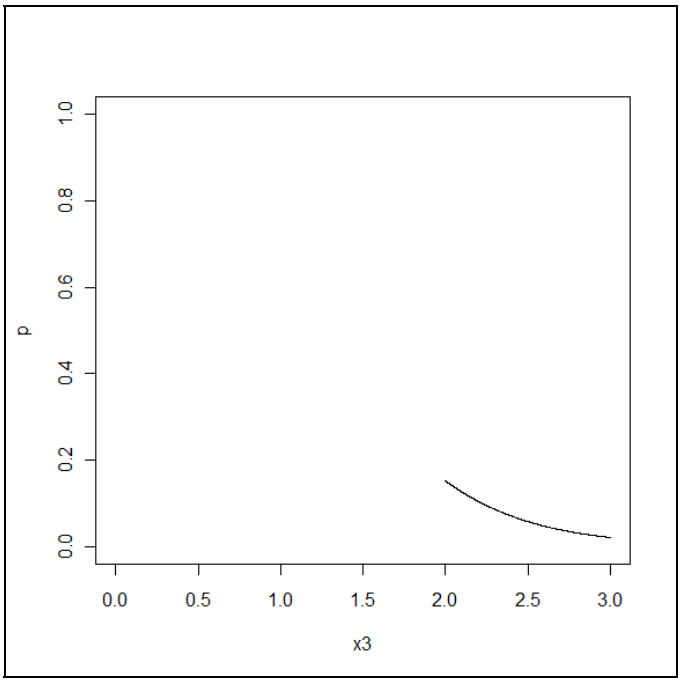

Figure 21. Plot of Model 2 made by varying x_3 over its range while setting all other variables to their nominal values

The variable with the largest effect on the probability the algorithm signals first are x_1 (probability of people getting infected), x_2 (probability going to the hospital for non-anthrax related flu), and x_3 (threshold). The results for x_1 (probability of people getting infected), x_2 (probability going to the hospital for non-anthrax related flu), and x_3 (threshold) are in the expected direction:

- As the probability of people getting infected (x_1) increases, we then expected that there would be more infected people going to the hospital that could be correctly diagnosed and thus the probability the algorithm signals first decreases,
- As the probability of going to the hospital for non-anthrax related flu (x_2) increases, the probability the algorithm signals first decreases to the point where $x_2 = 0.05$, and
- As the threshold(x_3) is increased, the probability the algorithm signals first decreases.

In this model, variables x_4 (daily increase in the probability an infected person will be correctly diagnosed) and x_5 (daily increase in the probability an infected person goes to the hospital) are not included. Therefore, the probability the algorithm signals first is not associated with x_4 (over the range considered: $1/21 \le x_4 \le 1/7$) and x_5 (over the range considered: $1/28 \le x_5 \le 1/14$).

The last step is to figure out which levels of the variables maximize and minimize the probability that the algorithm signals first. The probability the algorithm signals first is maximized ($\hat{p}$ = 0.999) at the boundaries for each of the variables: x_1 = 0.001, x_2 = 0.1, and x_3 = 2. On the other hand, the probability the algorithm signals first is minimized ($\hat{p}$ = 0) at x_1 = 0.1, x_2 = 0.069, and x_3 = 3. For both the maximization and minimization, since x_4 and x_5 are not in this model, therefore they can take on any values between $1/21 \le x_4 \le 1/7$ for x_4 and $1/28 \le x_5 \le 1/14$ for x_5.

Model adequacy checks examining the residuals are seen in Figure 22 and 23. There is no pattern in the residuals, therefore the constant variance and independent assumptions are met.

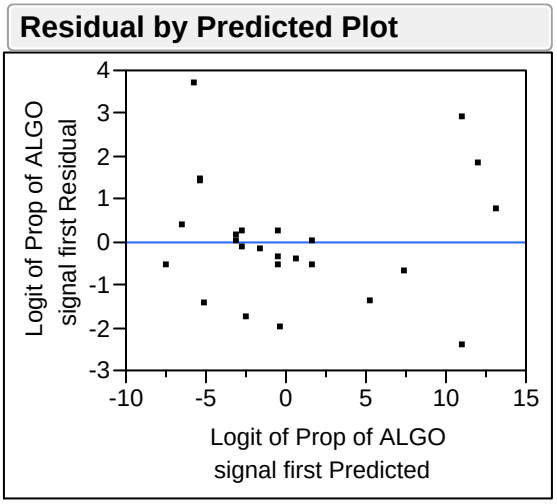

Figure 22. Plot of Model 2 residual by predicted plot

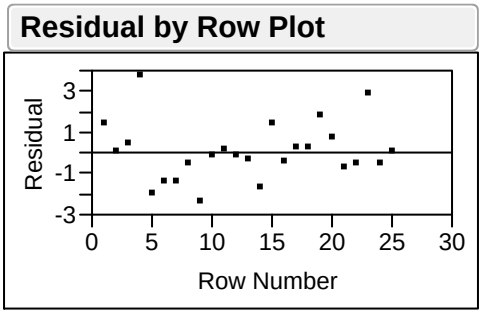

Figure 23. Plot of Model 2 residual by row plot

3. Comparisons of Model 1 and 2

In the comparisons of Model 1 and 2, only the main effects that are included in the models are analyzed. Table 8 shows the regression coefficients for Model 1 (exposed population of 1,000 people) and Model 2 (exposed population of 10,000 people). Both models' regression coefficients are consistent in their direction.

Model	β_0	β_1	β_2	β_3	β_4
1	12.0385	21.4731	-48.3301	-3.6478	-78.8509
2	5.879	33.419	-102.4512	-2.1464	n/a

Table 8. Model 1 and 2 regression coefficient comparisons

The magnitude of β_0, β_1, and β_2 decreases when going from Model 1 to Model 2, and the magnitude of β_3 increases. The probability the algorithm signals first is maximized at 0.996 and minimized at 0.027 for Model 1, while it is maximized at 0.999 and minimized at 0 for Model 2. Model 2 with an R square of 0.95 is a better regression line to fit the data than Model 1 with an R square of 0.92.

C. NUMBER OF DAYS TO ALGORITHM SIGNALING RESULTS

In the case of Model 1 and 2, the response variable is transformed into the logit for model fitting and analysis because probability needs to be constrained from 0 to 1. However, in the case of Model 3 and 4, it is not necessary to transform the number of days to algorithm signaling into the logit because number of days does not need to be constrained to the unit interval (though it does need to be non-negative). Table 9 shows the results of 25 simulation runs for both scenarios: one for an initial exposed population of 1,000 people (Model 3) and one for 10,000 exposed people (Model 4). Within each run, 10,000 replications are executed. The number of days to algorithm signaling is entered into JMP for model fitting and analysis.

Run #	Avg. number of days to algorithm signaling (n=1,000)	Avg. number of days to algorithm signaling (n=10,000)
1	1.5613	1.7861
2	4.9279	3.6729
3	1.4008	1.4348
4	4.7434	3.3637
5	1.8281	1.1991
6	3.4975	2.9137
7	2.0751	1.8571
8	1.4523	1.3333
9	3.1122	2.8658
10	5.4399	3.8886
11	2.9076	1.5246
12	4.6930	3.3234
13	4.7476	3.6536
14	2.3112	1.0576
15	4.7824	1.7403
16	4.9935	3.6251
17	5.7182	4.1356
18	6.2687	4.2740
19	2.9908	2.7474
20	2.6989	2.5368
21	3.0096	2.5387
22	4.4969	3.4865
23	2.8659	2.8623
24	4.2619	3.0838
25	2.8078	1.5743

Table 9. Number of days to algorithm signaling results (population size of 1,000 and 10,000)

1. Population Size of 1,000 (Model 3)

The number of days to algorithm signaling ranges from 1.4008 days (shortest time in run number 3) to 6.2687 days (longest time in run number 18). Run numbers 3 and 18 have the same probability a susceptible person goes to the hospital for non-anthrax related flu (x_2 = 0.1). They differ in the in the probability of transitioning from susceptible to infected state (x_1), and the threshold (x_3). In run number 3, the probability

is higher at 0.1 while it is at 0.001 for run number 18. The threshold in run number 3 is lower at 2.5 while the threshold for run number 18 is at 3. The daily increase in the probability an infected person will be correctly diagnosed (x_4) and an infected person goes to the hospital (x_5) in run number 3 are both lower than in run number 18. In run number 3, the daily increase in the probability an infected person will be correctly diagnosed is 1/7 and the daily increase in the probability that an infected person goes to the hospital 1/22. In run number 18, the daily increase in the probability an infected person will be correctly diagnosed is 1/21 and the daily increase in the probability that an infected person goes to the hospital is 1/28 days.

The model is fit in JMP using stepwise regression, regressing the estimated logit on the various simulation parameters. The results using OLS to fit the estimated number of days to algorithm signaling to the covariates are seen in Equation 8.

Model 3: 1,000 people exposed with quadratic and interaction terms (R^2=0.80)

$$\hat{y} = 6.2086 - 69.908x_1 - 10.795x_2 + 438.3063(x_1)^2 \qquad (8)$$

In order to graphically depict the effects of the variables with the largest effect (x_1 and x_2) on the number of days to algorithm signaling, Figure 24 and 25 shows the results for Model 3 where the other variables are set to their nominal values ($x_1 = x_2 = 0.05$, $x_3 = 2.5$, $x_4 = 1/14$, $x_5 = 1/21$) and then plot the estimated number of days to algorithm signaling as a function of the variables with the largest effect (x_1 and x_2).

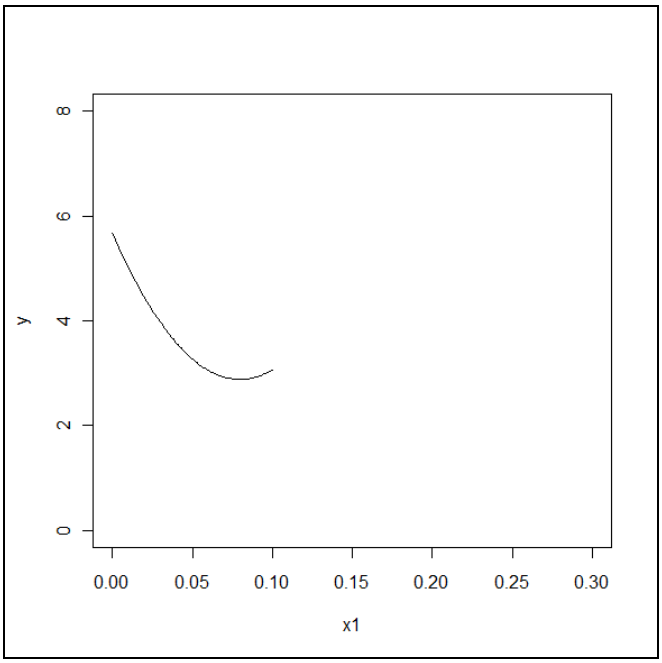

Figure 24. Plot of Model 3 made by varying x_1 over its range while setting all other variables to their nominal values

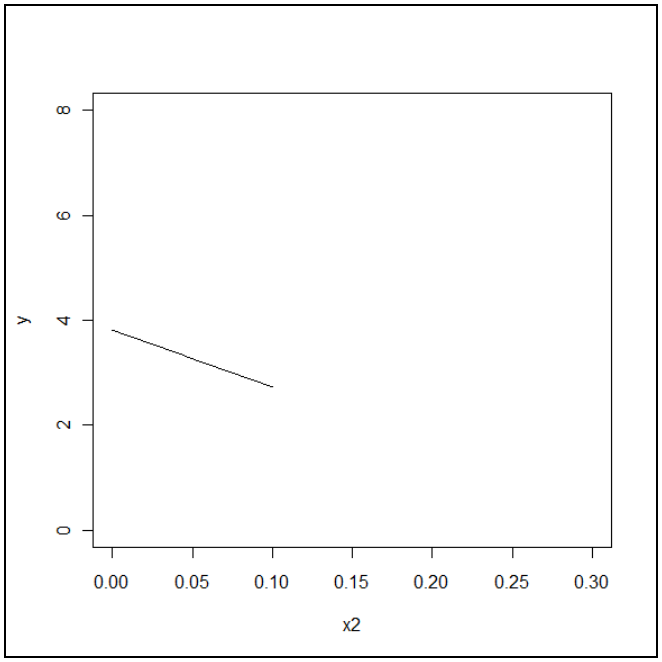

Figure 25. Plot of Model 3 made by varying x_2 over its range while setting all other variables to their nominal values

The variable with the largest effect on the number of days to algorithm signals first are x_1 (probability of people getting infected) and x_2 (probability going to the hospital for non-anthrax related flu). The results for x_1 (probability of people getting infected and x_2 (probability going to the hospital for non-anthrax related flu) are not in the expected direction:

- As the probability of people getting infected (x_1) increases, the probability the algorithm signals first decreases and thus the number of days to algorithm signals first increases, and
- As the probability of going to the hospital for non-anthrax related flu (x_2) increases, the probability the algorithm signals first decreases and thus the number of days to algorithm signals first increases.

In this model, variables x_3 (threshold), x_4 (daily increase in the probability an infected person will be correctly diagnosed) and x_5 (daily increase in the probability an infected person goes to the hospital) are not included. Therefore, the number of days to algorithm signals first is not associated with x_3 (over the range considered: $2 \le x_3 \le 3$), x_4 (over the range considered: $1/21 \le x_4 \le 1/7$) and x_5 (over the range considered: $1/28 \le x_5 \le 1/14$).

Determining which levels of the variables maximize and minimize the average time until the algorithm signals is the next step. The number of days to algorithm signals first is maximized ($\hat{y}$ = 6.13) at the boundaries for each of the variables: x_1 = 0.001 and x_2 = 0.001. On the other hand, the number of days to algorithm signals first is minimized ($\hat{y}$ = 2.34) at x_1 = 0.08 and x_2 = 0.1. For both the maximization and minimization, since x_3, x_4 and x_5 are not in this model, therefore they can take on any values between $2 \le x_3 \le 3$ for x_3, $1/21 \le x_4 \le 1/7$ for x_4, and $1/28 \le x_5 \le 1/14$ for x_5.

Model adequacy checks examining the residuals are seen in Figure 26 and 27. There is no pattern in the residuals, therefore the constant variance and independent assumptions are met.

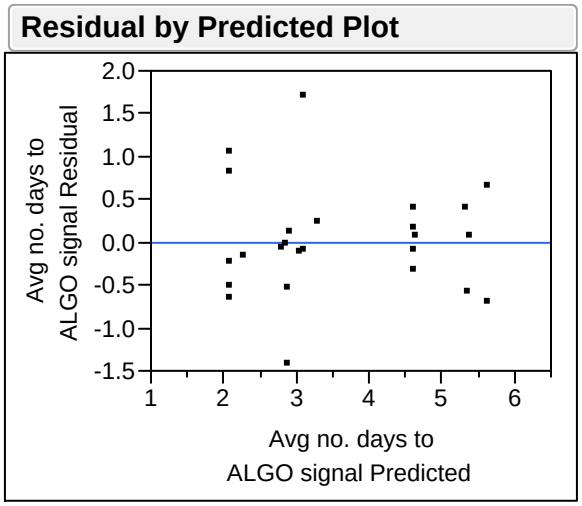

Figure 26. Plot of Model 3 residual by predicted plot

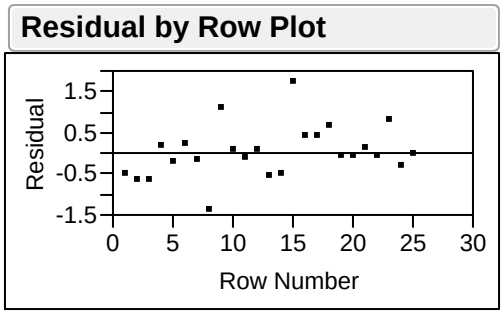

Figure 27. Plot of Model 3 residual by row plot

2. Population size of 10,000 (Model 4)

The number of days to algorithm signaling ranges from 1.0576 days (shortest time in run number 14) to 4.1356 days (longest time in run number 17). The probability of transitioning from susceptible to infected state (x_1) and the probability a susceptible person goes to the hospital for non-anthrax related flu (x_2) are both higher in run number 14 ($x_1 = 0.1$, $x_2 = 0.505$) compared to run number 17 ($x_1 = 0.001$, $x_2 = 0.001$). They have the same threshold of 3. The daily increase in the probability an infected person will be correctly diagnosed (x_4) and an infected person goes to the hospital (x_5) in run number 17 are both lower than in run number 14. In run number 17, the daily increase in the probability an infected person will be correctly diagnosed is 1/13 and the daily increase in

the probability that an infected person goes to the hospital 1/20. In run number 14, the daily increase in the probability an infected person will be correctly diagnosed is 1/15 days and the daily increase in the probability that an infected person goes to the hospital is 1/28.

The model is fit in JMP using stepwise regression, regressing the estimated logit on the various simulation parameters. The results using OLS to fit the estimated number of days to algorithm signaling to the covariates are seen in Equation 9.

Model 4: 10,000 people exposed with quadratic and interaction terms (R^2=0.98)

$$\begin{aligned}\hat{y} &= 2.5008 - 17.0909x_1 - 6.6243x_2 + 0.2167x_3 - 1.914x_4 \\ &\quad - 2.953x_5 - 119.045x_1x_2 - 6.5411x_1x_3 + 187.45x_1x_5 - 5.42x_2x_3 \\ &\quad + 255.361(x_1)^2 + 317.9626(x_2)^2\end{aligned} \tag{9}$$

In order to graphically depict the effects of the variables with the largest effect (x_1, x_2, x_3, and x_5) on the number of days to algorithm signaling, Figure 28 through 31 shows the results for Model 4 where the other variables are set to their nominal values (x_1 = x_2 = 0.05, x_3 = 2.5, x_4 = 1/14, x_5 = 1/21) and then plot the estimated number of days to algorithm signaling as a function of the variables with the largest effect (x_1, x_2, x_3, and x_5).

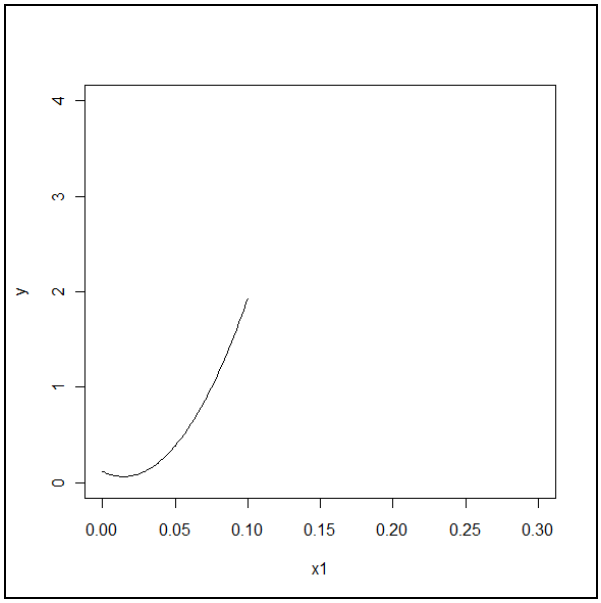

Figure 28. Plot of Model 4 made by varying x_1 over its range while setting all other variables to their nominal values

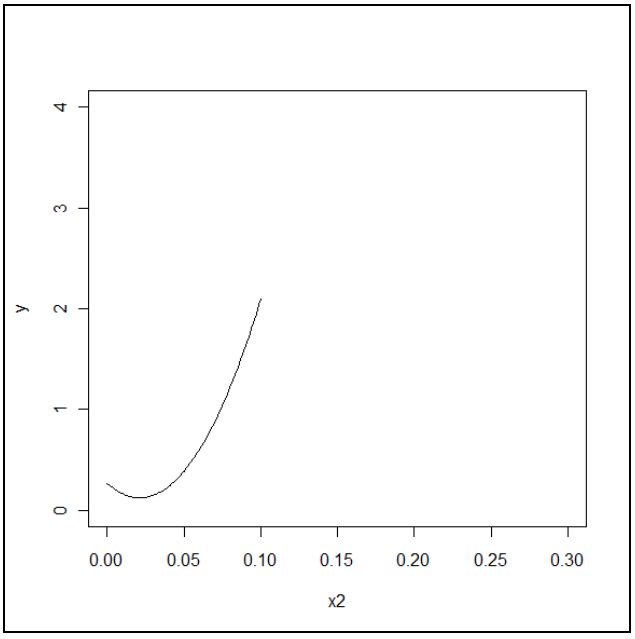

Figure 29. Plot of Model 4 made by varying x_2 over its range while setting all other variables to their nominal values

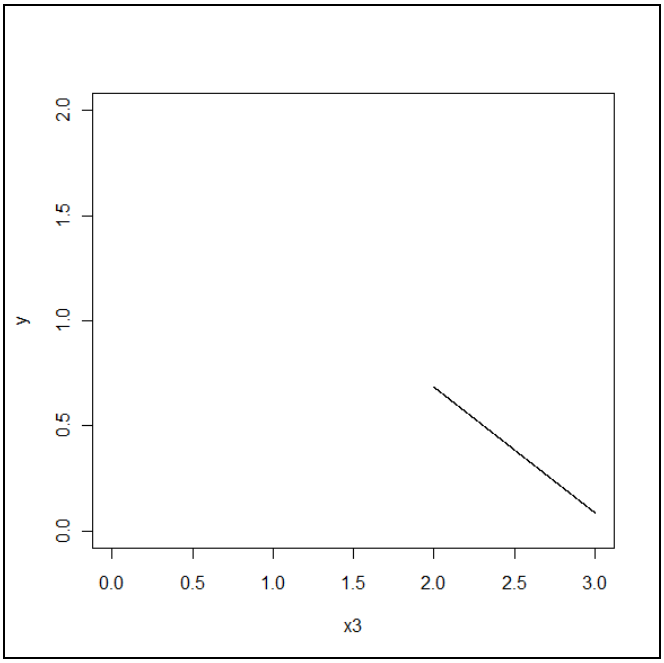

Figure 30. Plot of Model 4 made by varying x_3 over its range while setting all other variables to their nominal values

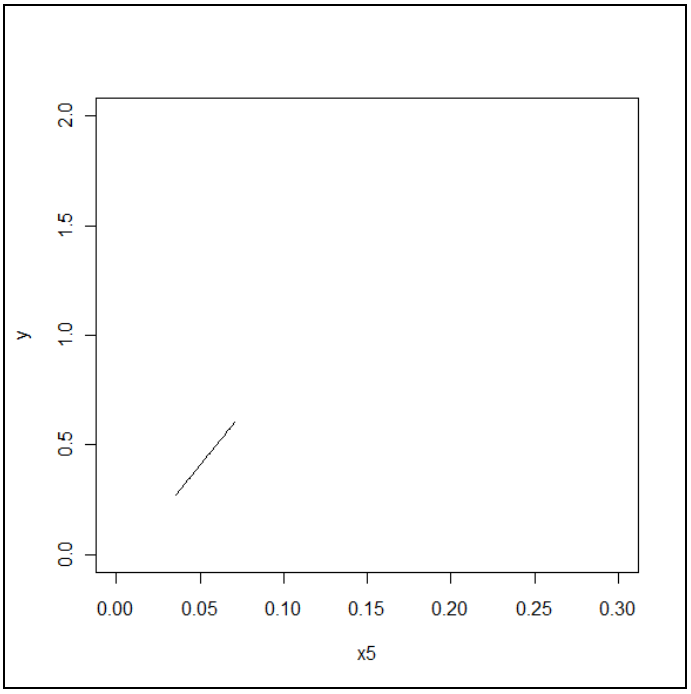

Figure 31. Plot of Model 4 made by varying x_5 over its range while setting all other variables to their nominal values

The variable with the largest effect on the number of days to algorithm signals first are x_1 (probability of people getting infected), x_2 (probability going to the hospital for non-anthrax related flu), x_3 (threshold), and x_5 (daily increase in the probability an infected person goes to the hospital). The results for x_1 (probability of people getting infected and x_2 (probability going to the hospital for non-anthrax related flu) and x_5 (daily increase in the probability an infected person goes to the hospital) are in the expected direction:

- As the probability of people getting infected (x_1) increases, the probability the algorithm signals first decreases and thus the number of days to algorithm signals first increases,
- As the probability of going to the hospital for non-anthrax related flu (x_2) increases, the probability the algorithm signals first decreases and thus the number of days to algorithm signals first increases, and

- As the daily increase in the probability an infected person goes to the hospital (x_5) increases, the probability the algorithm signals first decreases and thus the number of days to algorithm signals first increases.

Interestingly, Figure 18 shows that the threshold (x_3) is not in the expected direction. This is a surprising result, as we expected that:

- As the threshold (x_3) increases, the probability the algorithm signals first decreases and thus the number of days to algorithm signals first should increases.

In this model, variable x_4 (daily increase in the probability an infected person will be correctly diagnosed) is not included. Therefore, the number of days to algorithm signals first is not associated with x_4 (over the range considered: $1/21 \le x_4 \le 1/7$).

The last question is which levels of the variables maximize and minimize the number of days to algorithm signals first. The number of days to algorithm signals first is maximized ($\hat{y}$ = 4.68) at the boundaries for each of the variables: x_1 = 0.1, x_2 = 0.1, x_3 = 2, and x_5 = 1/14. On the other hand, the number of days to algorithm signals first is minimized ($\hat{y}$ = 0) at x_1 = 0.025, x_2 = 0.026, x_3 = 3, and x_5 = 1/28. For both the maximization and minimization, since x_4 is not in this model, therefore it can take on any values between $1/21 \le x_4 \le 1/7$.

Model adequacy checks examining the residuals are seen in Figures 32 and 33. There is no pattern in the residuals, therefore the constant variance and independent assumptions are met.

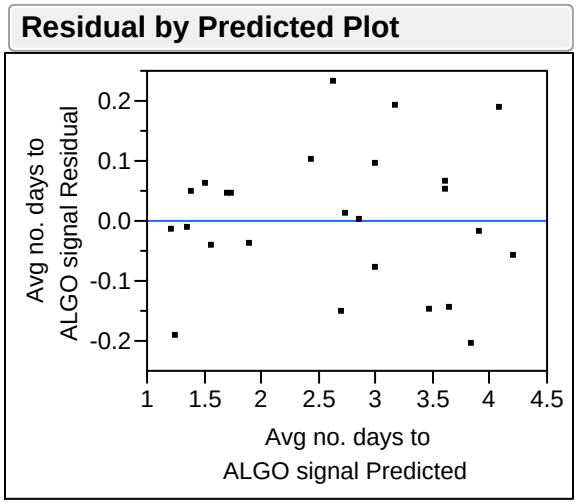

Figure 32. Plot of Model 4 residual by predicted plot

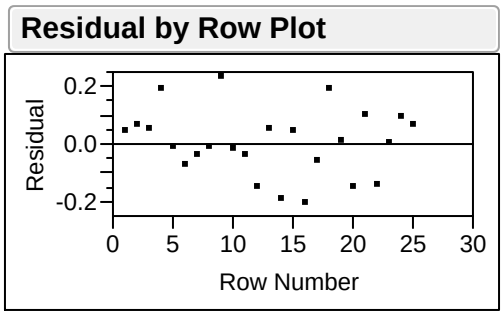

Figure 33. Plot of Model 4 residual by row plot

3. Comparisons of Model 3 and 4

In the comparisons of Model 3 and 4, only the main effects that are included in the model are being looked at. Table 10 shows the regression coefficients for Model 3 (exposed population of 1,000 people) and Model 4 (exposed population of 10,000 people). Both models' regression coefficients are consistent in their direction. However, Model 4 has more terms than Model 3 since it includes β_3, β_4, and β_5 in the model.

Model	β_0	β_1	β_2	β_3	β_4	β_5
3	6.2086	-69.908	-10.795	n/a	n/a	n/a
4	2.5008	-17.0909	-6.6243	0.2167	-1.914	-2.953

Table 10. Model 3 and 4 regression coefficient comparisons

The magnitude of β_0 decreases when going from Model 3 to 4, and the magnitude of β_1 and β_2 increases when going from Model 3 to 4. The number of days to algorithm signals first is maximized at 6.13 days and minimized at 2.34 days for Model 3, while it is maximized at 4.68 days and minimized at 0 for Model 4. Model 4 with an R square of 0.98 is a better regression line to fit the data than Model 3 with an R square of 0.8.

IV. CONCLUSIONS

A. BIOSURVEILLANCE IS USEFUL FOR EED

This research shows that biosurveillance statistical algorithms, such as the EARS C1, are useful in EED of a bioterror attack. The metrics used to determine the effectiveness of the EARS C1 algorithm (as seen in Table 11) are the probability it signals an anthrax outbreak first and the time it takes to do so. In the worst case scenarios, the probability the algorithm signals first is 13.04% for an exposed population of 1,000 people and it is 0.03% for an exposed population of 10,000 people. In the nominal case scenarios, the probability the algorithm signals first is 31.5% for an exposed population of 1,000 people and it is 0.3% for an exposed population of 10,000 people. Although these probabilities may seem low, note that whether the algorithm signaled first was quite situation dependent. And even in the worst case scenario for 1,000 people, the algorithm signaled first more than one time in 10. Thus, at the very least biosurveillance is an effective back-up to clinicians. On the other hand, there were scenarios in which the statistical algorithm almost always signaled first.

Furthermore, the EARS C1 algorithm does not take a long time to signal an anthrax outbreak. In the worst case scenarios, the longest time it takes for the algorithm to signal is 6.63 days for an exposed population of 1,000 people and 4.14 days for an exposed population of 10,000 people. In the nominal case scenarios, the time it takes for the algorithm to signal is 3.3 days for an exposed population of 1,000 people and 0.38 days for an exposed population of 10,000 people.

Population size	Min Prob. ALGO signals	Nominal Prob. ALGO signal	Max Prob. ALGO signal	Min days to ALGO signal	Nominal days to ALGO signal	Max days to ALGO signal
1,000	0.1304	0.315	0.9957	1.4	3.3	6.63
10,000	0.0003	.003	1	1.06	0.38	4.14

Table 11. Model 1 through Model 4 of the response variables results

The ideal algorithm maximizes the probability it signals first while minimizes the time it takes to signal. Table 12 gives the values of the parameters that maximize the probability the algorithm signals first. In the case of an exposed population of 1,000 people, x_5 is not in the model thus it can be any value between the specified ranges. In the case of an exposed population of 10,000 people, x_4 and x_5 are not in the model thus they can be any value between the specified ranges. The probability the algorithm signals first is maximized at 99.6% for an exposed population of 1,000 people and 99.9% for an exposed population of 10,000 people.

Population size	x_1	x_2	x_3	x_4	x_5
1,000	0.1	0.001	2	1/21	$1/28 \le x_5 \le 1/14$
10,000	0.001	0.1	2	$1/21 \le x_4 \le 1/7$	$1/28 \le x_5 \le 1/14$

Table 12. Values of the parameters that maximize the probability the algorithm signals first

Table 13 gives the values of the parameters that minimize the number of days it takes for an algorithm signal. In the case of an exposed population of 1,000 people, $x_3, x_4,$ and x_5 are not in the model thus they can be any value between the specified ranges. In the case of an exposed population of 10,000 people, x_4 is not in the model, thus it can be

any value between the specified ranges. The time it takes to signal is minimized at 2.34 days for an exposed population of 1,000 people and 0 day for an exposed population of 10,000 people.

Population size	x_1	x_2	x_3	x_4	x_5
1,000	0.08	0.1	$2 \le x_3 \le 3$	$1/21 \le x_4 \le 1/7$	$1/28 \le x_5 \le 1/14$
10,000	0.025	0.026	3	$1/21 \le x_4 \le 1/7$	1/28

Table 13. Values of the parameters that minimize the number of days to algorithm signal

The parameters with the largest effect on the probability the algorithm signals first are the probability of people getting infected (x_1), the probability of going to the hospital for non-anthrax related flu (x_2), the threshold (x_3), and the daily increase in the probability an infected person will be correctly diagnosed (x_4). An increase in the threshold and the transitional probabilities of people getting infected, going to the hospital for non-anthrax related flu and correct diagnosis by doctor result in a decrease in the probability the algorithm signals first, and thus an increase in the probability the doctor signals first. This finding is consistent with Professor Fricker's simulation results in the sense that as the higher the probability of correct diagnosis by doctor, the less likely the statistical algorithm is to signal first.

The parameters with the largest effect on the number of days to algorithm signal are the probability of people getting infected (x_1), the probability of going to the hospital for non-anthrax related flu (x_2), the threshold (x_3), and the daily increase in the probability an infected person goes to the hospital (x_5). An increase in the transitional probabilities of people getting infected, going to the hospital for non-anthrax related flu and an infected person goes to the hospital result in an increase in the time it takes for the algorithm to signal.

B. FUTURE RESEARCH OPPORTUNITIES

In this thesis, two exposed population sizes of 1,000 and 10,000 people are explored in the simulation model analysis. The results suggest a possibility of a population size effect in the sense that the larger the population size, the lower the probability of an algorithm to signal first. In order to better characterize the region where biosurveillance is useful (as seen in Figure 1), different population sizes should be evaluated. Additionally, the five simulation parameters were evaluated over a small range for their values. While these ranges were judged to be the most likely, it would be interesting to investigate the effects of a wider range for these parameters.

There are many biosurveillance statistical algorithms that can be implemented in the simulation model such as the EARS C2 and C3, the CUSUM, and the Shewhart. The simulation model in this thesis only implements the EARS C1 statistical algorithm. There could be interesting insights in comparing the performance among various statistical algorithms. Furthermore, while it is not necessary to model the probability the doctor signals first, since it is 1 minus the probability the algorithm signals first, the number of days to doctor signal can still be modeled to explore the effect of the variables that are significant.

APPENDIX A. OUTPUTS (POPULATION SIZE OF 1,000)

For a population size of 1,000 people, 25 simulation runs is executed. Each simulation run consists of 10,000 replications.

```
RUN #1:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  874.6673 +/-  0.3439
      Avg no. Infected:  43.8027 +/-  0.3289
      Avg no. At The Hospital:  81.5301 +/-  0.0493

      AVG NO. OF ALGORITHM SIGNALS:  0.1304 +/-  0.0066
      AVG NO. OF DOCTOR SIGNALS:  0.8696 +/-  0.0066

      AVG No. of Days from Susceptible to Algo Signal:   1.5613 +/-
0.0496
      AVG No. of Days from Susceptible to Doc Signal:   4.1611 +/-
0.0077

RUN #2:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  998.1784 +/-  0.0214
      Avg no. Infected:  0.9739 +/-  0.0189
      Avg no. At The Hospital:  0.8477 +/-  0.0057

      AVG NO. OF ALGORITHM SIGNALS:  0.9390 +/-  0.0047
      AVG NO. OF DOCTOR SIGNALS:  0.0610 +/-  0.0047

      AVG No. of Days from Susceptible to Algo Signal:   4.9279 +/-
0.0611
      AVG No. of Days from Susceptible to Doc Signal:   8.4033 +/-
0.1705

RUN #3:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  868.7775 +/-  0.2955
      Avg no. Infected:  49.8837 +/-  0.2884
      Avg no. At The Hospital:  81.3388 +/-  0.0478

      AVG NO. OF ALGORITHM SIGNALS:  0.0524 +/-  0.0044
      AVG NO. OF DOCTOR SIGNALS:  0.9476 +/-  0.0044

      AVG No. of Days from Susceptible to Algo Signal:   1.4008 +/-
0.0585
      AVG No. of Days from Susceptible to Doc Signal:   4.2956 +/-
0.0093

RUN #4:
Using 10000 independent replications, 95% CI for following measures as
followed:
```

```
      Avg no. Susceptible:  953.9723 +/-  0.0492
      Avg no. Infected:  1.4709 +/-  0.0172
      Avg no. At The Hospital:  44.5568 +/-  0.0399

      AVG NO. OF ALGORITHM SIGNALS:  0.2810 +/-  0.0088
      AVG NO. OF DOCTOR SIGNALS:  0.7190 +/-  0.0088

      AVG No. of Days from Susceptible to Algo Signal:   4.7434 +/-
0.1035
      AVG No. of Days from Susceptible to Doc Signal:   8.3513 +/-
0.0459


RUN #5:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  911.7260 +/-  0.4776
      Avg no. Infected:  45.4950 +/-  0.4670
      Avg no. At The Hospital:  42.7790 +/-  0.0382

      AVG NO. OF ALGORITHM SIGNALS:  0.2490 +/-  0.0085
      AVG NO. OF DOCTOR SIGNALS:  0.7510 +/-  0.0085

      AVG No. of Days from Susceptible to Algo Signal:   1.8281 +/-
0.0462
      AVG No. of Days from Susceptible to Doc Signal:   4.3655 +/-
0.0113

RUN #6:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  976.8183 +/-  0.2171
      Avg no. Infected:  22.2274 +/-  0.2128
      Avg no. At The Hospital:  0.9543 +/-  0.0073

      AVG NO. OF ALGORITHM SIGNALS:  0.9320 +/-  0.0049
      AVG NO. OF DOCTOR SIGNALS:  0.0680 +/-  0.0049

      AVG No. of Days from Susceptible to Algo Signal:   3.4975 +/-
0.0212
      AVG No. of Days from Susceptible to Doc Signal:   4.3574 +/-
0.0402

RUN #7:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  885.8269 +/-  0.1909
      Avg no. Infected:  31.8395 +/-  0.1815
      Avg no. At The Hospital:  82.3335 +/-  0.0465

      AVG NO. OF ALGORITHM SIGNALS:  0.0373 +/-  0.0037
      AVG NO. OF DOCTOR SIGNALS:  0.9627 +/-  0.0037

      AVG No. of Days from Susceptible to Algo Signal:   2.0751 +/-
0.1193
```

```
      AVG No. of Days from Susceptible to Doc Signal:   4.7316 +/-
0.0126
RUN #8:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  860.1158 +/-  0.3125
      Avg no. Infected:  58.5347 +/-  0.3082
      Avg no. At The Hospital:  81.3495 +/-  0.0464

      AVG NO. OF ALGORITHM SIGNALS:  0.0241 +/-  0.0030
      AVG NO. OF DOCTOR SIGNALS:  0.9759 +/-  0.0030

      AVG No. of Days from Susceptible to Algo Signal:   1.4523 +/-
0.1105
      AVG No. of Days from Susceptible to Doc Signal:   4.6403 +/-
0.0120


RUN #9:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  964.9003 +/-  0.2958
      Avg no. Infected:  34.1795 +/-  0.2923
      Avg no. At The Hospital:  0.9202 +/-  0.0069

      AVG NO. OF ALGORITHM SIGNALS:  0.9178 +/-  0.0054
      AVG NO. OF DOCTOR SIGNALS:  0.0822 +/-  0.0054

      AVG No. of Days from Susceptible to Algo Signal:   3.1122 +/-
0.0163
      AVG No. of Days from Susceptible to Doc Signal:   4.0231 +/-
0.0103
RUN #10:
Using 10000 independent replications, 95% CI for following measures as
followed:

      Avg no. Susceptible:  913.6131 +/-  0.0715
      Avg no. Infected:  1.6930 +/-  0.0220
      Avg no. At The Hospital:  84.6939 +/-  0.0577

      AVG NO. OF ALGORITHM SIGNALS:  0.3903 +/-  0.0096
      AVG NO. OF DOCTOR SIGNALS:  0.6097 +/-  0.0096

      AVG No. of Days from Susceptible to Algo Signal:   5.4399 +/-
0.1008
      AVG No. of Days from Susceptible to Doc Signal:   9.8614 +/-
0.0632

RUN #11:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  924.7160 +/-  0.2904
      Avg no. Infected:  31.7611 +/-  0.2738
      Avg no. At The Hospital:  43.5228 +/-  0.0392

      AVG NO. OF ALGORITHM SIGNALS:  0.2263 +/-  0.0082
```

AVG NO. OF DOCTOR SIGNALS: 0.7737 +/- 0.0082

AVG No. of Days from Susceptible to Algo Signal: 2.9076 +/- 0.0691
AVG No. of Days from Susceptible to Doc Signal: 4.8954 +/- 0.0156

RUN #12:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 913.8416 +/- 0.0576
Avg no. Infected: 1.3499 +/- 0.0141
Avg no. At The Hospital: 84.8086 +/- 0.0512

AVG NO. OF ALGORITHM SIGNALS: 0.1694 +/- 0.0074
AVG NO. OF DOCTOR SIGNALS: 0.8306 +/- 0.0074

AVG No. of Days from Susceptible to Algo Signal: 4.6930 +/- 0.1194
AVG No. of Days from Susceptible to Doc Signal: 7.9121 +/- 0.0399

RUN #13:
Avg no. Susceptible: 915.1090 +/- 0.0734
Avg no. Infected: 1.1578 +/- 0.0191
Avg no. At The Hospital: 83.7332 +/- 0.0611

AVG NO. OF ALGORITHM SIGNALS: 0.6307 +/- 0.0095
AVG NO. OF DOCTOR SIGNALS: 0.3693 +/- 0.0095

AVG No. of Days from Susceptible to Algo Signal: 4.7476 +/- 0.0710
AVG No. of Days from Susceptible to Doc Signal: 8.8979 +/- 0.0720

RUN #14:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 895.4357 +/- 0.3960
Avg no. Infected: 61.5230 +/- 0.3877
Avg no. At The Hospital: 43.0413 +/- 0.0364

AVG NO. OF ALGORITHM SIGNALS: 0.0739 +/- 0.0051
AVG NO. OF DOCTOR SIGNALS: 0.9261 +/- 0.0051

AVG No. of Days from Susceptible to Algo Signal: 2.3112 +/- 0.1147
AVG No. of Days from Susceptible to Doc Signal: 4.7250 +/- 0.0129

RUN #15:
Using 10000 independent replications, 95% CI for following measures as followed:

Avg no. Susceptible: 998.2334 +/- 0.0199
Avg no. Infected: 0.9150 +/- 0.0173
Avg no. At The Hospital: 0.8516 +/- 0.0058

AVG NO. OF ALGORITHM SIGNALS: 0.9353 +/- 0.0048
AVG NO. OF DOCTOR SIGNALS: 0.0647 +/- 0.0048

AVG No. of Days from Susceptible to Algo Signal: 4.7824 +/- 0.0574
AVG No. of Days from Susceptible to Doc Signal: 8.0618 +/- 0.1629

RUN #16:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 997.9773 +/- 0.0215
Avg no. Infected: 1.1488 +/- 0.0185
Avg no. At The Hospital: 0.8739 +/- 0.0059

AVG NO. OF ALGORITHM SIGNALS: 0.7409 +/- 0.0086
AVG NO. OF DOCTOR SIGNALS: 0.2591 +/- 0.0086

AVG No. of Days from Susceptible to Algo Signal: 4.9935 +/- 0.0649
AVG No. of Days from Susceptible to Doc Signal: 8.0587 +/- 0.0756

RUN #17:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 997.7239 +/- 0.0227
Avg no. Infected: 1.3561 +/- 0.0191
Avg no. At The Hospital: 0.9200 +/- 0.0062

AVG NO. OF ALGORITHM SIGNALS: 0.6810 +/- 0.0091
AVG NO. OF DOCTOR SIGNALS: 0.3190 +/- 0.0091

AVG No. of Days from Susceptible to Algo Signal: 5.7182 +/- 0.0708
AVG No. of Days from Susceptible to Doc Signal: 8.5502 +/- 0.0750

RUN #18:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 912.3599 +/- 0.0665
Avg no. Infected: 2.1806 +/- 0.0232
Avg no. At The Hospital: 85.4595 +/- 0.0525

AVG NO. OF ALGORITHM SIGNALS: 0.2356 +/- 0.0083
AVG NO. OF DOCTOR SIGNALS: 0.7644 +/- 0.0083

AVG No. of Days from Susceptible to Algo Signal: 6.2687 +/- 0.1449

AVG No. of Days from Susceptible to Doc Signal: 10.9104 +/- 0.0626

RUN #19:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 968.1928 +/- 0.3108
Avg no. Infected: 30.9146 +/- 0.3077
Avg no. At The Hospital: 0.8926 +/- 0.0066

AVG NO. OF ALGORITHM SIGNALS: 0.9845 +/- 0.0024
AVG NO. OF DOCTOR SIGNALS: 0.0155 +/- 0.0024

AVG No. of Days from Susceptible to Algo Signal: 2.9908 +/- 0.0174
AVG No. of Days from Susceptible to Doc Signal: 4.0516 +/- 0.0349

RUN #20:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 974.1262 +/- 0.2674
Avg no. Infected: 25.0157 +/- 0.2653
Avg no. At The Hospital: 0.8580 +/- 0.0063

AVG NO. OF ALGORITHM SIGNALS: 0.9957 +/- 0.0013
AVG NO. OF DOCTOR SIGNALS: 0.0043 +/- 0.0013

AVG No. of Days from Susceptible to Algo Signal: 2.6989 +/- 0.0162
AVG No. of Days from Susceptible to Doc Signal: 4.0000 +/- 0.0000

RUN #21:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 982.2369 +/- 0.2077
Avg no. Infected: 16.8805 +/- 0.2051
Avg no. At The Hospital: 0.8826 +/- 0.0064

AVG NO. OF ALGORITHM SIGNALS: 0.9694 +/- 0.0034
AVG NO. OF DOCTOR SIGNALS: 0.0306 +/- 0.0034

AVG No. of Days from Susceptible to Algo Signal: 3.0096 +/- 0.0218
AVG No. of Days from Susceptible to Doc Signal: 4.1046 +/- 0.0343

RUN #22:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 915.2453 +/- 0.0717
Avg no. Infected: 1.1097 +/- 0.0182
Avg no. At The Hospital: 83.6450 +/- 0.0604

```
      AVG NO. OF ALGORITHM SIGNALS:  0.6041 +/-  0.0096
      AVG NO. OF DOCTOR SIGNALS:  0.3959 +/-  0.0096

      AVG  No.  of  Days  from  Susceptible  to  Algo  Signal:   4.4969  +/-
0.0685
      AVG  No.  of  Days  from  Susceptible  to  Doc  Signal:   8.4233  +/-
0.0643

RUN #23:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  971.5566 +/-  0.2045
      Avg no. Infected:  27.5650 +/-  0.2021
      Avg no. At The Hospital:  0.8784 +/-  0.0064

      AVG NO. OF ALGORITHM SIGNALS:  0.9866 +/-  0.0023
      AVG NO. OF DOCTOR SIGNALS:  0.0134 +/-  0.0023

      AVG  No.  of  Days  from  Susceptible  to  Algo  Signal:   2.8659  +/-
0.0120
      AVG  No.  of  Days  from  Susceptible  to  Doc  Signal:   4.0000  +/-
0.0000

RUN #24:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  998.3674 +/-  0.0167
      Avg no. Infected:  0.7747 +/-  0.0140
      Avg no. At The Hospital:  0.8578 +/-  0.0058

      AVG NO. OF ALGORITHM SIGNALS:  0.8637 +/-  0.0067
      AVG NO. OF DOCTOR SIGNALS:  0.1363 +/-  0.0067

      AVG  No.  of  Days  from  Susceptible  to  Algo  Signal:   4.2619  +/-
0.0510
      AVG  No.  of  Days  from  Susceptible  to  Doc  Signal:   6.7850  +/-
0.0835

RUN #25:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  926.9914 +/-  0.2653
      Avg no. Infected:  29.4859 +/-  0.2480
      Avg no. At The Hospital:  43.5227 +/-  0.0397

      AVG NO. OF ALGORITHM SIGNALS:  0.2227 +/-  0.0082
      AVG NO. OF DOCTOR SIGNALS:  0.7773 +/-  0.0082

      AVG  No.  of  Days  from  Susceptible  to  Algo  Signal:   2.8078  +/-
0.0665
      AVG  No.  of  Days  from  Susceptible  to  Doc  Signal:   4.6914  +/-
0.0139
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. OUTPUTS (POPULATION SIZE OF 10,000)

For a population size of 10,000 people, 25 simulation runs is executed. Each simulation run consists of 10,000 replications.

```
RUN #1:

Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  8731.7465 +/-  1.0473
      Avg no. Infected:  451.5045 +/-  1.0276
      Avg no. At The Hospital:  816.7490 +/-  0.1477

      AVG NO. OF ALGORITHM SIGNALS:  0.0187 +/-  0.0027
      AVG NO. OF DOCTOR SIGNALS:  0.9813 +/-  0.0027

      AVG No. of Days from Susceptible to Algo Signal:   1.7861 +/-
0.0589
      AVG No. of Days from Susceptible to Doc Signal:   4.0000 +/-
0.0000


RUN #2:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9984.5573 +/-  0.1195
      Avg no. Infected:  6.4804 +/-  0.1087
      Avg no. At The Hospital:  8.9622 +/-  0.0205

      AVG NO. OF ALGORITHM SIGNALS:  0.8390 +/-  0.0072
      AVG NO. OF DOCTOR SIGNALS:  0.1610 +/-  0.0072

      AVG No. of Days from Susceptible to Algo Signal:   3.6729 +/-
0.0452
      AVG No. of Days from Susceptible to Doc Signal:   6.0739 +/-
0.0595


RUN #3:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  8725.6673 +/-  0.4530
      Avg no. Infected:  459.6378 +/-  0.4295
      Avg no. At The Hospital:  814.6949 +/-  0.1486

      AVG NO. OF ALGORITHM SIGNALS:  0.0023 +/-  0.0009
      AVG NO. OF DOCTOR SIGNALS:  0.9977 +/-  0.0009

      AVG No. of Days from Susceptible to Algo Signal:   1.4348 +/-
0.2192
      AVG No. of Days from Susceptible to Doc Signal:   4.0000 +/-
0.0000
```

```
RUN #4:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9548.2434 +/-  0.1962
      Avg no. Infected:  8.9145 +/-  0.0673
      Avg no. At The Hospital:  442.8421 +/-  0.1500

      AVG NO. OF ALGORITHM SIGNALS:  0.1086 +/-  0.0061
      AVG NO. OF DOCTOR SIGNALS:  0.8914 +/-  0.0061

      AVG No. of Days from Susceptible to Algo Signal:   3.3637 +/-
0.0951
      AVG No. of Days from Susceptible to Doc Signal:   5.5951 +/-
0.0195

RUN #5:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9126.8494 +/-  2.5781
      Avg no. Infected:  443.9956 +/-  2.5345
      Avg no. At The Hospital:  429.1550 +/-  0.1235

      AVG NO. OF ALGORITHM SIGNALS:  0.0844 +/-  0.0054
      AVG NO. OF DOCTOR SIGNALS:  0.9156 +/-  0.0054

      AVG No. of Days from Susceptible to Algo Signal:   1.1991 +/-
0.0272
      AVG No. of Days from Susceptible to Doc Signal:   4.0000 +/-
0.0000

RUN #6:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9844.1701 +/-  1.0238
      Avg no. Infected:  146.8568 +/-  1.0155
      Avg no. At The Hospital:  8.9732 +/-  0.0205

      AVG NO. OF ALGORITHM SIGNALS:  0.9788 +/-  0.0028
      AVG NO. OF DOCTOR SIGNALS:  0.0212 +/-  0.0028

      AVG No. of Days from Susceptible to Algo Signal:   2.9137 +/-
0.0112
      AVG No. of Days from Susceptible to Doc Signal:   4.0000 +/-
0.0000

RUN #7:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  8934.9395 +/-  0.3315
      Avg no. Infected:  241.5711 +/-  0.2994
      Avg no. At The Hospital:  823.4894 +/-  0.1491

      AVG NO. OF ALGORITHM SIGNALS:  0.0014 +/-  0.0007
      AVG NO. OF DOCTOR SIGNALS:  0.9986 +/-  0.0007
```

```
      AVG No. of Days from Susceptible to Algo Signal:   1.8571 +/-
0.3086
      AVG No. of Days from Susceptible to Doc Signal:   4.0097 +/-
0.0019

RUN #8:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  8724.7995 +/-  0.2571
      Avg no. Infected:  452.5076 +/-  0.2214
      Avg no. At The Hospital:  822.6929 +/-  0.1475

      AVG NO. OF ALGORITHM SIGNALS:  0.0003 +/-  0.0003
      AVG NO. OF DOCTOR SIGNALS:  0.9997 +/-  0.0003

      AVG No. of Days from Susceptible to Algo Signal:   1.3333 +/-
1.4342
      AVG No. of Days from Susceptible to Doc Signal:   4.0000 +/-
0.0000

RUN #9:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9721.7984 +/-  1.3892
      Avg no. Infected:  269.3565 +/-  1.3854
      Avg no. At The Hospital:  8.8451 +/-  0.0188

      AVG NO. OF ALGORITHM SIGNALS:  0.9998 +/-  0.0003
      AVG NO. OF DOCTOR SIGNALS:  0.0002 +/-  0.0003

      AVG No. of Days from Susceptible to Algo Signal:   2.8658 +/-
0.0092
      AVG No. of Days from Susceptible to Doc Signal:   4.0000 +/-
0.0000

RUN #10:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9145.1221 +/-  0.3014
      Avg no. Infected:  10.8172 +/-  0.0873
      Avg no. At The Hospital:  844.0607 +/-  0.2334

      AVG NO. OF ALGORITHM SIGNALS:  0.1383 +/-  0.0068
      AVG NO. OF DOCTOR SIGNALS:  0.8617 +/-  0.0068

      AVG No. of Days from Susceptible to Algo Signal:   3.8886 +/-
0.0873
      AVG No. of Days from Susceptible to Doc Signal:   6.4954 +/-
0.0261

RUN #11:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9320.5834 +/-  1.1333
      Avg no. Infected:  245.1088 +/-  1.0725
```

Avg no. At The Hospital: 434.3078 +/- 0.1305

AVG NO. OF ALGORITHM SIGNALS: 0.0488 +/- 0.0042
AVG NO. OF DOCTOR SIGNALS: 0.9512 +/- 0.0042

AVG No. of Days from Susceptible to Algo Signal: 1.5246 +/- 0.0758
AVG No. of Days from Susceptible to Doc Signal: 4.0354 +/- 0.0037

RUN #12:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9151.6558 +/- 0.2248
Avg no. Infected: 7.9654 +/- 0.0523
Avg no. At The Hospital: 840.3788 +/- 0.1921

AVG NO. OF ALGORITHM SIGNALS: 0.0538 +/- 0.0044
AVG NO. OF DOCTOR SIGNALS: 0.9462 +/- 0.0044

AVG No. of Days from Susceptible to Algo Signal: 3.3234 +/- 0.0971
AVG No. of Days from Susceptible to Doc Signal: 5.3208 +/- 0.0171

RUN #13:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9151.1031 +/- 0.3326
Avg no. Infected: 8.8764 +/- 0.0852
Avg no. At The Hospital: 840.0205 +/- 0.2633

AVG NO. OF ALGORITHM SIGNALS: 0.2771 +/- 0.0088
AVG NO. OF DOCTOR SIGNALS: 0.7229 +/- 0.0088

AVG No. of Days from Susceptible to Algo Signal: 3.6536 +/- 0.0570
AVG No. of Days from Susceptible to Doc Signal: 6.1353 +/- 0.0259

RUN #14:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9092.1364 +/- 1.2156
Avg no. Infected: 478.1780 +/- 1.1903
Avg no. At The Hospital: 429.6856 +/- 0.1163

AVG NO. OF ALGORITHM SIGNALS: 0.0139 +/- 0.0023
AVG NO. OF DOCTOR SIGNALS: 0.9861 +/- 0.0023

AVG No. of Days from Susceptible to Algo Signal: 1.0576 +/- 0.0437
AVG No. of Days from Susceptible to Doc Signal: 4.0092 +/- 0.0019

RUN #15:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 8721.8274 +/- 1.3847
Avg no. Infected: 464.5901 +/- 1.3847
Avg no. At The Hospital: 813.5825 +/- 0.1489

AVG NO. OF ALGORITHM SIGNALS: 0.0181 +/- 0.0026
AVG NO. OF DOCTOR SIGNALS: 0.9819 +/- 0.0026

AVG No. of Days from Susceptible to Algo Signal: 1.7403 +/- 0.0641
AVG No. of Days from Susceptible to Doc Signal: 4.0474 +/- 0.0042

RUN #16:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9983.4946 +/- 0.1002
Avg no. Infected: 7.4277 +/- 0.0903
Avg no. At The Hospital: 9.0777 +/- 0.0198

AVG NO. OF ALGORITHM SIGNALS: 0.5377 +/- 0.0098
AVG NO. OF DOCTOR SIGNALS: 0.4623 +/- 0.0098

AVG No. of Days from Susceptible to Algo Signal: 3.6251 +/- 0.0502
AVG No. of Days from Susceptible to Doc Signal: 5.5708 +/- 0.0280

RUN #17
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9981.8387 +/- 0.1007
Avg no. Infected: 8.8777 +/- 0.0898
Avg no. At The Hospital: 9.2835 +/- 0.0199

AVG NO. OF ALGORITHM SIGNALS: 0.4293 +/- 0.0097
AVG NO. OF DOCTOR SIGNALS: 0.5707 +/- 0.0097

AVG No. of Days from Susceptible to Algo Signal: 4.1356 +/- 0.0586
AVG No. of Days from Susceptible to Doc Signal: 5.8402 +/- 0.0272

RUN #18:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9140.2920 +/- 0.2857
Avg no. Infected: 12.5356 +/- 0.0916
Avg no. At The Hospital: 847.1724 +/- 0.2157

AVG NO. OF ALGORITHM SIGNALS: 0.0770 +/- 0.0052
AVG NO. OF DOCTOR SIGNALS: 0.9230 +/- 0.0052

AVG No. of Days from Susceptible to Algo Signal: 4.2740 +/- 0.1333
AVG No. of Days from Susceptible to Doc Signal: 6.9514 +/- 0.0279

RUN #19
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9739.9407 +/- 1.8247
Avg no. Infected: 251.2507 +/- 1.8204
Avg no. At The Hospital: 8.8085 +/- 0.0191

AVG NO. OF ALGORITHM SIGNALS: 1.0000 +/- 0.0000
AVG NO. OF DOCTOR SIGNALS: 0.0000 +/- 0.0000

AVG No. of Days from Susceptible to Algo Signal: 2.7474 +/- 0.0122
AVG No. of Days from Susceptible to Doc Signal: -7.0000 +/- ?

RUN #20:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9771.6956 +/- 2.2940
Avg no. Infected: 219.5363 +/- 2.2893
Avg no. At The Hospital: 8.7681 +/- 0.0193

AVG NO. OF ALGORITHM SIGNALS: 1.0000 +/- 0.0000
AVG NO. OF DOCTOR SIGNALS: 0.0000 +/- 0.0000

AVG No. of Days from Susceptible to Algo Signal: 2.5368 +/- 0.0155
AVG No. of Days from Susceptible to Doc Signal: -7.0000 +/- ?

RUN #21:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9877.8940 +/- 1.2423
Avg no. Infected: 113.2931 +/- 1.2367
Avg no. At The Hospital: 8.8129 +/- 0.0198

AVG NO. OF ALGORITHM SIGNALS: 0.9987 +/- 0.0007
AVG NO. OF DOCTOR SIGNALS: 0.0013 +/- 0.0007

AVG No. of Days from Susceptible to Algo Signal: 2.5387 +/- 0.0158
AVG No. of Days from Susceptible to Doc Signal: 4.0000 +/- 0.0000

RUN #22:
Using 10000 independent replications, 95% CI for following measures as followed:
Avg no. Susceptible: 9152.6061 +/- 0.3130
Avg no. Infected: 8.3418 +/- 0.0775
Avg no. At The Hospital: 839.0521 +/- 0.2517

```
      AVG NO. OF ALGORITHM SIGNALS:  0.2444 +/-  0.0084
      AVG NO. OF DOCTOR SIGNALS:  0.7556 +/-  0.0084

      AVG No. of Days from Susceptible to Algo Signal:   3.4865 +/-
0.0585
      AVG No. of Days from Susceptible to Doc Signal:   5.7992 +/-
0.0228

RUN #23:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9722.4944 +/-  1.3986
      Avg no. Infected:  268.6586 +/-  1.3949
      Avg no. At The Hospital:  8.8470 +/-  0.0188

      AVG NO. OF ALGORITHM SIGNALS:  1.0000 +/-  0.0000
      AVG NO. OF DOCTOR SIGNALS:  0.0000 +/-  0.0000

      AVG No. of Days from Susceptible to Algo Signal:   2.8623 +/-
0.0093
      AVG No. of Days from Susceptible to Doc Signal: -7.0000 +/- ?

RUN #24:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9986.1669 +/-  0.0848
      Avg no. Infected:  4.8968 +/-  0.0745
      Avg no. At The Hospital:  8.9364 +/-  0.0203

      AVG NO. OF ALGORITHM SIGNALS:  0.7421 +/-  0.0086
      AVG NO. OF DOCTOR SIGNALS:  0.2579 +/-  0.0086

      AVG No. of Days from Susceptible to Algo Signal:   3.0838 +/-
0.0377
      AVG No. of Days from Susceptible to Doc Signal:   4.9364 +/-
0.0293

RUN #25:
Using 10000 independent replications, 95% CI for following measures as
followed:
      Avg no. Susceptible:  9322.9228 +/-  0.9899
      Avg no. Infected:  242.4855 +/-  0.9332
      Avg no. At The Hospital:  434.5917 +/-  0.1270

      AVG NO. OF ALGORITHM SIGNALS:  0.0444 +/-  0.0040
      AVG NO. OF DOCTOR SIGNALS:  0.9556 +/-  0.0040

      AVG No. of Days from Susceptible to Algo Signal:   1.5743 +/-
0.0771
      AVG No. of Days from Susceptible to Doc Signal:   4.0078 +/-
0.0018
```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Atlas, R. M. (2002). Bio-terrorism: From thereat to reality. *Annual Review of Microbiology*, 56, 167–185.

Buckeridge, D. L. (2006). Evaluating detection of an inhalational Anthrax outbreak. *Emerging Infectious Diseases*, 12, 1942–1949.

Buss, A., (2010). Discrete event simulation modeling. Retrieved on February 2, 2011, from http://diana.nps.edu/oa3302/Handouts.html

CDC. (2010a). BioSense. Retrieved on February 17, 2011, from http://www.cdc.gov/biosense

CDC. (2010b). Early Aberration Reporting System. Retrieved on January 31, 2011, from http://www.bt.cdc.gov/surveillance/ears

Christopher, G. W., Cieslak, T. J., Pavlin, J. A., & Eitzen, E. M., Jr. (1997). Biological warfare. *The Journal of the American Medical Association*, 278, 412–417.

Fricker, R. D., Jr. (2011) Some Methodological Issues in Biosurveillance, *Statistics in Medicine*, **30**, 403–441.

Fricker, R. D., Jr., (2010). Introduction to statistical methods for biosurveillance (Draft). Retrieved on September 3, 2010, from http://faculty.nps.edu/rdfricke/OA4910.htm#book

Fricker, R. D., Jr., (2009). "Methodological Issues in Biosurveillance," short course presented at the Twelfth Biennial CDC/ATSDR Symposium on Statistical Methods, Decatur, GA, April 2009. Retrieved on November 15, 2010, from http://faculty.nps.edu/rdfricke/Biosurveillance.htm.

Fricker, R. D., Jr., Hegler, B. L., & Dunfee, D. A. (2008). Comparing biosurveillance detection methods: EARS' versus a CUSUM-based methodology. *Statistics in Medicine*, **27**, 3407-3429.

Fricker, R. D., Jr., & Rolka, H. (2006). Protecting against Biological Terrorism: Statistical Issues in Electronic Biosurveillance, *Chance*, **19**, 4–13.

Grey, M. R., & Spaeth, K. R. (2006). *The bioterorism sourcebook*. New York: McGraw-Hill.

Homeland Security. (2007). Homeland Security Presidential Directive 21: public health and medical preparedness. Retrieved on February 20, 2011 from http://www.dhs.gov/xabout/laws/gc_1219263961449.shtm#0

Hupert, N., Bearman, G. M., Mushlin, A. I., & Callahan, M. A. (2003). Accuracy of screening for inhalational anthrax after a bioterrorist attack. *Annals of Internal Medicine*, 139, 337–345.

Office of Technology Assessment. (1993). *Proliferation of Weapons of Mass Destruction.* Washington, DC: US Congress.

Sosin, D. M. (2003). Syndromic surveillance: the case for skillful investment view. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, 1, 247–253.

Zilinskas, R. A. (1997). Iraq's biological weapons. *The Journal of the American Medical Association*, 278, 418–424.

Zubay, G. (2005). *Agents of bioterrorism.* New York: Columbia, 131–169.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor Ron Fricker
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

4. Professor Rachel Silvestrini
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

5. Dr. Andrew Wiedlea
   Defense Threat Reduction Agency
   Ft. Belvoir, Virginia

6. Dr. Krista Hanni
   Monterey County Health Department
   Monterey, California

7. Mr. Matt Holm
   Defense Threat Reduction Agency
   Ft. Belvoir, Virginia