

MIASMA: A Medical Informatics Application for Systematic Microbiological Alerts

By

Randy Joseph Carnevale

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Informatics

December, 2011

Nashville, Tennessee

Approved:

Professor Randolph A. Miller

Professor Dario A. Giuse

Professor Thomas R. Talbot

Professor Cynthia S. Gadd

Professor S. Trent Rosenbloom

Professor William Schaffner

ACKNOWLEDGEMENTS

Financial support was provided by the National Library of Medicine Vanderbilt Biomedical Informatics training grant, T15 LM07450, Dr. Randolph Miller's National Library of Medicine grants, MOMENT (5R01 LM007995-06) and eTfor2 (5R01 LM010828-02), and my CDC Grant for Public Health Dissertation (1R36 CD000986-01). Without the assistance provided by these grants, this work would not have been possible.

My advisor, Dr. Randolph Miller, has been incredibly helpful and supportive through this and the preliminary projects. His expertise and knowledge, along with that of my other committee members, Dr. Thomas Talbot, Dr. Dario Giuse, Dr. Cynthia Gadd, Dr. Trent Rosenbloom, and Dr. William Schaffner, has helped me formulate and hone all the work that has been conducted.

Dr. Dario Giuse and David Staggs of the StarPanel team have provided me with a secure home for most of the software and databases developed through this project and with the necessary computing resources for keeping everything running. Ioana Danciu also provided me with access to the patient transfer data used in this study.

Support from members of the Vanderbilt Medical Center has also been critical in helping me refine MicroParse and to conduct the developmental study. From Infection Control, I owe a debt of thanks to Vicki Brinsko, Janet Szychowski, Lorrie Ingram, Kathie Wilkerson, Amy Dziewior, Tracy Louis, and Brian Nelsen, and from the Informatics Center's VIPER development team, I would like to thank Ted Anders and Tammy Tao. I also need to thank my infection control expert reviewers who spent a great

deal of time and effort looking through dozens and dozens of clusters: Dr. Thomas Talbot, Dr. William Schaffner, Dr. Titus Daniels, and Dr. Karen Bloch.

I would like to thank all the people in the Department of Biomedical Informatics who have helped along the way. Sophia Norella was a tremendous help in preparing and editing the dissertation manuscript, and has provided me with a great deal of emotional support. Rischelle Jenkins has managed so many aspects of my student life, and has been incredibly helpful along the way. Joyce Green has also been immensely helpful, putting in long hours with me during the grant writing process, being available to handle so many last-minute tasks, and assisting with my work in many other ways.

Lastly, I would like to thank my parents, Raymond and Elizabeth Carnevale. They have helped me not only throughout graduate school but through life in general, and their love and support has truly made the results of this work achievable for me.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS.....	ix
Chapter	
I. INTRODUCTION AND LITERATURE REVIEW	1
Introduction and Study Overview	1
Literature Review.....	1
Context and Definitions	1
Computer-assisted healthcare-associated infection monitoring.....	4
Electronic Syndromic Surveillance Systems	7
Background for Current Study.....	9
Statistical Process Control Algorithms: CUSUM and EWMA	10
Syndromic Surveillance Algorithms: Space-time scan statistic and WSARE...	12
II. BACKGROUND: THE MICROPARSE SYSTEM AND PROJECT	15
Overview of MicroParse	15
Introduction: Development and Validation of MicroParse.....	15
Methods: Development and Validation of MicroParse	16
Clinical Setting and Microbiology Data Source	16
Description of MicroParse	17
Parsing the Gram Stain and Culture Sections	20
Validation of MicroParse Data Capture Techniques	22
Extending the MicroParse Database: Non-narrative Results.....	22
Results: Development and Validation of MicroParse.....	23
Discussion: Development and Validation of MicroParse.....	24
Principal Findings of the MicroParse Validation Study	24
Institutional Use of MicroParse	24
MicroParse Post-validation Problems.....	25
Summary of MicroParse Project.....	27

III. BACKGROUND: ALGORITHM SELECTION – MIASMA DEVELOPMENTAL STUDY	28
Algorithm Analysis: Methods.....	29
MIASMA Developmental Algorithm Analysis: Setting.....	29
Algorithm Analysis: Goals of Developmental Study	30
Algorithm Analysis: Developmental Study Overview	30
Algorithm Analysis Phase 1: Algorithm Implementation and Execution	32
Algorithm Analysis Phase 2: Expert Review Process	33
Algorithm Analysis Phase 3: Parameter Tuning, Precision-Recall Analysis, Combined Tool Development, and Retrospective Evaluation.....	34
Algorithm Analysis: Results	36
Algorithm Analysis Phase 1: Algorithm Parameters in Developmental Study .	36
Algorithm Analysis Phase 2.1: Expert Review Results in Developmental Study	36
Algorithm Analysis Phase 2.2: Algorithm Performance during Developmental Study	39
Algorithm Analysis Phase 3.1: Parameter Adjustment Resulting from Developmental Study.....	40
Algorithm Analysis Phase 3.2: Scoring Metrics in Developmental Study.....	42
Algorithm Analysis Phase 3.3: Retrospective Evaluation of Combined Algorithms	44
Algorithm Analysis: Discussion of Developmental Study	44
Algorithm Analysis: Cluster Review for Developmental Study.....	44
Algorithm Analysis: System Performance and Ranking in Developmental Study	47
Algorithm Analysis: Developmental Study Limitations.....	48
Developmental Study Algorithm Analysis: Conclusion.....	50
IV. PROSPECTIVE MIASMA STUDY OVERVIEW	51
Prospective MIASMA Study Setting.....	51
Goals of Prospective MIASMA Evaluation	51
MIASMA System Design.....	52
System Configuration: Prospective Study	52
Prospective Pilot Study of MIASMA at VUMC	56
Data Analysis Methods – Prospective Evaluation.....	57
V. RESULTS OF PROSPECTIVE MIASMA EVALUATION	60
Alert Statistics.....	60
MIASMA Alert Categorization	60
Comparison of MIASMA Output to Manual Methods.....	64
Comparison of MIASMA Output to Public Health Records for Seasonal Illnesses	64
Investigation of Potential Outbreaks During Prospective Study Period Missed by MIASMA	66

Incorporating Patient Bed Location and Transfer Data	69
Expert Assessment of MIASMA System	69
VI. DISCUSSION OF PROSPECTIVE MIASMA EVALUATION.....	71
MIASMA Alert Quality	71
Algorithm Performance During the Prospective MIASMA Study.....	73
Seasonal Detection Performance	74
Adding Patient Bed Location Data to MIASMA.....	74
Expert Assessment of MIASMA’s Performance and Utility.....	75
MIASMA System Portability	76
MIASMA Prospective Study Limitations.....	77
VII.SYNOPSIS AND CONCLUSIONS	79
Summary	79
BIBLIOGRAPHY.....	80

LIST OF TABLES

Table	Page
1. Clusters considered as candidate outbreaks based on expert review	31
2. Percent agreement between reviewers (Cohen's kappa in parentheses)	37
3. Algorithm-detected candidate outbreaks by organism, cluster length in days, and location type.....	38
4. Cluster determination by algorithm	39
5. Algorithm performance stratified by cluster location (PPV in parentheses)	40
6. EWMA false alarms generated in 29 organism/unit combinations	41
7. Cluster detection by algorithm and expert categorization	61
8. Complete listing of detected clusters by affected site(s) and expert categorization during study period (Length given in days).....	62

LIST OF FIGURES

Figure	Page
1. Possible disease transmission mechanisms.....	2
2. Sample plain text microbiology report	17
3. Breaking culture/gram stain sections into component terms	20
4. Creating the MicroParse word trie.....	21
5. Matching text to coded phrases.....	21
6. Summary of methods used in Study Phase 2.....	32
7. Precision-recall measurements for individual algorithms; precision-recall curves for EWMA adjustments and initial scoring metric.....	42
8. Precision-recall curves for adjusted scoring metrics	44
9. Summary of the process followed during daily MIASMA execution	54
10. Main view of MIASMA cluster review webpage.....	55
11. Drill-down view from MIASMA cluster review webpage	56
12. Frequency of daily alert counts during MIASMA's 181-day study period.....	60
13. Weekly alerts generated by MIASMA during study period coded by expert designation	61
14. Comparison of TN State Department of Health and VUMC influenza results, 2010-2011. MIASMA alerts generated during weeks marked with bold border.....	65
15. RSV positive tests at VUMC during study period. MIASMA alerts generated during weeks marked with bold border.....	66
16. Test result data from Clostridium difficile in S44	68

LIST OF ABBREVIATIONS

CUSUM	cumulative sums
EWMA	exponentially weighted moving average
IP	infection preventionist
MIASMA	A Medical Informatics Application for Systematic Microbiological Alerts
MRSA	methicillin-resistant <i>Staphylococcus aureus</i>
STSS	space-time scan statistic
VUH	Vanderbilt University Hospital
VUMC	Vanderbilt University Medical Center
WSARE	What's Strange About Recent Events?

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Introduction and Study Overview

This PhD Dissertation project had as its objectives: (1) to develop MIASMA, a potentially open-source Medical Informatics Application for Systematic Microbiological Alerts that uses recently developed methods (e.g., from syndromic surveillance and from heuristic observations) to detect single-hospital outbreaks of both commonly occurring and rare bacterial species; (2) to deploy MIASMA in the Vanderbilt University Hospital (VUH) for use by the Department of Infection Control and Prevention; (3) to compare the alerting timeliness, positive predictive value, and sensitivity of MIASMA to current VUH infection control practices; and (4) to evaluate the utility of MIASMA when used to supplement current VUH infection control practices.

Literature Review

Context and Definitions

Healthcare-associated infections present a substantial national burden, affecting approximately 2 million U.S. patients annually at a rate of 5.69 infections per 100 hospital admissions.¹ These infections increase patients' lengths of stay and cause potentially preventable morbidity and mortality.² Such infections can spread from patient-to-patient through direct contact (e.g., via skin contact, use of contaminated

instruments, or contaminated injections), via aerosols, or via care providers.³ Outbreaks are defined as “an increase in occurrence of a complication or disease above the background rate.”⁴ As previously noted, outbreaks can occur as the result of several different modes of transmission. If multiple sources within the hospital exist (e.g., inadequate quality control during intravenous line placement or in subsequent line management), different patients may acquire varying infections from independent sources. Alternatively, patients may be exposed to a shared source, such as contaminated injectable radiological contrast dye used during invasive medical imaging.⁵ Furthermore, if the infection can spread from patient-to-patient, a single infected patient can transmit the contagion to others who may then themselves continue the chain of transmission. Figure 1 illustrates these three possible mechanisms by which an infection can spread.

When these transmissions continue to propagate, an outbreak results.

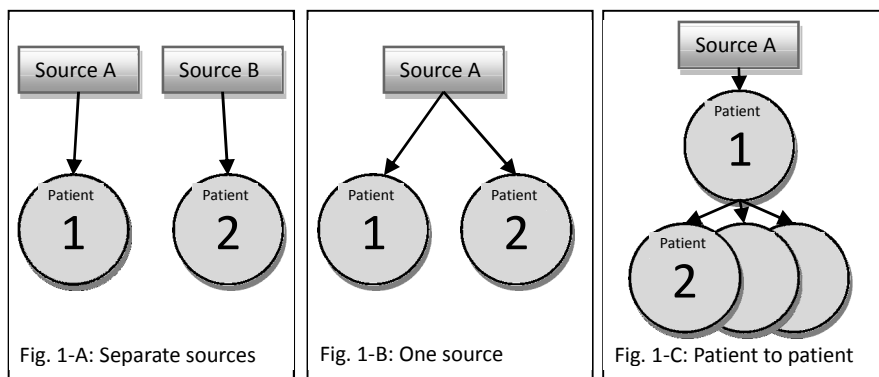


Figure 1: Possible disease transmission mechanisms

Abnormal infection rates that may indicate an outbreak vary widely, since background rates for infectious diseases can differ greatly. A single case of a rare

infection (such as pneumonic plague) may represent an underlying outbreak, whereas dozens of cases of infection with a common pathogen (such as *E. coli* urinary tract infections) may not. Outbreaks fall into two categories: clonal and non-clonal. A clonal outbreak occurs when progeny of a single organism spread to multiple patients. Non-clonal outbreaks typically occur when infection control techniques are suboptimal (e.g., improper hand washing). The resulting infections involve higher than usual rates of many different bacterial species. Non-clonal outbreaks are identifiable by detecting overall increases in infection rates on given hospital units. Clonal outbreaks, however, may remain unnoticed since the increase in infections by a single rarer species may not significantly affect the overall infection rate. Molecular fingerprinting techniques that allow the identification of an organism's genetic lineage based on highly variable sections of the microbial genome remain the gold standard for determining the clonality of bacterial isolates from different patients' cultures of the same species.⁶ Nevertheless, it is both more efficient and more cost effective to first screen for potential clonal outbreaks by comparing antibiotic sensitivity patterns for each bacterial species identified by cultures, since antibiotic sensitivities are routinely ordered for therapeutic guidance. When sensitivity patterns suggest that infected patients share a common pathogen, hospitals then may do more expensive genetic tests to determine clonality.⁷

To investigate and control potential outbreaks, hospital infection control staff confirm existing cases of disease, locate additional previously missed cases, and implement preventive measures to avoid further spread.⁸ Outbreak investigation combines the need to establish clonality of an infecting agent with the need to discern pseudoinfections and pseudooutbreaks from "true" outbreaks. Pseudoinfections are when

microorganisms (such as common skin contaminants) present in a stain or culture do not correspond to a clinical infection. Pseudo-outbreaks comprise a cluster of such pseudoinfections.⁴ Though pseudo-outbreaks are typically not important to investigate from a clinical perspective, they may be indicative of poor sample collection techniques. Thus, they are still important to detect and address.

Computer-assisted healthcare-associated infection monitoring

In the mid 1980's, Evans and colleagues at LDS Hospital in Salt Lake City developed one of the earliest published systems for automated detection of healthcare-associated infections, which also monitored antibiotic resistance patterns – the Computerized Infectious Disease Monitor, CIDM.⁹ To integrate other clinical data with information from the microbiology laboratory system, the Utah developers incorporated CIDM into LDS Hospital's HELP hospital information system. The CIDM ran daily and generated a variety of alerts for infection control personnel. Overall, the CIDM demonstrated the future potential of computer-supported outbreak surveillance: it detected infections as accurately as infection control professionals and made clinically useful antibiotic selection recommendations.¹⁰

In the early 1990's, at Barnes Hospital in St. Louis, Kahn et al. developed GERMWATCHER, a computerized expert system used to detect nosocomial infections.¹¹ The microbiology laboratory system at Barnes Hospital generated reports in a semi-structured format, using constrained, seemingly “natural” language that was generated using a limited terms dictionary. By leveraging the underlying Barnes microbiology terms dictionary, the researchers could use simple pattern matching to extract the

structured, standard-format portions of the report necessary for tracking nosocomial infections.

Following and building upon the work of such early pioneers, other investigators have developed and demonstrated the reliability of more complex systems for tracking healthcare-associated infections.¹² Some systems focused on tracking specific organisms (e.g., MRSA¹³); others on specific infection types (e.g., bloodstream infections¹⁴); and others on more general coverage for nosocomial infections of all types.^{15,16}

Systems with more limited coverage areas for infectious disease surveillance (e.g., detection constrained to a single hospital), have greater ability to use clinical reports to facilitate outbreak detection. For large-region surveillance, e.g., cities or states, the time costs associated with collecting and aggregating reports is prohibitive. At the University of Maryland Medical Center, Wright et al. developed a rule-based system that utilized data from clinical reports to help hospital infection control staff to generate automated “control charts” consisting of a bacterial species, a location, and an optional antibiotic resistance pattern. The control chart alerts notified users when characteristic patterns for specific pathogens were detected (e.g., when specific rare or dangerous organisms were detected, when pathogens had certain patterns of antibiotic resistances, or when a targeted pathogen it was isolated on a certain hospital unit).¹⁷ After configuring the control charts and analyzing retrospective data, the program was able to detect a number of outbreaks that the hospital infection control team had previously missed.¹⁷

Brossette et al. implemented a different outbreak detection approach, using data mining techniques to discover from past institutional culture results novel association rules for surveillance.¹⁸ Brossette’s method did not require “pre-existing” triggers that

manually designated what constituted a clinically important finding. When applying this method to one year of the local hospital's *Pseudomonas aeruginosa* antibiotic sensitivity results, Brossette's system found short-term and long-term shifts in resistance patterns.¹⁸

The World Health Organization (WHO) created WHONET, an algorithmic set of tools to perform retrospective analyses and derive antibiotic susceptibility trends.^{19,20} While the WHONET software package was designed primarily to aid in monitoring and managing antibiotic resistance in collaborative hospital networks, it could also be used within individual hospitals. However, the WHONET version available as of July 2011 does not yet include automated surveillance tools. Previously, a study integrating the SaTScan cluster detection software with WHONET demonstrated that standard, manual outbreak detection techniques missed potential *Shigella* outbreaks that SaTScan's automated techniques caught.²¹ More recently, a study used WHONET and SaTScan to attempt to locate outbreaks within hospitals.²² That study again found that WHONET/SaTScan detected a number of potential outbreaks that manual infection control methods failed to find, and that automated methods could potentially provide helpful guidance to hospital infection control staff.

At Regenstrief Institute in Indianapolis, Kho et al. made use of a city-wide health information exchange network to facilitate sharing data on patients with known positive MRSA cultures.²³ Each individual site maintained a registry of all MRSA-positive patients and shared the registry with the other three participating city hospitals through a standardized interface. After approximately one year, their registry-based system generated 2,698 admission alerts for patients with a known history of MRSA, with 19% of alerts arising from data shared by another institution.²⁴

Some commercial products include infection control applications that can be used for hospital-wide microbiological surveillance. Unfortunately, the commercial nature of such proprietary systems often limits the amount of information that their developers reveal through peer-reviewed publications. For example, Dr. Stanley Pestotnik and colleagues developed the Theradoc Expert System Platform based on the previously noted pioneering infection control research at LDS Hospital.²⁵ Yet technical details on Theradoc's operational algorithms and formal evaluations of that system's efficacy are not publically available. Similarly, technical information about the MedMined suite evolved from Dr. Stephen Brossette's data mining studies is not publically accessible.²⁶

A 2010 study deployed and evaluated DiversiLab, a system designed to aid infection control personnel in determining whether a cluster of bacterial cultures represents an outbreak or not. The Diversilab could provide simple molecular typing data for a number of bacterial species, e.g., from samples taken from patients suspected as victims of an outbreak.²⁷ DiversiLab readily aided in the identification of outbreaks of *Acinetobacter*, *S. maltophilia*, *Enterobacter cloacae* complex, *Klebsiella*, and *E. coli*, but was less useful for *P. aeruginosa*, *E. faecium*, and MRSA. DiversiLab thus provided a quick method for confirming some potential outbreaks found using epidemiologic data, allowing infection control staff to act more rapidly to stop further spread.

Electronic Syndromic Surveillance Systems

Following the 2001 anthrax attacks in the United States²⁸, fears of bioterrorism sparked interest in syndromic surveillance. Syndromic surveillance systems use a variety of algorithmic approaches to rapidly identify potential infectious disease outbreaks within

large geographic areas (e.g., cities, counties, or states). Such systems use pre-clinical data (e.g., sales records for health-related products purchased in pharmacies, or chief complaints of patients seen in Emergency Rooms) gathered from the relevant geographic region. A review of syndromic surveillance system evaluations²⁹ found that most early developers used purely temporal detection methods³⁰⁻³³, though some later incorporated spatial components³⁴⁻³⁶. Algorithms used for detection range from statistical methods developed originally for manufacturing process surveillance to specialized algorithms devised specifically for syndromic surveillance.³⁷ Systems such as Pittsburgh's RODS system³⁶ and Harvard's AEGIS system³⁸ have been implemented for statewide monitoring in Pennsylvania, Utah, and Massachusetts. The RODS system was used short-term for monitoring during the 2002 Salt Lake City Winter Olympics. During that short time interval, it generated two alerts, though fortunately neither corresponded to a public health issue.³⁹ The CDC's BioSense system⁴⁰ can monitor data on a national scale. It has been applied successfully for a variety of purposes, including influenza monitoring⁴¹ and the tracking of health effects of San Diego's 2007 wildfires.⁴²

Most syndromic surveillance systems do not make use of detailed clinical reports, such as culture results, because those data are neither in standard formats nor available rapidly or widely available enough in electronic format to be able to detect the "leading edge" of outbreaks in a large area.⁴³ For example, Eurosentinel, a large manual disease surveillance project conducted in Europe, was only able to make use of clinical reports after a weekly update from participating physicians.^{39,43} Furthermore, an analysis of the utility of syndromic surveillance found that the positive predictive value and specificity were too low when used with clinical data due to the lag time in procuring such data.⁴⁰

Syndromic surveillance methods have been successfully applied to smaller geographical regions as well, with one study demonstrating the ability to detect lower respiratory infection clusters within an individual city while using national retrospective data.⁴⁴

With the maturation of the field of syndromic surveillance, investigators have recently turned to fine-tuning existing system models. More current studies have focused on reducing false alarms,⁴⁵ optimizing the public health response to surveillance alerts,^{46,47} standardizing syndrome definitions,⁴⁸ and determining optimal parameters for frequently used syndromic surveillance algorithms.⁴⁹

Work on specialized syndromic surveillance projects has also expanded. Several recent studies have tried to detect clusters of individual diseases and disease groups (e.g., influenza,⁵⁰⁻⁵⁶ tuberculosis,⁵⁷ and sexually-transmitted illnesses⁵⁸). Others have instead attempted to determine the environmental causes for certain increases in symptoms,⁵⁹⁻⁶³ including high-profile events such as the May 2011 Icelandic volcanic ash cloud.^{64,65}

Since the mid-2000's, syndromic surveillance algorithm development has focused primarily on building on Martin Kulldorff's original space-time scan statistic (STSS) developed originally in 1997.⁶⁶ Kulldorff and others have continued to improve the initial algorithm and make its spatial search capabilities more flexible.^{67,68} Others have simply drawn inspiration from its design and have designed competing algorithms.^{69,70}

Background for Current Study

Past approaches to automated methods of hospital outbreak detection fall into two categories: active and passive surveillance. Active surveillance approaches use decision support algorithms to automatically inform infection control staff of suspicious infectious

disease patterns that require further attention. Passive surveillance approaches provide tools that simply aggregate or display relevant information in a more usable and manipulable electronic format for infection control staff to review upon their own initiative, allowing them to better detect interesting patterns “manually”.

The current study used four algorithms previously applied to regional syndromic surveillance to serve as screening tools for actively detecting potential clonal hospital outbreaks – individually and in combination. Two of these aberrancy detection algorithms originated in manufacturing quality control: cumulative sums (CUSUM) and exponentially weighted moving average (EWMA). The other two came from syndromic surveillance research: space-time scan statistic (STSS) and What’s Strange About Recent Events (WSARE).

Statistical Process Control Algorithms: CUSUM and EWMA

Statistical process control originated in 1931, when Walter Shewhart of Bell Laboratories first described control chart methodologies to monitor manufacturing processes.⁷¹ Statistical process control algorithms use previous data to estimate future values, including the mean and reasonable upper and lower limits. If actual future measurements fall within the predicted limits, the process is “under control.” New measurements that fall outside the calculated control limits may indicate that a noteworthy change has occurred in the underlying process. The simplest statistical process control algorithms set upper and lower limits as a multiple of the previously measured standard deviation and plot each new measurement against these limits. While

this approach provides a method easy enough to plot manually on a graph, it does not effectively detect small shifts in the mean.⁷²

CUSUM, the first algorithm deployed in the current study, is calculated by taking the cumulative summation of the difference between each measured value \bar{x} and the estimated in-control mean $\hat{\mu}_0$ ⁷²:

$$S_m = \sum_{i=1}^m (\bar{x}_i - \hat{\mu}_0)$$

In a process that is under control, each measured value \bar{x} should be reasonably close to the mean (e.g., within 2-3 standard deviations). Thus, a plot of each calculated value of S_m should be centered at zero with small fluctuations up or down. When calculating upper and lower bounds for S_m , methods that increase the bounds over time (“V-mask” methods) have historically provided greater sensitivity to small shifts in the mean and decreased impact from older measurements as compared to traditional control charts.^{73,74}

Another approach to improving Shewhart’s original control charts, the exponentially weighted moving average statistic (EWMA), directly incorporates exponentially decreasing weights applied successively to old values, thus providing a measurement less affected by random noise than CUSUM. EWMA is recursively defined as:

$$EWMA_t = \lambda Y_t + (1 - \lambda)EWMA_{t-1}$$

where $EWMA_0$ is the historical mean, Y_t is the measurement at time t , and λ is the decay rate of past measurements, with $0 < \lambda \leq 1.72$. At $\lambda=1$, the EWMA formula matches the Shewhart control chart formula. Optimal λ values vary depending on the problem domain, but empirically, values between 0.2 and 0.3 have provided good performance in manufacturing.^{72,75}

The typical upper and lower bounds for EWMA are similar to those used in Shewhart's control charts, and are given by $EWMA_0 \pm (k s_{ewma})$ with standard deviation s_{ewma} and factor k depending on the problem domain.⁷⁵ The value of λ affects the variance of the EWMA statistic and thus the limits, as the estimated variance is given by:

$$s^2_{ewma} = \left(\frac{\lambda}{2 - \lambda} \right) s^2$$

where s^2 is the historical variance. Though more difficult to calculate, EWMA charts have the benefit of being more sensitive to small shifts in the mean than Shewhart's control charts while still being easy to interpret graphically.

Syndromic Surveillance Algorithms: Space-time scan statistic and WSARE

Martin Kulldorff first introduced the space-time scan statistic in 1997, and later provided a reference implementation, via the SaTScan system.⁶⁶ At that time, most syndromic surveillance researchers used purely temporal disease cluster detection methods, including the algorithms used in statistical process control.^{29,37} The STSS algorithm incorporates spatial information into its detection as well to attempt to improve detection over a large geographic area. It uses a two-stage process. First, STSS searches

the study area for the circular region most likely to comprise a disease cluster, assuming the spread of the disease follows either a Bernoulli model or a Poisson model. Second, it estimates the statistical significance of the cluster using Monte Carlo simulation. Complete details regarding the STSS algorithm appear in Kulldorff's publications.^{34,66,79}

Many studies have employed STSS with success, including those observing commonly occurring infectious diseases,³⁴ emerging infectious diseases,⁷⁶ and cancer incidence.^{77,78} The STSS approach has played a central role in many regional studies using WHONET,^{21,80,81} and in a recent study, investigators found that SaTScan could successfully bolster standard infection control practices in a hospital setting as well.²²

As STSS addressed the growing need for incorporating spatial data, WSARE addressed the growing need for a cluster detection algorithm that could incorporate multidimensional data (e.g., gender, age, and location in addition to disease status).^{29,30,37} The WSARE approach first constructs a Bayesian network model based on the problem domain's historical data. It then uses the Bayesian network to find the single "best" clustering rule for the given day and estimates a p-value using Benjamini and Hochberg's False Discovery Rate method⁸² to adjust for the multiple hypothesis tests.³⁰ Because the underlying Bayesian model can include a node for each data element, WSARE easily incorporates multidimensional data. For example, if the data include gender, zip code, and influenza diagnoses, WSARE could in theory detect an increase in influenza across the study region, an increase in influenza in women region-wide, or an increase in influenza in one specific zip code. The primary use of WSARE has occurred in conjunction with the RODS public health surveillance system⁸³ -- both for temporary short term monitoring of the 2002 Winter Olympics³⁹ and for long-term public health

surveillance of the state of Pennsylvania.⁸⁴ Complete details of the WSARE algorithm appear in Wong et al.³⁰

CHAPTER II

BACKGROUND: THE MICROPARSE SYSTEM AND PROJECT

Overview of MicroParse

The MicroParse project had as its objectives: (1) to provide Vanderbilt University Hospital (VUH) with computerized tools for monitoring microbiological data; (2) to provide the VUH Infection Control Service with tools to help monitor and track infection-relevant patient-related data such as culture results, hospital location, current orders, and contact precautions status. Prior to the MicroParse study, VUH microbiology data were only available in plain text (i.e., not formally structured) format from the microbiology laboratory system as individual culture reports or as a single patient's lab study results. Clinicians viewed microbiology study results for a particular patient through integration of the microbiology-result-containing laboratory system with VUH's electronic health record system ("StarPanel"). Access was limited to viewing the text of one microbiology test report at a time.

Introduction: Development and Validation of MicroParse

Providing VUH with automated tools for monitoring microbiological data first required an accurate source of microbiological data. As of 2005, VUH used a proprietary microbiology lab system (Triple G[®]) that did not allow access to its underlying database, thus making direct access to the microbiology data impossible. The plain text reports supplied by the microbiology lab system to physicians provided the only easily accessible

method of output. For computerized tools to make use of the plain text reports, however, another tool had to first parse the reports into a coded format. The MicroParse project aimed to provide the parsing functionality that created structured data outputs.

Methods: Development and Validation of MicroParse

Clinical Setting and Microbiology Data Source

Vanderbilt University Hospital at the time of MicroParse development in 2005 was a 650-bed academic medical facility located in Nashville, TN. Its microbiology lab system processed nearly 20,000 unique microbiology culture and test reports per month. The proprietary software underlying Triple G[®] generated microbiology reports only in a human-readable format with variable structure. That made report parsing (by computer algorithms) to identify pathogen names and other characteristics less than straightforward.

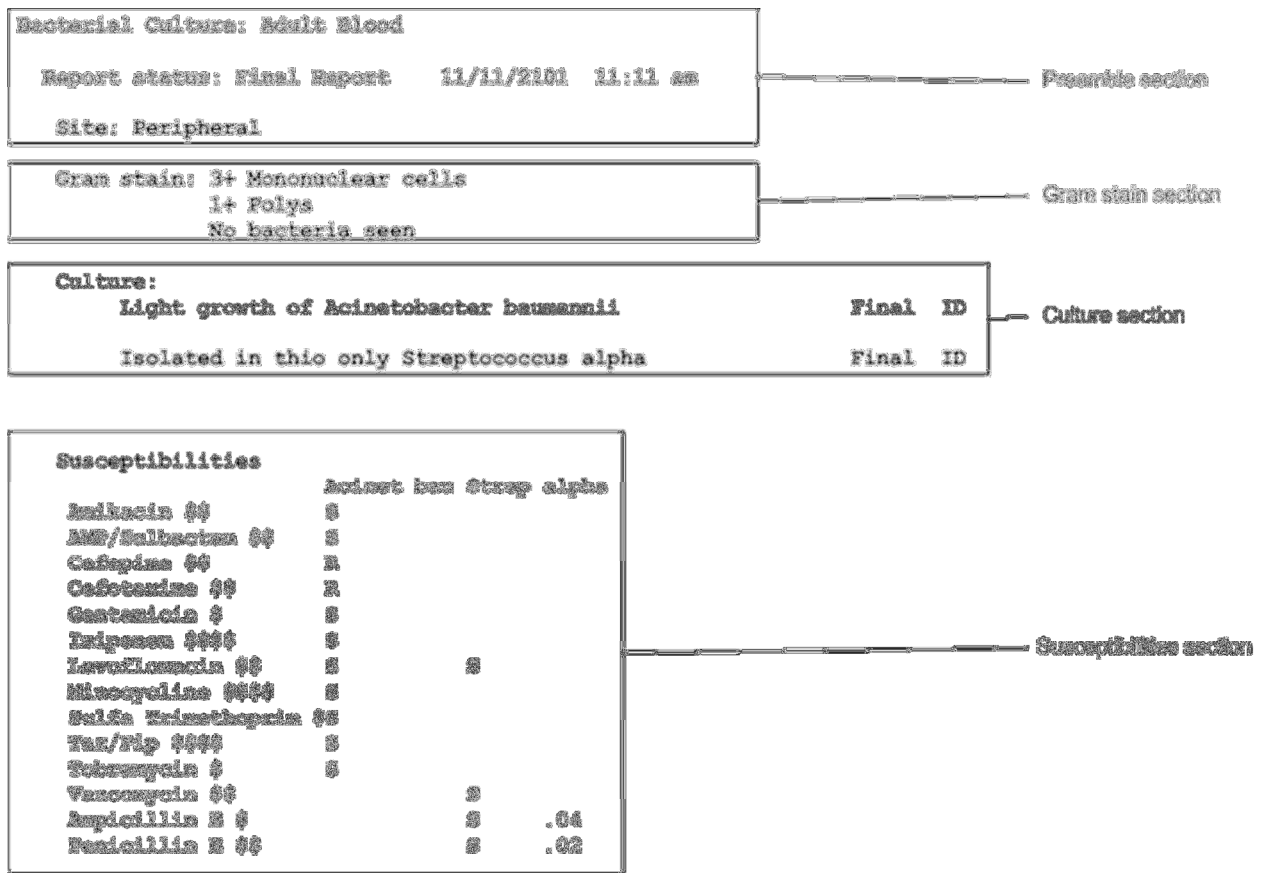


Figure 2: Sample plain text microbiology report

Description of MicroParse

The project created a parsing program, MicroParse, to process the microbiology text reports into usable microorganism-related data. To the present time, for purposes of security, confidentiality, and convenience of data access, MicroParse runs on the protected set of machines within the cluster of servers dedicated to StarPanel, VUH's electronic health record system. The StarPanel team configured their system to feed the plain text microbiology reports to MicroParse every 10 minutes. Then, MicroParse processed each new report, and in turn passed structured data back to StarPanel for storage.

MicroParse first decomposes each plain text report into 4 sections: preamble, Gram stain, culture, and susceptibilities (Figure 2). The preamble contains information about the culture result, including the culture category (e.g., blood, CSF, urine), the report time and date, the report status (i.e., preliminary or final), and the site from which the specimen was taken (e.g., arm wound, bone marrow). Because the preamble tends to follow a fairly specific order with common terms, this information is easily recognized using Perl-compatible regular expressions.⁸⁵ For example, to extract the report's status, MicroParse uses:

```
/Report status:([a-zA-Z ]+)/
```

The Gram stain and culture sections were more difficult to parse since they more closely approximated natural language. However, text from these sections comes primarily from the dictionary of microbiology terms defined within Triple G (“VUH Microbiology Thesaurus”) stored within the microbiology lab system. To generate reports, lab technicians select finding codes based on the results of the test or culture; the lab system then enters a standardized phrase into the report. However, technicians may also include free text in the reports, and some of the phrases share words in common, making recognition of the original coded terms difficult. Fortunately, unlike most other data within the microbiology lab system, the microbiology terms dictionary is externally accessible. This allowed MicroParse to use an externalized copy of the VUH Microbiology Thesaurus as an aid to parsing reports.

The susceptibilities section of the textual VUH Microbiology reports contains a table that, in its top row, indicates an abbreviation for each isolated bacterium, and on

subsequent rows, indicated only by column position, the results of testing each organism's growth in the presence of various antibiotics. The antibiotics for which susceptibilities were tested are named in the first (leftmost) column of the table rows (except the first row). The table columns are generally fixed-width fields, though complications can arise. For example, the abbreviated names often run together in the first row and the abbreviations are occasionally inconsistent (e.g., nonspecific coagulase-negative staphylococcus can appear with any of 6 different column headings). Without consistency in the column names, demarcating the column breaks can be difficult. Also, when the microbiology laboratory provides minimum inhibitory concentrations for a given isolated bacterium, unpredictable changes occur in the column alignments. When multiple bacteria grow from a single culture and their sensitivities are presented as side-by-side columns in a report, it is often the case that not all organisms were tested against all antibiotics, so the absence of testing is indicated by blank fields (extra spaces) within the table columns – further complicating the parsing task.

After processing a given culture report, MicroParse stored the information in a MySQL database, also located within the StarPanel machine cluster. Figure 3 shows an example of the parsed fields for one of the lines in the report shown in Figure 2. In this case, MicroParse stored the codes `IITO` (“isolated in thio only”) and `STRALP` (“streptococcus alpha”), and labeled their identification status as “final.” The database then interfaced with StarPanel to provide information about organisms identified in the report to other programs within VUMC.

Parsing the Gram Stain and Culture Sections

Much of the text found in both the culture and Gram stain sections was drawn directly from the VUH Microbiology Thesaurus. For example, to identify an isolated bacterium in the culture section, it was common to see a term from the categories QUANT (quantity) and FIDORG (final identified organism) (Figure 3).

Line in report:		
Isolated in this only	Streptococcus alpha	Final ID

Terms dictionary:		
CODE	TEXT	CATEGORY

ITTO	Isolated in this only	QUANT

STRALP	Streptococcus alpha	FIDORG

Figure 3: Breaking culture/gram stain sections into component terms

The Thesaurus allowed MicroParse to process the Gram stain and culture results. The Gram stain segment of the report tends to be straightforward, as nearly all lines consist of a STAINQTY term followed a STAINDESC term, making the parsing process simple. The culture section often contains more complex information, however.

To parse the culture section, MicroParse first processes the VUH Microbiology Thesaurus into a word trie.⁸⁶ MicroParse first breaks each phrase in the terms dictionary into its constituent words. Starting with the first word in the phrase, MicroParse then builds the trie from these words, storing the valid partial phrases along with an indication of where completed phrases end (Figure 4).

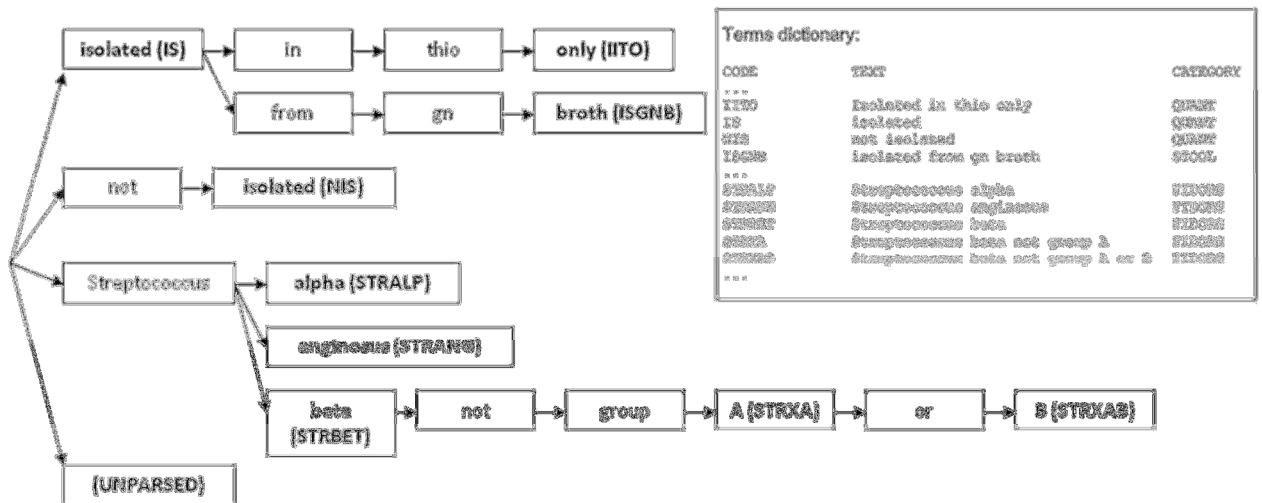


Figure 4: Creating the MicroParse word trie

Once MicroParse has created the trie as a reference source, it is ready to process the text in the culture section. The culture section is split into paragraphs based on line breaks, and for each paragraph, MicroParse searches the trie for all possible matching phrases. It then chooses as an encoding the combination of phrases that maximizes the number of parsed words while minimizing the number of phrases used (see Figure 5). A placeholder entry in the trie, UNPARSED, captures all phrases MicroParse encounters that do not have any valid matches in the dictionary.

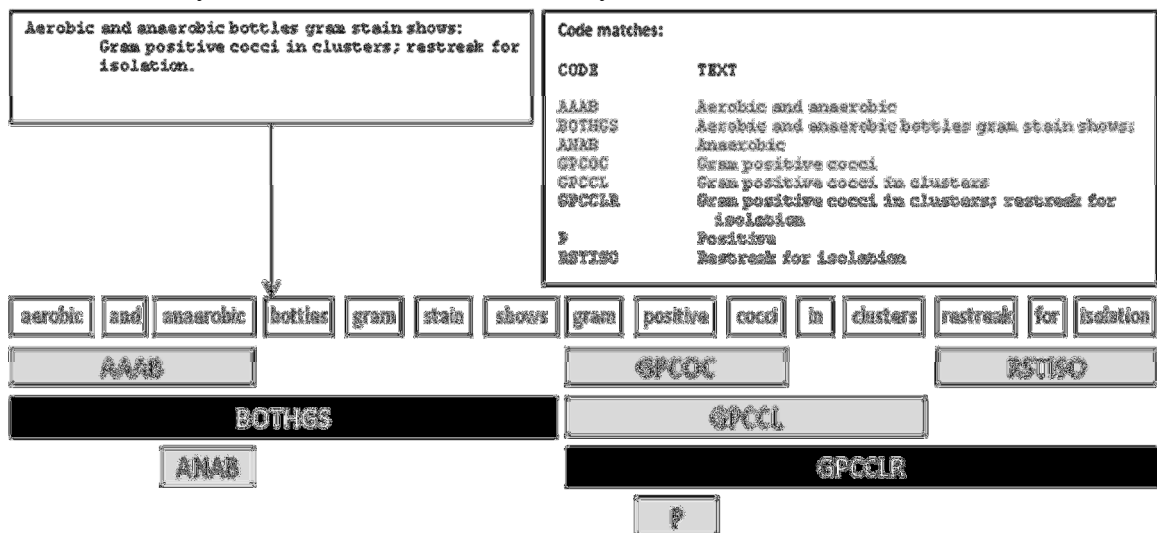


Figure 5: Matching text to coded phrases.

MicroParse finds all possible matches to encoded phrases within the block of text and selects the combination of codes that leaves the fewest possible words unparsed and uses the least number of phrases. In this example, MicroParse matches all words using two codes, BOTHGS and GPCCLR.

Validation of MicroParse Data Capture Techniques

To confirm that MicroParse properly parsed reports and retrieved the clinically relevant information from them, the MicroParse project underwent a validation study. MicroParse retrieved all parsed VUH culture reports from 3 days for analysis: Saturday, January 6, 2007, Monday, January 15, 2007, and Friday, January 19, 2007. Project members also acquired a complete data dump of all microbiology reports issued on those three days directly from the microbiology laboratory system to confirm that MicroParse did not incorrectly alter the text of reports or miss any reports.

Taking this information, a computer script matched the reports to records in the MicroParse database. Dr. Thomas Talbot, an Infectious Disease expert, reviewed all of the matched records from the above-mentioned sample to confirm that the information stored in the MicroParse database accurately reflected all relevant content found in the original microbiology lab report.

Extending the MicroParse Database: Non-narrative Results

As MicroParse gradually garnered increasing use within VUH's clinical practice, feedback from infection control staff indicated that including all available microbiological lab data in the MicroParse database would provide a more complete

picture of hospital infections than just the narrative reports. Clinical project members reviewed a complete list of all test types that the VUH lab system processed and determined which could be relevant for determining patients with bacterial, viral, or fungal infections. These non-narrative results consisted of an HL7 message giving the test name (e.g., a PCR or DFA test result for a specific pathogen), internal test ID code, and a test result. Project members reviewed past results for each of the test types to determine how to identify a positive or negative result for each individual test and constructed Perl-compatible regular expressions⁸⁵ to identify them. Project members then updated the VUH Microbiology Thesaurus to include additional non-“official” codes for positive and negative non-narrative results, allowing users of the MicroParse database to retrieve parsed results from non-narrative reports in the same format as the narrative reports.

Results: Development and Validation of MicroParse

As of July 2011, MicroParse handled approximately 1,050 VUH microbiology reports per day. It is able to process and store reports at a peak rate of approximately 150 reports per second, yielding a theoretical limit of over 10,000,000 reports per day. As of July 18, 2011, the MicroParse database contained 2.6 million reports, using 2.4GB of disk space. During the validation study of MicroParse, the expert reviewer, Dr. Talbot, found parsing errors with 17 reports out of 1,895 reviewed reports (0.9%), with 11 errors being immediately fixable and promptly corrected generically.

Discussion: Development and Validation of MicroParse

Principal Findings of the MicroParse Validation Study

During its validation study, MicroParse performed well. MicroParse processed and stored reports very quickly and scales well even if the number of reports processed by VUH increases significantly. Most of the issues encountered by the expert reviewer took little effort to fix. The problem with linking reports based on free text references within one of the reports presented a more difficult obstacle. It would require more sophisticated parsing techniques to potentially extract the information contained in the references (e.g., “E coli – sensitivities same as most recent previous culture”).

Institutional Use of MicroParse

MicroParse has provided VUH clinicians and informaticians with new opportunities. Early pilot MicroParse application projects included an MRSA/VRE tracking dashboard and automated antibiogram generation capabilities. MicroParse has also provided a number of investigational studies and real-time monitoring systems with microbiology result data that would have previously required manual electronic chart reviews, enabling previously costly and time-consuming activities to be automated.

The VUH Enterprise Data Warehouse also has received data in near real-time from MicroParse. This has further allowed other institutional projects to take advantage of the MicroParse dataset. The largest such project, named VIPER, provides VUH infection preventionists with a summary of all microbiological findings in the hospital

with the option to review more thoroughly data on relevant on individual organisms or patients.

MicroParse Post-validation Problems

Though the initial validation study found few problems with MicroParse's parsing techniques, over time, other issues became apparent. The new issues largely arose from the lab system's dynamic nature. Since the validation study in 2007, the lab system has seen many changes, including a large update from the system's vendor, which required some changes to the parsing process. In addition, VUH lab technicians sporadically add new entries to the VUH Microbiology Thesaurus, modify the names of antibiotics within the system, or modify the organism "short names" used in the Susceptibilities section of the narrative reports. Such changes lead to incorrect parses or incorrect linkages by MicroParse across reports on organism names or antibiotics if MicroParse is not kept up to date.

Furthermore, project members added an organism tagging table to the MicroParse database mapping VUH Microbiology Thesaurus terms to bacterial, viral, and fungal species. This table allowed users to search for positive or negative reports for specific organisms. However, occasional free text reports or reports from external labs imported into the VUH lab system can cause MicroParse to wrongly declare a positive finding when there is none. For example, without adding additional entries to the Microbiology Thesaurus, MicroParse would process the phrase "negative for *Bordetella pertussis*" as the codes NEGATIVE, UNPARSED, (for) and BORDPE (*Bordetella pertussis*) and a later search for positive reports for *B. pertussis* would incorrectly retrieve this report. This

problem could be addressed with more sophisticated natural language processing (NLP) techniques. However, as it is a relatively rare issue, adding NLP could potentially cause new unforeseen problems.

Lastly, though adding the previously mentioned non-narrative tests has benefited MicroParse's users, the list of tests, the associated positive/negative regular expressions, and the dummy Thesaurus entries all require updating over time as new tests become available to clinicians to keep MicroParse from missing relevant organism data. This creates a minor technical issue since a clinical user set to update the table would struggle with constructing the regular expressions and a technical user would struggle with identifying the new tests that require inclusion. It also creates an organizational issue in maintaining the linkage to the StarPanel servers to ensure that the correct tests are reaching MicroParse. Because of the mechanism used to send reports from StarPanel to MicroParse, updating the list of tests currently requires a change to a core StarPanel Perl module, necessitating coordination with the StarPanel team.

Despite these issues, however, MicroParse provides a valuable data resource to VUH. Maintenance and operation of MicroParse involves low personnel and computing resource requirements. Eventually, VUH plans to adopt a new lab management system that allows encoded export of results, thus potentially making MicroParse obsolete. Nevertheless, given the cost and labor requirements necessary to make such an upgrade, and the likelihood that at least some (e.g., outside laboratory) reports will be unstructured as free text entered by a laboratory technician, MicroParse will continue to have utility for the foreseeable future.

Summary of MicroParse Project

The MicroParse project provided VUH clinicians and staff with new access to microbiological data that has been since used to improve patient care. The MicroParse tool allows its users to search the microbiology results database flexibly, facilitating a number of approaches to monitoring microbiological data.

CHAPTER III

BACKGROUND: ALGORITHM SELECTION – MIASMA DEVELOPMENTAL STUDY

The MIASMA developmental study evaluated the ability of four aberrancy detection algorithms to function as a screening tool for recognizing potentially clonal microbiological outbreaks including identification by non-culture results. The goals were to do so earlier and more effectively than previous manual methods. Successful MIASMA automated alerting might allow infection control staff to intervene sooner, control outbreaks earlier, and potentially prevent further transmission. This developmental study also determined if the targeted automated surveillance methods could achieve better performance than manual surveillance methods on specific, narrow, objective measures, recognizing that in other situations, manual methods might be more effective. By determining which outbreak characteristics each automated method most easily and accurately detected, the study provided insight into how to apply future automated systems in actual practice settings.

Portions of the following text sections have been adapted from “Evaluating the utility of syndromic surveillance algorithms for screening to detect potentially clonal hospital infection outbreaks” by Carnevale, Talbot, Schaffner, Bloch, Daniels, and Miller, *JAMIA* 18(4), July 2011.⁸⁷

Algorithm Analysis: Methods

MIASMA Developmental Algorithm Analysis: Setting

The developmental study was conducted at Vanderbilt University using data from the Vanderbilt University Hospital (VUH), The Vanderbilt Clinic, and the Monroe Carell Jr. Children's Hospital at Vanderbilt. The Vanderbilt Institutional Review Board approved the study design prior to conducting the study.

Data for this study was obtained using MicroParse, a computer program previously described in Chapter II that receives all microbiology reports from the Vanderbilt Medical Center microbiology laboratory via the electronic medical record (EMR) system (see Chapter II). MicroParse analyzes and encodes the "free text" natural language culture result output from the microbiology reports and stores the coded results corresponding to organism results and antibiotic sensitivities in a database. The current study extracted "positive" microbiological cultures (preliminary, final, and final with antibiotic sensitivities) and related tests (e.g., PCR and DFA tests) from MicroParse. "Positive" tests that indicated the presence of a potential pathogen were then forwarded for analysis in the current project using the cluster detection algorithms. Similar MicroParse output data were available to VUH Infection Control personnel during their real-time outbreak detection efforts.

Algorithm Analysis: Goals of Developmental Study

The MIASMA developmental study measured the performance characteristics of four existing aberrancy detection algorithms using individual “positive” microbial identification instances at VUH. A secondary goal was to determine how one might later combine the algorithms into a single, optimal, future detection system. Thus, the developmental study required accurate measurement of each algorithm’s performance metrics individually and in various combinations.

Algorithm Analysis: Developmental Study Overview

The developmental study evaluated four algorithms, including two custom implementations (CUSUM⁷² and EWMA⁷²) and two reference implementations (WSARE³⁰ and Kulldorff’s space-time scan statistic⁶⁶; SaTScan). The de-identified dataset included daily case counts for each organism taken from all microbiologic culture data collected from 2001 through 2006 from inpatient units, outpatient clinics, and emergency rooms. It included only the first result of a given culture type (i.e., organism and sensitivity pattern) for each patient on each unit to avoid giving extra weight to multiple serial cultures of the same organism from the same patient.

The developmental study comprised three phases. Phase 1 implemented the four aberrancy detection algorithms using the hospital-derived retrospective microbiologic culture data, producing a list of potential past outbreak clusters. For review purposes, the PI also developed a web-based tool for displaying the relevant microbiological data for each cluster for Infection Control experts’ review.

In developmental study Phase 2, four Vanderbilt University School of Medicine Infectious Diseases faculty members, who were blinded to algorithm source, reviewed the algorithm-generated suspected clusters and categorized them as probable, possible, or non-outbreaks. This was accomplished using the web-based tool, and with it, experts also assessed whether or not the cluster would have merited investigation had they been aware of the cluster at the time. The developmental study labeled expert-reviewed clusters deemed to be probable outbreaks or possible outbreaks that merited further investigation as “candidate outbreaks.” Conversely, possible outbreaks or non-outbreaks not meriting investigation were called false positives. The study excluded clusters that experts labeled non-outbreaks that merited investigation (e.g., a cluster suggesting poor sample collection techniques) from analysis to avoid unfairly penalizing or rewarding individual algorithms. Table 1 summarizes the cluster classification process. The list of reviewed clusters also included three confirmed outbreaks that had been independently identified previously by the hospital’s infection control staff. Figure 6 summarizes the process followed for this phase.

Table 1: Clusters considered as candidate outbreaks based on expert review

	Probable outbreak	Possible outbreak	Non-outbreak
Would investigate	Candidate	Candidate	<i>Exclude</i>
No investigation necessary	Candidate	False positive	False positive

In developmental study Phase 3, project members empirically used the Phase 2 results as feedback to adjust configuration parameters associated with each algorithm and investigated additional methods for combining the algorithms’ output into a single

outbreak detection screening tool. The investigators then carried out a 6-month retrospective evaluation of the new system.

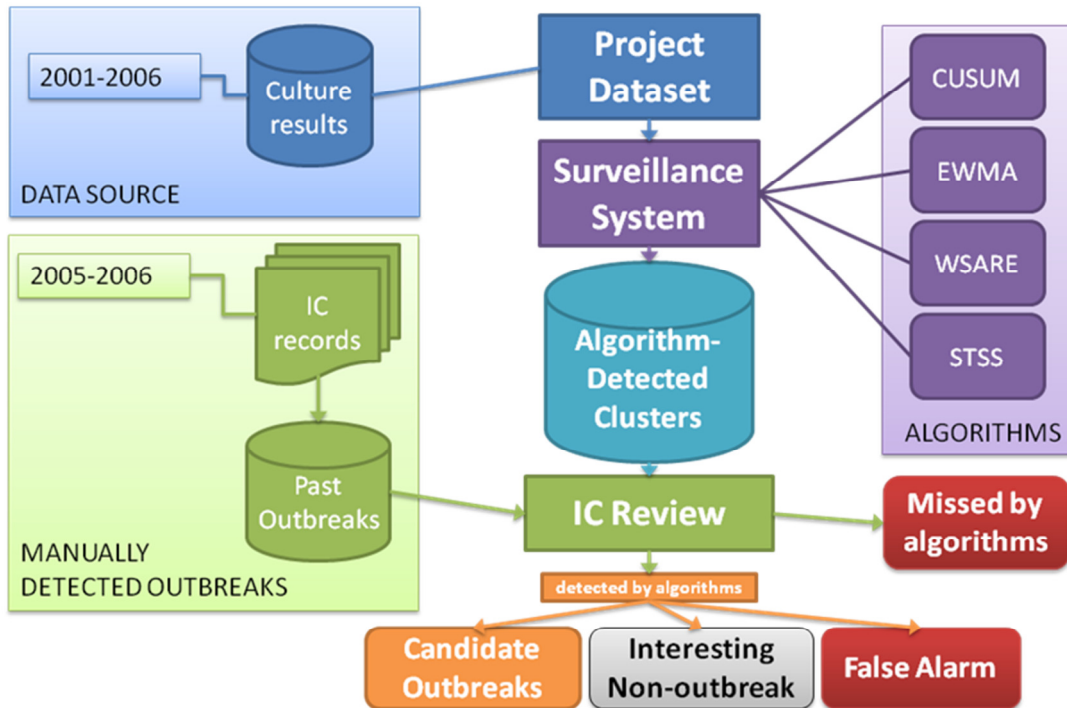


Figure 6: Summary of methods used in Study Phase 2

Algorithm Analysis Phase 1: Algorithm Implementation and Execution

The developmental study configured each algorithm to identify clusters of positive cultures from daily case-culture counts for each organism – both for individual hospital units and across the entire institution. The study divided the culture dataset into three parts. The first set (1 year; 1/1/2001-12/31/2001) provided historical “seed” data for each algorithm. The second set (3 years; 1/1/2002-12/31/2004) served as a testing set for tuning the parameters of each algorithm and designing the review module before study initiation. This second set also provided additional historical baseline data for the final review. The third set (2 years; 1/1/2005-12/31/2006) provided the testing data for the

study Phase 2 expert review. The study converted output from each of the four study algorithms into a common format to prevent the reviewers from identifying which algorithm had generated a given cluster.

Algorithm Analysis Phase 2: Expert Review Process

The developmental study assigned two of the four expert reviewers to examine each algorithm-identified potential cluster independently. Discordant assessments were resolved by submitting each to a “tiebreaker” reviewer randomly selected from the two reviewers who had not previously evaluated the cluster. To calibrate the reliability of the tiebreaking opinions, the study also presented the tiebreak reviewers with several randomly chosen clusters on which the first two reviewers’ determinations agreed (either as “candidates” or not).

The developmental study supplemented the list of candidate outbreaks identified by the review process (as defined above) with three infection control-investigated clusters (IC clusters) that had been independently characterized previously by the hospital’s infection control staff. These three consisted of disease clusters subjected to genetic or serologic testing during the study time period.

Following the clinicians’ reviews, the study calculated the sensitivity and positive predictive value (recall and precision) for each cluster identification algorithm based on the “consensus” classifications (by two or three reviewers, per protocol) of suspected outbreaks and IC clusters. The study compared the individual algorithms’ performance statistics pairwise using McNemar’s test.

Algorithm Analysis Phase 3: Parameter Tuning, Precision-Recall Analysis, Combined Tool Development, and Retrospective Evaluation

In developmental study Phase 3, the project empirically analyzed the effects of varying algorithm parameters on each algorithm's ability to identify Phase 2 expert-labeled candidate outbreaks. The study also explored potential methods of combining the individual algorithms with additional heuristic data to produce better candidate outbreak identification than obtained by the individual algorithms per se.

A first approach was to adjust parameters for whichever customizable algorithm that demonstrated better performance in Phase 2 (CUSUM or EWMA) to detect as many of the candidate outbreaks as possible. For each of the expert-identified candidate outbreak clusters, the study calculated k , the minimum threshold at which the chosen algorithm would generate an alert for the outbreak, using varying decay rates λ (0.05, 0.07, 0.1, 0.15, 0.2, 0.25, and 0.3). Project members recorded the number of additional alerts that would also have triggered at the given value of k . Based on these measurements, the study determined the optimal value of λ and generated precision-recall curves for varying values of k when using the optimized algorithm.

The developmental study also explored methods of combining the output from the four original algorithms using various scoring metrics by which the resulting clusters could be ranked. A first step attempted to order the clusters by their previously measured value of k . Project members then made additional adjustments to the rank weights regarding several features identified as potentially important by the expert Infectious Disease faculty reviewers during the Phase 2 review, including hospital location type (inpatient vs. outpatient) and primary culture source type (urine, blood, wound, etc.).

The developmental study also examined the potential for not “alerting” for clusters comprised of organisms with substantially different antibiotic susceptibilities. This approach had the potential to eliminate noise due to clusters comprised of different clones from the same bacterial species. This analysis focused on clusters for which sensitivity results were available for at least 50% of their component cultures. Project members developed an algorithm that calculated for each cluster an antibiotic susceptibility variability score by summing the number of individual antibiotic sensitivity result pairwise differences within the cluster and weighting the overall result by the number of cultures within the cluster having each of the compared patterns. The resulting score thus represented the average number of differing antibiotic sensitivities between each pair of bacterial isolates. This filtering method, applied to the output of the individual screening algorithms, allowed the analysis to exclude clusters not meeting empirically derived uniformity limits (i.e., those that appeared to be non-clonal based on sufficiently varied culture sensitivities) while still allowing the system to detect potentially clonal clusters that had mutated only slightly in their antibiotic resistance over the course of the outbreak. A final best-case heuristic combination of all of these new methods comprised the Phase 3 combined detection system.

With the above-described adjustments in place, Phase 3 of the developmental study concluded by conducting a brief retrospective validation of the combined outbreak detection system’s recall. The system was run using new data from 1/1/2010-6/30/2010 and the resulting clusters were compared to the list of confirmed outbreaks that had been previously discovered by hospital infection control staff using manual methods for that time period.

Algorithm Analysis: Results

Algorithm Analysis Phase 1: Algorithm Parameters in Developmental Study

Using the first and second datasets, the MIASMA developmental study empirically adjusted the parameters for each algorithm. For EWMA, the project team set a decay rate $\lambda = 0.3$ and an alerting threshold $k = 5$. For CUSUM, the project team used a V-mask for determining the alerting threshold with a daily rise of 3 times the standard deviation of the CUSUM statistic for each particular organism. SaTScan was executed using its purely temporal Poisson model, and WSARE was executed using Fisher's exact scoring metric with 100 randomizations for each day.

Algorithm Analysis Phase 2.1: Expert Review Results in Developmental Study

For institution-wide microbial data covering the two-year developmental study period, the four outbreak detection algorithms collectively generated a total of 257 alerts (CUSUM: 114, EWMA: 66, SaTScan: 21, WSARE: 56). To present alerts more efficiently to clinical expert reviewers, the study combined any computer-generated alerts with start and stop dates differing by fewer than two days into one single alert. As a result, six alerts detected by two algorithms and one alert detected by three algorithms were combined to form the final review list of 249 cluster alerts.

Table 2: Percent agreement between reviewers (Cohen's kappa in parentheses)

Reviewer A	Reviewer B	% Agreement
1	2	86% (0.22)
1	3	81% (0.47)
1	4	88% (0.48)
2	3	85% (0.49)
2	4	88% (0.38)
3	4	79% (0.11)

Pairwise percent agreement on the expert-reviewed clusters ranged from 79% to 88% with Cohen's kappa ranging from 0.11 to 0.49 (Table 2). Overall, reviewers agreed on their determinations for 210 of the 249 alerts, with 17 (8.1%) deemed candidate outbreaks.

For the 39 clusters on which the pair of initial reviewer assessments disagreed, the study assigned a randomly selected third reviewer. Of the 39, the third reviewer deemed nine (23%) to be candidate outbreaks. In addition, for calibration determination, the third reviewer also rated six randomly selected candidate outbreaks (where the two initial reviewers agreed the cluster was a potential outbreak) and six randomly selected false alarms (where the reviewers had agreed the cluster was not an outbreak). The third reviewer agreed with the first two reviewers on all six of the false alarms. However, for the six pairwise-agreed-upon candidate outbreaks, the third expert reviewer only agreed with the initial experts' judgment once (17%).

Table 3: Algorithm-detected candidate outbreaks by organism, cluster length in days, and location type

Organism	Days	Location
<i>Acinetobacter baumannii</i>	1	inpatient
<i>Shigella sonnei</i>	1	outpatient
<i>Enterobacter sakazakii</i>	1	housewide
<i>Acinetobacter species</i>	1	inpatient
<i>Acinetobacter species</i>	1	outpatient
<i>Acinetobacter baumannii</i>	1	inpatient
Enterovirus	3	housewide
<i>Staphylococcus aureus</i>	3	outpatient
Enterovirus	4	housewide
<i>Moraxella catarrhalis</i>	5	inpatient
<i>Clostridium perfringens</i>	6	housewide
<i>Mycobacterium avium complex</i>	6	housewide
<i>Acinetobacter baumannii</i>	7	inpatient
<i>Shigella sonnei</i>	9	outpatient
<i>Salmonella serotype mbandaka</i>	13	housewide
<i>Proteus mirabilis</i>	17	outpatient
<i>Plesiomonas shigelloides</i>	18	housewide
RSV	25	housewide
<i>Shigella sonnei</i>	30	housewide
<i>Mycobacterium avium complex</i>	61	housewide
Diphtheroids species	63	housewide
Enterovirus	64	housewide
<i>Candida albicans</i>	67	inpatient
<i>Shigella sonnei</i>	88	housewide
Diphtheroids species	89	housewide
RSV	140	housewide

The hospital infection control service had previously identified five suspected outbreak clusters during the developmental study period. Those clusters were not detected by any of the algorithms as originally configured for the Phase 1 study. Of the five, 2 have been excluded methodologically from consideration by the study analysis. In one, the lab assay for the involved organism, *C. difficile*, was not included in the study data input since the dataset only included organisms identified by microbiological culturing and thus *C. difficile* antigen test results could not be processed by the detection

algorithms. In the other, the outbreak spanned several months and began prior to the beginning of the study period. The developmental study “gold standard” outbreak dataset therefore contained 29 candidate outbreaks: 17 from the initial expert consensus review, 9 from the second expert conflict-resolving review, and 3 from the infection control archival data. Table 3 shows the 26 candidate outbreaks detected by the algorithms.

Algorithm Analysis Phase 2.2: Algorithm Performance during Developmental Study

For the four evaluated algorithms, positive predictive value relative to the expert-determined gold standard ranged from 5.3% to 29%, with sensitivities ranging from 0.21 to 0.31. Table 4 shows individual results for each algorithm. The PI performed pairwise comparisons of each algorithm’s performance using McNemar’s exact test. The differences in sensitivity were not sufficient to reject the null hypothesis that the algorithms had identical performance. For positive predictive value, CUSUM was significantly lower than all other algorithms ($p < 0.001$ in all comparisons), and EWMA and WSARE were significantly lower than SaTScan ($p < 0.001$ for each).

Table 4: Cluster determination by algorithm

	Candidate	Non-candidate	PPV	Sensitivity
CUSUM	6	108	5.3%	0.21
EWMA	9	57	14%	0.31
SaTScan	6	15	29%	0.21
WSARE	7	49	13%	0.24

Stratifying the analysis by location type (hospital-wide clusters and inpatient units as inpatient; clinics and emergency rooms as outpatient) demonstrated that clusters from

inpatient locations were much more likely to be considered candidate outbreaks than clusters from outpatient locations (inpatient: 21/120 clusters vs. outpatient: 5/129 clusters; chi-square $p=0.002$). Table 5 shows results and positive predictive values for each algorithm stratified by inpatient/outpatient location.

Table 5: Algorithm performance stratified by cluster location (PPV in parentheses)

	Outpatient	Inpatient
CUSUM	0/55 (0%)	6/59 (10.2%)
EWMA	3/30 (10%)	6/36 (16.7%)
SaTScan	0/0	6/21 (28.6%)
WSARE	2/46 (4.3%)	5/10 (50%)

Algorithm Analysis Phase 3.1: Parameter Adjustment Resulting from Developmental Study

As EWMA yielded both better PPV and sensitivity than CUSUM, the developmental study team adjusted EWMA's parameters in Phase 3. Testing various possible decay rates for EWMA parameters indicated that higher decay rates tended to decrease false positive rates as shown in Table 6, as was suggested by past research in the domain.⁴⁹ For the organism/unit combinations given in the 29 system-detectable candidate outbreaks, EWMA's λ parameter set to 0.05 yielded 32 potential false alarms vs. 8 false alarms at $\lambda = 0.3$. Using $\lambda = 0.3$ and $k=5$, EWMA detected 24 of the 29 candidate outbreaks including the 3 infection-control-confirmed outbreaks, but with 629 false positive alerts. Upon increasing the EWMA k value to 6, the system detected 23 of 29 gold standard candidate outbreaks, with 467 false alarms, but excluded one of the

confirmed outbreaks. Recall-precision analysis yielded an area under the curve (AUC) of 0.127. Figure 7 shows the recall-precision curve for the adjusted EWMA algorithm along with the initial performance of the four unadjusted algorithms.

Table 6: EWMA false alarms generated in 29 organism/unit combinations

λ	False positive alerts
0.05	32
0.07	24
0.1	21
0.15	13
0.2	9
0.25	9
0.3	8

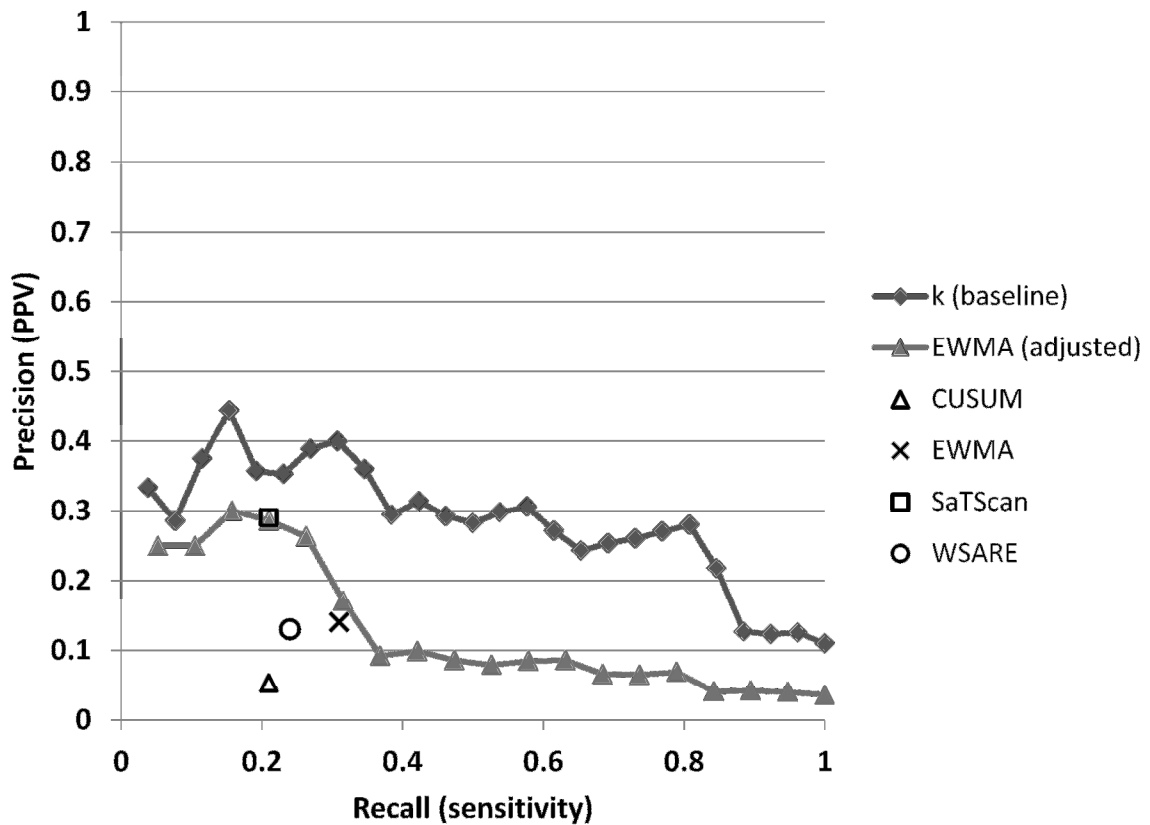


Figure 7: Precision-recall measurements for individual algorithms; precision-recall curves for EWMA adjustments and initial scoring metric

Algorithm Analysis Phase 3.2: Scoring Metrics in Developmental Study

Using the minimum alerting threshold k as the initial ranking metric to sort the original list of 249 clusters generated by the four algorithms yielded an AUC of 0.28. Figure 7 shows the recall-precision curve for this initial metric, with the curve for the adjusted EWMA and points for each of the individual algorithms.

To investigate whether primary culture specimen type (e.g., blood, urine, wound, etc.) could help to separate clinically significant clusters from less important ones, project members developed an algorithm that labeled each cluster by specimen type if more than 50% of the cultures in a given cluster shared a common source. A chi-square test compared that specimen type to all other cultures independent of source type. The only statistically significant relationship this analysis identified was that urine cultures were less reliable indicators of clusters than other culture sites (2.0% of urine vs. 13% non-urine; $p=0.03$). After adjusting the ranking metric downward for clusters of urine cultures, the k-sorted precision-recall AUC improved from 0.28 to 0.36. As observed in Phase 2, clusters in inpatient locations were more likely to produce candidate outbreaks than clusters in outpatient units. After increasing the ranking metric for inpatient clusters, the AUC rose from 0.36 to 0.49.

Project members calculated antibiotic susceptibility difference scores for the 165 clusters that met the 50% criterion, including 6 of the 19 candidate outbreaks. Antibiotic susceptibility difference scores ranged from 0 to 138 in the false alarm clusters and from 0 to 2.7 in the candidate outbreaks. Based on these results, project members generated new precision-recall curves after eliminating all clusters with similarity scores greater than a conservative threshold of 5 and an aggressive threshold of 3. These adjustments increased the precision-recall AUC from 0.49 to 0.53 for the conservative threshold and to 0.55 for the aggressive threshold. Precision-recall curves for each of these adjustments are shown in Figure 8.

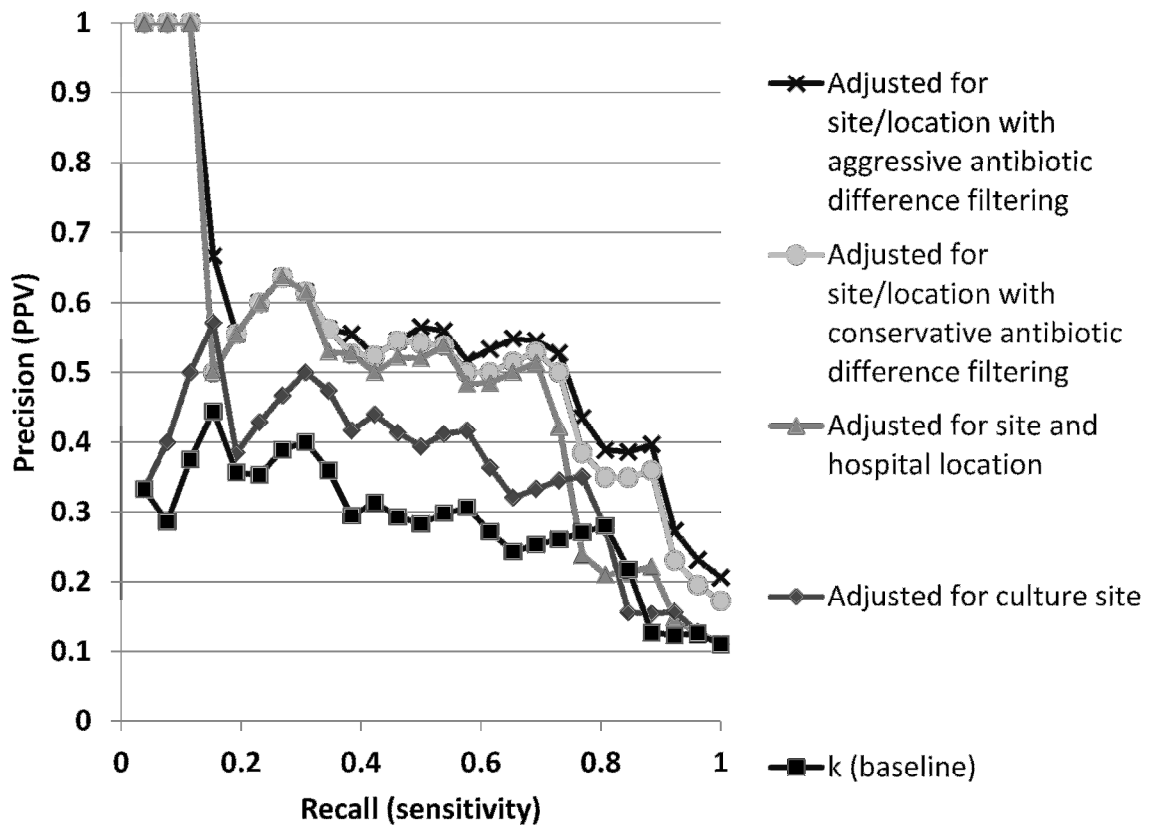


Figure 8: Precision-recall curves for adjusted scoring metrics

Algorithm Analysis: Discussion

The exploratory MIASMA developmental study attempted to determine whether one or more aberrancy detection algorithms might be adapted to screening for potentially clonal hospital outbreak detection. Because each algorithm produced a list of “interesting” suspect clusters substantially different from the others, an ideal system in this setting would consist of multiple algorithms working together.

Algorithm Analysis: Cluster Review for Developmental Study

Few candidate outbreaks were detected by more than one algorithm. In addition, each candidate algorithm varied greatly in the nature of the detected clusters. To some

extent, the initial algorithm tuning parameters that were used in the developmental study may have pre-determined these results. Since the design of each algorithm focused on more immediate detection of outbreaks for syndromic surveillance, WSARE and SaTScan favored very short-term clusters – most candidate outbreaks for which they generated alerts lasted 1-3 days. By contrast, EWMA included some short-term clusters, but alerted on many more clusters with durations in the 1-2 week range. Finally, CUSUM showed the greatest variation, detecting a few short-duration clusters, but also some quite long ones that spanned multiple months. These lengthy “clusters” tended to be uninteresting to the reviewers, since they did not seem to represent an outbreak with a single source.

Analysis of the expert review process demonstrated the degree of subjectivity in determining which clusters were potentially interesting. The first round of reviews only managed moderate levels of inter-rater agreement as shown in Table 2. Because the overall prevalence of true positive clusters was relatively low, measured values of Cohen’s kappa were low despite a high percentage of agreement between reviewers. The low kappa suggests that despite having similar training and using similar review criteria, the expert reviewers disagreed fairly often, and that constructing a true gold standard is not possible. In the second round “tiebreaker” reviews, the third reviewer only agreed with the initial reviews on 17% of the “seed” candidate outbreaks. By contrast, when the third reviewer examined clusters for which one of the two had designated it as a candidate cluster and the other original reviewer did not, the third reviewer designated the cluster as a candidate 23% of the time.

The low reviewer agreement suggests that an ideal hospital outbreak detection screening tool should favor sensitivity over positive predictive value since experts may disagree on which clusters merit further investigation. This strategy is further supported by standard infection control practice: in a prospective study, further investigation including molecular typing to confirm clonality would have followed for each of the potentially interesting clusters. Because such investigation will easily distinguish true positives from false positives, it is more important that the detection system acts as a “screening test” that does not produce many false negatives.

After the developmental study review process, the infection control experts suggested a number of ways to improve potential MIASMA detection and determination processes. First, a few simple rules could significantly reduce the false positive rate. For example, certain culture types (e.g., urine), certain organisms (e.g., coagulase-negative *Staphylococcus*), and certain hospital units (e.g., emergency department) do not often correspond to outbreaks, and thus the system should require a higher outbreak alerting threshold in such situations. Alternatively, rare and dangerous organisms (e.g., *Bacillus anthracis*) should trigger an alert for a single case. More granular patient unit groupings would also be helpful. In some of the hospital-wide clusters, the reviewers suspected there could be something of interest occurring within a subset of the involved hospital units, such as two or three geographically adjacent units, or two units that frequently exchanged patients (e.g., the general surgery floor and the surgical ICU). Lastly, the reviewers noted that some of the identified clusters might productively serve purposes other than outbreak detection. For example, organisms suggesting foodborne illness or other outpatient issues (e.g., RSV) might not be worth investigating as a hospital-

associated outbreak, but informing outpatient clinics of the rising prevalence in the general case mix might allow the clinics to more accurately diagnose or rapidly treat additional patients with similar symptoms.

Algorithm Analysis: System Performance and Ranking in Developmental Study

The lack of consensus regarding alerts generated by the four algorithms, and the excessive false positive rate for the parameter-adjusted EWMA system suggested that none of the four algorithms evaluated could solely provide a reliable alerting mechanism. Thus, to create a functionally useful MIASMA alerting system for hospital infection control purposes, some algorithmic combination technique that leveraged the relative strengths of each individual algorithm would likely provide the best overall system. In addition, based on the expert raters' debriefing comments, some heuristic rules not present in any of the systems might beneficially impact the combined MIASMA system.

Prior to the developmental study's data analysis, the expert reviewers stated that performance goals for a useful generic outbreak detection system that infection control services might use in clinical practice. They stated that such a system would require at least 50% positive predictive value at 0.9 sensitivity, and at least 0.25 sensitivity at 75% positive predictive value. By ranking the combined list of clusters using the adjusted scoring metric and eliminating clusters with dissimilar antibiotic susceptibilities, the developmental study was able to achieve a 40% positive predictive value up to a sensitivity of 0.9 and a sensitivity of approximately 0.15 at a positive predictive value of 75%. While these results did not attain the targeted performance levels, it suggested that further improvements might be able to reach the targeted levels of performance. By

incorporating the experts' post-study advice and by addressing the developmental study's limitations in the next MIASMA detection system design iteration, it might be possible to further reduce the false positive alerts while maintaining a good true positive detection rate and potentially reach the stated goals as described in the next chapter.

Algorithm Analysis: Developmental Study Limitations

The subjectivity of the review process led to an imperfect “gold standard” list of candidate outbreaks in the developmental study. The gold standard list could easily have missed some true outbreaks due to reviewer disagreement on what constituted a candidate cluster. Furthermore, the selection of algorithms for the study did not include the newest syndromic surveillance methods available by the end of the study⁸⁸⁻⁹⁰ and the parameter tuning required to implement each of the four algorithms may not have been optimal. The result was that true outbreak clusters in the developmental study may have been omitted from the algorithms' output lists before ever being seen by the reviewers. That none of infection control service independently-verified outbreaks during the developmental study period appeared on the combined output list of the four algorithms suggests that suboptimal detection at the algorithmic level was likely a factor in the study.

The culture results dataset used to generate the alerts also embodied potential methodological flaws. The developmental study used only the first result for a given organism/patient/unit combination in the dataset. While this approach prevented spurious alerts for multiple consecutive positive cultures on the same patient, it may have been too conservative overall. For example, a patient with *E. coli* cultures in January 2005 and

January 2006 would only be included in 2005, though it is unlikely that the patient's infection lasted a full year. Additional errors may also arise from the system's lack of information about changes within the hospital over time. For example, in late 2005 (approximately halfway through the study period), the burn patient intensive care unit was relocated to another geographic ward, so new patient-organism-location clusters that previously would have been suppressed as duplicate cultures were not suppressed since they were reported from a "different" geographic unit. In addition, some clusters were simply a result of increased surveillance for certain organisms or an increase in a hospital unit's size or number of patient-days as the study did not adjust for increases in patient-bed-days.

The adjustment for antibiotic sensitivity similarity was somewhat crude. For example, if an algorithm detected a cluster made up two distinct clones with widely differing sensitivities, the resulting average difference between the two could be large enough to eliminate the cluster from further consideration. Ideally, available antibiotic sensitivity data should have been included earlier in the detection process.

Lastly, the performance of the system on retrospective datasets was not a guarantee of similar future performance. Because the review process was time consuming for the reviewers and the number of expected candidate outbreaks was limited, the resulting parameter adjustments have not been validated extensively. The "optimal" alerting thresholds determined in the developmental study may have been overfitted to the then-current data. Nevertheless, the six month retrospective evaluation demonstrated that the resulting system was able to detect all outbreaks confirmed by hospital infection control staff during that time period.

Algorithm Analysis: Conclusion

The developmental study explored the potential for a syndromic-surveillance-based approach to screening for potentially clonal inpatient infectious disease outbreaks. Each of the four aberrancy detection algorithms that the study examined had different performance characteristics that limited its individual applicability to the problem at hand. However, by combining the output from each algorithm and then sorting and filtering the possible clusters that the algorithms collectively identified -- based on additional heuristic data that the algorithms cannot easily incorporate -- the developmental study created a prototypic combined screening tool that demonstrated better potential to be clinically useful for hospital outbreak detection than any of the individual algorithms. Thus, while in-hospital outbreak surveillance presented different challenges than those faced by regional syndromic surveillance, the algorithms developed for syndromic surveillance might eventually be adapted to the inpatient screening setting. Further, more formal evaluation of such combined systems should occur.

CHAPTER IV

PROSPECTIVE MIASMA STUDY OVERVIEW

Prospective MIASMA Study Setting

The project conducted a prospective study of the refined MIASMA algorithms previously described in Chapter III from November 2010 through April 2011 at Vanderbilt University using data from the Vanderbilt University Hospital, The Vanderbilt Clinic, and the Monroe Carell Jr. Children's Hospital at Vanderbilt. The Vanderbilt Institutional Review Board approved the study design prior to its inception. The study team obtained data using MicroParse, the previously described (Chapter II) computer program that receives all microbiology reports from the Vanderbilt Medical Center microbiology laboratory via the electronic medical record system.

Goals of Prospective MIASMA Evaluation

Based on the observation that simple detection methods applied to hospital data had already shown promising results,^{13,91} the prospective MIASMA study began with the presumption that new, more advanced approaches to computer-assisted hospital infection control could potentially improve patient outcomes. By supplementing manual hospital infection control practices with an automated outbreak alerting system, the project attempted to achieve more rapid outbreak detection and more efficient alerting for outbreaks (i.e., fewer outbreaks missed with fewer "false positive" alerts that consume resources to investigate).

MIASMA System Design

The prospective evaluation study team designed the final MIASMA algorithms based on lessons learned during the previous developmental study (Chapter III).⁸⁷ As that research demonstrated that no single syndromic surveillance algorithm of the four tested provided adequate sensitivity, the final MIASMA version accepts positive culture count data and microbiology-related non-culture laboratory test results as input to all four developmental-stage algorithms. The final MIASMA algorithm also employs a rules engine in parallel with the four component algorithms to help generate a list of potential disease clusters and to then filter the list to heuristically remove clusters unlikely to correspond to a clonal outbreak. New potential clusters in the list are then added to MIASMA's database to present to infection control staff for subsequent review.

System Configuration: Prospective Study

The process the MIASMA system followed during the study is summarized in Figure 9. Each day, the MIASMA system reviewed the past 180 days of microbiological test results (cultures and lab test results) of augmented MicroParse data (i.e., culture results plus microbiology-related laboratory test results tests for various organisms). Prior to sending the data to each of the four cluster detection algorithms (CUSUM, EWMA, SaTScan, and WSARE) and the rules engine, MIASMA preprocessed it to properly format the data for each algorithm as input. The MIASMA preprocessing also identified hospital inpatient unit groupings that could enable CUSUM and EWMA to locate outbreaks occurring in a related units that often shared staff and/or patients.

Based on the results of the Chapter III developmental study, and in an attempt to reduce false positive rates, the MIASMA study team modified the custom algorithms (CUSUM and EWMA) to increase the threshold needed to generate an alert. The study team also implemented a rules engine algorithm, based on a classification from local Infectious Disease experts of the likelihood of each type of organism to participate in a “significant” outbreak (a rough measure of how many patient-culture dyads might be needed to trigger a cluster alert). This heuristic approach helped to ensure that MIASMA would generate alerts in the case of a single positive culture for extremely virulent organisms (e.g., smallpox or anthrax) or for instances of unusually dangerous antibiotic susceptibility patterns (e.g., vancomycin-resistant *Staphylococcus aureus*)

As in the developmental study, each MIASMA component algorithm returned a list of suspicious clusters of infections. Each cluster was characterized by an organism name, start and end dates, and a hospital location. Then MIASMA combined the list of results from each algorithm to create its daily “operating list” of clusters. Because the data each algorithm used each day largely overlapped with the data used in the previous 179 days, the current day’s operating list typically contained a number of clusters MIASMA had already identified in an earlier execution. Thus, MIASMA would first compare the operating list to the database table containing the previously suspected clonal outbreaks. Duplicate entries were then linked to the existing database entry and dropped from the current day’s operating list.

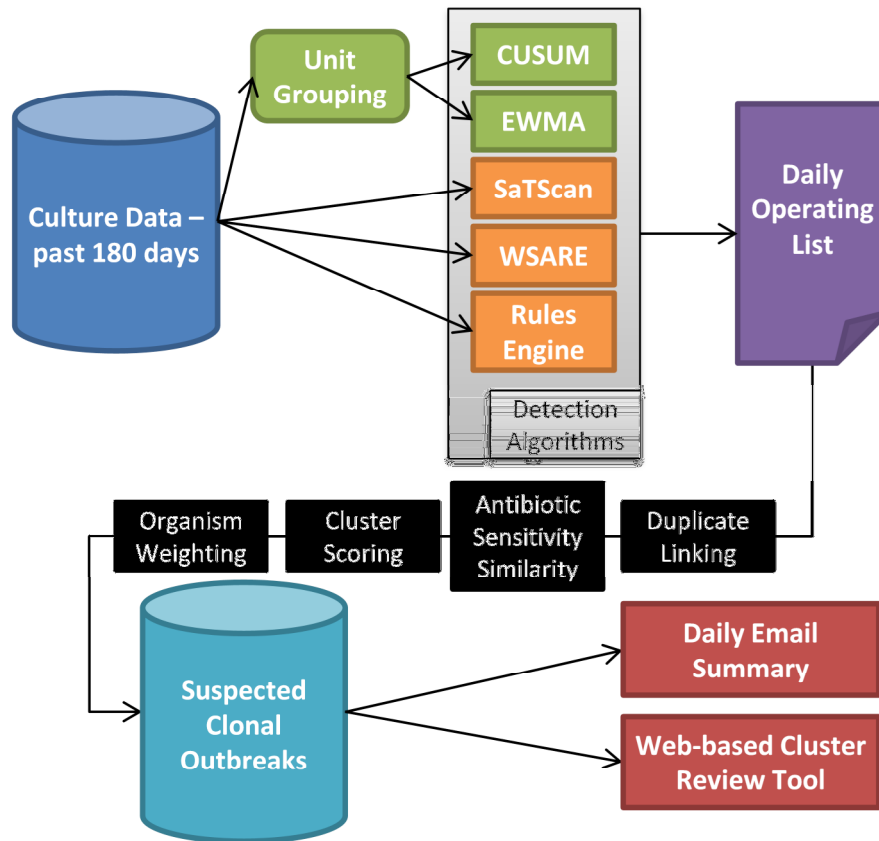


Figure 9: Summary of the process followed during daily MIASMA execution

MIASMA then calculated heuristic scores used to filter and rank each cluster. First, for clusters with antibiotic sensitivity information available, it calculated an antibiotic sensitivity similarity score as described previously in Chapter III. Clusters not meeting an empirically derived internal similarity cutpoint were then removed from the operating list. For example, if four *E. coli* cultures had widely disparate sensitivity patterns, the cluster containing them would be dropped due to presumed non-clonality. Next, MIASMA scored clusters using the EWMA-based metric described in Chapter III. MIASMA then assigns score weights according to the organism type by increasing the scores of clusters of more dangerous or virulent organisms and decreasing the scores for less pathogenic organisms and for organisms commonly resulting from contamination

during sample collection. At this point, clusters with scores below an empirically-derived score cutpoint are also excluded from the operating list. MIASMA then creates a new database entry for the current day for each remaining cluster.

During the prospective evaluation period, on afternoon of the same day as specific cluster generation, MIASMA sent the ID/Infection Control expert an email summary of the daily additions to the database (if any) with a link to a password-protected supporting data review webpage. The review webpage listed all newly detected clusters with basic information about each, including the location in the hospital where the cluster occurred, the number of patients with positive microbiological results, and the start, end, and detection dates of the cluster. The MIASMA tracking web page also provided a text box for note entry and a dropdown box for each cluster, allowing the infection control expert, based on a deeper analysis of available data, to categorize the cluster as a probable hospital-based outbreak, a probable community-based outbreak, a pseudooutbreak, or a false alarm. Figure 10 displays a screenshot of the MIASMA cluster review webpage.

bug	unit	cases	startdate	enddate	score	firstdetected	algo	status	notes
Haemophilus influenzae	<input type="text"/>	17	2011-01-14	2011-02-02	1.34	2011-02-16	satscan	Potential nosocomial outbreak	4 cases on <input type="text"/> but all of different sensitivity patterns
Influenza virus type a	<input type="text"/>	100	2011-01-11	2011-02-08	3.045	2011-02-07	satscan	Potential nosocomial outbreak	2 cases on <input type="text"/> end one on <input type="text"/> - Investigated: All collec
RSV	<input type="text"/>	122	2011-01-10	2011-02-08	0.975	2011-02-07	satscan	Potential nosocomial outbreak	2 cases on <input type="text"/> - Investigated: both collected day of admi
Pseudomonas aeruginosa	<input type="text"/>	2	2011-01-21	2011-01-21	3.88	2011-01-24	wsare	Potential nosocomial outbreak	No further action
Enterococcus faecium	<input type="text"/>	2	2010-12-25	2010-12-25	1.88	2010-12-30	wsare	Potential nosocomial outbreak	
Staphylococcus (coagulase negative)	<input type="text"/>	2	2011-03-30	2011-03-30	0	2011-04-06	wsare		

[View all clusters \(includes those no longer needing action\)](#)
[Outbreak Investigation Tracking tool](#)

Figure 10: Main view of MIASMA cluster review webpage

If the user desired more information about a cluster, each row on the review page contained a link that the reviewer could follow to a more detailed “drill-down” view of each cluster. The drill-down page included a summary of the antibiotic resistance patterns observed in the cluster, unit counts for house-wide clusters or clusters from a group of units, a graphical view of the weekly positive result counts, and the full text of all relevant microbiological results. See Figure 11 for an example of this display.

Reports found: 2

Report count may not match case count from table since duplicate reports for the same patient can be included in the table

abx string	Count
...R.....S.....S.....S.....R.	1
R.....S.....I.....R.	1

MRN	Unit	Report count
		1
		1

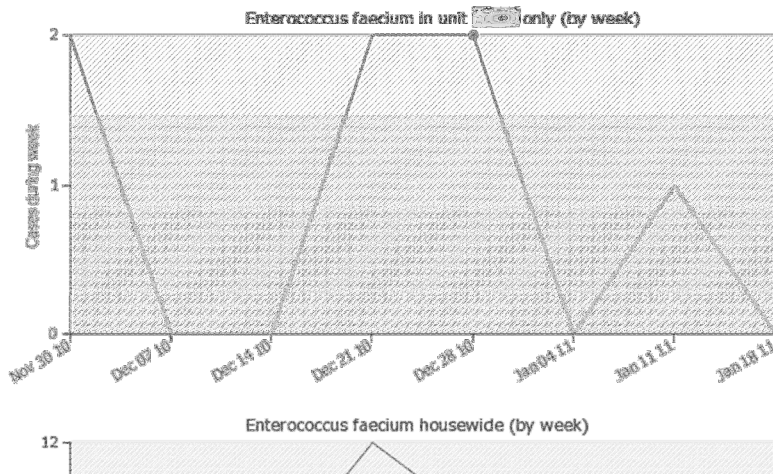


Figure 11: Drill-down view from MIASMA cluster review webpage

Prospective Pilot Study of MIASMA at VUMC

To evaluate MIASMA’s utility as a supplement to traditional infection control practices, the project team implemented MIASMA in the Vanderbilt University Medical Center Department of Infection Control and Prevention. The evaluation study comprised

a two-phase rollout. In Phase 1, before general availability of the algorithm output, only the PI monitored MIASMA output to evaluate its detection timeliness; no alerts were issued to clinicians. In Phase 2, MIASMA was fully implemented as intended, with near real-time alerts to Infection Control clinicians.

The prospective Phase 2 MIASMA evaluation covered the six months spanning November 2010 through April 2011. During this period, MIASMA sent its daily email summaries to VUMC's chief hospital epidemiologist for review and to the PI for evaluation study monitoring purposes. The chief hospital epidemiologist then used the MIASMA review webpage to gather basic information about any new clusters, investigate the clusters using the data available from the MIASMA review page and any other desired sources (e.g., the VUMC electronic medical record system or data collected by the VUMC infection preventionists), and then classify the clusters using the MIASMA review webpage. He reviewed each MIASMA alert on the day that it occurred and rated each as a false alarm, pseudo-outbreak, possible community outbreak, or possible nosocomial outbreak and could later revisit his determinations as more information became available. During the study period (late January 2011), MicroParse began to receive results from additional non-bacterial culture microbiological tests, including influenza and RSV testing.

Data Analysis Methods – Prospective Evaluation

The study team compared MIASMA's timeliness and accuracy of detecting outbreaks to traditional infection control methods. During both phases of the prospective study, the study team recorded data for any clusters found by either manual methods or

by MIASMA (or by both). Nosocomial cluster detections vs. false alarm rates were compared for each pair of algorithms using a chi-squared test.

For any outbreaks missed by MIASMA but detected by standard infection control methods, the study conducted a thorough investigation to determine the cause of MIASMA's failure to alert. The study team located all implicated culture and test results to find whether they had been included in MIASMA's input and investigated methods for locating any missing results. These included incorporating patient bed location data prior to the culture being taken, and methods for estimating when culture-negative cases might have occurred, based on patterns in physician order entry data.

After observing the benefits of including patient bed location and transfer data in locating additional patients that might be part of a cluster, the study investigated the utility of incorporating such patient location and transfer data (i.e., a history of bed locations the patient occupied) within MIASMA. In this separate analysis, for each positive culture result, MIASMA retrieved any available patient bed location data from 5 days prior to the positive result and added a "dummy" positive result for each bed location the patient had occupied at that time. The modified MIASMA algorithm then filtered duplicate positives, as it had done in normal operation, to prevent multiple counting of results for a single patient, which could skew results. MIASMA then ran twice on June 18, 2011: once with the standard data feed and once with the patient bed transfer data included.

To investigate MIASMA's performance at locating "expected" outbreaks, the study team analyzed its ability to detect seasonal illnesses (influenza A, influenza B, and RSV). The Tennessee State Department of Health publishes weekly positive test counts

for these illnesses. The study team compared the data collected by the Health Department to the data observed by MIASMA to roughly estimate how much of a timeliness advantage MIASMA could achieve since the normal rise of each illness mimics the pattern seen in an outbreak.

Lastly, at the end of the prospective study period, the PI informally interviewed the chief epidemiologist to solicit his perception of the system's performance and utility in everyday infection control practice.

CHAPTER V

RESULTS OF PROSPECTIVE MIASMA EVALUATION

Alert Statistics

During the 181-day prospective study period, MIASMA generated 78 alerts. On 136 days, (75%) there were no alerts, and 36 days (20%) had one alert. Figure 12 displays the complete data.

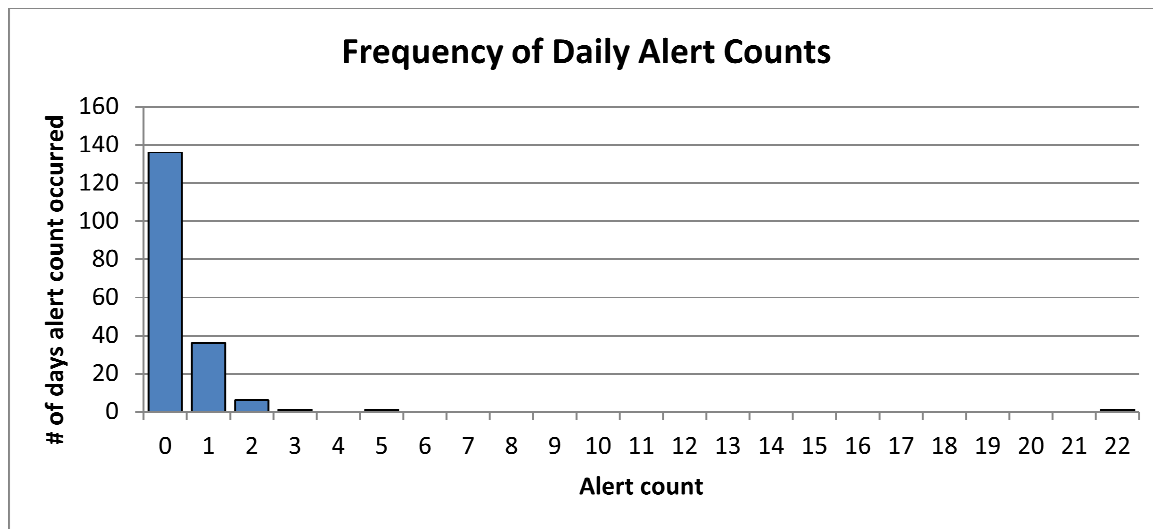


Figure 12: Frequency of daily alert counts during MIASMA's 181-day study period

MIASMA Alert Categorization

Of the 78 alerts, the Infection Control expert rated 3 (4%) as possible nosocomial outbreaks, 51 (65%) as possible community outbreaks, 6 (8%) as pseudo-outbreaks, and 18 (23%) as false alarms. Figure 13 displays the alerts generated each week as coded by the epidemiologist's designation. Table 7 displays the counts for each designation by the

detecting algorithm. There were no significant pairwise differences among the four algorithms, though the comparison of WSARE to STSS was nearly significant in favor of WSARE ($p=0.051$). Table 8 provides a complete listing of all detected clusters during the study period.

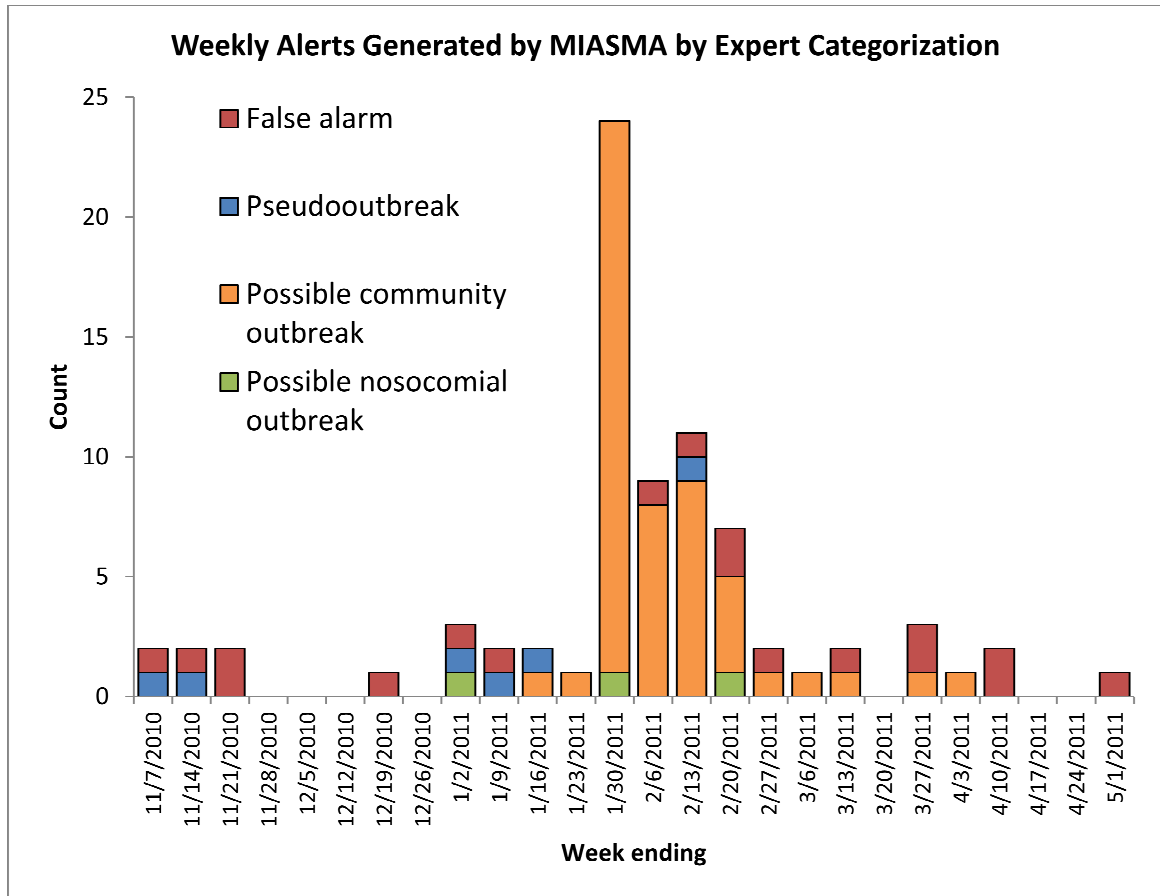


Figure 13: Weekly alerts generated by MIASMA during study period coded by expert designation

Table 7: Cluster detection by algorithm and expert categorization

	Nosocomial	Community	Pseudo	False alarm
CUSUM	0	12	0	0
EWMA	0	0	0	3
STSS	1	18	6	13
WSARE	2	31	2	3

Table 8: Complete listing of detected clusters by affected site(s) and expert categorization during study period (Length given in days)

Organism	Unit	Cases	Length	Expert Determination
Enterococcus faecium	11NM	2	1	Nosocomial
Pseudomonas aeruginosa	8N	2	1	Nosocomial
Haemophilus influenzae	all	17	20	Nosocomial
Streptococcus pyogenes	ED/P	3	1	Community
RSV	all	6	21	Community
Influenza virus type a	all	85	27	Community
Influenza virus type b	all	125	28	Community
RSV	all	101	28	Community
Influenza virus type a	all	15	1	Community
Influenza virus type a	ED/P	6	1	Community
Influenza virus type a	EMER	7	1	Community
RSV	ED/P	12	1	Community
RSV	ED/P	9	1	Community
Influenza virus type a	ED/P	3	1	Community
RSV	ED/P	9	1	Community
Influenza virus type a	ED/P	5	1	Community
Influenza virus type a	all	8	1	Community
Streptococcus pyogenes	ED/P	15	1	Community
Influenza virus type b	ED/P	9	1	Community
Influenza virus type b	all	8	1	Community
Influenza virus type b	EMER	2	1	Community
Influenza virus type b	ED/P	5	1	Community
RSV	ED/P	6	1	Community
Influenza virus type b	all	6	1	Community
RSV	ED/P	3	1	Community
Influenza virus type b	ED/P	2	1	Community
Streptococcus pyogenes	ED/P	17	1	Community
Influenza virus type b	EMER	3	1	Community
RSV	all	105	28	Community
Influenza virus type a	all	97	28	Community
RSV	ED/P	11	1	Community
RSV	all	105	27	Community
Influenza virus type a	all	3	1	Community
Influenza virus type a	ED/P	7	1	Community
Influenza virus type a	all	95	28	Community
RSV	all	117	28	Community
Influenza virus type a	all	100	27	Community
RSV	all	122	28	Community
RSV	ED/P	12	1	Community

Table 8 (continued): Complete listing of detected clusters by affected site(s) and expert categorization during study period (Length given in days)

Organism	Unit	Cases	Length	Expert Determination
Influenza virus type b	all	122	28	Community
RSV	all	111	21	Community
Influenza virus type a	all	11	1	Community
Streptococcus pyogenes	ED/P	177	44	Community
Streptococcus pyogenes	ED/P	206	53	Community
Influenza virus type a	all	108	27	Community
Influenza virus type a	all	112	27	Community
Influenza virus type a	ED/P	4	1	Community
Influenza virus type a	ED/P	7	1	Community
Influenza virus type a	EMER	2	1	Community
RSV	all	149	27	Community
RSV	all	157	28	Community
RSV	ED/P	196	43	Community
Influenza virus type a	ED/P	61	30	Community
RSV	ED/P	268	72	Community
Lactobacillus species	all	36	25	Pseudo
Lactobacillus species	all	42	28	Pseudo
Acid-fast bacillus	all	2	2	Pseudo
Bacteroides thetaiotaomicron	all	2	1	Pseudo
Lactobacillus species	all	11	2	Pseudo
Streptococcus pneumoniae	all	11	7	Pseudo
Coag-negative Staphylococcus	PBIL	2	2	False alarm
Coag-negative Staphylococcus	7C	2	7	False alarm
Enterobacter aerogenes	all	15	15	False alarm
Burkholderia cepacia	all	2	2	False alarm
Lactobacillus species	EMER	2	1	False alarm
Gram positive cocci	all	6	4	False alarm
Trichophyton rubrum	all	4	15	False alarm
Epstein Barr virus	all	8	6	False alarm
Diphtheroids species	all	12	4	False alarm
Chlamydia trachomatis	all	13	2	False alarm
Aspergillus fumigatus	all	5	9	False alarm
Streptococcus intermedius	all	4	4	False alarm
Candida species	all	4	2	False alarm
Hepatitis C	all	40	23	False alarm
Citrobacter koseri	all	8	8	False alarm
Coag-negative Staphylococcus	OVUH	2	1	False alarm
Enterococcus species	all	10	17	False alarm
Gram positive	ADL2	3	1	False alarm

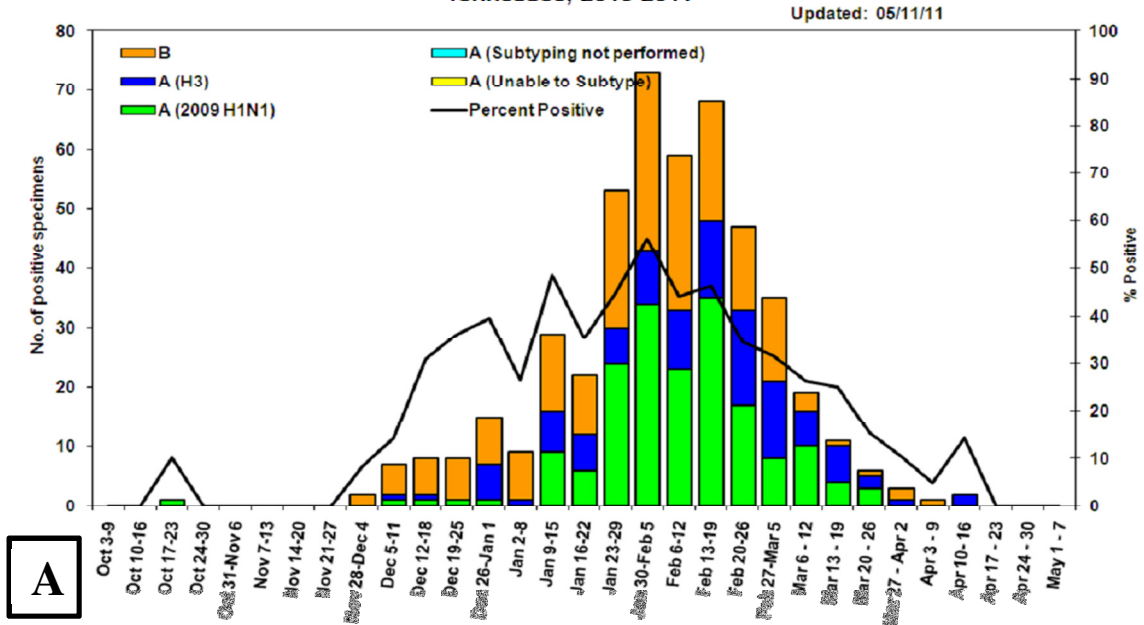
Comparison of MIASMA Output to Manual Methods

During both phases of the prospective evaluation study, there was no overlap in suspected bacterial cluster identification between MIASMA and the traditional infection control methods. Thus, no direct comparisons measuring timeliness of detection could occur.

Comparison of MIASMA Output to Public Health Records for Seasonal Illnesses

The Tennessee State Department of Health publishes weekly positive test counts for influenza A and influenza B. Figure 14A shows the retrospective graph published by the TN Department of Health on May 7, 2011.⁹² Figure 14B shows a comparable graph generated from MIASMA's source data with dates marked when MIASMA generated at least 1 alert. Figure 15 shows data from the same time period for RSV. State data for RSV were not available because the testing instrument used at the state labs was being serviced.

Week (Oct. 3, 2010 - May 7, 2011)
**Influenza Positive Tests Submitted to TN Dept. of Health Laboratory Services,
 Tennessee, 2010-2011**



**Influenza Positive Tests from
 Vanderbilt University Medical Center, 2010-2011**

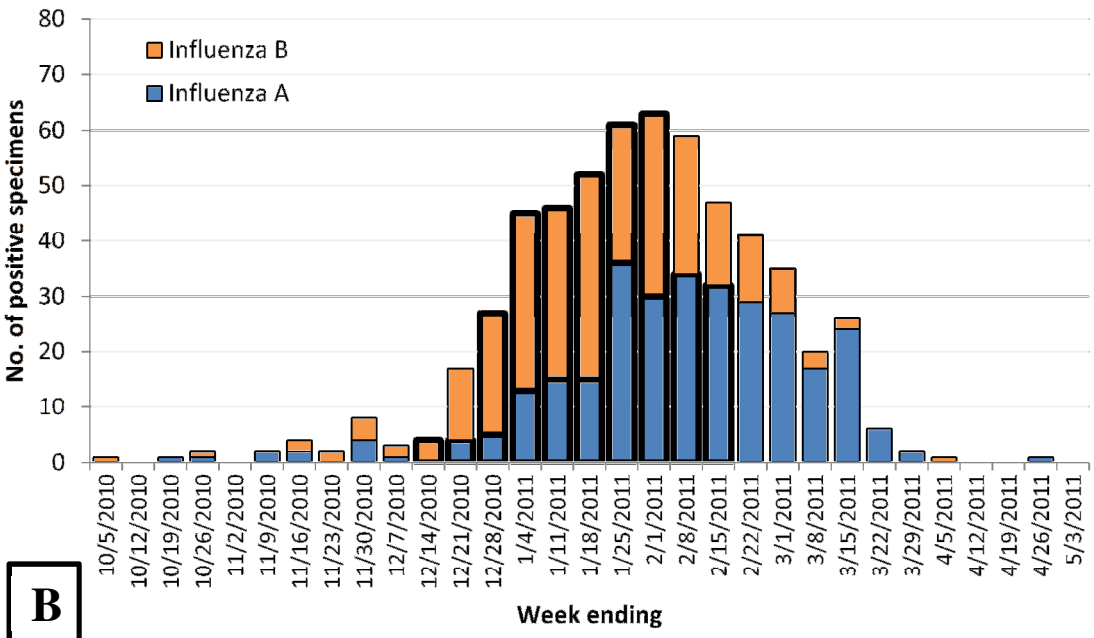


Figure 14: Comparison of TN State Department of Health and VUMC influenza results, 2010-2011. MIASMA alerts generated during weeks marked with bold border.

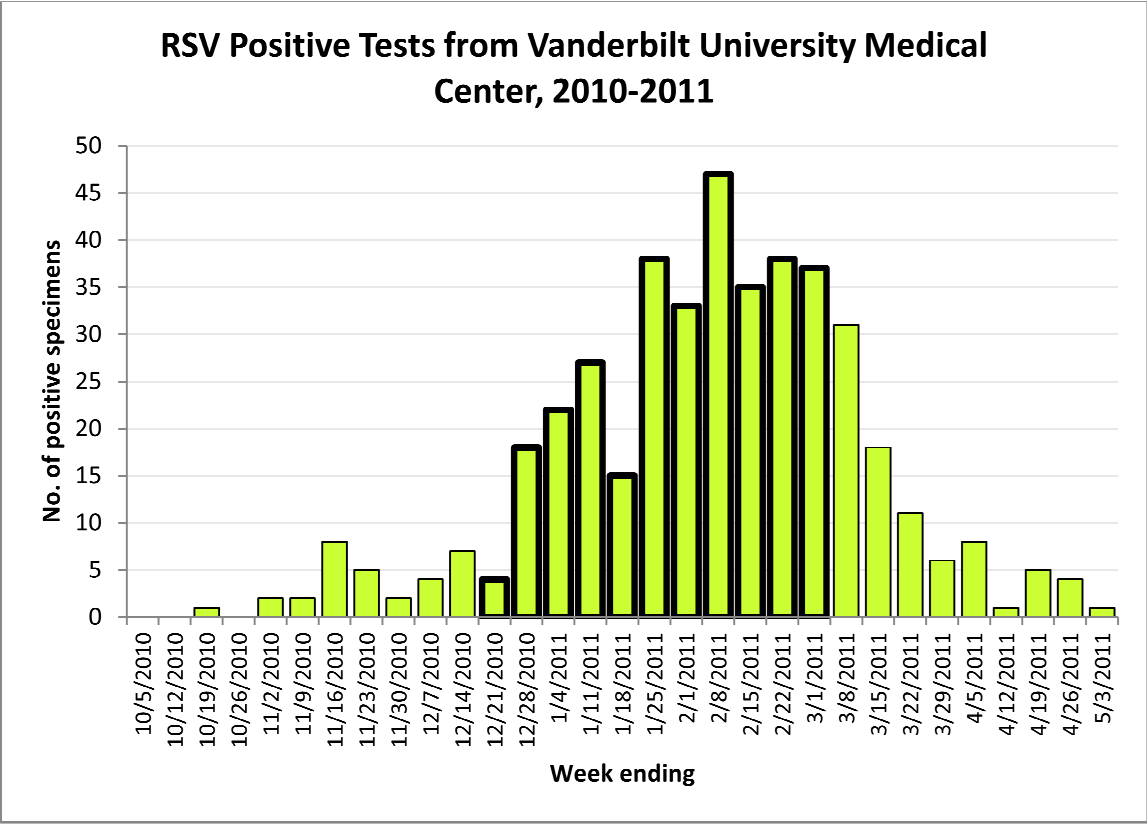


Figure 15: RSV positive tests at VUMC during study period. MIASMA alerts generated during weeks marked with bold border.

Investigation of Potential Outbreaks During Prospective Study Period Missed by
MIASMA

During the prospective study period, a single outbreak confirmed by the Infection Control Service occurred. It was an outbreak of *Clostridium difficile* associated with patients who had been housed previously on S44, a general surgery unit. Infection control staff located a total of 12 suspected patients with *C. difficile* during a 1 month span, with 6 culture-positive cases and 6 clinically diagnosed cases, some of whom had negative culture results. Investigation of the MIASMA dataset identified 4 of the 6 culture-positive results on the affected unit. Locating the remaining 2 culture-positive patients required

the incorporation of a separate database containing patient bed location information, as the patients had been moved to new bed locations prior to their positive culture results. Figure 16A shows the overall *C. difficile* incidence. Overall, the incidence during the month increased by slightly over 1 additional positive culture per week during the outbreak period (1.25/week during the outbreak period vs. 0.11/week normally).

Since some of the implicated patients were culture-negative, the PI also explored the effects of including tests negative for *C. difficile* as a “partial positive” test since a physician presumably ordered the test suspecting potential *C. difficile* infection. Figure 16B shows the results of including negative tests as $\frac{1}{4}$ of a positive test. When using the weighted score, the increase during the outbreak period was slightly over 1 positive test per week greater than the non-outbreak periods (1.75/week during the outbreak period vs 0.57 normally).

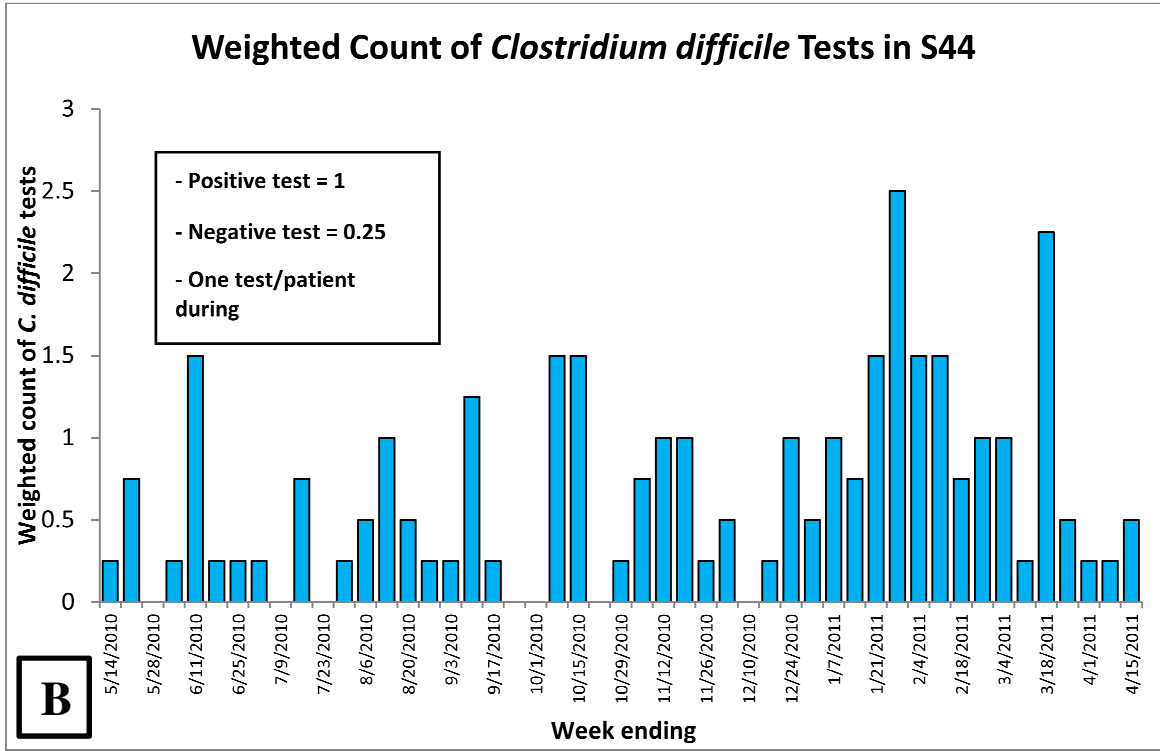
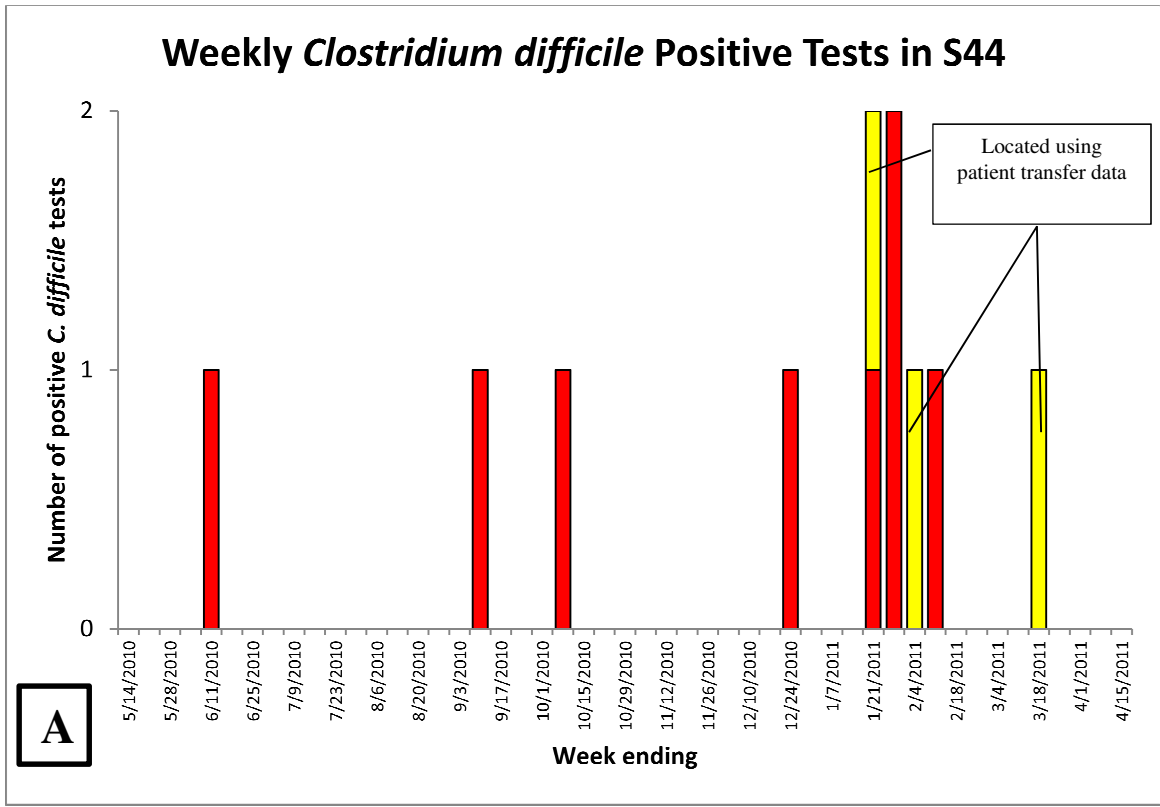


Figure 16: Test result data from *Clostridium difficile* in S44

Incorporating Patient Bed Location and Transfer Data

The patient bed transfer-enriched dataset was approximately 25% larger than the standard dataset. Overall, the algorithms found 90 clusters on the standard data and 89 using the bed transfer data. However, the clusters found using each dataset were quite different: 46 clusters were found using both datasets, while 44 were found using the standard dataset only and 43 using the transfer-enriched dataset only.

Of the 90 clusters found using the standard dataset, 61 were filtered using the standard MIASMA suppression techniques previously described and were never viewed by the epidemiologist; the remaining 29 were duplicates of previously detected clusters. Those 29 consisted of 20 community outbreaks, 7 false alarms, and 2 pseudo-outbreaks. When using the transfer data, MIASMA no longer detected 11 of the community outbreaks, 2 of the false alarms, or either of the pseudo-outbreaks. A chi-square test showed no difference between the designations of the clusters detected using both datasets versus the standard dataset alone ($p=0.326$).

Expert Assessment of MIASMA System

VUMC's chief epidemiologist stated that MIASMA was best described a "safety net:" it would not be able to replace regular surveillance, but it did provide a backup to help prevent standard practices from missing potential outbreaks. Furthermore, MIASMA's increased activity during influenza and RSV season provided an unexpected value to the chief epidemiologist as these data helped inform institutional policies that were implemented once the community incidence of these infections had increased.

These data also helped guide clinicians regarding probability of these infections in patients presenting with respiratory symptoms.

CHAPTER VI

DISCUSSION OF PROSPECTIVE MIASMA EVALUATION

MIASMA Alert Quality

From the chief epidemiologist's feedback, the volume of the alerts was deemed reasonable in terms of workload, with only alerts occurring only 1 day in every 4. In addition, there were no more than two false alarms or pseudo-outbreaks in any given week. This helped prevent alert fatigue. Of note, during late January and early February, the alerts were dominated by clusters of community-based seasonal illnesses (influenza, RSV, and Group A *Streptococcus*). The volume of alerts could easily be reduced by more restrictive methods of eliminating duplicate clusters. For example, in this study, by only allowing one alert for a given unit/organism combination in a given month if the first cluster is deemed to be community-based, MIASMA would have generated only 37 alerts, fewer than half of the 78 actually generated during the study.

On one day (1/29/2011), MIASMA generated 22 of the 78 alerts based on cluster detections made primarily by WSARE. This anomaly was the result of corrections made to the underlying culture reporting data. Previously, influenza antigen tests, RSV antigen tests, and Group A *Streptococcus* probes were not properly included in the MIASMA input data. Once those test results were incorporated into MIASMA's input dataset, WSARE detected clusters of each of these pathogens though all hospital wards and in the pediatric and adult emergency departments.

Unfortunately, MIASMA could not locate the only Infection Control Service confirmed outbreak during the study period. Since half of the patients involved in that outbreak were located by the infection preventionists only based on a clinical symptomatic diagnosis of *Clostridium difficile*, MIASMA's lab-and-culture-based approach was at a disadvantage. Observing the incidence of *C. difficile* in S44 over time, there was clearly an increase of positive tests during the month of the outbreak. However, the increase was only slight (approximately 1 additional positive test result per week), and if the detection algorithms were set to be sensitive enough to detect such a cluster, the number of false positive alerts for other organisms would likely increase. Similarly, if one were to incorporate the negative *C. difficile* tests as partial positives, one would observe a similar small increase over the baseline, but the increase would not be sufficient to trigger an alert from any of the algorithms.

One alternative method of detecting such a cluster would be to expand the MIASMA rules engine to make better use of domain knowledge. During the study, MIASMA simply used the rules engine to alert in the event of a single positive culture for extremely virulent organisms (e.g., smallpox or anthrax) or unusually dangerous antibiotic susceptibility patterns (e.g., vancomycin-resistant *Staphylococcus aureus*). However, the engine could very easily trigger an alert for any desired incidence threshold. For example, the outbreak in question could have been located by firing an alert any time 4 or more positive *C. difficile* tests occur on any specific unit within a 2 week period. By incorporating this and similar rules as suggested by the chief epidemiologist, MIASMA could potentially detect additional relevant clusters without adding many false alarms.

Algorithm Performance During the Prospective MIASMA Study

None of the algorithms were significantly better than the others, though it might have been possible to show that WSARE outperformed STSS if the study had used a larger sample size. A borderline significant difference was observed, as noted. The study was insufficiently powered to find a difference of such a small magnitude. Nevertheless, this result supports the finding from Chapter III showing that no one of the implemented component surveillance algorithms was able to sufficiently detect all desirable clusters in this problem domain.

Because the two custom-implemented algorithms (i.e., “tuneable” parameters were specifically adjusted by the study team for EWMA and CUSUM) could be scaled back to prevent excessive alert generation, they were much less prolific in generating alerts than were WSARE and STSS (3 and 12 alerts for EWMA and CUSUM respectively vs. 38 each for WSARE and STSS). However, this may not have been ideal, as CUSUM identified only community outbreaks during the study and EWMA only false alarms, suggesting that the filtering rules MIASMA applied to reduce the number of alerts generated by these two algorithms may not have been effective at “distilling” the signal from the noise within them.

Though newer scan statistic algorithms than STSS have been developed,⁶⁷⁻⁷⁰ their focus is primarily on making the STSS spatial detection methods more flexible. Since this study used spatial information in the form of hospital unit groupings that were based on staff and patient movement rather than geographic spatial information, however, the newer algorithmic improvements probably would not be useful. Thus, as STSS has been

much more extensively used and validated than newer algorithms, it was likely an adequate or better choice for MIASMA than one of the newer algorithms.

Seasonal Detection Performance

The community-based clusters for which MIASMA generated alerts tended to serve as noise when trying to detect hospital-based outbreaks. However, information about the community influenza and RSV incidence proved useful to VUMC's chief epidemiologist as noted above. The previously mentioned methods of eliminating the excess alerts might have diminished this unexpected MIASMA benefit. Furthermore, it is clear from this study that MIASMA cannot replace infection control staff, but it can likely provide a valuable supplement to make surveillance efforts more comprehensive. Thus, it may not be desirable to remove the influenza and RSV alerts.

Comparing MIASMA's influenza data to the TN State Department of Health shows that MIASMA achieved an approximately 1 week lead time in detection versus the State Department of Health on the initial rise in influenza. Initial MIASMA influenza alerts occurred very early. Coupled with the fact that the State Department of Health only issues reports once weekly, MIASMA had in effect up to a 2 week advantage in identifying the start of influenza season.

Adding Patient Bed Location Data to MIASMA

Incorporating historical patient bed location data into MIASMA's input stream had an unexpectedly large effect on the resulting clusters detected. Prior to testing the system with the bed location data included, the study team expected to simply see the

same clusters as with the standard non-dynamic bed data with possibly a few additional detected clusters since the transfer dataset contained all the same results plus the extra data resulting from the prior bed locations. However, some clusters also disappeared after adding bed locations. Presumably, this effect resulted from an increase in the baseline rates of infection in some units. This increase could subsequently make the spike of infections during the previously detected cluster appear less abnormal. The clusters eliminated by including the patient transfer data were not significantly different by designation, however, so it seems unlikely that incorporating the transfer data significantly improved the quality of the generated alerts unless the newly added clusters were particularly likely to be deemed outbreaks. A larger data sample encompassing longer observation periods is necessary to determine objectively what utility the dynamic patient location data might add to MIASMA.

Other methods of incorporating dynamic bed location information might improve the overall performance. For example, incorporating information about incubation times and contagious periods for individual pathogens would allow MIASMA to be more discriminating in retrieving relevant patient bed location data for only the times when a given patient would likely have contracted the disease or when he or she could pass the disease on to others.

Expert Assessment of MIASMA's Performance and Utility

As the volume of cultures and lab tests conducted during clinical practice grows, MIASMA's functionality as a potential safety net increases in importance. The role of VUMC infection preventionists (IPs) has recently shifted from looking at all cultures

serially to a more specialized approach where each IP is responsible for cultures that fall under a common rubric (e.g., only reviewing bloodstream infections in catheterized patients). This specialized model thus makes it more likely that the IPs might miss an outbreak that does not fall into one of the predetermined categories. The current prospective study thus corroborates past findings regarding automated surveillance for outbreaks in the hospital setting⁹¹ that found that with whole-house surveillance becoming less practical for large hospitals, automated surveillance can provide a useful supplement to standard practice.

The MIASMA study delivered an unexpected benefit through the information gained by the frequent alerts for influenza and RSV. In past years, RSV and influenza data were not routinely collected by the IPs to be passed on to the chief epidemiologist. Thus, he had not been able to provide advice on the timing of such epidemics to inquiring clinicians. However, with MIASMA regularly keeping him informed of the influenza and RSV activity within the hospital, emergency departments, and clinics, he was able to give accurate and timely information.

MIASMA System Portability

The PI plans to make MIASMA and the custom implementations of the public domain EWMA and CUSUM algorithms freely available for not-for-profit use under a Simplified BSD license (or FreeBSD license).⁹³ The MIASMA algorithms require only a PHP installation to run. The WSARE and STSS implementations are available from their original authors under free use licenses. However, as they are only available as binaries, they can only be used on platforms for which their original authors have compiled them.

Thus, STSS is available via the SaTScan download page⁹⁴ for Windows, Max OSX, and Linux, and WSARE is available from Carnegie Mellon University's Auton Lab download page⁹⁵ for Windows and Linux only. For other platforms, users can easily configure MIASMA to exclude any unusable component algorithms.

Currently, MIASMA is configured to work with the MicroParse MySQL database only. However, the code that queries the database is separated from the program logic, allowing users to relatively easily modify those portions of the system. Because MIASMA's requirements for data structuring are relatively simple, only requiring a list of positive cultures labeled with name, date, and location, it is relatively easy to adapt the system to other database engines (e.g., PostgreSQL or Oracle) and schemas. The database tables containing the hospital unit groupings and organism names would need to be modified to reflect the local nomenclature to make use of all of MIASMA's features, however.

MIASMA Prospective Study Limitations

The largest limitation of the pilot prospective MIASMA study was the small number of outbreaks (one) that occurred during the study period. Ideally, the study could have been conducted across multiple sites and for a longer duration to allow MIASMA to detect more outbreaks of differing types. With only a single "true" outbreak occurring, and in that instance, with half of the implicated patients being culture-negative, it was difficult to definitively determine whether MIASMA was truly useful in practice.

On a related note, the dearth of true outbreaks made comparing MIASMA's performance to that of manual methods on timeliness and accuracy infeasible. Again, this

could have been solved by conducting the study across multiple sites for a longer time. That would allow the study to find situations where both MIASMA and the local infection control staff detected the same outbreaks. Extending the study would also allow better evaluation of techniques for locating the *Clostridium difficile* outbreak missed during the study period.

With the potential for patient bed location transfer data to be useful adjunct information to MIASMA's constituent algorithms, further analysis of the impact of incorporating bed location transfer data could help strengthen the MIASMA system. Ideally, a new study would have mirrored the prospective study design using bed location transfer data and had all resulting clusters classified by an expert. The end result would be a complete picture of how the detected clusters compared using each of the datasets and a much clearer idea of the effects of including patient bed location data.

CHAPTER VII

SYNOPSIS AND CONCLUSIONS

Summary

The MIASMA project developed, deployed and evaluated MIASMA, a system that uses recently developed methods (e.g., from syndromic surveillance and from heuristic observations) to detect single-hospital outbreaks of both commonly occurring and rare bacterial species. Because there were relatively few outbreaks during the study period, further research would be necessary to determine the degree to which MIASMA assisted hospital staff. However, MIASMA successfully supplemented VUMC's standard hospital outbreak detection practices.

BIBLIOGRAPHY

1. Haley RW, Culver DH, White JW, Morgan WM, Emori TG. The nationwide nosocomial infection rate. A new need for vital statistics. *Am J Epidemiol*. 1985;121(2):159-67.
2. Mylotte JM, Graham R, Kahler L, Young BL, Goodnough S. Impact of nosocomial infection on length of stay and functional improvement among patients admitted to an acute rehabilitation unit. *Infect Control Hosp Epidemiol*. 2001;22(2):83-7.
3. Sébille V, Chevret S, Valleron AJ. Modeling the spread of resistant nosocomial pathogens in an intensive-care unit. *Infect Control Hosp Epidemiol*. 1997;18(2):84-92.
4. Mayhall CG. *Hospital Epidemiology and Infection Control (Hospital Epidemiology & Infection Control)*. Third Edition. Lippincott Williams & Wilkins; 2004.
5. Cohen AL, Ridpath A, Noble-Wang J, et al. Outbreak of *Serratia marcescens* bloodstream and central nervous system infections after interventional pain management procedures. *Clin J Pain*. 2008;24(5):374-380.
6. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proceedings of the National Academy of Sciences*. 1993;90(10):4384 -4388.
7. Tenover FC, Arbeit R, Archer G, et al. Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. *J. Clin. Microbiol*. 1994;32(2):407-415.
8. Anon. *APIC Text of Infection Control And Epidemiology*. Association for Professionals in Infection; 2005.
9. Evans RS, Gardner RM, Bush AR, et al. Development of a Computerized Infectious-Disease Monitor (Cidm). *Computers and Biomedical Research*. 1985;18(2):103-113.
10. Evans RS, Larsen RA, Burke JP, et al. Computer surveillance of hospital-acquired infections and antibiotic use. *Jama*. 1986;256(8):1007-11.
11. Kahn MG, Steib SA, Fraser VJ, Dunagan WC. An expert system for culture-based infection control surveillance. *Proc Annu Symp Comput Appl Med Care*. 1993:171-5.
12. Bouam S, Girou E, Brun-Buisson C, Karadimas H, Lepage E. An intranet-based automated system for the surveillance of nosocomial infections: Prospective validation compared with physicians self-reports. *Infection Control and Hospital Epidemiology*. 2003;24(1):51-55.

13. Sagel U, Mikolajczyk RT, Krämer A. Using mandatory data collection on multiresistant bacteria for internal surveillance in a hospital. *Methods Inf Med.* 2004;43(5):483-485.
14. Bellini C, Petignat C, Francioli P, et al. Comparison of automated strategies for surveillance of nosocomial Bacteremia. *Infection Control and Hospital Epidemiology.* 2007;28(9):1030-1035.
15. Brossette SE, Hacek DM, Gavin PJ, et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection - The future of infection control surveillance? *American Journal of Clinical Pathology.* 2006;125(1):34-39.
16. Doherty J, Noirod LA, Mayfield J, et al. Implementing GermWatcher, an enterprise infection control application. *AMIA Annu Symp Proc.* 2006:209-13.
17. Wright MO, Perencevich EN, Novak C, et al. Preliminary assessment of an automated surveillance system for infection control. *Infection Control and Hospital Epidemiology.* 2004;25(4):325-332.
18. Brossette SE, Sprague AP, Hardin JM, et al. Association rules and data mining in hospital infection control and public health surveillance. *Journal of the American Medical Informatics Association.* 1998;5(4):373-381.
19. Stelling JM, O'Brien TF. Surveillance of antimicrobial resistance: the WHONET program. *Clin Infect Dis.* 1997;24 Suppl 1:S157-68.
20. Anon. WHO | WHONET Software. Available at: <http://www.who.int/drugresistance/whonetsoftware/en/>. Accessed July 19, 2011.
21. Stelling J, Yih K, Galas M, et al. Detection of Shigella Outbreaks in Argentina Using WHONET and SaTScan. Available at: <http://www.isdsjournal.org/articles/3272.pdf>. Accessed July 19, 2011.
22. Huang SS, Yokoe DS, Stelling J, et al. Automated detection of infectious disease outbreaks in hospitals: a retrospective cohort study. *PLoS Med.* 2010;7(2):e1000238.
23. Kho AN, Dexter P, Lemmon L, et al. Connecting the dots: creation of an electronic regional infection control network. *Stud Health Technol Inform.* 2007;129(Pt 1):213-7.
24. Kho AN, Lemmon L, Dexter PR, Doebbeling B. An operational citywide electronic infection control network: results from the first year. *AMIA Annu Symp Proc.* 2008:1222.
25. Anon. Research – TheraDoc. Available at: <http://www.theradoc.com/about-us/find/research/>. Accessed July 19, 2011.
26. Anon. Data Mining Surveillance Service - CareFusion. Available at: <http://www.carefusion.com/medical-products/infection-prevention/surveillance-analytics/medminded-data-mining-surveillance-service.aspx>. Accessed July 19, 2011.

27. Fluit AC, Terlingen AM, Andriessen L, et al. Evaluation of the DiversiLab system for detection of hospital outbreaks of infections by different bacterial species. *J. Clin. Microbiol.* 2010;48(11):3979-3989.
28. Jernigan JA, Stephens DS, Ashford DA, et al. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerg Infect Dis.* 2001;7(6):933-44.
29. Buckeridge DL. Outbreak detection through automated surveillance: A review of the determinants of detection. *Journal of Biomedical Informatics.* 2007;40(4):370-379.
30. Wong WK, Moore A, Cooper G, Wagner M. What's strange about recent events (WSARE): An algorithm for the early detection of disease outbreaks. *Journal of Machine Learning Research.* 2005;6:1961-1998.
31. Brown SM, Benneyan JC, Theobald DA, et al. Binary cumulative sums and moving averages in nosocomial infection cluster detection. *Emerging Infectious Diseases.* 2002;8(12):1426-1432.
32. Heisterkamp SH, Dekkers ALM, Heijne JCM. Automated detection of infectious disease outbreaks: hierarchical time series models. *Statistics in Medicine.* 2006;25(24):4179-4196.
33. Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. *Bmc Public Health.* 2001;1.
34. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *Plos Medicine.* 2005;2(3):216-224.
35. Lewis MD, Pavlin JA, Mansfield JL, et al. Disease outbreak detection system using syndromic data in the greater Washington DC area. *American Journal of Preventive Medicine.* 2002;23(3):180-186.
36. Kahn MG, Bailey TC, Steib SA, Fraser VJ, Dunagan WC. Statistical process control methods for expert system performance monitoring. *J Am Med Inform Assoc.* 1996;3(4):258-269.
37. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. *Journal of Biomedical Informatics.* 2005;38(2):99-113.
38. Reis BY, Kirby C, Hadden LE, et al. AEGIS: A Robust and Scalable Real-time Public Health Surveillance System. *J Am Med Inform Assoc.* 2007;14(5):581-588.
39. Gesteland PH, Gardner RM, Tsui F-C, et al. Automated Syndromic Surveillance for the 2002 Winter Olympics. *J Am Med Inform Assoc.* 2003;10(6):547-554.

40. Bradley CA, Rolka H, Walker D, Loonsk J. BioSense: implementation of a National Early Event Detection and Situational Awareness System. *MMWR Morb Mortal Wkly Rep.* 2005;54 Suppl:11-9.
41. Brownstein JS, Mandl KD. Reengineering real time outbreak detection systems for influenza epidemic monitoring. *AMIA Annu Symp Proc.* 2006:866.
42. Anon. Monitoring health effects of wildfires using the biosense system--San Diego County, California, October 2007. *MMWR Morb Mortal Wkly Rep.* 2008;57(27):741-4.
43. Bravata DM, McDonald KM, Smith WM, et al. Systematic Review: Surveillance Systems for Early Detection of Bioterrorism-Related Diseases. *Ann Intern Med.* 2004;140(11):910-922.
44. van den Wijngaard CC, van Asten L, van Pelt W, et al. Syndromic surveillance for local outbreaks of lower-respiratory infections: would it work? *PLoS ONE.* 2010;5(4):e10406.
45. Peter W, Najmi AH, Burkom HS. Reducing false alarms in syndromic surveillance. *Stat Med.* 2011;30(14):1665-1677.
46. Izadi M, Buckeridge DL. Optimizing the response to surveillance alerts in automated surveillance systems. *Stat Med.* 2011;30(5):442-454.
47. Uscher-Pines L, Babin SM, Farrell CL, et al. Research priorities for syndromic surveillance systems response: consensus development using nominal group technique. *J Public Health Manag Pract.* 2010;16(6):529-534.
48. Chapman WW, Dowling JN, Baer A, et al. Developing syndrome definitions based on consensus and current use. *J Am Med Inform Assoc.* 2010;17(5):595-601.
49. Wang X, Zeng D, Seale H, et al. Comparing early outbreak detection algorithms based on their optimized parameter values. *J Biomed Inform.* 2010;43(1):97-103.
50. Harcourt SE, Smith GE, Elliot AJ, et al. Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK. *Epidemiol Infect.* 2011:1-6.
51. Moore K, Black J, Rowe S, Franklin L. Syndromic surveillance for influenza in two hospital emergency departments. Relationships between ICD-10 codes and notified cases, before and during a pandemic. *BMC Public Health.* 2011;11:338.
52. Griffin BA, Jain AK, Davies-Cole J, et al. Early detection of influenza outbreaks using the DC Department of Health's syndromic surveillance system. *BMC Public Health.* 2009;9:483.

53. Carlson SJ, Dalton CB, Tuyl FA, et al. Flutracking surveillance: comparing 2007 New South Wales results with laboratory confirmed influenza notifications. *Commun. Dis. Intell.* 2009;33(3):323-327.
54. Kwan-Gett TS, Baer A, Duchin JS. Spring 2009 H1N1 influenza outbreak in King County, Washington. *Disaster Med Public Health Prep.* 2009;3 Suppl 2:S109-116.
55. Chan EH, Tamblyn R, Charland KML, Buckeridge DL. Outpatient physician billing data for age and setting specific syndromic surveillance of influenza-like illnesses. *J Biomed Inform.* 2011;44(2):221-228.
56. Rabadan R, Calman N, Hripcsak G. Next generation syndromic surveillance: molecular epidemiology, electronic health records and the pandemic Influenza A (H1N1) virus. *PLoS Curr.* 2009;1:RRN1012.
57. Zhou X, Ye J, Feng Y. Tuberculosis Surveillance by Analyzing Google Trends. *IEEE Trans Biomed Eng.* 2011. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21435969>. Accessed July 14, 2011.
58. Kabbash IA, Al-Mazroa MA, Memish ZA. Evaluation of syndromic management of sexually transmitted infections in Saudi Arabia. *J Infect Public Health.* 2011;4(2):73-79.
59. Rappold AG, Stone SL, Cascio WE, et al. Peat Bog Wildfire Smoke Exposure in Rural North Carolina Is Associated with Cardio-Pulmonary Emergency Department Visits Assessed Through Syndromic Surveillance. *Environ Health Perspect.* 2011. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21705297>. Accessed July 14, 2011.
60. Perry AG, Korenberg MJ, Hall GG, Moore KM. Modeling and syndromic surveillance for estimating weather-induced heat-related illness. *J Environ Public Health.* 2011;2011:750236.
61. Jossieran L, Caillère N, Brun-Ney D, et al. Syndromic surveillance and heat wave morbidity: a pilot study based on emergency departments in France. *BMC Med Inform Decis Mak.* 2009;9:14.
62. Smith S, Elliot AJ, Mallaghan C, et al. Value of syndromic surveillance in monitoring a focal waterborne outbreak due to an unusual *Cryptosporidium* genotype in Northamptonshire, United Kingdom, June - July 2008. *Euro Surveill.* 2010;15(33):19643.
63. Jossieran L, Fouillet A, Caillère N, et al. Assessment of a syndromic surveillance system based on morbidity data: results from the Oscour network during a heat wave. *PLoS ONE.* 2010;5(8):e11984.
64. Elliot AJ, Singh N, Loveridge P, et al. Syndromic surveillance to assess the potential public health impact of the Icelandic volcanic ash plume across the United Kingdom, April 2010. *Euro Surveill.* 2010;15(23). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20546694>. Accessed July 19, 2011.

65. Brand H, Krafft T. The Icelandic ash cloud and other erupting health threats: what role for syndromic surveillance? *Eur J Public Health*. 2010;20(4):367-368.
66. Kulldorff M. A spatial scan statistic. *Communications in Statistics-Theory and Methods*. 1997;26(6):1481-1496.
67. Tango T, Takahashi K, Kohriyama K. A space-time scan statistic for detecting emerging outbreaks. *Biometrics*. 2011;67(1):106-115.
68. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *Int J Health Geogr*. 2008;7:14.
69. Que J, Tsui F-C. Rank-based spatial clustering: an algorithm for rapid outbreak detection. *J Am Med Inform Assoc*. 2011;18(3):218-224.
70. Jiang X, Cooper GF. A Bayesian spatio-temporal method for disease outbreak detection. *J Am Med Inform Assoc*. 2010;17(4):462-471.
71. Shewhart W. *Economic control of quality of manufactured product*. New York: D. Van Nostrand Company Inc. 1931.
72. National Institute of Standards and Technology. NIST/SEMATECH e-Handbook of Statistical Methods. 2008. Available at: <http://www.itl.nist.gov/div898/handbook/index.htm>.
73. Rowlands RJ, Nix ABJ, Abdollahian MA, Kemp KW. Snub-Nosed V-Mask Control Schemes. *Journal of the Royal Statistical Society. Series D (The Statistician)*. 1982;31(2):133-142.
74. Lucas JM. A Modified "V" Mask Control Scheme. *Technometrics*. 1973;15(4):833-847.
75. Lucas JM, Saccucci MS, Robert V. Baxley J, et al. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*. 1990;32(1):1-29.
76. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V. Dead bird clusters as an early warning system for West Nile virus activity. *Emerging Infect. Dis*. 2003;9(6):641-646.
77. Kulldorff M. Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 2001;164(1):61-72.
78. Kulldorff M, Feuer EJ, Miller BA, Freedma LS. Breast Cancer Clusters in the Northeast United States: A Geographic Analysis. *Am. J. Epidemiol*. 1997;146(2):161-170.

79. Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. *Computational Statistics & Data Analysis*. 2003;42(4):665-684.
80. Stelling J, Yih WK, Galas M, et al. Automated use of WHONET and SaTScan to detect outbreaks of *Shigella* spp. using antimicrobial resistance phenotypes. *Epidemiol. Infect.* 2010;138(6):873-883.
81. Pelecanos AM, Ryan PA, Gatton ML. Outbreak detection algorithms for seasonal disease data: a case study using Ross River virus disease. *BMC Med Inform Decis Mak.* 2010;10:74.
82. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995;57(1):289-300.
83. Espino JU, Wagner M, Szczepaniak C, et al. Removing a barrier to computer-based outbreak and disease surveillance--the RODS Open Source Project. *MMWR Morb Mortal Wkly Rep.* 2004;53 Suppl:32-9.
84. Tsui F-C, Espino JU, Dato VM, et al. Technical Description of RODS: A Real-time Public Health Surveillance System. *J Am Med Inform Assoc.* 2003;10(5):399-408.
85. Anon. PCRE - Perl Compatible Regular Expressions. Available at: <http://www.pcre.org/>. Accessed July 19, 2011.
86. Fredkin E. Trie memory. *Commun. ACM.* 1960;3(9):490-499.
87. Carnevale RJ, Talbot TR, Schaffner W, et al. Evaluating the utility of syndromic surveillance algorithms for screening to detect potentially clonal hospital infection outbreaks. *J Am Med Inform Assoc.* 2011;18(4):466-472.
88. Mnatsakanyan ZR, Burkom HS, Coberly JS, Lombardo JS. Bayesian Information Fusion Networks for Biosurveillance Applications. *Journal of the American Medical Informatics Association.* 2009;16(6):855-863.
89. Jiang X, Cooper GF. A Recursive Algorithm for Spatial Cluster Detection. *AMIA Annu Symp Proc.* 2007;2007:369-373.
90. Que J, Tsui F-C. A Multi-level Spatial Clustering Algorithm for Detection of Disease Outbreaks. *AMIA Annu Symp Proc.* 2008;2008:611-615.
91. Hacek DM, Cordell RL, Noskin GA, Peterson LR. Computer-assisted surveillance for detecting clonal outbreaks of nosocomial infection. *Journal of Clinical Microbiology.* 2004;42(3):1170-1175.
92. Anon. Tennessee Department of Health: Influenza-Like Illness Surveillance Summary Archive. Available at: http://health.state.tn.us/TNflu_report_archive.htm. Accessed June 30, 2011.

93. Anon. The FreeBSD Copyright. Available at:
<http://www.freebsd.org/copyright/freebsd-license.html>. Accessed July 19, 2011.
94. Anon. SaTScan - Download for Windows. Available at:
http://www.satscan.org/download_satscan.html. Accessed July 19, 2011.
95. Anon. Software. Available at:
<http://www.autonlab.org/autonweb/downloads/software.html>. Accessed July 19, 2011.