

University of Louisville

## ThinkIR: The University of Louisville's Institutional Repository

---

Electronic Theses and Dissertations

---

5-2008

### Reliability and validity of EMS dispatch code-based categorization of emergency patients for syndromic surveillance.

Matthew Raymond Groenewold  
*University of Louisville*

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

---

#### Recommended Citation

Groenewold, Matthew Raymond, "Reliability and validity of EMS dispatch code-based categorization of emergency patients for syndromic surveillance." (2008). *Electronic Theses and Dissertations*. Paper 536. <https://doi.org/10.18297/etd/536>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

RELIABILITY AND VALIDITY OF EMS DISPATCH CODE-BASED  
CATEGORIZATION OF EMERGENCY PATIENTS FOR SYNDROMIC  
SURVEILLANCE

By

Matthew Raymond Groenewold  
B.A., Indiana University Southeast, 1997  
M.S.P.H., University of Louisville, 2002

A Dissertation  
Submitted to the Faculty of the  
Graduate School of the University of Louisville  
in partial Fulfillment of the Requirements  
for the Degree of

Doctor of Philosophy

School of Public Health and Information Sciences  
University of Louisville  
Louisville, Kentucky

May 2008

RELIABILITY AND VALIDITY OF EMS DISPATCH CODE-BASED  
CATEGORIZATION OF EMERGENCY PATIENTS FOR SYNDROMIC  
SURVEILLANCE

By

Matthew Raymond Groenewold  
B.A., Indiana University Southeast, 1997  
M.S.P.H., University of Louisville, 2002

A Dissertation Approved on

March 28, 2008

by the following Dissertation Committee:

---

Susan B. Muldoon, Ph.D., M.P.H., Dissertation Director

---

Frank D. Groves, M.D., M.P.H.

---

Adewale Troutman, M.D., M.P.H.

---

Robert R. Jacobs, Ph.D.

---

David J. Tollerud, M.D., M.P.H.

## **DEDICATION**

This dissertation is dedicated to my family, in particular my wife, Melissa. She has had as much of a hand in my accomplishments as me and has sacrificed more than I have to get me to this point.

To my son, Levi, as well, whom I can only hope to inspire as much as he inspires me.

I am also grateful to my parents, Mike and Gail Groenewold, for all of the opportunities they have provided me.

## **ACKNOWLEDGEMENTS**

I would like to thank my academic advisor, Dr. Susan Muldoon, for her guidance over the years. I would also like to thank Dr. Frank Groves for his helpful advice and comments on the earlier drafts of this dissertation, as well as the rest of the members of my dissertation committee for their time and interest in my work. I would also like to acknowledge the inspiration, guidance and encouragement I received from my professor and friend, Dr. Tim Aldrich.

## ABSTRACT

### RELIABILITY AND VALIDITY OF EMS DISPATCH CODE-BASED CATEGORIZATION OF EMERGENCY PATIENTS FOR SYNDROMIC SURVEILLANCE

March 28, 2008

A retrospective study involving the secondary analysis of public health surveillance records was undertaken to characterize the reliability and validity of an EMS dispatch data-based scheme for assigning emergency patients to surveillance syndromes in relation to two other schemes, one based on hospital ED clinicians' manual categorization according to patients' chief complaint and clinical presentation, and one based on ICD-9 coded hospital ED diagnoses.

Comparisons of a sample of individual emergency patients' syndrome assignments according to the EMS versus each of the two hospital categorization schemes were made by matching EMS run records to their corresponding emergency department patient encounter records. This new, linked dataset was analyzed to assess the level of agreement beyond chance between the three possible pairs of syndrome categorization schemes in assigning patients to a respiratory or non-respiratory syndrome and to a gastrointestinal or non-gastrointestinal syndrome. Cohen's kappa statistics were used to measure chance-adjusted agreement between categorization schemes (raters). Z-tests and a chi-square-like test based on the variance of the kappa statistic were used

to test the equivalence of kappa coefficients across syndromes, population subgroups and pairs of syndrome assignment schemes.

The sensitivity, specificity, predictive value positive and predictive value negative of EMS dispatch and chief complaint-based categorization schemes were also calculated, using the ICD-9-coded ED diagnosis-based categorization scheme as the criterion standard. Comparisons of all performance characteristic (i.e. sensitivity, specificity, predictive value positive and predictive value negative) values were made across categorization schemes and surveillance syndromes to determine whether they were significantly different.

The use of EMS dispatch codes for assigning emergency patients to surveillance syndromes was found to have limited but statistically significant reliability in relation to more commonly used syndrome grouping methods based on chief complaints or ICD-9 coded ED diagnoses. The reliability of EMS-based syndrome assignment varied significantly by syndrome, age group and comparison rater. When ICD-9 coded ED diagnosis-based grouping is taken as the criterion standard of syndrome definition, the validity of EMS-based syndrome assignment was limited but comparable to chief complaint-based assignment. The validity of EMS-based syndrome assignment varied significantly by syndrome.

## TABLE OF CONTENTS

	PAGE
DEDICATION.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF TABLES.....	ix
INTRODUCTION .....	1
Syndromic Surveillance.....	2
EMS-Based Syndromic Surveillance.....	9
Louisville Metro Syndromic Surveillance Systems .....	12
Evaluation of Syndromic Surveillance Systems .....	17
METHODS AND MATERIALS .....	27
Statistical Analysis Plan .....	28
Setting.....	30
Data .....	32
Linkage of Hospital and EMS Records.....	33
Measurement of Interrater Agreement .....	33
Comparison of Kappa Statistics .....	35
Sensitivity, Specificity and Predictive Value .....	36
Comparison of Performance Characteristics.....	37
Power and Sample Size Considerations.....	38
RESULTS .....	41



Summary.....65  
DISCUSSION.....69  
LIMITATIONS .....82  
CONCLUSIONS.....52  
REFERENCES .....87  
CURRICULUM VITAE .....102

## LIST OF TABLES

TABLE	PAGE
1. Description of LMPHW Surveillance Syndromes: Patients Assigned Based on Chief Complaint Clinicians' Initial Impression .....	14
2. CDC-Defined Surveillance Syndromes Based on Consensus Groupings of ICD-9 Diagnostic Codes.....	15
3. EMS Dispatch Categories (Syndromes) Routinely Monitored by LMPHW .....	16
4. Summary of the Results of Five Studies Evaluating the Reliability and Validity of Surveillance Syndromes .....	23
5. Categorization of a Series of Emergency Patients to Two Surveillance Syndromes by EMS Dispatch Code and by Chief Complaint .....	43
6. Categorization of a Series of Emergency Patients to Two Surveillance Syndromes by EMS Dispatch Code and by ED Diagnosis .....	44
7. Categorization of a Series of Emergency Patients to Two Surveillance Syndromes by Chief Complaint and by ED Diagnosis.....	45
8. Respiratory and Gastrointestinal Syndrome Assignment (Trait Prevalence) by Three Categorization Schemes (Raters) and Significance of Tests for Marginal Homogeneity for Scheme Pairings .....	46
9. Pairwise Agreement Between Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories) .....	48

10. Comparison of Gender-Stratified Estimates of the Kappa Coefficient for Three Pairings of Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories) .....	49
11. Comparison of Age-Stratified Estimates of the Kappa Coefficient for Three Pairings of Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories) .....	50
12. Comparison Across Syndromes of Kappa Estimates for Three Pairs of Three Different Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories).....	51
13. Comparison Across Rater Pairs of Kappa Estimates for Pairs of Categorization Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes .....	53
14. Post Hoc Comparisons of Kappa Estimates for Pairs of Categorization Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes .....	54
15. Agreement of Three Schemes for Assigning Emergency Patients to Surveillance Syndromes Using Fleiss' Kappa for Multiple Raters .....	56
16. Agreement of Three Schemes for Assigning Emergency Patients to Surveillance Syndromes Using Gwet's AC1 Statistic .....	56
17. Performance Characteristics of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to Surveillance Syndromes, Compared to a Diagnosis-Based Criterion Standard .....	57

18. Comparison Across Syndromes of Performance Characteristics of an EMS Dispatch Code-Based Scheme for Assigning Emergency Patients to Surveillance Syndromes.....	59
19. Comparison Across Syndromes of Performance Characteristics of a Hospital ED Chief Complaint-Based Scheme for Assigning Emergency Patients to Surveillance Syndromes.....	60
20. Comparison of the Sensitivity and Specificity of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Respiratory Surveillance Syndrome.	62
21. Comparison of the Positive and Negative Predictive Values of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Respiratory Surveillance Syndrome.....	63
22. Comparison of the Sensitivity and Specificity of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Gastrointestinal Surveillance Syndrome .....	64
23. Comparison of the Positive and Negative Predictive Values of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Gastrointestinal Surveillance Syndrome.....	65
24. Comparison of Levels of Agreement (Kappa) Between Syndrome Categorization Methods Reported in the Literature with the Current Study...	77
25. Comparison of the Sensitivities of Various Schemes for Assigning Emergency Patients to Respiratory or Influenza-Like Illness (ILI) Syndromes Reported in the Literature with the Current Study.....	79

## INTRODUCTION

Concern over the possibility of large-scale bioterrorism has increased greatly since the anthrax attacks of September and October 2001. (1,2) In response to this concern, as well as concerns over public health crises caused by newly emerging or reemerging infectious threats such as West Nile virus, (3) SARS, (4) human H5N1-type influenza infections (5) and multi-drug resistant infections, (6) the public health community has sought to develop and deploy new, nontraditional surveillance methodologies that could provide early warning of such events. (7-10) Such an early warning system would alert authorities to the existence of an outbreak very early in its course, allowing for more rapid intervention.

There are at least two main reasons why traditional surveillance systems, exemplified by passive, diagnosis-based disease reporting, are generally not considered suitable for early outbreak detection. First, they are not timely enough to allow for the implementation of control measures when they would be most effective in limiting morbidity and mortality. Disease-reporting systems rely on a diagnosis, which generally requires laboratory confirmation. Because some lab tests or cultures require days or even weeks to become positive, reliance on lab-confirmed, diagnosis-based disease reporting could result in a delayed response to an outbreak. (7,11)

Second, they are thought to be overly specific and to lack the sensitivity that would likely be required to detect intentional outbreaks or those of unknown origin. (11,12) Practitioner awareness of diseases caused by the most likely potential bioterrorism agents and of emerging infectious diseases is limited. Additionally, many of these diseases have nonspecific prodromes—characterized, for example, by fever, chills, malaise and myalgia—that are similar to those of other, more common illnesses. Consequently, affected people presenting to healthcare providers may initially be misdiagnosed. Further, data from disease-reporting systems is often incomplete. Practitioners do not always report reportable diseases and often do not order the tests required for laboratory confirmation of diagnoses, either because they do not suspect the disease or because the test is deemed unnecessary in the context of the clinical care of an individual patient. (11,12)

### **Syndromic Surveillance**

Defined variously as “surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response” (13) and “the ongoing, systematic collection, analysis, interpretation and application of real-time (or near-real-time) indicators of diseases and outbreaks that allow for their detection before public health authorities would otherwise note them,” (14) syndromic surveillance has emerged as the most common alternative to traditional surveillance systems for early outbreak detection. (15-21) Historically, syndromic surveillance has been used as

a tool for case identification and management, particularly for sexually transmitted infections in developing countries or other settings where a lack of resources makes laboratory confirmation impractical or impossible. (22-25) In more recent years, however, its application to the early detection of outbreaks has been increasingly explored and adopted. (15-21,26-28)

During the 1990's, the convergence of two factors served to substantially shift the paradigm for public health surveillance from one of passive, highly specific, diagnosis-based and often paper-based systems intended mostly for monitoring secular disease trends toward one of automated, highly sensitive, syndrome-based systems intended primarily for the early recognition of disease outbreaks. (7,8,11,27,29) The first was the rapid proliferation of advancements in information technology, making possible the rapid and efficient capture, transfer, storage and analysis of large amounts of data from a variety of new as well as traditional sources and moving the standard of timeliness closer to real-time. (11,27,30-33) The second was the perception of a generally increased threat of biological and chemical terrorism, (34-36) coupled with anxiety over emerging and reemerging diseases, (37-39) making the very early detection of intentionally caused or naturally occurring epidemics of paramount concern. (7)

This paradigm shift was exemplified at the time by national-level efforts to develop a standardized informatics infrastructure for health data (33) and to move from a patchwork approach to surveillance to a more integrated one in National Electronic Disease Surveillance System. (40) Research efforts were also directed toward the development of important aspects of syndromic

surveillance such as syndrome definition and validation (11) with an eye toward the eventual implementation of a national-level epidemic early warning system with a nation-wide (or, at least, nationally representative) data catchment. (28) In its Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE), the DoD at least partially realized this goal for military treatment facilities. (41,42)

Prior to September 11, 2001, the syndromic approach had been adapted to outbreak detection on a localized level as well. Temporary or “drop-in” syndromic systems were used to provide enhanced surveillance capabilities for high profile events such as the Olympic Games in Atlanta in 1996, (27) the meeting of the World Trade Organization (WTO) in Seattle and the Super Bowl in Miami in 1999, (27) the Democratic and Republican National Conventions in 2000 (27) and the Sydney Olympics (43) and the G8 Summit in Fukuoka and Miyazaki, Japan, in 2000. (44) Immediately following the September 11<sup>th</sup> terrorist attacks, drop-in syndromic surveillance systems were established in New York City in anticipation of secondary, biological attacks. (45,46)

The events of September and October 2001 stimulated the widespread and urgent adoption of syndromic surveillance methodologies by local, (17,19,27,36,47) state (27,47) and national (27,28) public health jurisdictions. This explosion of activity had the effect of, on the one hand, making syndromic surveillance systems much more common and, driven by the perception of necessity, of producing a diversity of innovative approaches to data sources, collection and transmission, as well as statistical methodologies. (8,9,48) On the



other hand, it has produced a patchwork of disconnected, often localized systems surveilling different syndromes using different measures and analytical approaches rather than the standardized, nationwide system that had once been envisioned. (33) A notable exception to this is the Centers for Disease Control and Prevention (CDC) BioSense system. (28) Currently, the use of syndromic surveillance has been sufficiently institutionalized to warrant an annual national conference (49) and an online journal (50) dedicated to the subject.

The hallmark of the syndromic approach to public health surveillance is the use of nonspecific, often pre-diagnostic indicators as the observations on which surveillance is performed. These systems are intended to identify cases, not of specifically diagnosed diseases, but which occur in relatively broad, predefined categories of symptomology, referred to as syndromes. (16,48)

The processing of syndromic surveillance data for outbreak detection has been described as consisting of four methodological stages: the syndrome grouping stage, the modeling stage, the detection stage and the alert stage. (51,52) In this first stage, the syndrome grouping stage, individual observations from the source data stream are assigned to particular groups or syndromes, according to a set of implicit or explicit categorization rules. As constructs, each of these syndromes is intended to comprise the constellation of observations that would capture—if not all—then at least the majority of presentations by which the conditions it is intended to represent might manifest themselves.

The way in which the syndrome grouping stage is operationalized depends on the diseases one wishes to detect and the source data stream

available for surveillance (i.e. the types of observations available). Currently, the syndromes most commonly under surveillance are designed to capture the majority of initial clinical presentations associated with potential bioterrorism agents. (8-11) However, they have also been designed to monitor food-borne disease outbreaks, (53) heat-related morbidity (54) and influenza and influenza-like illnesses (ILI). (55-57)

In terms of data sources, collection and categorization, some syndromic surveillance systems, particularly short-duration, event-based systems, require providers to manually input syndrome data, which is labor intensive and represents a significant burden on the participating institutions. (45,46,59) Such systems are difficult to maintain on a 24 hour per day, seven day per week basis and are generally not sustainable for long periods of time. (27) Increasingly, however, syndromic surveillance systems make use of data from nontraditional sources, including existing sources of data that are routinely collected for other purposes and which are often available more quickly, frequently and easily than traditional surveillance data, allowing them to operate in real or near-real time. (32,33,57,59) This surveillance methodology is, therefore, considered more likely to be able to detect the occurrence of cases of bioterrorism-associated disease in their prodromal phase, when they would otherwise be indistinguishable from other, more common diseases. (60)

Some syndromic surveillance systems make use of hospital admissions data (58,59) or clinical data provided by health plans reflecting urgent care, clinic, primary care provider or other outpatient, ambulatory care setting encounters.

(32,55,56,61-67) Data for most syndromic surveillance systems, however, come from hospital emergency departments. (24,68-75) Patients from all of these settings must be categorized into syndromes either manually by a clinician, typically a triage nurse or ED physician, (45,46,59) or automatically, using preexisting electronic ED data such as free-text descriptions of chief complaint. (65,68,70,71,74,75) The automated use of electronic triage logs or other chief complaint data may be facilitated by computerized natural language processors, which map key words and phrases within text strings to syndromic categories using Bayesian or probabilistic algorithms. (71,76,77)

Although some definitions of syndromic surveillance focus on the prediagnostic nature of the data that are used, a number of systems actually make use of ICD-9 diagnostic codes to categorize patients into syndromes, particularly from ambulatory care settings, including emergency departments. (68,77-80) These data may be more reliable than other, prediagnostic types, and, if they are available early (e.g. at the time of release from the ED), they may also be considered timely enough to serve as the basis for an early warning system. (51,81,82) Indeed, in an effort to standardize the definition of syndromes that make use of them, a CDC working group published consensus groupings of ICD-9 codes that could be considered to constitute various syndromes for bioterrorism surveillance. (83)

The characteristic adaptation of syndromic surveillance to early outbreak detection—the application of statistical algorithms to the number of cases (or events) occurring within the syndromes being surveilled in order to detect

significant upward departures from expected incidence and to notify public health officials of the need for follow-up investigation of such signals—comprises the last three methodological stages in the processing of syndromic data. (51,52) In the second stage—the modeling stage—expected or baseline incidence within each syndrome is established based on historical data. This may be as simple as averaging historical event counts over some specified period prior to the point of analysis or it may involve sophisticated time series models that take account of trend and seasonality in the data. (84,85)

The next stage is the detection stage, in which daily (or some other time interval, e.g. hourly, shift or weekly) observed syndrome counts are compared against expected frequencies and a statistical determination is made regarding the significance of the difference between the two values. There are a wide variety of statistical methods for detecting aberrations in time series data and many have been applied to syndromic surveillance, including methods from the fields of epidemiology, statistical process control, signal processing and data mining. (86) Examples include the use of spatial and temporal scan statistics (87-92) and other models of interpoint distance distribution, (93) epidemic thresholds, (94) multi-item gamma Poisson shrinkage estimation (95) and cusum and other control charts. (94,96-98)

In the final, or alert, stage, the magnitude or significance (or both) of any nonrandom signals in the data is compared against a preset alert threshold to evaluate whether further scrutiny of the cluster of observations that comprised the signal is warranted. (51,52) If, upon further scrutiny, which may include

additional analyses and the collection of additional data, a judgment may be made that an actual outbreak is occurring, resulting in the initiation further epidemiological field investigation and outbreak countermeasures. (99-101)

## **EMS-Based Syndromic Surveillance**

While the majority of syndromic surveillance systems use clinical, patient encounter data from ambulatory care settings, there are a number of surveillance systems for early outbreak detection that make use of non-clinical data to monitor the frequency of certain categories of health services or other health-related events. Examples include systems based on data from Medicaid prescriptions, (102) the sale of over the counter pharmaceuticals and certain medical items, (103-109) calls to poison control centers (110,111) and other medical call lines (112-115), school absenteeism, (116) and Emergency Medical Services (EMS) runs. (117-119) Such non-clinical data are generally used because, for one reason or another, they are more readily available to surveillance system operators or researchers than clinical data and are typically taken to be proxy measures of the occurrence of actual, clinical events. For example, one published account has described the use of hospital parking facility volume as a proxy for patient visits. (120)

Because of their reliably close correlation with actual, clinical patient encounters, their focus on cases of acute illness, their relatively easy accessibility and their rapid availability in relation to the clinical event they are intended to represent, EMS dispatch data have often been seen as a logical

choice of nontraditional, non-clinical, proxy data upon which to perform syndromic surveillance. While, currently, there are comparatively few articles in the peer-reviewed literature that describe or refer specifically to EMS-based surveillance systems, a review of abstracts submitted to a national conference on syndromic surveillance held in 2003 revealed that 10% (n = 6) cited the use of EMS data for syndromic surveillance. (121) Additionally, the widespread use (currently used in 73 North American jurisdictions) of commercially available software designed specifically for EMS-based surveillance speaks to this methodology's general acceptability in the surveillance community. (122)

In New York City, the local health department had implemented a syndromic approach to bioterrorism surveillance using EMS dispatch data three years before the attacks of 2001. The system monitored the daily volume of certain types of ambulance request calls intended to be representative of influenza-like illnesses. (117) After 2001, EMS-based surveillance systems were established in other jurisdictions in the United States (121,122); Canada (122,123) and Europe. (118)

EMS-based systems offer several potential advantages over other sources of syndromic surveillance data. First, EMS-based systems can take advantage of the preexisting informatics infrastructure that exists in the form of emergency dispatch and communication systems. (124) These systems not only make possible the automatic, electronic collection, aggregation and transmission of data, they also provide a ready-made method for syndrome categorization in the nature-of-call-based dispatch codes that are routinely assigned to ambulance

runs. (117,124) Also, the EMS data stream may be easier for public health authorities to access than hospital data streams, particularly when EMS services and/or their dispatch centers are government operated.

Second, EMS systems can offer high-percentage monitoring coverage of a population in a well-defined catchment area, at least for acute morbidity. This is particularly true when the population is served by a single EMS provider. Additionally, the administrative boundaries of EMS and local public health agencies are often congruent (e.g. city, county), which can make data analysis and interpretation easier.

Finally, monitoring requests for EMS service brings the surveillance activity closer in time to the events it is intended to detect (i.e. onset of symptoms), thereby potentially providing earlier warning of an outbreak and greater lead time for the implementation of countermeasures. This advantage is complemented by the fact the EMS services and their dispatch centers operate on a 24 hour basis, making real-time data analysis possible.

EMS-based surveillance systems also have potential disadvantages. These include the fact that patients usually call on EMS in case of severe, acute symptoms, which would tend to exclude clinical presentations characteristic of the early, prodromal phases of bioterrorism-related diseases. Also, the validity of using dispatch codes as proxy syndrome definitions has not been well established.

## **Louisville Metro Syndromic Surveillance Systems**

Like many state and local health departments, the Louisville Metro Department of Public Health and Wellness (LMPHW)—then the Louisville Metro Health Department—first began looking at syndromic surveillance in response to the events of September 11, 2001 and the subsequent anthrax attacks in September and October of that year. Also like many other health departments, it first implemented syndromic surveillance as an event-based, “drop-in” system intended to provide enhanced public health protection during a specific period when the risk of or vulnerability to biological attack was perceived to be increased. For Louisville, that period was the weeks surrounding the 2002 Kentucky Derby and Kentucky Derby Festival. (125)

During this four-week period, LMPHW collected and analyzed emergency department encounter data from 12 local hospitals, categorized into seven predetermined syndromes. (125) During the same period, LMPHW also began surveilling data streams from other non-traditional sources, including the analysis of coroner case and EMS dispatch data. (126) While the coroner and EMS data-based surveillance systems have operated continuously since their inception, the ED data-based system continued to be operated as an event-based, drop-in system focused around the Kentucky Derby Festival during 2003 and 2004. During that period, participation on the part of local hospitals steadily declined. In 2005, however, LMPHW implemented a new, continuously operating syndromic surveillance system based on the analysis of ED patient encounter data from five sentinel hospitals. (126)



Currently, LMPHW continuously monitors community patterns of acute morbidity by surveilling hospital ED and EMS dispatch data. The emergency department data come from the five Louisville hospitals of the Norton Healthcare system: Audubon, Kosair Children's, Norton, Southwest and Suburban. The EMS dispatch data are provided electronically to the health department in real time by MetroSafe—the Louisville Metro emergency communications center—and include records of all requests for emergency ambulance service responded to by Louisville Metro Emergency Medical Services (LMEMS), a municipal, third-service emergency ambulance provider serving Louisville-Jefferson County.

(126)

At each of the five participating emergency departments, all patients are manually assigned by a clinician to one seven locally defined syndromes (Table 1), based on their chief complaint and clinical presentation. These data are transmitted to the health department electronically on a daily basis. Two of these hospitals, Norton and Kosair Children's hospitals, also provide ICD-9 coded diagnoses. These data are further used by the health department to separately assign each patient from these two hospitals to one of ten CDC-defined (83) syndromes (Table 2). (126)

**Table 1.****Description of LMPHW Surveillance Syndromes: Patients Assigned Based on Chief Complaint Clinicians' Initial Impression**

<b>Syndrome</b>	<b>Description</b>
Cardiac	Patient whose clinical presentation is suggestive of cardiac involvement including patients with chest pain (radiating or non-radiating), arrhythmia, hypotension without accompanying evidence of hypovolemia, etc.
Gastro-Intestinal (GI)	Patient with clinical presentation suggestive of GI tract involvement including patients with recent history or signs and symptoms of severe diarrhea, either watery or bloody or with accompanying signs or symptoms of dehydration (e.g. poor skin turgor, thirst, hypotension etc.) as well as patients with nausea and vomiting etc.
Infectious Disease (ID)	Patient whose clinical presentation is suggestive of an infectious process but whose primary symptomology is not readily referable to one of the organ systems in the other syndromes. This would exclude pneumonia with respiratory distress (which would be Resp) or gastro-enteritis accompanied by fever (which would be GI) but would include patients with high fever, myalgia, headache and/or sore throat, weakness, prostration and/or listlessness, arthralgia, lymphadenopathy and cough as well as other Influenza-like illness (ILI) presentations. This would also include patients with fever and concurrent finding of skin lesions, especially vesicles or pustules as well as patients with septicemia or other systemic infections.
Neurological (Neuro)	Patient whose clinical presentation is suggestive of nervous system involvement including seizures, paralysis (including flaccid paralysis), paraesthesia, hemiparesis etc. This would also include patients with signs and symptoms suggestive of CNS infection such as: fever with intense headache, stiff neck and/or altered level of consciousness.
Psychiatric (Psych)	Patient whose clinical presentation is suggestive of a psychiatric or psychological disorder.
Respiratory (Resp)	Patient whose clinical presentation is suggestive of respiratory system involvement including dyspnea, respiratory distress, abnormal breath/lung sounds (including stridor), intercostal or suprasternal retractions, non-cardiogenic pulmonary edema (ARDS), cyanosis without evidence of inadequate circulation or perfusion, pneumonia, etc.
Other	Patient whose clinical presentation is not consistent with any of the above syndromes. For example, trauma, obstetric, gynecologic or ophthalmologic presentations, obvious drug overdose, apparent metabolic disorders, genitourinary complaints (including UTI), etc.

**Table 2.**

**CDC-Defined Surveillance Syndromes Based on Consensus Groupings of ICD-9 Diagnostic Codes**

<u>Syndrome Description</u>
Botulism-Like Syndrome
Fever
Gastrointestinal (Upper and Lower)
Hemorrhagic Illness
Lesion
Lymphadenopathy
Rash
Respiratory
Severe Illness/Death

Individual EMS runs are categorized for syndromic analysis on the basis of the ProQA emergency medical dispatch code, an alphanumeric code that indicates the nature of the call and is assigned to each run by the dispatcher using a standardized protocol. ProQA (127) is a commercially available emergency medical dispatch (EMD) software package that guides dispatchers through structured caller interrogation and instruction procedures based on Medical Priority Dispatch System (MPDS) incident type/chief complaint-centered protocols promulgated by the National Academy of Emergency Medical Dispatch (NAEMD). (128) Following the ProQA prompts, the dispatcher ultimately determines the appropriate incident response level and structure, which is associated with a dispatch determinant code. LMPHW routinely monitors the number of runs occurring in eight of 34 major dispatch categories defined by the first two characters of the ProQA dispatch code (Table 3).

Table 3.

**EMS Dispatch Categories (Syndromes) Routinely Monitored by LMPHW**

ProQA CODE	DESCRIPTION
01	<b>ABDOMINAL PAIN / PROBLEMS</b>
02	ALLERGIES (REACTIONS) / ENVENOMATIONS
03	ANIMAL BITES / ATTACKS
04	ASSAULT / SEXUAL ASSAULT
05	BACK PAIN (NON-TRAUMATIC OR NON-RECENT)
06	<b>BREATHING PROBLEMS</b>
07	BURNS (SCALDS) / EXPLOSION
08	CARBON MONOXIDE / INHALATION / HAZMAT
09	<b>CARDIAC OR RESPIRATORY ARREST / DEATH</b>
10	<b>CHEST PAIN</b>
11	CHOKING
12	<b>CONVULSIONS / SEIZURES</b>
13	DIABETIC PROBLEMS
14	DROWNING (NEAR) / DIVING / SCUBA ACCIDENT
15	ELECTROCUTION / LIGHTNING
16	EYE PROBLEMS / INJURIES
17	FALLS
18	HEADACHE
19	HEART PROBLEMS / A.I.C.D.
20	HEAT / COLD EXPOSURE
21	HEMORRHAGE / LACERATIONS
22	INDUSTRIAL / MACHINERY ACCIDENTS
23	OVERDOSE / POISONING (INGESTION)
24	PREGNANCY / CHILDBIRTH / MISCARRIAGE
25	PSYCHIATRIC / ABNORMAL BEHAVIOR / SUICIDE ATTEMPT
26	<b>SICK PERSON (SPECIFIC DIAGNOSIS)</b>
27	STAB / GUNSHOT / PENETRATING TRAUMA
28	STROKE / CVA
29	TRAFFIC / TRANSPORTATION ACCIDENTS
30	TRAUMATIC INJURIES
31	<b>UNCONSCIOUS / FAINTING (NEAR)</b>
32	<b>UNKNOWN PROBLEM (MAN DOWN)</b>
33	TRANSFER / INTERFACILITY / PALLIATIVE CARE
50	MVA

Categories in **BOLD** are routinely monitored by LMPHW

Syndromic data from each of the three sources—hospital clinician, ICD-9 coded diagnosis and EMS dispatch code—are analyzed separately and comprise three distinct surveillance systems. Because they surveil overlapping but

nevertheless different populations and categories of health events, operationalized in different ways, the three systems provide separate perspectives on acute morbidity patterns and are intended to be complementary. (126)

All three systems use the same signal detection algorithm. A one-sided, positive cusum analysis is conducted for event counts within each syndrome to detect significant departures from expected incidence on a daily basis. (126) When an alert is signaled by the algorithm, additional, follow-up analyses of the cluster of health events associated with the signal are carried out. (126)

A surveillance alert in itself, however, is not considered to be a positive indicator of the existence of an outbreak. Rather, it serves to indicate that the occurrence of health events within a particular category warrants further scrutiny and closer monitoring. Results of the follow-up analyses and other information, including concomitant alerts in other surveillance systems and clinical or other, non-syndromic public health evidence of the existence of an outbreak, are considered together when deciding whether to initiate an epidemiological field investigation or other public health measures. (126)

### **Evaluation of Syndromic Surveillance Systems**

The widespread use of syndromic surveillance systems for early outbreak detection and response is a relatively recent development, having been stimulated by the anthrax attacks of 2001 and a generally increased perception of threat of bioterrorism. The usefulness of such systems for this purpose has yet to

be definitively established, however. (14,26,121,129,130) Because the development and operation of syndromic surveillance systems can involve significant expense in terms of both financial and other resources, and because the investigation of false alarms that might be generated by such systems can also be costly, research is needed that will help establish, not only the utility of syndromic surveillance broadly, but also the relative value of different data sources and analytic approaches. (14,26,121,129,130)

The need to evaluate the performance of syndromic surveillance systems had been recognized early in the course of its widespread deployment. (14,26,121,129) However, cases of bioterrorism related illness have been extremely rare in the United States, notwithstanding the anthrax attacks of 2001. In particular, the types of incidents that syndromic surveillance was originally designed to detect—large-scale outbreaks resulting from widespread exposure to a biological agent—have not, it can be argued, occurred at all. This lack of authentic outbreak data for reference has complicated the evaluation of the detection capabilities of syndromic surveillance systems. (52) As a result, researchers have either examined authentic data sets for signals of naturally occurring outbreaks that can be independently confirmed (52) or, more commonly, have relied on computer simulations, (52,131-133) both of which have their own limitations.

Consistent with general guidelines for the evaluation of surveillance systems, (134) the CDC has published a provisional framework for evaluating surveillance systems that are employed for early outbreak detection. (135) The

framework outlines five main areas that should be addressed in a comprehensive evaluation of an early outbreak surveillance system: a detailed description of the system, outbreak detection, system experience and conclusions and recommendations. It is the outbreak detection aspect of system evaluation, however, which has received the most emphasis and attention.

The authors of the framework and others have pointed out that the key features of a useful syndromic surveillance system are timeliness and adequate performance characteristics (sensitivity, specificity, PVP and PVN) for outbreak detection. Systems that are not either timely or sensitive enough may result in delayed outbreak detection and consequent increases in morbidity and mortality. Systems that lack adequate specificity, however, may lead to costly and inefficient investigations of false alarms. Consequently, the majority of evaluative research on syndromic surveillance has focused either on the timely availability of syndromic data from novel sources or on validating the many statistical approaches that are used to distinguish the signal of an outbreak from the background noise of natural variation in the data. When these studies provide quantitative data on system performance characteristics, therefore, they do so in terms of outbreak detection.

Validation of a syndromic signal detection method, however, requires that the observations that comprise a signal have been accurately assigned to the category in which they are being analyzed. In other words, the syndrome grouping method must be valid. In most studies, particularly simulation studies,

however, this assumption is accepted axiomatically, without prior validation of the methods used to define or construct the syndromes.

Such validation of syndromes as constructs, especially in terms of the relative reliability of different methodologies for assigning individual patients to or grouping data into syndrome categories has received little research attention in relation to the variety of syndromes being employed and the different ways in which they are operationalized. The ability of a syndromic surveillance system to detect an outbreak of a particular type of disease is dependent on its ability to reliably and accurately classify cases of clinical illness into the syndrome that has been constructed to represent that type of disease. (51) That is to say, the meaningfulness of a statistical signal within a syndrome depends on the relationship between the observable characteristics of a particular disease and the set of observations that are understood to constitute that syndrome. Additionally, the comparability of information from syndromic surveillance systems that utilize different syndrome categorization methodologies depends on the consistency of the meaning of the syndrome construct across methodologies.

Of the published studies that have compared syndrome categorization methods, nearly all have compared chief complaint and ICD-9 diagnostic code-based schemes. (52,77,79,81,82,136-138) One study included an additional comparison with a classification scheme based on a combination of the two. (52) Additionally, one study compared categorization schemes based on chief complaints, diagnostic codes and manual assignment by clinicians. (82)



While the objective of one of these studies was to validate a comprehensive set of syndromes, (81) those most often evaluated were respiratory and gastrointestinal syndromes. All of these studies quantified the performance of the categorization methods in terms of either agreement between the different methods, using the kappa statistic, (81,82) or in terms of sensitivity and predictive value positive (PVP) versus a criterion standard of either syndrome definition (68) or outbreak detection. (52)

Generally speaking, ICD-9 diagnostic code-based syndrome grouping methods were found to be more accurate and reliable, if not more timely, than chief complaint-based methods, however estimates of kappa and sensitivity varied considerably both between and within the studies. For example, Fleischuaer et al. reported kappa estimates of 0.28, 0.59 and 0.33 and of 0.70, 0.63 and 0.71 for the respiratory and gastrointestinal syndromes respectively across their three comparisons, (82) while Begier et al. reported chance adjusted measures of agreement between chief complaint and ICD-9 diagnosis code-based methods of 0.68 for both the respiratory and gastrointestinal syndromes. (81) Beitel et al. reported criterion standard-based sensitivities of 0.47 for a chief complaint based syndrome grouping method for a respiratory syndrome versus 0.56 and 0.87 respectively for diagnosis-based upper and lower respiratory infection syndromes in a pediatric setting. (68) They reported a sensitivity of 0.72 for a classification scheme based on a combination of chief complaint and diagnosis code data. (68) Reis and Mandl reported respiratory syndrome outbreak detection sensitivities of 0.26, 0.28 and 0.34 for classification methods

based on chief complaint, ICD-9 diagnosis code and a combination of both, respectively for one hospital and 0.36, 0.39 and 0.47 for another. (51)

Only one published study has attempted to validate a syndrome categorization scheme based on EMS dispatch data. The performance of the EMS-based categorization scheme was not compared to another classification method. Rather, the criterion standard for syndrome assignment was based on a post hoc, retrospective, chart review-based classification by a blinded panel of physicians. The syndrome being evaluated was an influenza-like illness (ILI) syndrome based on a combination of four call types (i.e. dispatch codes). The authors reported a syndrome definition sensitivity of 0.58 and a PVP of 0.22 in relation to the criterion standard. (139) A summary of the results of the syndromic surveillance evaluation studies most relevant to the current analysis is given in Table 4.

**Table 4.**

**Summary of the Results of Five Studies Evaluating the Reliability and Validity of Surveillance Syndromes**

<b>Study</b>	<b>Syndromes evaluated</b>	<b>Data Sources/Syndrome Grouping Methods Compared</b>	<b>Metrics Used</b>	<b>Reported values</b>
Begier et al., 2003	Death, Sepsis, Rash, Respiratory, Gastrointestinal, Unspecified Infection, Neurologic, Other, Overall System	Chief Complaint, ICD-9 Coded Diagnosis	Kappa	0.63, 0.11, 0.58, 0.68, 0.68, 0.42, 0.09, 0.66, 0.64
Beitel et al., 2004	Respiratory	Chief Complaint, ICD-9 Coded Diagnosis, Combination of the Two	Sensitivity (vs. criterion standard based on Hx, PE, lab results)	Chief Complaint: 0.47 ICD-9 (upper resp): 0.56 ICD-9 (lower resp): 0.87 Combination: 0.72
Fleischauer et al., 2004	Respiratory, Gastrointestinal, Rare Syndromes, None, Overall System	Clinical Impression Recorded on Surveillance Form, Retrospective Classification of Chief Complaint, ICD-9 Coded Diagnosis	Kappa	Surveillance Form vs. Chief Complaint: 0.28, 0.70, 0.24, 0.43, 0.48 Surveillance Form vs. ICD-9: 0.59, 0.63, 0.31, 0.52, 0.55 Chief Complaint vs. ICD-9: 0.33, 0.71, 0.19, 0.50, 0.52
Greenko et al., 2003	Influenza-Like Illness	EMS Dispatch Codes	Sensitivity, PVP (vs. criterion standard based on retrospective review of medical record)	0.58, 0.22
Reis and Mandl, 2004	Respiratory	Chief Complaint, ICD-9 Coded Diagnoses, Both	Sensitivity (for outbreak detection, based on simulated outbreaks)	Hospital 1--Chief Complaint: 0.26 ICD-9: 0.28 Both: 0.34 Hospital 2--Chief Complaint: 0.36, ICD-9: 0.39 Both: 0.47

Here, I am interested in the degree to which the categorization of emergency patients for syndromic surveillance based on EMS dispatch codes agrees with their syndrome assignment upon arrival in the hospital emergency department, based both on manual clinician assignment and on diagnoses, which is considered the criterion standard. For this study, I have taken advantage of the fact that, for a particular subset of emergency patients whose encounter data are captured by LMPHW's syndromic surveillance systems—those who are transported by LMEMS ambulance to either Norton or Kosair Children's hospitals, their individual syndrome assignments under each of the three categorization schemes employed by LMPHW are available.

Using these data, I propose characterize the validity of an EMS dispatch code-based scheme for assigning emergency patients to surveillance syndromes by quantifying the degree to which it agrees with chief complaint and ICD-9 diagnosis code-based schemes and by describing its performance characteristics (i.e. sensitivity, specificity, predictive value positive and predictive value negative) with respect to a criterion standard based on syndrome categorization according to CDC-defined ICD-9 diagnosis code groups. I further propose to test these measurements for significant differences across comparison settings, syndromes and population subgroups.

The validity of syndromic grouping methods is important because they have been shown to substantially affect the performance characteristics of the overall surveillance system. (51,77,79,81) Additionally, identifying systematic differences between categorization methods based on different data sources can

inform decisions about which methods should be used and can inform the interpretation of information from different surveillance systems. In particular, agreement between ambulance dispatch based and other syndromic grouping methods is important when considering the comparability of information from these different syndromic surveillance systems.

This is especially relevant in states like Kentucky where the expansion of syndromic surveillance is considered to be of importance, but where many local health departments do not have data sharing agreements with local hospitals, lack the informatics infrastructure required for data transfer or have no hospital within their jurisdictions. For these local, often rural, health departments, ambulance dispatch data may be the only viable data stream on which to perform syndromic surveillance.

Using data from LMPHW's syndromic surveillance systems, this study will address the following research questions:

1. Does assignment of emergency patients to surveillance syndromes using a categorization scheme based on EMS dispatch data agree with syndrome categorization according to chief complaint and ED diagnosis-based schemes?
2. Do the chance-adjusted levels of agreement between schemes for assigning patients to surveillance syndromes differ by population subgroup?
3. Do the chance-adjusted levels of agreement between schemes for assigning patients to surveillance syndromes differ by syndrome?

4. Do the chance-adjusted levels of agreement between schemes for assigning patients to surveillance syndromes differ across comparisons?
5. Do schemes for assigning patients to surveillance syndromes based on EMS, chief complaint and ICD-9 diagnosis code data agree using multi-rater measures of agreement?
6. Do the performance characteristics of pre-diagnostic schemes for assigning patients to surveillance syndromes differ by assignment scheme or by syndrome?

## METHODS AND MATERIALS

A retrospective study involving the secondary analysis of public health surveillance records was undertaken to characterize the reliability and validity of an EMS dispatch data-based (EMS) scheme for assigning emergency patients to surveillance syndromes in relation to two other schemes, one based on hospital ED clinicians' manual categorization according to patients' chief complaint and clinical presentation (CC), and one based on ICD-9 coded hospital ED diagnoses (ICD-9). Comparisons of a sample of individual emergency patients' syndrome assignments according to the prehospital versus each of the two hospital categorization schemes were made by matching EMS run records to their corresponding emergency department patient encounter records.

Because the three categorization schemes do not use the same sets of syndromes, comparisons were made based only on the two syndromes that they each have in common, respiratory (Resp) and gastrointestinal (GI). In the EMS dispatch-based scheme, these syndromes were operationalized as runs assigned the "Breathing Problems" or "Abdominal Pain/Problems" ProQA dispatch codes, based on the nature of the 911 call.

This study was approved by the University of Louisville Institutional Review Board prior to initiation.

## Statistical Analysis Plan

The following specific hypotheses were tested to answer the general research questions posed by this study. Wherever applicable, these hypotheses were tested separately for each of three comparisons: EMS vs. CC, EMS vs. ICD-9 and CC vs. ICD-9 for both the respiratory and gastrointestinal syndromes.

- Hypothesis 1: Syndrome assignment of emergency patients using one categorization scheme will agree, beyond chance levels, with syndrome assignment of the same patients using another categorization scheme.

$$H_0: \kappa_{(\text{EMS vs. CC/EMS vs. ICD-9/CC vs. ICD-9}) (\text{Resp/GI})} = 0$$

$$H_1: \kappa_{(\text{EMS vs. CC/EMS vs. ICD-9/CC vs. ICD-9}) (\text{Resp/GI})} \neq 0$$

- Hypothesis 2.1: The chance-adjusted level of agreement between the schemes for assigning emergency patients to surveillance syndromes is different for females than it is for males.

$$H_0: \kappa_{\text{Female}} = \kappa_{\text{Male}}$$

$$H_1: \kappa_{\text{Female}} \neq \kappa_{\text{Male}}$$

- Hypothesis 2.2: The chance-adjusted level of agreement between the schemes for assigning emergency patients to surveillance syndromes differs across age groups.

$$H_0: \kappa_{\text{Age Group 1}} = \kappa_{\text{Age Group 2}} = \kappa_{\text{Age Group 3}} = \kappa_{\text{Age Group 4}}$$

$$H_1: \kappa_{\text{Age Group 1}} \neq \kappa_{\text{Age Group 2}} \neq \kappa_{\text{Age Group 3}} \neq \kappa_{\text{Age Group 4}}$$

- Hypothesis 3: The chance-adjusted level of agreement between the schemes for assigning emergency patients to surveillance syndromes is different for different syndromes.



$$H_0: \kappa_{\text{Resp}} = \kappa_{\text{GI}}$$

$$H_1: \kappa_{\text{Resp}} \neq \kappa_{\text{GI}}$$

- Hypothesis 4.1: The chance-adjusted levels of agreement between the three possible pairs of the three schemes (i.e. EMS vs. Hosp, EMS vs. CDC, Hosp vs. CDC) for assigning emergency patients to surveillance syndromes differ significantly in global comparisons.

$$H_0: \kappa_{\text{EMS vs. CC}} = \kappa_{\text{EMS vs. ICD-9}} = \kappa_{\text{CC vs. ICD-9}}$$

$$H_1: \kappa_{\text{EMS vs. CC}} \neq \kappa_{\text{EMS vs. ICD-9}} \neq \kappa_{\text{CC vs. ICD-9}}$$

- Hypothesis 4.2: The chance-adjusted levels of agreement between the three possible pairings of the three schemes for assigning emergency patients to surveillance syndromes differ significantly in pairwise, post-hoc comparisons.

$$H_0: \kappa_{\text{EMS vs. CC}} = \kappa_{\text{EMS vs. ICD-9}}, \text{ etc.}$$

$$H_1: \kappa_{\text{EMS vs. CC}} \neq \kappa_{\text{EMS vs. ICD-9}}, \text{ etc.}$$

- Hypothesis 5.1: The three schemes for assigning emergency patients to surveillance syndromes will agree beyond chance levels using Fleiss' extension of kappa for multiple raters.

$$H_0: \kappa_{\text{Fleiss}} = 0$$

$$H_1: \kappa_{\text{Fleiss}} \neq 0$$

- Hypothesis 5.2: The three schemes for assigning emergency patients to surveillance syndromes will agree beyond chance levels using Gwet's  $AC_1$  statistic.

$$H_0: AC_1 = 0$$

$$H_1: AC_1 \neq 0$$

- Hypothesis 6.1: The performance characteristics of the pre-diagnostic categorization schemes will differ significantly by surveillance syndrome.

$$H_0: \pi_{(\text{Sens/Spec/PVP/PVN Resp})} = \pi_{(\text{Sens/Spec/PVP/PVN GI})}$$

$$H_1: \pi_{(\text{Sens/Spec/PVP/PVN Resp})} \neq \pi_{(\text{Sens/Spec/PVP/PVN GI})}$$

- Hypothesis 6.2: The performance characteristics of the EMS categorization scheme will differ significantly from those of the chief complaint scheme.

$$H_0: \pi_{(\text{Sens/Spec/PVP/PVN EMS})} = \pi_{(\text{Sens/Spec/PVP/PVN CC})}$$

$$H_1: \pi_{(\text{Sens/Spec/PVP/PVN EMS})} \neq \pi_{(\text{Sens/Spec/PVP/PVN CC})}$$

## Setting

The context of this study was three syndromic surveillance systems operated by the LMPHW. This study was conducted as part of an ongoing, long-term effort on the part of LMPHW to comprehensively evaluate its early outbreak detection capacity, including aspects of its syndromic surveillance program. The LMPHW is an agency of the Louisville/Jefferson County Metro Government. (140) As the result of a 2003 city-county merger, the Louisville/Jefferson County Metro Government is the successor to the former governments of the City of Louisville and Jefferson County. (141)

With an estimated population of 701,500 in 2006, of which approximately 76% are non-Latino whites, 19% are non-Latino African American, 2% are Latino, and 3% are of some other race, Louisville/Jefferson County is the largest

municipality in Kentucky and (according to Louisville's mayor and Metro Council) the sixteenth largest city in the United States. The community includes a mix of urban, suburban and rural areas covering some 386 square miles. (142)

Of Louisville Metro's 12 acute care hospitals, five were operated by Norton Healthcare, Inc. at the time of this study. Norton Healthcare is not-for-profit hospital and health services system based in Louisville and claiming a 45% share of the local healthcare market. (143) All five Norton hospital emergency departments submitted data to the LMPHW for syndromic surveillance during the study period. However, only data from two of them, Norton Hospital and Kosair Children's Hospital, which provided ICD-9 coded diagnoses, were included in this study. Norton Hospital is 586-bed tertiary care and teaching facility. (144) Kosair Children's Hospital is a 263-bed pediatric tertiary care, research and teaching facility. (145) Both are located in downtown Louisville.

Established in 2005, LMEMS is the successor organization to the former Jefferson County EMS and Louisville Fire and Rescue, Division of EMS, which were abolished as a result of the merger of the City of Louisville and Jefferson County. With a staff of approximately 91 paramedics and 140 EMTs fielding 25 ambulances and 10 paramedic response cars, LMEMS provides around the clock advanced life support (ALS) prehospital emergency medical services for 96% of the population of Jefferson County. (146,147) In 2007, LMEMS responded to 94,597 ambulance requests. (148)

In Louisville Metro, all 911 ambulance requests are received by and all EMS responses are dispatched by MetroSafe. MetroSafe is the consolidated

dispatch and emergency communications center for Louisville Metro Police, Fire and EMS who, along with the Louisville/Jefferson County Emergency Management Agency (LMEMA), jointly operate the facility. EMS runs are dispatched using a computer assisted dispatch (CAD) system running the ProQA emergency medical dispatch (EMD) software. (149)

## **Data**

This descriptive study of the reliability and validity of three different patient categorization schemes for syndromic surveillance made use of hospital emergency department and EMS dispatch data that had been previously provided to the LMPHW. The hospital data originally came from Norton and Kosair Children's hospitals. The EMS dispatch data originally came from LMEMS via MetroSafe. These data are routinely provided to LMPHW for the purpose of public health surveillance.

Hospital ED surveillance records for all patients  $\geq$  one year of age that were seen in the emergency departments of Norton Hospital and Kosair Children's Hospital between July 1<sup>st</sup>, 2006 and March 31<sup>st</sup>, 2007 were included in the study. EMS run records for all patients transported by LMEMS to Norton and Kosair hospitals during the same time period were also included.

Hospital ED surveillance records contained the following data items: arrival date, arrival time, hospital clinician's syndrome categorization, gender, date of birth, age, zip code and ICD-9 coded diagnosis. The ED diagnosis was used to assign each patient to one of the CDC-defined surveillance syndromes.

EMS run records contained the following fields that were relevant to this study: unit number, dispatch date, dispatch time, dispatch code, destination hospital and ED arrival time. Additionally, EMS run records contained a free-text field for the dispatcher's comments. The dispatcher's comments routinely included the age and gender of the patient, which were extracted by parsing the text string and recorded as separate variables.

### **Linkage of Hospital and EMS Records**

EMS run records for patients transported to Norton and Kosair Children's hospitals were matched to the patients' corresponding hospital emergency department records. Because the EMS run and emergency department data available to the LMPHW did not contain a common unique identifier field, the records were matched on the basis of age, gender, date and time fields. EMS run records were matched to hospital ED records if the date, age and gender fields were equal and if the ED arrival time field in the EMS run record was within  $\pm 15$  minutes of the ED arrival time field in the hospital record. Unmatched hospital and EMS records were excluded from the analysis.

### **Measurement of Interrater Agreement**

Each of the new records resulting from the successful linkage of records from the hospital emergency department and ambulance run data sets contained the individual patient's syndrome assignment according to each of the three categorization schemes. For each of the schemes, two new dichotomous

variables were created reflecting patients' assignment or non-assignment to each of the two surveillance syndromes analyzed in the study. That is, each patient was categorized as being assigned to the respiratory or to a non-respiratory syndrome and to the gastrointestinal or a non-gastrointestinal syndrome under each of the three schemes, with the non-respiratory and non-gastrointestinal syndromes representing the collapsing of all of the other syndromes into one.

This new, linked dataset was analyzed to assess the level of agreement beyond chance between the three possible pairs of syndrome categorization schemes in assigning patients to the respiratory or non-respiratory syndrome and to the gastrointestinal or non-gastrointestinal syndrome. Cohen's kappa statistics (150) were used to measure chance-adjusted agreement between categorization schemes (raters) and were calculated, along with their standard errors, using SPSS v. 11.0. Ninety-five percent confidence intervals around each value of kappa were calculated using MS Excel 2003 according to the following equation:

$$95\% \text{ C.I.} = k \pm 1.96 \text{ SE}(k)$$

Overall levels of agreement between the three categorization schemes were measured using two generalized versions of kappa for multiple raters, one developed by Fleiss (151,152) and one, the AC1 statistic, developed by Gwet. (153) Both statistics were calculated with SAS v. 9.1 using a macro developed by Gwet. (154)

## Comparison of Kappa Statistics

Z-tests and a chi-square-like test based on the variance of the kappa statistic were used to test the equivalence of kappa coefficients across syndromes, population subgroups and pairs of syndrome assignment schemes. Because kappa is normally distributed and because their variances can be estimated, two values of kappa may be compared to see if they are significantly different. In this study, pairwise comparisons were based on the Z-test using the following equation:

$$Z = \frac{|k_1 - k_2|}{\sqrt{\text{var}(k_1) + \text{var}(k_2)}} \quad (152)$$

Global tests for multiple equal values of kappa (i.e. across population subgroups and across multiple rater pairs) were based on the chi-square distribution using the following equation proposed by Fleiss for that purpose:

$$\chi^2_{\text{Equal}\kappa} = \sum_{m=1}^g \frac{(k_m - k^{\text{"common"}})^2}{\text{var}(k_m)} \quad (152)$$

With  $g - 1$  degrees of freedom where there are a total of  $g$  different estimated values of kappa,  $m$ .

For comparisons across population subgroups, the "common" kappa was taken to be the overall kappa for the full sample. For comparisons across the different rater pairs, the "common" kappa was calculated as per Fleiss:

$$\text{"common"} \kappa = \frac{\sum_{m=1}^g \frac{k_m}{\text{var}(k_m)}}{\sum_{m=1}^g \frac{1}{\text{var}(k_m)}} \quad (152)$$

These calculations and determination of associated p-values were carried out using MS Excel 2003.

Statistical significance for all hypothesis tests was set at  $p < 0.05$  with one exception. In the case of multiple post-hoc comparisons between the three possible pairs of kappa values following a global test across three kappa estimates, the following Bonferroni-corrected threshold for significance was used:  $\alpha = 0.05(1/n)$ , where  $n$  is the number of post hoc comparisons (i.e.  $\alpha = 0.05[1/3] = 0.017$ ). (155)

### **Sensitivity, Specificity and Predictive Value**

The sensitivity, specificity, predictive value positive (PVP) and predictive value negative (PVN) of EMS dispatch and chief complaint-based categorization schemes were also calculated, using the ICD-9-coded ED diagnosis-based categorization scheme as the criterion standard. Within each syndrome, for both categorization schemes, a true positive (TP) was defined as a patient who was assigned by that scheme to the particular surveillance syndrome—either respiratory or gastrointestinal—and was also assigned to that syndrome by the ED diagnosis-based categorization scheme. A false positive (FP) was defined as a patient who was assigned to the syndrome by the scheme being assessed, but not by the diagnosis-based scheme. A true negative (TN) was a patient who was not assigned to the syndrome of interest (i.e. was assigned to either the non-respiratory or non-gastrointestinal category) by either the scheme being assessed or the criterion standard scheme. A false negative (FN) was a patient



who was not assigned to the syndrome of interest by the scheme being assessed, but was assigned to the syndrome by the criterion standard scheme.

Sensitivity was calculated as the number of true positives divided by the total number of patients assigned to the syndrome of interest by the diagnosis-based scheme ( $TP/[TP + FN]$ ). Specificity was calculated as the number of true negatives divided by the total number of patients who were not assigned to the syndrome of interest by the criterion standard ( $TN/[TN + FP]$ ). PVP was calculated as the number of true positives divided by the total number of patients assigned to the syndrome of interest by the scheme being assessed ( $TP/[TP + FP]$ ). PVN was calculated as the number of true negatives divided by the total number of patients not assigned to the syndrome of interest by the scheme being assessed ( $TN/[TN + FN]$ ).

Sensitivity, specificity and predictive value positive and negative calculations were performed in MS Excel 2003.

### **Comparison of Performance Characteristics**

Comparisons of all performance characteristic (i.e. sensitivity, specificity, PVP and PVN) values were made across categorization schemes and surveillance syndromes to determine whether they were significantly different. Comparisons of all performance characteristics across syndromes and of PVP and PVN across categorization schemes were made in MS Excel 2003 using the Z test for independent proportions.

Because proportions were measured within the same groups of subjects as defined by the criterion standard classifier (i.e. “true” positive and “true” negative) in sensitivity and specificity comparisons across categorization schemes, McNemar’s test for paired proportions (156) was used to determine the significance of these differences.

### **Power and Sample Size Considerations**

The sample size required to generate acceptable levels of power when calculating kappa is a consideration that has often been overlooked in reliability studies, perhaps because kappa is frequently seen simply as a descriptive statistic, rather than one for statistical inference. Donner and Eliasziw have produced exact power contours that display the number of subjects and observers that would be required to achieve eighty percent power for various effect sizes. (157) However, as has been pointed out, the kappa coefficient is not recommended as a test statistic for testing the null hypothesis of randomness. (158,159) Power calculations, therefore, are not strictly relevant for the production of a single measure of interrater agreement, even when one wishes to infer the population parameter from a sample.

What is relevant are the accuracy and precision of these estimates, which are affected by sample size. The stability of such estimates can be determined by the width of their corresponding confidence intervals. Sample sizes, therefore, can be considered sufficiently large in so far as they produce acceptably narrow

ninety-five percent confidence intervals. In this study, ninety-five percent confidence intervals are reported for all kappa estimates.

Kappa is well suited to hypothesis tests of the equivalence of different estimates of kappa, across population subgroups or testing methods for example. (158-161) The level of power that can be achieved by a particular sample size is, therefore, directly relevant to such comparisons. Donner has produced tables indicating the sample size required to achieve eighty and ninety percent power for hypothesis tests of the equivalence of different estimates of the kappa coefficient. (161) According to these tables, a sample size of 1082 subjects would be sufficient to achieve ninety percent power to detect a difference of size 0.20 in estimates of kappa if the prevalence of the trait (in this study, the proportion of cases assigned to the syndrome of interest) were as low as 0.10, with alpha equal to 0.05. A sample size of 808 would be sufficient to achieve eighty percent power.

According to the method of Schlesselman, (162) the sample size required to detect a significant difference between two proportions,  $p_1$  and  $p_2$ , hypothesized to be 0.30 and 0.40 respectively (for example), in two groups of equal size with 80% power and with  $\alpha = 0.05$ , is 356 observations per group.

In cases where the groups are of unequal size, Fleiss' generalization (152) of the method of Casagrande (163) can be used to determine the required sample size. So, for example, to detect the difference between two proportions,  $p_1$  and  $p_2$ , hypothesized to be 0.30 and 0.40 respectively, with 80% power and

with  $\alpha = 0.05$  where the number of observations for  $p_2$  is half that of  $p_1$ , the required sample size is  $n_1$  (590) +  $n_2$  (295) or 885 total observations.

According to the method of Dupont, (164) 352 pairs of observations would be sufficient to achieve 80% power to detect a significant difference between two hypothesized proportions, 0.30 and 0.40, using McNemar's test, assuming  $\alpha = 0.05$  and a correlation coefficient for failure between paired subjects of  $\phi = 0.01$  (i.e. tending to maximize  $n$ ).

The above sample size calculations were carried out using version 2.1.30 (February 2003) of the PS Power and Sample Size Calculations computer program created by Dupont and Plummer. (165)

Of course, to the extent that they involve fewer observations, comparisons between population subgroups in stratified analyses will achieve lesser levels of statistical power and will be less able to detect smaller effect sizes than comparisons involving the total sample. However, while the results of stratified analyses are presented here, they are not, with two specific exceptions (hypotheses 2.1 and 2.2), a major focus of this study.

## RESULTS

During the study period, July 1, 2006 through March 31, 2007, there were 45,947 patient visits to Norton and Kosair Children's hospitals, 19,014 to Norton and 26,933 to Kosair. All but 33 of the hospital ED surveillance records for these visits were included in the study. All of the records that were excluded were missing either gender ( $n = 29$ ) or age ( $n = 4$ ) data. The remaining 45,917 records reflected an accessible population with a median age of 14 yrs (mean = 22, SD = 21.83, range 1 – 104), and that was 54.10% ( $n = 24,833$ ) female. The median age of the 19,006 Norton patients was 40 years (mean = 43.15, SD = 18.73). Sixty-five point two percent ( $n = 12,398$ ) were female. The median age of the 26,911 Kosair patients was 6 years (mean = 7.15, SD = 5.36). Forty-six point two percent ( $n = 12,435$ ) were female.

During the same period, LMEMS logged 5,575 patient transports to Norton and Kosair Children's hospitals. The median age of these patients was 28 years (mean = 34.63, SD = 25.00). Sixty-two point seven percent ( $n = 3,496$ ) were female. Two thousand two hundred fifty-seven (40.48%) of these ambulance run records were successfully matched to corresponding hospital ED surveillance records. The median age of this analytic cohort was 39 years (mean = 41.51, SD = 23.36). Sixty-six point nine percent ( $n = 1509$ ) of the patients were female. Compared to the unmatched records (median age = 22 yrs, female = 59.9%) the

analytic cohort was disproportionately female ( $\chi^2 = 27.93$ ,  $p < 0.001$ ) and, comparing median ages using the Wilcoxon rank-sum test, significantly older ( $W = 8145640$ ,  $p < 0.001$ ).

Tables 5 – 7 show the counts of concordant and discordant patient categorizations for both the respiratory and gastrointestinal syndromes for each of the three comparisons of categorization scheme pairs, reflecting the following levels of raw (i.e. not adjusted for chance) agreement: The EMS dispatch code and chief complaint-based schemes agreed 83% of the time in assigning patients to the respiratory syndrome and 88% of the time in assigning them to the gastrointestinal syndrome. The EMS dispatch code and ED diagnosis-based schemes agreed 73% of the time for the respiratory syndrome and 87% of the time for the gastrointestinal syndrome. The chief complaint and ED diagnosis-based schemes 81% and 88% of the time respectively for the respiratory and gastrointestinal syndromes.

**Table 5.**

**Categorization of a Series of Emergency Patients to Two Surveillance Syndromes by EMS Dispatch Code and by Chief Complaint**

Categorization by Chief Complaint (CC)						
Categorization by Dispatch Code (EMS)	Respiratory			Gastrointestinal		
	Syndrome	Non- Syndrome	Total	Syndrome	Non- Syndrome	Total
	Syndrome	210	319	529	86	93
Non- Syndrome	64	1664	1728	176	1902	2078
<b>Total</b>	274	1983	2257	262	1995	2257

Table 6.

**Categorization of a Series of Emergency Patients to Two Surveillance Syndromes by EMS Dispatch Code and by ED Diagnosis**

Categorization by ED Dx (ICD-9)						
Categorization by Dispatch Code (EMS)	Respiratory			Gastrointestinal		
	Syndrome	Non- Syndrome	Total	Syndrome	Non- Syndrome	Total
	Syndrome	170	359	529	85	94
Non- Syndrome	243	1485	1728	204	1874	2078
<b>Total</b>	413	1844	2257	289	1968	2257



Table 7.

**Categorization of a Series of Emergency Patients to Two Surveillance Syndromes by Chief Complaint and by ED Diagnosis**

		Categorization by ED Dx (ICD-9)					
		Respiratory			Gastrointestinal		
Categorization by Chief Complaint (CC)		Syndrome	Non-Syndrome	Total	Syndrome	Non-Syndrome	Total
	Syndrome		129	145	274	142	120
Non-Syndrome		284	1699	1983	147	1848	1995
<b>Total</b>		413	1844	2257	289	1968	2257

Calculated values of Cohen's kappa are sensitive to both trait prevalence and marginal homogeneity. Table 7 shows the proportion of analytic cohort cases assigned to the respiratory and gastrointestinal syndromes (i.e. the trait prevalence) by each of the three syndrome categorization schemes and the significance of the results of McNemar's test for homogeneity of the marginal proportions for the 2 x 2 comparisons of each of the possible pairings of syndrome categorization schemes for the total sample as well as stratified by gender and age group.

Table 8.

**Respiratory and Gastrointestinal Syndrome Assignment (Trait Prevalence) by Three Categorization Schemes (Raters) and Significance of Tests for Marginal Homogeneity for Scheme Pairings**

Resp Syndrome	n	Trait Prevalence (%) <sup>1</sup>			P of Marginal Homogeneity <sup>2</sup>		
		EMS <sup>3</sup>	CC <sup>4</sup>	ICD-9 <sup>5</sup>	EMS vs CC	EMS vs ICD-9	CC vs ICD-9
<b>Total</b>	2257	23.44	12.14	18.30	<0.001	<0.001	<0.001
<b>Females</b>	1509	23.99	12.59	17.83	<0.001	<0.001	<0.001
<b>Males</b>	748	22.33	11.23	19.25	<0.001	0.119	<0.001
<b>&lt;18 yrs</b>	499	21.04	10.62	15.63	<0.001	0.004	0.001
<b>18-39 yrs</b>	644	17.39	8.23	15.06	<0.001	0.225	<0.001
<b>40-59 yrs</b>	551	26.50	13.43	23.23	<0.001	0.222	<0.001
<b>60+ yrs</b>	563	29.48	16.70	19.54	<0.001	<0.001	0.208
<b>GI Syndrome</b>							
<b>Total</b>	2257	7.93	11.61	12.80	<0.001	<0.001	0.111
<b>Females</b>	1509	9.87	13.32	15.18	<0.001	<0.001	0.060
<b>Males</b>	748	4.01	8.16	8.02	<0.001	<0.001	1.000
<b>&lt;18 yrs</b>	499	6.61	9.22	7.62	0.066	0.522	0.256
<b>18-39 yrs</b>	644	12.11	14.44	15.37	0.188	0.057	0.594
<b>40-59 yrs</b>	551	6.53	10.89	12.34	0.001	<0.001	0.396
<b>60+ yrs</b>	563	5.68	11.19	14.92	<0.001	<0.001	0.019

Table 9 shows the estimates of Cohen's kappa coefficients and their ninety-five percent confidence intervals for each of the three possible pairings of syndrome categorization schemes by gender and age group as well as for the total sample for both the respiratory and gastrointestinal surveillance syndromes. Examination of the confidence intervals reveals that, with one exception, they were all positive and excluded the null value, zero. This means that, in all but one instance, agreement exceeded chance at the ninety-five percent confidence level. The exception was that, for forty to fifty-nine year olds, the ambulance

dispatch and ED diagnosis-based schemes did not agree beyond chance levels in assigning patients to the respiratory syndrome.

While the kappa coefficients were statistically significant in nearly every case (versus the null hypothesis of randomness), effect sizes—the estimated degree of agreement—were modest. According to Landis and Koch (1966), values of kappa < 0.00 are interpreted as “poor,” 0.00 – 0.20 as “slight,” 0.21 – 0.40 as “fair,” 0.41 – 0.60 as “moderate,” 0.61 – 0.80 as “substantial” and 0.81 -0.99 as “near perfect” agreement. Here, the estimated values of kappa ranged from “slight” to, at best, “moderate.”

For the respiratory syndrome, the ambulance dispatch code and ED diagnosis-based schemes exhibited the least agreement for the full sample ( $k = 0.196$ , “slight” agreement), while the ambulance dispatch and hospital ED clinician-based schemes exhibited the strongest agreement ( $k = 0.432$ , “moderate”). As with the respiratory syndrome, the ambulance dispatch code and ED diagnosis-based schemes showed the weakest agreement for the gastrointestinal syndrome ( $k = 0.294$ , “fair”). However, the hospital ED clinician and ED diagnosis-based schemes showed the strongest agreement for the gastrointestinal syndrome ( $k = 0.448$ , “moderate”).

Table 9.

**Pairwise Agreement Between Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories)**

	n	EMS <sup>2</sup> vs CC <sup>3</sup>			EMS vs ICD-9 <sup>4</sup>			CC vs ICD-9		
		Kappa	95% C.I.		Kappa	95% C.I.		Kappa	95% C.I.	
Respiratory Syndrome <sup>1</sup>			LCL	UCL		LCL	UCL		LCL	UCL
<b>Total</b>	2257	0.432	0.387	0.477	0.196	0.151	0.241	0.269	0.218	0.320
<b>Females</b>	1509	0.453	0.398	0.508	0.197	0.142	0.252	0.272	0.209	0.335
<b>Males</b>	748	0.386	0.304	0.468	0.193	0.113	0.273	0.264	0.178	0.350
<b>&lt; 18 yrs</b>	499	0.425	0.323	0.527	0.447	0.347	0.547	0.502	0.390	0.614
<b>18 - 39 yrs</b>	644	0.311	0.213	0.409	0.241	0.147	0.335	0.418	0.314	0.522
<b>40 - 59 yrs</b>	551	0.414	0.326	0.502	0.059	-0.027	0.145	0.093	0.005	0.181
<b>60+ yrs</b>	563	0.521	0.443	0.599	0.091	0.007	0.175	0.151	0.057	0.245
<b>GI Syndrome<sup>5</sup></b>										
<b>Total</b>	2257	0.327	0.266	0.388	0.294	0.235	0.353	0.448	0.393	0.503
<b>Females</b>	1509	0.336	0.265	0.407	0.309	0.242	0.376	0.442	0.377	0.507
<b>Males</b>	748	0.268	0.143	0.393	0.202	0.082	0.322	0.425	0.309	0.541
<b>&lt; 18 yrs</b>	499	0.410	0.267	0.553	0.409	0.258	0.560	0.506	0.371	0.641
<b>18 - 39 yrs</b>	644	0.239	0.139	0.339	0.275	0.175	0.375	0.461	0.365	0.557
<b>40 - 59 yrs</b>	551	0.387	0.258	0.516	0.306	0.183	0.429	0.399	0.283	0.515
<b>60+ yrs</b>	563	0.328	0.201	0.455	0.230	0.120	0.340	0.431	0.323	0.539

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome
2. Patients categorized by EMS Dispatch Code
3. Patients categorized by hospital emergency department clinician based on chief complaint/presentation
4. Patients categorized by ICD-9 coded emergency department diagnosis using CDC-defined syndrome groupings
5. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome

For both respiratory and gastrointestinal syndrome assignment, kappa estimates for females did not differ significantly from those of males for any of the pairs of categorization schemes. Table 10 shows gender-stratified kappa estimates and their variances, as well as the chi-square and associated p-values for female-male comparisons.

**Table 10.**

**Comparison of Gender-Stratified Estimates of the Kappa Coefficient for Three Pairings of Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories)**

Respiratory Syndrome <sup>1</sup>	Females (n=1509)		Males (n=748)		Chi-Square <sup>2</sup>	P-Value
	Kappa	Var	Kappa	Var		
EMS <sup>3</sup> vs CC <sup>4</sup>	0.453	0.00078	0.386	0.00176	1.762	0.184
EMS vs ICD-9 <sup>5</sup>	0.197	0.00078	0.193	0.00168	0.007	0.935
CC vs ICD-9	0.272	0.00102	0.264	0.00194	0.022	0.883
GI Syndrome <sup>6</sup>						
EMS vs CC	0.336	0.00130	0.268	0.00410	0.912	0.339
EMS vs ICD-9	0.309	0.00116	0.202	0.00372	2.469	0.116
CC vs ICD-9	0.442	0.00109	0.425	0.00348	0.185	0.667

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome

2. Degrees of freedom = 1

3. Patients categorized by EMS Dispatch Code

4. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

5. Patients categorized by ICD-9 coded emergency department diagnosis using CDC-defined syndrome groupings

6. Patients categorized as GI Syndrome or Non-GI Syndrome

For respiratory syndrome assignment, kappa estimates varied significantly across age groups for each of the three possible pairings of categorization schemes. For gastrointestinal syndrome assignment, they did not. Table 11 shows age-stratified kappa estimates and their variances, as well as chi-square statistics and associated p-values for comparisons across age groups.

**Table 11.**

**Comparison of Age-Stratified Estimates of the Kappa Coefficient for Three Pairings of Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories)**

Respiratory Syndrome <sup>1</sup>	< 18 yrs (n=499)		18 - 39 yrs (n=644)		40 - 59 yrs (n=551)		60+ yrs (n=563)		Chi-Square <sup>2</sup>	P-Value
	Kappa	Var	Kappa	Var	Kappa	Var	Kappa	Var		
EMS <sup>3</sup> vs CC <sup>4</sup>	0.425	0.003	0.311	0.003	0.414	0.002	0.521	0.002	10.985	0.012
EMS vs ICD-9 <sup>5</sup>	0.447	0.003	0.241	0.002	0.059	0.002	0.091	0.002	40.758	<0.001
CC vs ICD-9	0.502	0.003	0.418	0.003	0.093	0.002	0.151	0.002	45.953	<0.001
<b>GI Syndrome<sup>6</sup></b>										
EMS vs CC	0.410	0.005	0.239	0.003	0.387	0.004	0.328	0.004	5.097	0.165
EMS vs ICD-9	0.409	0.006	0.275	0.003	0.306	0.004	0.230	0.003	3.712	0.294
CC vs ICD-9	0.506	0.005	0.461	0.002	0.399	0.003	0.431	0.003	1.562	0.668

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome

2. Degrees of freedom = 3

3. Patients categorized by EMS Dispatch Code

4. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

5. Patients categorized by ICD-9 coded emergency department diagnosis using CDC-defined syndrome groupings

6. Patients categorized as GI Syndrome or Non-GI Syndrome

Kappa estimates for each of the three possible pairs of the three categorization schemes differed significantly by syndrome; however, the direction of the difference was not consistent. The EMS dispatch code and chief complaint-based schemes exhibited significantly higher levels of agreement in assigning patients to the respiratory syndrome than in assigning patients to the gastrointestinal syndrome. Both the ambulance dispatch and hospital ED clinician-based schemes, however, exhibited significantly lower levels of agreement with the ED diagnosis-based scheme when assigning patients to the respiratory syndrome than when assigning patients to the gastrointestinal syndrome. Table 12 shows kappa estimates and their variances for each of the three pairs of categorization schemes by syndrome, as well as Z scores and their associated p-values for comparisons across syndromes.

**Table 12.**

**Comparison Across Syndromes of Kappa Estimates for Three Pairs of Three Different Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories)**

	Respiratory Syndrome <sup>1</sup>		GI Syndrome <sup>2</sup>		Z	P-Value
	Kappa	Var	Kappa	Var		
<b>EMS<sup>3</sup> vs CC<sup>4</sup></b>	0.432	0.00053	0.327	0.00096	2.720	0.007
<b>EMS vs ICD-9<sup>5</sup></b>	0.196	0.00053	0.294	0.00090	2.592	0.010
<b>CC vs ICD-9</b>	0.269	0.00068	0.448	0.00078	4.685	<0.001

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome
2. Patients categorized as GI Syndrome or Non-GI Syndrome
3. Patients categorized by EMS Dispatch Code
4. Patients categorized by hospital emergency department clinician based on chief complaint/presentation
5. Patients categorized by ICD-9 coded emergency department diagnosis using CDC-defined syndrome groupings

Global comparisons of the equivalence of kappa estimates across the three possible pairs of categorization schemes revealed significant differences for both the respiratory and gastrointestinal syndromes. Pairwise post hoc analyses conducted to identify specifically which categorization scheme pairings differed from the others suggested that there were significant differences between all of the pairs within both syndromes, with one exception. In assigning patients to the gastrointestinal syndrome, the kappa estimate for the EMS and chief complaint-based schemes did not differ significantly from that of the EMS and ED diagnosis-based schemes.

Table 13 shows the results of global chi-square tests for differences among the three categorization scheme pairs for both surveillance syndromes. Table 14 shows the results of post-hoc Z tests for the difference between pairs of kappa estimates for specific pairings of categorization schemes.



**Table 13.**

**Comparison Across Rater Pairs of Kappa Estimates for Pairs of Categorization Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes**

Resp Syndrome <sup>1</sup>		Kappa	Var	"Common" k <sup>2</sup>	Chi-Square <sup>3</sup>	P-Value
	EMS <sup>4</sup> vs CC <sup>5</sup>	0.432	0.00053			
	EMS vs ICD-9 <sup>6</sup>	0.196	0.00053			
	CC vs ICD-9	0.269	0.00068			
				0.301	54.796	<0.001
GI Syndrome <sup>7</sup>						
	EMS vs CC	0.327	0.00096			
	EMS vs ICD-9	0.294	0.00090			
	CC vs ICD-9	0.448	0.00078			
				0.361	15.845	<0.001

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome

2. i.e. the expected proportion, calculated as per Fleiss (152)

3. Degrees of Freedom = 3

4. Patients categorized by EMS Dispatch Code

5. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

6. Patients categorized by ICD-9 coded emergency department diagnosis using CDC-defined syndrome groupings

7. Patients categorized as Gastrointestinal Syndrome or Non-Gastrointestinal Syndrome

**Table 14.**

**Post Hoc Comparisons of Kappa Estimates for Pairs of Categorization Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes**

Resp Syndrome <sup>1</sup>	Kappa	Kappa			Z	P <sup>5</sup>
		EMS <sup>2</sup> vs ICD-9 <sup>3</sup>	CC <sup>4</sup> vs ICD-9	EMS vs CC		
EMS vs CC	0.432	0.196	***	***	7.256	<0.001
EMS vs ICD-9	0.196	***	0.269	***	2.103	0.035
CC vs ICD-9	0.269	***	***	0.432	4.696	<0.001

  

GI Syndrome <sup>6</sup>	Kappa	Kappa			Z	P <sup>5</sup>
		EMS <sup>2</sup> vs ICD-9 <sup>3</sup>	CC <sup>4</sup> vs ICD-9	EMS vs CC		
EMS vs CC	0.327	0.294	***	***	0.765	0.444
EMS vs ICD-9	0.294	***	0.448	***	3.753	<0.001
CC vs ICD-9	0.448	***	***	0.327	2.897	0.004

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome
2. Patients categorized by EMS Dispatch Code
3. Patients categorized by ICD-9 coded emergency department diagnosis using CDC-defined syndrome groupings
4. Patients categorized by hospital emergency department clinician based on chief complaint/presentation
5. Bonferroni-corrected threshold for significance = 0.017
6. Patients categorized as Gastrointestinal Syndrome or Non-Gastrointestinal Syndrome

In comparisons using Fleiss' extension of kappa for multiple raters, the three categorization schemes exhibited levels agreement that were significantly greater than chance for both of the surveillance syndromes. The effects sizes for the respiratory and gastrointestinal syndromes were 0.291 and 0.360 respectively, both of which would be considered "fair," according to Landis and Koch's scheme.

Using Gwet's AC1 statistic, the three categorization schemes exhibited greater than chance levels of agreement for the gastrointestinal syndrome only. The effect size in that case was considerably larger than was estimated by the

kappa statistic, however. The AC1 statistic for the gastrointestinal syndrome was 0.847, which would be interpreted as “near perfect” according to Landis and Koch.

Table 15 shows the kappa estimates and their ninety-five percent confidence intervals for the three categorization schemes by syndrome. Table 16 shows the AC1 statistics as well as results of Z tests for significance.

**Table 15.**

**Agreement of Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories) Using Fleiss' Kappa for Multiple Raters**

Respiratory Syndrome <sup>1</sup>	Kappa	SE	95% C.I.	
			LCL	UCL
Non-Syndrome <sup>2</sup>	0.291	0.069	0.156	0.426
Syndrome <sup>3</sup>	0.291	0.030	0.233	0.350
Overall <sup>4</sup>	0.291	0.024	0.245	0.338
<b>GI Syndrome<sup>5</sup></b>				
Non-Syndrome	0.360	0.090	0.183	0.537
Syndrome	0.360	0.031	0.300	0.420
Overall	0.360	0.033	0.295	0.425

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome
2. Agreement for non-assignment to syndrome
3. Agreement for assignment to syndrome
4. Overall agreement for syndrome assignment and non-assignment
5. Patients categorized as GI Syndrome or Non-GI Syndrome

**Table 16.**

**Agreement of Three Schemes (Raters) for Assigning Emergency Patients to Surveillance Syndromes (Categories) Using Gwet's AC1 Statistic**

Respiratory Syndrome <sup>1</sup>	AC1	SE	Z	P
Non-Syndrome <sup>2</sup>	0.820	0.459	1.787	0.037
Syndrome <sup>3</sup>	0.176	0.111	1.581	0.057
Overall <sup>4</sup>	0.704	0.459	1.535	0.062
<b>GI Syndrome<sup>5</sup></b>				
Non-Syndrome	0.915	0.445	2.057	0.020
Syndrome	0.293	0.084	3.479	<0.001
Overall	0.847	0.443	1.912	0.028

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome
2. Agreement for non-assignment to syndrome
3. Agreement for assignment to syndrome
4. Overall agreement for syndrome assignment and non-assignment
5. Patients categorized as GI Syndrome or Non-GI Syndrome

Table 17 shows the sensitivity, specificity, predictive value positive and predictive value negative for each of the two pre-diagnostic (i.e. EMS dispatch and chief complaint-based) categorization schemes, as compared to the ED diagnosis-based scheme—taken as the criterion standard—by surveillance syndrome and population subgroups.

**Table 17.**

**Performance Characteristics of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to Surveillance Syndromes, Compared to a Diagnosis-Based Criterion Standard**

	Resp Syndrome <sup>1</sup>				GI Syndrome <sup>2</sup>			
	Sens <sup>3</sup>	Spec <sup>4</sup>	PVP <sup>5</sup>	PVN <sup>6</sup>	Sens	Spec	PVP	PVN
<b>EMS<sup>7</sup></b>								
<b>Total</b>	0.412	0.805	0.321	0.859	0.294	0.952	0.475	0.902
<b>Males</b>	0.389	0.816	0.335	0.849	0.183	0.972	0.367	0.932
<b>Females</b>	0.424	0.800	0.315	0.865	0.323	0.941	0.497	0.886
<b>&lt; 18 Yrs</b>	0.641	0.869	0.476	0.929	0.421	0.963	0.485	0.953
<b>18 – 39 Yrs</b>	0.392	0.865	0.339	0.889	0.333	0.917	0.423	0.883
<b>40 – 59 Yrs</b>	0.313	0.749	0.274	0.783	0.279	0.965	0.528	0.905
<b>60 + Yrs</b>	0.382	0.726	0.253	0.829	0.202	0.969	0.531	0.874
<b>CC<sup>8</sup></b>								
<b>Total</b>	0.312	0.921	0.471	0.857	0.491	0.939	0.542	0.926
<b>Males</b>	0.292	0.930	0.500	0.846	0.500	0.955	0.492	0.956
<b>Females</b>	0.323	0.917	0.458	0.862	0.489	0.930	0.557	0.911
<b>&lt; 18 Yrs</b>	0.474	0.962	0.698	0.908	0.605	0.950	0.500	0.967
<b>18 – 39 Yrs</b>	0.371	0.969	0.679	0.897	0.525	0.925	0.559	0.915
<b>40 – 59 Yrs</b>	0.195	0.884	0.338	0.784	0.441	0.938	0.500	0.923
<b>60 + Yrs</b>	0.282	0.861	0.330	0.832	0.440	0.946	0.587	0.906

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome
2. Patients categorized as GI Syndrome or Non-GI Syndrome
3. Sensitivity
4. Specificity
5. Predictive Value Positive
6. Predictive Value Negative
7. Patients categorized by EMS Dispatch Code
8. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

Tables 18 and 19 show the results of statistical comparisons of the performance characteristics of the two pre-diagnostic categorization schemes across surveillance syndromes. The differences across syndromes in performance characteristics for the ambulance dispatch code-based categorization scheme were all significant for the total sample. The direction of the difference, however, was not consistent. The scheme was more sensitive when assigning patients to the respiratory syndrome than when assigning them to the gastrointestinal syndrome. Specificity, predictive value positive and predictive value negative were all significantly higher for the gastrointestinal syndrome than for the respiratory syndrome.

For the chief complaint-based categorization scheme, the directions of the differences in performance characteristics across syndromes were consistent for the total sample. They were all higher for the gastrointestinal syndrome. In the case of predictive value positive, however, the difference was not statistically significant. For all of the other characteristics, the differences were significant.

Table 18.

**Comparison Across Syndromes of Performance Characteristics of an EMS Dispatch Code-Based Scheme for Assigning Emergency Patients to Surveillance Syndromes**

		Resp <sup>1</sup>		GI <sup>2</sup>		Z	P-Value
		n	p	n	p		
Sens <sup>3</sup>	Total	413	0.412	289	0.294	3.186	0.001
	Females	269	0.424	229	0.323	2.309	0.021
	Males	144	0.389	60	0.183	2.848	0.004
	<18 yrs	78	0.641	38	0.421	2.245	0.025
	18-39 yrs	97	0.392	99	0.333	0.851	0.395
	40-59 yrs	128	0.313	68	0.279	0.481	0.631
	60+ yrs	110	0.382	84	0.202	2.692	0.007
	Spec <sup>4</sup>	Total	1844	0.805	1968	0.952	14.009
Females		1240	0.800	1280	0.941	10.616	<0.001
Males		604	0.816	688	0.972	9.310	<0.001
<18 yrs		421	0.869	461	0.963	5.080	<0.001
18-39 yrs		547	0.865	545	0.917	2.795	0.005
40-59 yrs		423	0.749	483	0.965	9.443	<0.001
60+ yrs		453	0.726	479	0.969	10.383	<0.001
PVP <sup>5</sup>		Total	529	0.321	179	0.475	3.698
	Females	362	0.315	149	0.497	3.872	<0.001
	Males	167	0.335	30	0.367	0.334	0.739
	<18 yrs	105	0.476	33	0.485	0.087	0.931
	18-39 yrs	112	0.339	78	0.423	1.174	0.240
	40-59 yrs	146	0.274	36	0.528	2.914	0.004
	60+ yrs	166	0.253	32	0.531	3.151	0.002
	PVN <sup>6</sup>	Total	1728	0.859	2078	0.902	4.050
Females		1147	0.865	1360	0.886	1.604	0.109
Males		581	0.849	718	0.932	4.855	<0.001
<18 yrs		394	0.929	466	0.953	1.490	0.136
18-39 yrs		532	0.889	566	0.883	0.297	0.766
40-59 yrs		405	0.783	515	0.905	5.166	<0.001
60+ yrs		397	0.829	531	0.874	1.928	0.054

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome

2. Patients categorized as GI Syndrome or Non-GI Syndrome

3. Sensitivity

4. Specificity

5. Predictive Value Positive

6. Predictive Value Negative

Table 19.

**Comparison Across Syndromes of Performance Characteristics of a Hospital ED Chief Complaint-Based Scheme for Assigning Emergency Patients to Surveillance Syndromes**

		Resp <sup>1</sup>		GI <sup>2</sup>		Z	P-Value
		n	p	n	p		
<b>Sens<sup>3</sup></b>	<b>Total</b>	413	0.312	289	0.491	4.794	<0.001
	<b>Females</b>	269	0.323	229	0.489	3.762	<0.001
	<b>Males</b>	144	0.292	60	0.500	2.837	0.005
	<b>&lt;18 yrs</b>	78	0.474	38	0.605	1.324	0.185
	<b>18-39 yrs</b>	97	0.371	99	0.525	2.169	0.030
	<b>40-59 yrs</b>	128	0.195	68	0.441	3.647	<0.001
	<b>60+ yrs</b>	110	0.282	84	0.440	2.295	0.022
	<b>Spec<sup>4</sup></b>	<b>Total</b>	1844	0.921	1968	0.939	2.142
<b>Females</b>		1240	0.917	1280	0.930	1.280	0.200
<b>Males</b>		604	0.930	688	0.955	1.901	0.057
<b>&lt;18 yrs</b>		421	0.962	461	0.950	0.858	0.391
<b>18-39 yrs</b>		547	0.969	545	0.925	3.253	0.001
<b>40-59 yrs</b>		423	0.884	483	0.938	2.860	0.004
<b>60+ yrs</b>		453	0.861	479	0.946	4.402	<0.001
<b>PVP<sup>5</sup></b>		<b>Total</b>	274	0.471	262	0.542	1.648
	<b>Females</b>	190	0.458	201	0.557	1.963	0.050
	<b>Males</b>	84	0.500	61	0.492	0.097	0.922
	<b>&lt;18 yrs</b>	53	0.698	46	0.500	2.012	0.044
	<b>18-39 yrs</b>	53	0.679	93	0.559	1.426	0.154
	<b>40-59 yrs</b>	74	0.338	60	0.500	1.898	0.058
	<b>60+ yrs</b>	94	0.330	63	0.587	3.192	0.001
	<b>PVN<sup>6</sup></b>	<b>Total</b>	1983	0.857	1995	0.926	7.055
<b>Females</b>		1319	0.862	1308	0.911	3.916	<0.001
<b>Males</b>		664	0.846	687	0.956	6.804	<0.001
<b>&lt;18 yrs</b>		446	0.908	453	0.967	3.648	<0.001
<b>18-39 yrs</b>		591	0.897	551	0.915	1.034	0.301
<b>40-59 yrs</b>		477	0.784	491	0.923	6.109	<0.001
<b>60+ yrs</b>		469	0.832	500	0.906	3.443	0.001

1. Patients categorized as Respiratory Syndrome or Non-Respiratory Syndrome

2. Patients categorized as GI Syndrome or Non-GI Syndrome

3. Sensitivity

4. Specificity

5. Predictive Value Positive

6. Predictive Value Negative



Tables 20 through 23 show the results of comparisons of the performance characteristics of the ambulance dispatch code-based categorization scheme with the hospital ED clinician-based scheme for both the respiratory and the gastrointestinal surveillance syndromes. In assigning emergency patients to the respiratory syndrome, there were significant differences for the full sample in sensitivity, specificity, and predictive value positive between the two categorization schemes. The ambulance dispatch code-based scheme exhibited the greater sensitivity, while the hospital ED clinician-based scheme exhibited greater specificity and predictive value positive. The difference between the predictive values negative between the two schemes was not statistically significant.

In assigning emergency patients to the gastrointestinal syndrome, the two categorization schemes differed significantly in sensitivity, specificity and predictive value negative for the full sample. The ambulance dispatch code-based scheme exhibited the greater specificity while the hospital ED clinician-based scheme exhibited greater sensitivity and predictive value negative. The difference between the predictive values positive between the two schemes was not statistically significant.

Table 20.

**Comparison of the Sensitivity and Specificity of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Respiratory Surveillance Syndrome**

		n	EMS <sup>1</sup>	CC <sup>2</sup>	P-Value <sup>3</sup>
<b>Sens<sup>4</sup></b>	<b>Total</b>	413	0.412	0.312	<0.001
	<b>Females</b>	269	0.424	0.323	0.001
	<b>Males</b>	144	0.389	0.292	0.038
	<b>&lt;18 yrs</b>	78	0.641	0.474	0.019
	<b>18-39 yrs</b>	97	0.392	0.371	0.851
	<b>40-59 yrs</b>	128	0.313	0.195	0.006
	<b>60+ yrs</b>	110	0.382	0.282	0.013
	<b>Spec<sup>5</sup></b>	<b>Total</b>	1844	0.805	0.921
<b>Females</b>		1240	0.800	0.917	<0.001
<b>Males</b>		604	0.816	0.930	<0.001
<b>&lt;18 yrs</b>		421	0.869	0.962	<0.001
<b>18-39 yrs</b>		547	0.865	0.969	<0.001
<b>40-59 yrs</b>		423	0.749	0.884	<0.001
<b>60+ yrs</b>		453	0.726	0.861	<0.001

1. Patients categorized by EMS Dispatch Code
2. Patients categorized by hospital emergency department clinician based on chief complaint/presentation
3. Based on McNemar's test for paired proportions, using the binomial distribution
4. Sensitivity
5. Specificity

Table 21.

**Comparison of the Positive and Negative Predictive Values of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Respiratory Surveillance Syndrome**

		EMS <sup>1</sup>		CC <sup>2</sup>		Z	P-Value
		n	p	n	p		
<b>PVP<sup>3</sup></b>	<b>Total</b>	529	0.321	274	0.471	4.153	<0.001
	<b>Females</b>	362	0.315	190	0.458	3.317	0.001
	<b>Males</b>	167	0.335	84	0.500	2.523	0.012
	<b>&lt;18 yrs</b>	105	0.476	53	0.698	2.648	0.008
	<b>18-39 yrs</b>	112	0.339	53	0.679	4.100	<0.001
	<b>40-59 yrs</b>	146	0.274	74	0.338	0.981	0.327
	<b>60+ yrs</b>	166	0.253	94	0.330	1.324	0.186
<b>PVN<sup>4</sup></b>	<b>Total</b>	1728	0.859	1983	0.857	0.226	0.821
	<b>Females</b>	1147	0.865	1319	0.862	0.205	0.837
	<b>Males</b>	581	0.849	664	0.846	0.105	0.916
	<b>&lt;18 yrs</b>	394	0.929	446	0.908	1.099	0.272
	<b>18-39 yrs</b>	532	0.889	591	0.897	0.416	0.677
	<b>40-59 yrs</b>	405	0.783	477	0.784	0.049	0.961
	<b>60+ yrs</b>	397	0.829	469	0.832	0.111	0.912

1. Patients categorized by EMS Dispatch Code

2. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

3. Predictive Value Positive

4. Predictive Value Negative

Table 22.

**Comparison of the Sensitivity and Specificity of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Gastrointestinal Surveillance Syndrome**

		<b>n</b>	<b>EMS<sup>1</sup></b>	<b>CC<sup>2</sup></b>	<b>P-Value<sup>3</sup></b>
<b>Sens<sup>4</sup></b>	<b>Total</b>	289	0.294	0.491	<0.001
	<b>Females</b>	229	0.323	0.489	<0.001
	<b>Males</b>	60	0.183	0.500	<0.001
	<b>&lt;18 yrs</b>	38	0.421	0.605	0.118
	<b>18-39 yrs</b>	99	0.333	0.525	0.014
	<b>40-59 yrs</b>	68	0.279	0.441	0.035
	<b>60+ yrs</b>	84	0.202	0.440	0.001
<b>Spec<sup>5</sup></b>	<b>Total</b>	1968	0.952	0.939	0.036
	<b>Females</b>	1280	0.941	0.930	0.206
	<b>Males</b>	688	0.972	0.955	0.065
	<b>&lt;18 yrs</b>	461	0.963	0.950	0.345
	<b>18-39 yrs</b>	545	0.917	0.925	0.694
	<b>40-59 yrs</b>	483	0.965	0.938	0.029
	<b>60+ yrs</b>	479	0.969	0.946	0.043

1. Patients categorized by EMS Dispatch Code

2. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

3. Based on McNemar's test for paired proportions, using the binomial distribution

4. Sensitivity

5. Specificity

**Table 23.**

**Comparison of the Positive and Negative Predictive Values of Two Pre-Diagnostic Schemes for Assigning Emergency Patients to a Gastrointestinal Surveillance Syndrome**

		EMS <sup>1</sup>		CC <sup>2</sup>		Z	P-Value
		n	p	n	p		
PVP <sup>3</sup>	<b>Total</b>	179	0.475	262	0.542	1.385	0.166
	<b>Females</b>	149	0.497	201	0.557	1.123	0.262
	<b>Males</b>	30	0.367	61	0.492	1.128	0.259
	<b>&lt;18 yrs</b>	33	0.485	46	0.500	0.133	0.894
	<b>18-39 yrs</b>	78	0.423	93	0.559	1.772	0.076
	<b>40-59 yrs</b>	36	0.528	60	0.500	0.264	0.792
	<b>60+ yrs</b>	32	0.531	63	0.587	0.521	0.602
PVN <sup>4</sup>	<b>Total</b>	2078	0.902	1995	0.926	2.784	0.005
	<b>Females</b>	1360	0.886	1308	0.911	2.093	0.036
	<b>Males</b>	718	0.932	687	0.956	1.999	0.046
	<b>&lt;18 yrs</b>	466	0.953	453	0.967	1.087	0.277
	<b>18-39 yrs</b>	566	0.883	551	0.915	1.735	0.083
	<b>40-59 yrs</b>	515	0.905	491	0.923	1.001	0.317
	<b>60+ yrs</b>	531	0.874	500	0.906	1.647	0.100

1. Patients categorized by EMS Dispatch Code

2. Patients categorized by hospital emergency department clinician based on chief complaint/presentation

3. Predictive Value Positive

4. Predictive Value Negative

**Summary**

In summary, this study produced the following results in the full sample for the hypotheses tested.

- Hypothesis 1: Syndrome assignment of emergency patients using one categorization scheme will agree, beyond chance levels, with syndrome assignment of the same patients using another categorization scheme.

- This hypothesis was *supported* for each of the three comparisons between pairings of categorization schemes for both the respiratory and gastrointestinal surveillance syndromes.
- Hypothesis 2.1: The chance-adjusted level of agreement between the schemes for assigning emergency patients to surveillance syndromes is different for females than it is for males.
  - This hypothesis was *not supported* for any of the three comparisons between pairings of categorization schemes for either the respiratory or gastrointestinal syndromes.
- Hypothesis 2.2: The chance-adjusted level of agreement between the schemes for assigning emergency patients to surveillance syndromes differs across age groups.
  - This hypothesis was *supported* for each of the three comparisons between pairings of categorization schemes for the respiratory syndrome. This hypothesis was *not supported* for any of the three comparisons between pairings of categorization schemes for the gastrointestinal syndrome.
- Hypothesis 3: The chance-adjusted level of agreement between the schemes for assigning emergency patients to surveillance syndromes is different for different syndromes.
  - This hypothesis was *supported* for all three of the comparisons between pairs of categorization schemes.

- Hypothesis 4.1: The chance-adjusted levels of agreement between the three possible pairs of the three schemes (i.e. EMS vs. Hosp, EMS vs. CDC, Hosp vs. CDC) for assigning emergency patients to surveillance syndromes differ significantly in global comparisons.
  - This hypothesis was *supported* for both the respiratory and gastrointestinal syndromes.
  
- Hypothesis 4.2: The chance-adjusted levels of agreement between the three possible pairings of the three schemes for assigning emergency patients to surveillance syndromes differ significantly in pairwise, post-hoc comparisons.
  - For the respiratory syndrome, this hypothesis was *supported* for two of three comparisons; it was *not supported* for the comparison between the EMS vs. ICD-9 and CC vs. ICD-9 kappa estimates. For the gastrointestinal syndrome, this hypothesis was *supported* for two of three comparisons; it was *not supported* for the comparison between the EMS vs. CC and EMS vs. ICD-9 kappa estimates.
  
- Hypothesis 5.1: The three schemes for assigning emergency patients to surveillance syndromes will agree beyond chance levels using Fleiss' extension of kappa for multiple raters.
  - This hypothesis was supported for both the respiratory and gastrointestinal syndromes.

- Hypothesis 5.2: The three schemes for assigning emergency patients to surveillance syndromes will agree beyond chance levels using Gwet's AC<sub>1</sub> statistic.
  - This hypothesis was not supported for the respiratory syndrome; it was supported for the gastrointestinal syndrome.
- Hypothesis 6.1: The performance characteristics of the pre-diagnostic categorization schemes will differ significantly by surveillance syndrome.
  - For the EMS-based categorization scheme, this hypothesis was *supported* for all four performance characteristics. For the chief complaint-based categorization scheme, this hypothesis was *supported* for sensitivity, specificity and predictive value negative; it was *not supported* for predictive value positive.
- Hypothesis 6.2: The performance characteristics of the EMS categorization scheme will differ significantly from those of the chief complaint scheme.
  - For the respiratory syndrome, this hypothesis was *supported* for sensitivity, specificity and predictive value positive; it was *not supported* for predictive value negative. For the gastrointestinal syndrome, this hypothesis was *supported* for sensitivity, specificity and predictive value negative; it was *not supported* for predictive value positive.



## DISCUSSION

This study began by posing six research questions:

1. *Does assignment of emergency patients to surveillance syndromes using a categorization scheme based on EMS dispatch data agree with syndrome categorization according to chief complaint and ED diagnosis-based schemes?*

Yes. EMS dispatch-based syndrome assignment did agree with both chief complaint and ED diagnosis-based assignment beyond chance levels, but not strongly. For the respiratory syndrome, EMS-based assignment exhibited moderate agreement with chief complaint-based assignment and slight agreement with diagnosis-based assignment. For the gastrointestinal syndrome, EMS-based assignment exhibited fair agreement with both chief complaint and diagnosis-based assignment. In comparison, chief complaint and diagnosis-based assignment exhibited fair agreement for the respiratory syndrome and moderate agreement for the gastrointestinal syndrome.

These levels of agreement are generally lower than those reported in other studies which have compared agreement between syndrome classifications based on different data sources. (81, 82) The inconsistency across syndromes of the relative strengths of agreement between categorization schemes makes interpretation difficult, but is consistent with the results of Fleischauer et al. (82),

who also found the highest levels of agreement to be for different rater pairs in respiratory and gastrointestinal syndromes.

2. *Do the chance-adjusted levels of agreement between schemes for assigning patients to surveillance syndromes differ by population subgroup?*

Yes and no. The answer here depended both on the variable used to subdivide the population (i.e. gender, age) and on the surveillance syndrome being considered. For each of the three pairs of categorization schemes, levels of agreement did not differ significantly by gender for either syndrome. Nor did the levels of agreement between any of the three pairs of schemes differ significantly across age groups in the case of the gastrointestinal syndrome. In the case of the respiratory syndrome, however, levels of agreement did differ significantly across age groups for each of the three pairs of categorization schemes.

Again, these results are consistent with the results of Fleischauer et al. (82) They found no gender-specific differences in agreement for either respiratory or gastrointestinal syndromes. They also did find age-specific differences in agreement for the respiratory syndrome but not for the gastrointestinal syndrome.

3. *Do the chance-adjusted levels of agreement between schemes for assigning patients to surveillance syndromes differ by syndrome?*

Yes. Levels of agreement were significantly different for the respiratory and gastrointestinal syndromes in each of the three comparisons between

categorization scheme pairs. However, the direction of the difference was not the same in each case. EMS and chief complaint-based assignment agreed more strongly for the respiratory syndrome, while EMS versus diagnosis-based assignment and chief complaint versus diagnosis-based assignment agreed more strongly for the gastrointestinal syndrome.

Consistent with other studies that have found considerable variation in the levels of agreement between categorization schemes across syndrome categories, (81,82) these results suggest that the reliability of any given syndrome grouping method should be considered separately for each syndrome.

*4. Do the chance-adjusted levels of agreement between schemes for assigning patients to surveillance syndromes differ across comparisons?*

Yes. The levels of agreement between categorization scheme pairs differed significantly across the three comparisons for both the respiratory and gastrointestinal syndromes. Post hoc comparisons revealed significant differences between kappa estimates for two of three comparisons in both syndromes. In the respiratory syndrome, the estimated kappa for EMS versus diagnosis-based syndrome assignment did not differ significantly from that of chief complaint versus diagnosis-based assignment. In the gastrointestinal syndrome, the estimated kappa for EMS versus chief complaint-based syndrome assignment did not differ significantly from that of EMS versus diagnosis-based assignment.

In their comparison of the levels of agreement between syndrome categorization schemes based on clinicians' initial impression recorded on a

surveillance form, a retrospective, blinded classification of chief complaints by physicians and ICD-9-coded ED diagnosis, Fleischauer et al. also found significant differences between rater pair comparisons. (82) In their study as well, the direction of these differences were inconsistent across syndromes.

Systematic (i.e. non-random) differences in the levels of agreement between categorization schemes based on different data sources across different rater pair comparisons are especially relevant not only to the interpretation of signals from surveillance systems that make use of different data sources, but also to informing the decision of those planning the implementation of new syndromic surveillance systems on which data sources to use. For example, this study showed that, for respiratory syndrome surveillance, EMS dispatch code-based syndrome assignment agrees significantly more strongly with chief complaint-based assignment than with ED diagnosis-based assignment. If, hypothetically, public health officials were conducting emergency department-based syndromic surveillance in one county and planned to expand syndromic surveillance into an adjacent county without a hospital (or perhaps they did not have access to data from that county's hospital) using EMS dispatch data, they might opt to assign patients from the first county's hospital ED on the basis of chief complaint rather than diagnosis because the construct of a respiratory illness syndrome would be more similarly operationalized and, therefore, comparable.

Conversely, if public health officials were operating EMS and ED-based systems in the same county which were meant to be supplementary, they might

opt to group the ED data on the basis of diagnosis, knowing that the two systems operationalize the respiratory syndrome construct differently and, together, might capture more of the cases intended to be captured by the syndrome. Working together in that way, the systems might complement one another and increase the sensitivity of the overall surveillance program more than using two more redundant operationalizations of the respiratory syndrome would.

5. *Do schemes for assigning patients to surveillance syndromes based on EMS, chief complaint and ICD-9 diagnosis code data agree using multi-rater measures of agreement?*

Yes and no. The three categorization schemes agree beyond chance levels for both syndromes according to Fleiss' multi-rater kappa. However, in both cases, the level of agreement was only fair. The AC1 statistic, on the other hand, suggested much higher levels of chance-adjusted agreement among the three schemes for the gastrointestinal syndrome. In the case of the respiratory syndrome, however, the three schemes did not agree beyond chance levels.

Currently, there are no published studies involving validation or any other evaluation of syndromic surveillance systems that make use of multi-rater measures of agreement. One reason for this is undoubtedly the simple fact that there have been few published validations of syndromic categorization methods generally, and even fewer that assess inter-rater agreement between methods. It is likely, however, that another reason is that it is not clear how to interpret agreement (or non-agreement) among multiple categorization schemes simultaneously in the context of syndromic surveillance. Still, when there are

more than two raters being considered (i.e. EMS, CC, ICD-9), the most appropriate comparison is a simultaneous one using a multi-rater measure of agreement, rather than an attempt to average the kappa values across all possible rater pairs.

Fleiss' generalized kappa for multiple raters gave results that are generally in line with the levels of agreement in pairwise rater comparisons in this study. The AC1 statistic, however, suggests much better chance-adjusted levels of agreement, even though they were not statistically significant for the respiratory syndrome. It is tempting, therefore, to look at this relatively new method as an attractive alternative for analyzing these data, even in the case of paired rater comparisons, especially since it is supposed to be less sensitive to trait prevalence and more robust to violations of the assumption of equal marginal probabilities (both problems with these data) than kappa statistics. However, since the AC1 statistic is so new, so unfamiliar and has received so little use and scrutiny, it cannot currently be accepted as the standard, and is regarded here as provisional.

6. *Do the performance characteristics of pre-diagnostic schemes for assigning patients to surveillance syndromes differ by categorization scheme or by syndrome?*

Yes. Significant differences existed between performance characteristics in seven out of eight comparisons (4 characteristics x 2 syndromes) across categorization schemes and in six out of eight comparisons (4 characteristics x 2 categorization schemes) across syndromes. There was little consistency,

however, in the direction of these differences in either cross-syndrome or cross-rater comparisons.

Sensitivity and PVP are seen as the most important performance characteristics to a system's detection capability (134) and, while EMS-based assignment was significantly more sensitive for the respiratory syndrome and chief complaint-based assignment was significantly more sensitive for the gastrointestinal syndrome, sensitivities and PVP for both assignment schemes were quite low, identifying less than half of the cases in each syndrome as defined by ED diagnosis. Specificities and PVN, on the other hand, were generally high. And, while this fact does not help to improve the detection capabilities (either case or outbreak) of the system, it could help provide some reassurance that outbreaks have not occurred or are not occurring.

PVP and PVN are sensitive to trait prevalence, which, as defined by ED diagnosis, is rather low in these data. When prevalence is very low, PVP approaches zero and PVN approaches 100%. The effect of low prevalence is seen in this study, where PVP is low, indicating that only half or fewer of the cases identified would actually belong in the syndrome to which they were assigned, and PVN is quite high, which could provide false reassurance if considered in isolation.

It should be pointed out that the performance characteristics reported here are for case (i.e. a patient that ought to be included in a syndrome) identification, not outbreak detection, although outbreak detection sensitivity does rely on the ability of the system to detect cases. It should also be noted that the performance

characteristics reported are based on the assumption that ED diagnosis-based syndrome assignment is the appropriate criterion standard. Although this assumption is reasonable since ED diagnosis is the most proximal medical judgment available to the system for all patients and the final judgment for many, if not most, patients, it may or may not be true.

Taken together, the answers to these research questions indicate that surveillance syndrome assignment based on EMS dispatch data has limited reliability in relation to other data sources, that the degree of reliability depends on the syndrome under surveillance and the age group being surveilled and that this method's reliability varies significantly depending on how the comparison syndrome construct is operationalized—that is, whether it is based on chief complaint or ED diagnosis. They further indicate that EMS-based syndrome assignment has limited validity—in that it has low sensitivity and PVP for identifying respiratory and gastrointestinal syndrome cases—and that its validity varies by syndrome.

This does not mean that EMS-based syndromic surveillance is not worthwhile or that it should not be undertaken. First, while low, the measures of reliability reported here for EMS-based syndrome categorization are not altogether dissimilar from those sometimes found for syndrome assignment based on other, more commonly used methods. In this study, for example, while significantly greater than that of EMS-based categorization, chief complaint-based categorization exhibited only fair agreement with ED diagnosis-based categorization for the respiratory syndrome and moderate agreement for the



gastrointestinal syndrome. Fleischauer et al. as well reported respiratory syndrome reliability measures for pairwise comparisons of three categorization methods (initial clinical impression, retrospective review of chief complaint and ED diagnosis) that were, at best, moderate ( $k = 0.59$  for initial clinical impression vs. ED diagnosis) and, more often, fair ( $k = 0.28$  for initial clinical impression vs. retrospective review of chief complaint and  $k = 0.33$  for retrospective review of chief complaint vs. ED diagnosis). (82) Table 24 shows the levels of agreement between various syndrome categorization methods reported in the literature in comparison to the results of this study.

**Table 24**

**Comparison of Levels of Agreement (Kappa) Between Syndrome Categorization Methods Reported in the Literature with the Current Study**

Syndrome	Comparison	Begier et al., 2003	Fleischauer et al., 2004	Current Study
Resp	ICD-9 vs CC <sup>1</sup>	0.68	0.33	0.27
	SF vs CC <sup>2</sup>	***	0.28	***
	SF vs ICD-9 <sup>3</sup>	***	0.59	***
	EMS vs CC <sup>4</sup>	***	***	0.43
	EMS vs ICD-9 <sup>5</sup>	***	***	0.20
GI	ICD-9 vs CC	0.68	0.71	0.45
	SF vs CC	***	0.70	***
	SF vs ICD-9	***	0.63	***
	EMS vs CC	***	***	0.33
	EMS vs ICD-9	***	***	0.29

1. Categorization based on ICD-9 coded ED diagnosis vs. categorization by chief complaint
2. Categorization based on clinical impression recorded on surveillance form vs. categorization by retrospective classification of chief complaint
3. Categorization based on clinical impression recorded on surveillance form vs. categorization by ICD-9 coded ED diagnosis
4. Categorization based on EMS dispatch code vs. categorization by hospital emergency department clinician based on chief complaint/presentation
5. Categorization based on EMS dispatch code vs. categorization based on ICD-9 coded ED diagnosis

In this study, validity measures for chief complaint-based syndrome assignment in comparison to an ED diagnosis-based criterion standard were even more comparable to those of EMS-based assignment than were reliability measures. For example, chief complaint and EMS-based categorization exhibited very similar sensitivities, albeit for opposite syndromes. While the chief complaint-based scheme was significantly more sensitive than the EMS-based scheme for the gastrointestinal syndrome, the EMS-based scheme was significantly more sensitive for the respiratory syndrome. In Greenko et al.'s evaluation of an EMS-based categorization scheme for an influenza-like illness syndrome, sensitivity and PVP were also low, 58% and 22% respectively. (139) Table 25 shows the sensitivities for various syndrome assignment schemes reported in the literature in comparison to the results of the current study.

**Table 25**

**Comparison of the Sensitivities of Various Schemes for Assigning Emergency Patients to Respiratory or Influenza-Like Illness (ILI) Syndromes Reported in the Literature with the Current Study**

Comparison	Categorization			
	CC <sup>1</sup>	ICD-9 <sup>2</sup>	Combo <sup>3</sup>	EMS <sup>4</sup>
Criterion standard of respiratory syndrome definition based on retrospective classification according to Hx, PE, Labs <sup>5</sup>				
Respiratory Syndrome	0.47	***	0.72	***
Upper Respiratory Syndrome	***	0.56	***	***
Lower Respiratory Syndrome	***	0.87	***	***
Detection of simulated outbreak <sup>6</sup>				
Hospital 1 Respiratory Syndrome	0.26	0.28	0.34	***
Hospital 2 Respiratory Syndrome	0.36	0.39	0.47	***
Criterion standard of ILI syndrome definition based on retrospective review of medical records <sup>7</sup>	***	***	***	0.58
Criterion standard of respiratory syndrome definition based on ICD-9 coded ED diagnosis <sup>8</sup>	0.31	***	***	0.41

1. Categorization based on chief complaint

2. Categorization based on ICD-9 coded ED diagnosis

3. Categorization based on combination of chief complaint and ICD-9 coded ED diagnosis

4. Categorization based on EMS dispatch code

5. From Beitel et al., 2004

6. From Reis and Mandl, 2004

7. From Greenko et al., 2003

8. From the current study

Second, the primary value of an evaluation such as this is not in establishing whether reliability or validity measures exceed a particular, arbitrary threshold of sufficiency, but rather in simply quantifying those attributes so as to inform the operation of the surveillance system and the interpretation of the information it produces. Indeed, while CDC guidelines suggest the closely related metrics—sensitivity and PVP—are the preferred framework for quantifying the outbreak detection capabilities of syndromic surveillance systems, because acceptable levels of precision will likely vary from jurisdiction to jurisdiction

depending upon the perceived likelihood of an outbreak, benefits of early detection and likely costs of investigating false alarms, specific targets for these metrics are not given. (134)

Third, and perhaps most important, because syndromic surveillance systems in general, not just EMS-based systems, are increasingly perceived to suffer from poor—or at best, inconsistent—outbreak detection capability, the justification for their continued operation has increasingly been restated in terms of their ability to provide what has been called “health situational awareness” and less in terms of outbreak detection. (167) The CDC has defined health situational awareness as “the ability to utilize detailed, real-time health data to confirm, refute and provide an effective response to the existence of an outbreak. It is also used to monitor an outbreak’s magnitude, geography, rate of change and life cycle.” (168) In its recommendations to 911 emergency call centers for developing protocols related to pandemic influenza preparedness, the US National Highway and Traffic Safety Administration (NHTSA) specifically calls for the use of EMS dispatch data for both syndromic surveillance and as a health situational awareness tool. (169)

As a health situational awareness tool, the strength of the correlation between the proxy indicators monitored by a syndromic surveillance system and the actual incidence of outbreak cases is less important than in the case of early outbreak detection. This is because the observed number of events has direct meaning as an indicator of a “ground truth” regarding the demand for health services during an outbreak, rather than just an indirect or proxy meaning as

potential indicators of outbreak cases in an attempt to detect the leading edge of an increase in such cases. In other words, they provide a broad contextual awareness of the local morbidity and mortality background against which the public health significance of specific health events can be viewed.

## LIMITATIONS

The results of this study are subject to certain limitations arising from both the data collection and analysis methodologies. First, the fact that only about 41% of eligible EMS runs could be matched to their corresponding hospital records and the fact the records that were matched were disproportionately female and older when compared to the records that had to be excluded from the study pose a significant challenge to the generalizability of the results of the study. Probabilistic linkage methods have been used to match EMS run records to both hospital (170) and state trauma registry (171) records and have achieved much higher proportions of successful matches than the exact linkage methods used here. However, in those cases many more data fields were available for matching.

Second, while the kappa statistic is currently still considered the most accepted measure of interrater agreement for binary and categorical data, it is a decidedly imperfect and oft criticized statistic that poses problems both generally and in the particular context of these data. Speaking generally, Uebersax (172) has summarized some of the reasons that statisticians have become increasingly wary of kappa as a reliability measure:

1. Kappa is not really a chance-corrected measure of agreement.

2. Kappa is an omnibus index of agreement. It does not make distinctions among various types and sources of disagreement.
3. Kappa is influenced by trait prevalence (distribution) and base-rates.
4. Kappa may be low even though there are high levels of (raw) agreement and even though individual ratings are accurate.
5. Kappa requires that two raters/procedures use the same rating categories.

In terms of this study specifically, reasons three four and five pose particular challenges. As Table 8 shows, trait prevalence was low in the study data and, in most cases, the marginal rates (i.e. rater categorization rates) were significantly different. As Tables 5 – 7 show, raw agreement in these data was quite high. Taken together, these facts suggest the possibility that the kappa statistic may understate the true level of interrater agreement in this study.

One possible solution to these problems would be to use an alternate measurement of agreement that is less sensitive to such violations. In this study, this was done provisionally, using the AC1 statistic developed by Gwet, and the results were considerably different than those achieved using Cohen's kappa. This suggests the AC1 statistic may be a more useful measure of reliability for syndromic surveillance data that are affected by low trait prevalence and disparate rater categorization rates. However, the AC1 statistic has problems of its own, not least of which is the fact that its calculation is rather onerous and few available computer statistical packages are programmed to calculate it. More importantly, the AC1 statistic has not yet gained general acceptability in the statistical community as a standard measure of interrater reliability. In the future,

greater use and exposure may or may not reveal problems inherent in the AC1 statistic, too. Until then, it must be used with caution. (173)



## CONCLUSIONS

The use of EMS dispatch codes for assigning emergency patients to surveillance syndromes has limited but statistically significant reliability in relation to more commonly used syndrome grouping methods based on chief complaints or ICD-9 coded ED diagnoses. The reliability of EMS-based syndrome assignment varies significantly by syndrome, age group and comparison rater. When ICD-9 coded ED diagnosis-based grouping is taken as the criterion standard of syndrome definition, the validity of EMS-based syndrome assignment is limited but comparable to chief complaint-based assignment. The validity of EMS-based syndrome assignment varies significantly by syndrome.

Knowledge of differences in reliability and validity of EMS-based syndrome assignment across syndromes and age groups may be used to identify population groups and disease outcomes for which EMS-based syndromic surveillance might best be employed.

Due to its limited reliability and validity, EMS-based syndromic surveillance is not recommended as a primary early outbreak detection method. Used as an adjuvant in conjunction with other syndromic and conventional methods, however, EMS-based syndromic surveillance may be useful for outbreak corroboration and for providing health situational awareness.

Future evaluations of the reliability of surveillance syndrome categorization methods should carefully consider the use of alternatives to the kappa statistic such as the AC1 statistic, especially when trait prevalence is low and the base rates of syndrome categorization differ significantly between raters.

## REFERENCES

1. Centers for Disease Control and Prevention. Update: Investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. *Morb Mortal Wkly Rep.* 2001;50(41):889-93.
2. Josefson D. US fear of bioterrorism spreads as anthrax cases increase. *BMJ.* 2001;323(7318):877-8.
3. Centers for Disease Control and Prevention. Outbreak of West Nile-like viral encephalitis—New York, 1999. *Morb Mortal Wkly Rep.* 1999;48(38):845-9.
4. Centers for Disease Control and Prevention. Outbreak of severe acute respiratory syndrome—worldwide, 2003. *Morb Mortal Wkly Rep.* 2003;52(11):226-8.
5. Larkin M. Avian flu: sites seek to respond and reassure. *Lancet Infect Dis.* 2005;5(3):129-31.
6. Spellberg B, Guidos R, Gilbert D, et al. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Disease Society of America. *Clin Infect Dis.* 2008;46(2):155-64.
7. Wagner MM, Tsui FC, Espino JU, et al. The emerging science of very early detection of disease outbreaks. *J Public Health Management Practice.* 2001;7(6):51-9.
8. Pavlin JA, Mostashari F, Kortepeter MG, et al. Innovative surveillance methods for rapid detection of disease outbreaks and bioterrorism: results of an interagency workshop on health indicator surveillance. *Am J Public Health.* 2003;93:1230-5.
9. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism related diseases. *Ann Intern Med.* 2004;140:910-22.
10. Green MS, Kaufman Z. Surveillance for early detection and monitoring of infectious disease outbreaks associated with bioterrorism. *Isr Med Assoc J.* 2002;4(7):503-6.

11. Zelicoff A, Brillman J, Forslund DW, et al. The rapid syndrome validation project. *Proc Am Med Inform Assoc.* 2001;Ann Symp:771-5.
12. Quenel P, Dab W, Hannoun C, Cohen JM. Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics. *Int J Epidemiol.* 1994;32:849-55.
13. Centers for Disease Control and Prevention. Annotated Bibliography for Syndromic Surveillance. [online] 2006 [accessed 2008 March 13]. Available from: [http://www.cdc.gov/ncphi/diss/hndss/files/Syn\\_Surv\\_biblio.pdf](http://www.cdc.gov/ncphi/diss/hndss/files/Syn_Surv_biblio.pdf)
14. Sosin, D.M. (2003). Syndromic Surveillance: The Case for Skillful Investment View, Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 1, 247-53.
15. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. Syndromic surveillance and bioterrorism-related epidemics. *Emerg Infect Dis.* 2003;9:1197-1204.
16. Henning KJ. Syndromic Surveillance. In Institute of Medicine 2003. *Microbial Threats to Health: Emergence, Detection and Response.* Ed. Smolinski MS, Hamburg MA, Lederberg J. The National Academies Press, Washington DC. 309-50.
17. Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis.* 2004;10:858-64.
18. Henning KJ. What is syndromic surveillance. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):18-22.
19. Mostashari F, Hartman J. Syndromic surveillance: a local perspective. *J Urban Health* 2003;80(2 Suppl 1):i1-7.
20. Goodwin T, Noji E. Syndromic Surveillance. *Euro J Emerg Med.* 2004;11(1):1-2.
21. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. *J Am Med Inform Assoc.* 2004;11(2):141-50.
22. Grosskurth H, Mosha F, Todd J, et al. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomized controlled trial. *Lancet.* 1995;346:530-6.

23. Harrison A, Wilkinson D, Lurie M, Connolly AM, Karim SA. Improving quality of sexually transmitted disease case management in rural South Africa. *AIDS*. 1998;12(17):2329-35.
24. Matteelli A, Kassa A, Gerbase A, et al. Passive sentinel surveillance system for sexually transmitted diseases in primary healthcare sites in Ethiopia, 1991-3. *Sex Transm Infect*. 2000;76(2):131-3.
25. La Ruche G, Ladner J, Lattier R, Djeha D, Louise D, Coulibaly IM. Surveillance of STD syndromes: contributing to the STD programme in Cote d'Ivoire. *Health Policy Plan*. 2000;15(4):441-6.
26. Syndromic Surveillance. *Healthcare Hazard Management Monitor*. 2003;16(8):1-6.
27. Koo D. Leveraging syndromic surveillance. *J Public Health Manag Pract*. 2005;11(3):181-3.
28. Bradley CA, Rolka H, Walker D, Loonsk J. BioSense: Implementation of a national early event detection and situational awareness system. *Morb Mortal Wkly Rep*. 2005;54(Suppl):11-9.
29. Koo D, Parrish RG. The changing healthcare information infrastructure in the United States: opportunities for a new approach to public health surveillance. In: Teutsch SM, Churchill RE, eds. *Principles and practice of Public health Surveillance*. 2<sup>nd</sup> ed. New York: Oxford University Press; 2000:76-94.
30. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. *J Am Med Inform Assoc*. 2002;9(2):105-15.
31. Teich JM, Wagner MM, Mackenzie CF, Schafer KO. The informatics response in disaster, terrorism, and war. *J Am Med Inform Assoc*. 2002;9(2):97-104.
32. Platt R, Bocchino C, Caldwell B, et al. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. *J Urban Health*. 2003;80(Suppl 1):i25-i31.
33. Barthell EN, Aronsky D, Cochrane DG, Cable G, Stair T. The Frontlines of Medicine Project progress report: standardized communication of emergency department triage data for syndromic surveillance. *Ann Emerg Med*. 2004;44(3):247-52.

34. Grabenstein J, Downs K, Dotson D. Extraordinary infections: a focus on bioterrorism. *J Am Pharm Assoc.* 2000;40(Suppl 1):S36-S37.
35. Leach DL, Ryan DG. Biological weapons: preparing for the worst. *MLO Med Lab Obs.* 2000;32(9):26-9.
36. Fraser MR, Brown DL. Bioterrorism preparedness and local public health agencies: building response capacity. *Public Health Rep.* 2000;115(4):329-30.
37. Centers for Disease Control and Prevention. Addressing emerging infectious disease threats: a prevention strategy for the United States. Executive summary. *MMWR Recomm Rep.* 1994;43(RR-5):1-18.
38. Hughes JM, La Montagne JR. Emerging infectious diseases. *J Infect Dis.* 1994;170(2):263-4.
39. Bryan RT, Pinner RW, Berkelman RL. Emerging infectious diseases in the United States, Improved surveillance, a requisite for prevention. *Ann N Y Acad Sci.* 1994;740:346-61.
40. National Electronic Disease Surveillance System Working Group. National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine. *J Public Health Manag Pract.* 2001;7(6):43-50.
41. Burkom HS, Elbert Y, Feldman A, Lin J. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):67-73.
42. Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):159-65.
43. Jorm LR, Thackway SV, Churches TR, Hills MW. Watching the games: public health surveillance for the Sydney 2000 Olympic Games. *J Epidemiol Community Health.* 2003;57(2):102-8.
44. Matsui T, Takahashi H, Ohyama T, et al. An evaluation of syndromic surveillance for the G8 summit in Miyazaki and Fukuoka, 2000. *J Japanese Assoc Infect Dis.* 2002;76(3):161-6.
45. Centers for Disease Control and Prevention. Syndromic surveillance for bioterrorism following the attacks on the World Trade Center. *Morb Mortal Wkly Rep.* 2002;51(Special Issue):13-5.

46. Das D, Weiss D, Mostashari F, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. *J Urban Health*. 2003;80(Suppl 1):i76-i88.
47. Hopkins RS. Design and operation of state and local infectious disease surveillance systems. *J Public Health Manag Pract*. 2005;11(3):184-90.
48. Fienberg SE, Shmueli G. Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Stat Med*. 2005;24:513-29.
49. International Society for Disease Surveillance. *Syndromic.org*. [online] 2008 [accessed 2008 March 14]. Available from: <http://www.syndromic.org>.
50. International Society for Disease Surveillance. *Advances in Disease Surveillance*. [online] 2008 [accessed 2008 March 14]. Available from: <http://www.isdsjournal.org>.
51. Reis BY, Mandl KD. Syndromic surveillance: the effect of syndrome grouping on model accuracy and outbreak detection. *Ann Emerg Med*. 2004;44:235-41.
52. Mandl KD, Reis BY, Cassa C. Measuring outbreak-detection performance by using controlled feature set simulations. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):130-6.
53. Wethington H, Bartlett P. The RUSick2 foodborne disease forum for syndromic surveillance. *Emerg Infect Dis*. 2004;10(3):401-5.
54. Leonardis GS, Hajat S, Kovats RS, Smith GE, Cooper D, Gerard E. Syndromic surveillance use to detect the early effects of heat waves: an analysis of NHS direct data in England. *Soz Praventivmed*. 2006;51(4):194-201.
55. Jones NF, Marshall R. Evaluation of an electronic general-practitioner-based syndromic surveillance system – Auckland, New Zealand, 2000-2001. *MMWR*. 2004;53(Suppl 1):173-8.
56. Miller B, Kassenborg H, Dunsmuir W, et al. Syndromic surveillance for influenzalike illness in an Ambulatory Care Network. *Emerg Infect Dis*. 2004;10(10):1806-11.
57. Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity of bioterrorism and influenza surveillance. *Ann Emerg Med*. 2003;41(4):447-52.

58. Hadler JL, Siniscalchi A, et al. Hospital admissions syndromic surveillance—Connecticut, October 2001—June 2004. *Morb Mortal Wkly Rep.* 2005;54(Suppl 1):169-73.
59. 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(9). [online] 2001 [accessed 2008 March 14]. Available from: <http://www.biomedcentral.com/1471-2458/1/9>.
60. Zeng X, Wagner M. Modeling the effects of epidemics on routinely collected data. *Proc AMIA Annu Symp.* 2001:781-5.
61. Lazarus R, Kleinman K, Dashevsky I, et al. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerg Infect Dis.* 2002;8(8):753-60.
62. Yih WK, Caldwell B, Harmon R, et al. National bioterrorism syndromic surveillance demonstration program. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):43-6.
63. Hammond L, Papadopoulos S, Johnson C, MaWhinney S, Nelson B, Todd, JK. Use of an internet-based community surveillance network to predict seasonal communicable disease morbidity. *Pediatrics.* 2002;109(3):414-8.
64. Henry JV, Magruder S, Snyder M. Comparison of office visit and nurse advice hotline data for syndromic surveillance – Baltimore – Washington, D.C., Metropolitan Area, 2002. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1): 112-6.
65. Lober WB, Trigg LJ, Karras BT, et al. Syndromic surveillance for bioterrorism using automated collection of computerized discharge diagnoses. *J Urban Health.* 2003;80(Suppl 1):i97-i106.
66. Brinsfield KH, Bunn JE, Barry MA, McKenna V, Dyer KS, Sulis C. Using volume-based surveillance for an outbreak early warning system. *Acad Emerg Med.* 2001;8:492.
67. Mundorff MB, Gesteland P, Haddad M, Rolfs R. Syndromic surveillance using chief complaints from urgent-care facilities during the Salt Lake 2002 Olympic Winter Games. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):254.
68. Beitel AJ, Olson KL, Reis BY, Mandl KD. Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Ped Emerg Care.* 2004;20(6):355-60.



69. Cochrane DG. Perspective of an emergency physician group as a data provider for syndromic surveillance. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):209-14.
70. Hirshon JM. The rationale for developing public health surveillance systems based on emergency department data. *Acad Emerg Med.* 2000;7(12):1428-32.
71. Mikosz CA, Silva J, Black S, Gibbs G, Cardenas I. Comparison of two major emergency department-based free-text chief-complaint coding systems. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):101-5.
72. Townes JM, Kohn MA, Southwick KL, et al. Investigation of an electronic emergency department information system as a data source for respiratory syndrome surveillance. *J Public Health Manag Pract.* 2004;10(4):299-307.
73. Foldy S, Biedrzycki PA, Barthell EN, et al. Syndromic surveillance using regional emergency medicine internet. *Ann Emerg Med.* 2004;44(3):242-6.
74. Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity of bioterrorism and influenza surveillance. *Ann Emerg Med.* 2003;41(4):447-52.
75. Travers DA, Waller A, Haas SW, et al. Emergency department data for bioterrorism surveillance: electronic data availability, timeliness, sources and standards. *AMIA Annu Symp Proc.* 2003;2003:664-8.
76. Chapman WW, Christensen LM, Wagner MM, Haug PJ, Ivanov O, Dowling JN, Olszewski RT. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artificial Intelligence in Medicine.* 2005;33:1-10.
77. Ivanov O, Wagner MM, Chapman WW, Olszewski RT. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. *Proc AMIA Symp.* 2002;:345-9.
78. Tsui F, Wagner M, Dato V, Chang C. Value of ICD-9-Coded Chief Complaints for Detection of Epidemics. *Proc AMIA Symp.* 2001;:711-5.
79. Espino, J.U., Wagner, M.M. Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. *Proc AMIA Symp.* 2001;:164-8.
80. Day FC, Schriger DL, La M. Automated linking of free-text complaints to reason-for-visit categories and *International Classification of Diseases* diagnoses

in emergency department patient record databases. *Ann Emerg Med*. 2004;43(3):401-9.

81. Begier EM, Sockwell D, Branch LM, et al. The National Capitol Region's emergency syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? *Emerg Infect Dis*. 2003;9(3):393-6.

82. Fleischauer AT, Silk BJ, Schumacher M, et al. The validity of chief complaint and discharge diagnosis in emergency department-based syndromic surveillance. *Acad Emerg Med*. 2004;11:1262-7.

83. Centers for Disease Control and Prevention. Syndrome definitions for diseases associated with critical bioterrorism-associated agents. [online] 2003 [accessed 2007 May 31]. Available from: <http://www.bt.cdc.gov/surveillance/syndromedef/>.

84. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak*. 2003;3:2.

85. Nobre FF, Monteiro ABS, Telles PL, Williamson GD. Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology.

86. Burkom HS. Development, adaptation, and assessment of alerting algorithms for biosurveillance. *Johns Hopkins APL Technical Digest*. 2003;24(4):335-42.

87. Burkom HS. Biosurveillance applying scan statistics with multiple, disparate data sources. *J Urban Health*. 2003;80(Suppl 1):i57-i65.

88. Kleinman KP, Abrams AM, Kulldorff M, Platt R. A model-adjusted space-time scan statistic with an application to syndromic surveillance. *Epidemiol Infect*. 2005;133:409-19.

89. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med*. 2005;2(3):216-24.

90. Wallenstein S, Naus J. Scan statistics for temporal surveillance for biologic terrorism. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):74-8.

91. Duczmal L, Buckeridge D. Using modified spatial scan statistic to improve detection of disease outbreak when exposure occurs in the workplace—Virginia 2004. *Morb Mortal Wkly Rep*. 2005;54(Suppl 1):187.

92. Neill DB, Moore A. Fast grid-based scan statistic for detection of significant spatial disease clusters. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):255.
93. Bonetti M, Olson KL, Mandl KD, Pagano M. Parametric models for interpoint distances and their use in biosurveillance. *Proceedings of the American Statistical Association, Biometrics Section.* Alexandria, VA: American Statistical Association, 2003.
94. Groenewold MR. Comparison of two signal detection methods in a coroner-based system for near real-time mortality surveillance. *Public Health Rep.* 2007;122:521-30.
95. Rolka H, Bracy D, Russell C, Fram D, Ball R. Using simulation to assess the sensitivity and specificity of a signal detection tool for multidimensional public health surveillance data. *Statist Med.* 2005;24(4):551-62.
96. O'Brien SJ, Christie P. Do CuSums have a role in routine communicable disease surveillance? *Public Health.* 1997;111(4):255-8.
97. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM RSPRT charts for monitoring in medical contexts. *Statist Methods Med Res.* 2003;12(2):147-70.
98. Hanslik T, Boelle PY, Flahault A. The control chart: an epidemiological tool for public health monitoring. *Public Health.* 2001;115:227-81.
99. Steiner-Sichel L, Greenko J, Heffernan R, Layton M, Weiss D. Field investigations of emergency department syndromic surveillance signals—New York City. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):184-9.
100. Terry W, Ostrowsk B, Huang A. Should we be worried? Investigation of signals generated by an electronic syndromic surveillance system—Westchester County, New York. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):190-5.
101. Duchin JS. Epidemiological response to syndromic surveillance signals. *J Urban Health.* 2003;80(2) Supplement 1:i115-i116.
102. Chen JH, Schmit K, et al. Use of Medicaid prescription data for syndromic surveillance—New York. *Morb Mortal Wkly Rep.* 2005;54(Suppl 1):31-4.
103. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proceedings of the National Academy of Sciences of the United States of America.* 2002;99(8):5237-40.

104. Magruder SF, Lewis SH, Najmi A, Florio E. Progress in understanding and using over-the-counter pharmaceuticals for syndromic surveillance. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):117-22.
105. Pavlin JA, Murdock P, Elbert E, et al. Conducting population behavioral health surveillance by using automated diagnostic and pharmacy data systems. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):166-72.
106. Najmi AH, Magruder SF. Estimation of hospital emergency room data using otc pharmaceutical sales and least mean squares filters. *BMC Med Inform Decis Mak.* 2004;4:5.
107. Wagner MM, Tsui F-C, Hogan W, et al. National Retail Data Monitor for public health surveillance. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):40-2.
108. Edge VL, Pollari F, Lim G, et al. Syndromic surveillance of gastrointestinal illness using pharmacy over-the-counter sales. A retrospective study of waterborne outbreaks in Saskatchewan and Ontario. *Can J Public Health.* 2004;95(6):446-50.
109. Das D, Mostashari F, Weiss D, Balter S, Heffernan R. Monitoring over-the-counter pharmacy sales for early outbreak detection – New York City, August 2001 – September 2003. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):235.
110. Derby MP, McNally J, Ranger-Moore J, et al. Poison control center-based syndromic surveillance for foodborne illness. *Morb Mortal Wkly Rep.* 2005;54(Suppl):31-4.
111. Abubakar I, Leonardi GS, Edwards N, Herriott N. Inter-rater agreement in defining chemical incidents at the National Poisons Information Service, London. *J Epidemiol Community Health.* 2004;58:718-22.
112. Cooper DL, Smith G, Baker M, et al. National symptom surveillance using calls to a telephone health advice service – United Kingdom, December 2001 – February 2003. *Morb Mortal Wkly Rep.* 2004;53(Suppl 1):179-83.
113. Harcourt SE, Smith GE, Hollyoak V, et al. Can calls to NHS Direct be used for syndromic surveillance? *Commun Dis Public Health.* 2001;4:178-88.
114. Nicoll A, Smith G, Cooper D, Chinemana F, Gerard E. The public health value of syndromic surveillance: calls to a national health help-line (NHS Direct). *Euro J Public Health.* 2004;14(4) Supplement 1:68-9.
115. Rodman JS, Frost F, Jakubowski W. Using nurse hot line calls for disease surveillance. *Emerg Infect Dis.* 1998;4(2):329-32.

116. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection, New York City. *BMC Public Health*. 2005;5:105.
117. Mostashari F, Fine A, Das D, Adams J, Layton M. Use of ambulance dispatch data as an early warning system for influenza like illness, New York City. *J Urban Health*. 2003;80(2 Suppl 1):i43-9.
118. Bork KH, Klein BM, Molbak K, Trautner S, Petersen UB, Heegaard E. Surveillance of ambulance dispatch data as a tool for early warning. *Eurosurveillance*. 2006;11(12):229-33.
119. Dockrey MR, Trigg LJ, Lober WB. An information systems for 911 dispatch monitoring system and analysis. *Proc AMIA 2002 Annu Symp*:: 1008.
120. Stetson D, Bloom R, Crary D, Cheng K, McClellan G. Temporal correlation of nontraditional and traditional evidence of a natural outbreak. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):259.
121. Sosin DM, DeThomasis J. Evaluation challenges for syndromic surveillance—making incremental progress. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):125-9.
122. Stout Solutions, LLC. FirstWatch: real-time early warning system. [online] 2008 [Accessed 16 March 2008]. Available from: <http://www.firstwatch.net>.
123. MacDonald RD, Farr B, Neill M, et al. An emergency medical services transfer authorization center in response to the Toronto severe acute respiratory syndrome outbreak. *Prehosp Emerg Care*. 2004;8(2):223-31.
124. Busko J. EMS and medical surveillance: is there a role for us in separating the signal from the noise? *Emerg Med Serv*. [online] 2008 [Accessed 26 February 2008]. Available from: <http://www.emsresponder.com/Emergency-Medical-Services/EMS-and-medical-surveillance.html>.
125. Carrico R, Goss L. Syndromic surveillance: hospital emergency department participation during the Kentucky Derby Festival. *Disaster Manag Response*. 2005;3(3):73-9.
126. Louisville Metro Public Health and Wellness, Office of Emergency and Public Health Preparedness. Draft syndromic surveillance standard operating guidelines. Louisville (KY): Louisville Metro Public Health and Wellness; 2006 Jun. 23 p.

127. Priority Dispatch. EMD Medical ProQA. [online] 2007 [Accessed 17 March 2008]. Available from:  
<http://www.prioritydispatch.net/index.php?a=products&b=medproqa>.
128. NAEMD. National Academies of Emergency Medical Dispatch. [online] 2008 [Accessed 17 March 2008]. Available from:  
<http://www.emergencydispatch.org>.
129. Reingold A. If syndromic surveillance is the answer, what is the question? *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*. 2003;1(2):1-5.
130. Stoto MA, Schonlau M, Mariano LT. Syndromic surveillance: is it worth the effort? *Chance*. 2004;17(1):19-24.
131. Buckeridge DL, Burkom H, Moore A, Pavlin J, Cutchis P, Hogan W. Evaluation of syndromic surveillance systems – design of an epidemic simulation model. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):137-43.
132. Hutwagner LC, Thompson WW, Seeman GM, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Statist Med*. 2005;24:543-50.
133. Kleinman KP, Abrams A, et al. Simulations for assessing statistical methods of biologic terrorism surveillance. *Morb Mortal Wkly Rep*. 2005;54(Suppl 1):101-8.
134. Centers for Disease Control and Prevention. Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group. *MMWR Recomm Rep*. 2001;50(No. RR-13).
135. Centers for Disease Control and Prevention. Framework for evaluating public health surveillance systems for early detection of outbreaks; recommendations from the CDC Working Group. *MMWR Recomm Rep*. 2004;53(No. RR-5):1-13.
136. Cochrane DG, Allegra J, Rothman J. Physician's choice of charting template versus ICD-9 code – agreement between two syndromic surveillance methods using emergency department electronic medical records. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):234.
137. Mocny M, Cochrane D, Allegra J, et al. Improving agreement between two algorithms for biosurveillance of respiratory disease in the emergency department – chief complaint and ICD-9 code. *Morb Mortal Wkly Rep*. 2004;53(Suppl 1):253.

138. Raman S, Levin J, Hall D, Frey K. Using categorization of reason-for-visit strings as the basis for an outbreak detection system—Minnesota, 2002-2003. *Morb Mortal Wkly Rep.* 2005;54(Suppl 1):200.
139. Greenko J, Mostashari F, Fine A, Layton M. Clinical evaluation of the Emergency Medical Services (EMS) ambulance dispatch-based syndromic surveillance system, New York City. *J Urban Health.* 2003;80(Suppl 1):i50-i56.
140. Louisville Metro Public Health and Wellness. Health and Wellness. [online] 2008 [Accessed 17 March 2008]. Available from: <http://www.louisvilleky.gov/Health>.
141. Louisville/Jefferson County Metro Government. The Official Guide to Louisville Metro Government. [online] 2008 [Accessed 17 March 2008]. Available from: <http://louisvilleky.gov/NR/rdonlyres/F3D49926-B4FE-4516-95D8-DD5CF86186BA/0/governmentguide.pdf>.
142. Louisville Metro Public Health and Wellness, Office of Policy Planning and Evaluation. Health Status Assessment Report 2007. [online] 2007 [Accessed 17 March 2008]. Available from: [http://www.louisvilleky.gov/NR/rdonlyres/24260BFF-D353-437C-9A82-3EB59FD37BC2/0/HSAR\\_2007.pdf](http://www.louisvilleky.gov/NR/rdonlyres/24260BFF-D353-437C-9A82-3EB59FD37BC2/0/HSAR_2007.pdf).
143. Norton Healthcare. About Norton Healthcare. [online] 2008 [Accessed 17 March 2008]. Available from: <http://www.nortonhealthcare.com/about>.
144. Kentucky Cabinet for Health and Family Services, Division of Certificate of Need. Inventory of Kentucky Health Facilities, Health Services and Major Medical Equipment. [online] 2007 [Accessed 17 March 2008]. Available from: <http://chfs.ky.gov/NR/rdonlyres/D243B5CB-3310-4DE4-A9D1-0E1ACF8DC2BE/0/inventory.xls>.
145. Norton Healthcare. Kosair Children's Hospital. [online] 2008 [Accessed 17 March 2008]. Available from: <http://www.nortonhealthcare.com/locations/hospitals/kosair/index.aspx>.
146. Louisville Metro Emergency Medical Services. EMS. [online] 2008 [Accessed 17 March 2008]. Available from: <http://www.louisvilleky.gov/EMS>.
147. Garza M. Louisville Slugger: one city's miraculous EMS transformation. *J Emerg Med Serv.* 2006;31(8):56-61.
148. EMS Runs [Internet]. Louisville: Louisville/Jefferson County Metro Government. 2002- [Accessed 17 March 2008]. Available from: <http://reports.metronet.gov/ReportManager/Pages/Report.aspx?ItemPath=%2fHealthReports%2fEmsRuns>.

149. Louisville MetroSafe. MetroSafe. [online] 2008 [Accessed 17 March 2008]. Available from: <http://www.louisvilleky.gov/MetroSafe>.
150. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37–46.
151. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971;76:378-82.
152. Fleiss JL. *Statistical Methods for Rates and Proportions.* Wiley: New York, 1981.
153. Gwet, K. *Handbook of Inter-Rater Reliability: How to measure the level of agreement between two or multiple raters.* Stataxis Publishing Company: Gaithersburg, MD, 2001.
154. Gwet K. INTER\_RATER macro. [online] 2002 [Accessed 17 March 2008]. Available from: [http://www.stataxis.com/files/sas/Inter\\_Rater.txt](http://www.stataxis.com/files/sas/Inter_Rater.txt).
155. Abdi H. Bonferroni and Sidak corrections for multiple comparisons. In: Salkind NJ, editor. *Encyclopedia of Measurement and Statistics.* Thousand Oaks, CA: Sage; 2007.
156. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika.* 1947;12:153-7.
157. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med.* 1987;6:441-8.
158. Kraemer HC, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Statist Med.* 2002;21:2109-29.
159. Tooth LR, Ottenbacher KJ. The  $\kappa$  statistic in rehabilitation research: an examination. *Arch Phys Med Rehabil.* 2004;85:1371-6.
160. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37(5):360-3.
161. Donner A. Sample size requirements for the comparison of two or more coefficients of inter-observer agreement. *Statist Med.* 1998;17:1157-68.
162. Schlesselman JJ. *Case-Control Studies: Design, Conduct, Analysis.* Oxford: New York, 1982.



163. Casagrande JT, Pike MC, Smith PG. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics*. 1978;34:483-6.
164. Dupont WD. Power calculations for matched case-control studies. *Biometrics*. 1988;44:1157-68.
165. Dupont WD, Plummer WD. Power and sample size calculations: a review and computer program. *Controlled Clin Trials*. 1990;11:116-28.
166. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
167. McManus J, Huebner K, Scheulen J. The science of surge: Detection and situational awareness. *Acad Emerg Med*. 2006;13:1179-82.
168. Centers for Disease Control and Prevention. Early event detection and health situational awareness. [online] 2007 [Accessed 26 March 2008]. Available from: <http://www.cdc.gov/biosense/publichealth.htm#early>.
169. Working Group on 9-1-1 Readiness for Pandemic Influenza. Preparing for pandemic influenza: recommendations for protocol development for 9-1-1 personnel and public safety answering points (PSAPs). Washington DC: National Highway and Traffic Safety Administration (US); 2007. Appendix C: Using 9-1-1 and EMS data for situational awareness and surveillance; p. 46-52.
170. Dean JM, Vernon DD, Cook L, Nechodom P, Reading J, Suruda A. Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: a potential tool for evaluation of emergency medical services. *Ann Emerg Med*. 2001;37(6):616-26.
171. Newgard CD. Validation of probabilistic linkage to match de-identified ambulance records to a state trauma registry. *Acad Emerg Med*. 2006;13:69-75.
172. Uebersax J. Kappa Coefficients: A Critical Appraisal. [online] 2002 [Accessed 23 March 2008]. Available from: <http://www.ourworld.compuserve.com/homepages/jsuebersax/kappa.htm>.
173. Blood E, Spratt KF. Disagreement on agreement: two alternative agreement coefficients. Paper 186-2007. Paper presented at: SAS Global Forum 2007. 2007 Apr 16-19; Orlando, FL, USA.

## **CURRICULUM VITAE**

**Matthew Groenewold, PhD (Candidate), MSPH**  
**1853 Gresham Rd**  
**Louisville, KY 40205**

### **PERSONAL INFORMATION**

Country of Citizenship: United States  
Home Phone Number: (502) 468-3507  
Work Phone Number: (502) 574-5292  
Fax Number: (502) 457-6202  
E-mail Address: matt.groenewold@louisvilleky.gov

### **EDUCATION**

1/03-5/08 – University of Louisville School of Public Health and Information Sciences, Louisville, KY – PhD (Epidemiology) [Currently ABD, degree expected 5/08, admitted to candidacy 3/05]  
8/01-12/02 – University of Louisville School of Public Health and Information Sciences, Louisville, KY – MSPH (Epidemiology)  
8/01-6/02 – University of Louisville School of Public Health and Information Sciences, Louisville, KY – Graduate Certificate (Clinical Investigation Sciences)  
8/94-5/97 – Indiana University Southeast, New Albany, IN – BA (Biology)  
8/83-6/87 – Columbus North High School, Columbus, IN

### **PROFESSIONAL TRAINING**

7/02 – Epidemiology Rapid Response Team Training, Kentucky Department for Public Health, Frankfort, KY  
8/00 – WMD-Haz Mat Technician (COBRA) Training, US DOJ, Office of Justice Programs, Center for Domestic Preparedness, Anniston, AL  
6/99 – Basic Environmental Crime Investigations, Southern Environmental Enforcement Network, Atlanta, GA  
2/99 – Emergency Response to Hazardous Materials Incidents, Haz Mat Technician Training, US EPA and Tetra Tech NUS, Inc., Charlestown, IN  
1/93-1/94 – Methodist Hospital Paramedic Training Program, Methodist Hospital, Indianapolis, IN  
5/90 – Emergency Medical Technician Training, Maryland Fire and Rescue Institute, College Park, MD  
1/90-3/90 – Gunner's Mate "A" School, USCG Reserve Training Center, Yorktown, VA

2/89 – Maritime Law Enforcement Training, Boarding Officers Course, First Coast Guard District Operations Training Team, Otis ANG, MA  
7/88-9/88 – Coast Guard Recruit Training, USCG Training Center, Cape May, NJ

### **WORK/VOLUNTEER EXPERIENCE**

8/06-Present-Adjunct Instructor of Health and Sports Sciences: Responsible for instructing undergraduate and graduate courses in biostatistics, epidemiology, health promotion and public health. Prepare, administer and grade lectures, tests and assignments. – University of Louisville, College of Education and Human Development, Louisville, KY.

7/03-Present - Epidemiologist: Assigned to the Office of Emergency and Public Health Preparedness. Responsible for planning and implementing epidemiological surveillance systems for bioterrorism and other events affecting public health. Design and conduct epidemiological studies. Conduct epidemiological field investigations. – Louisville Metro Department of Public Health and Wellness, Louisville, KY.

7/02-7/03 - Health Education Specialist II: Assigned to Childhood Lead Poisoning Prevention Program. Responsible for developing and preparing health education materials and programs. Conduct health education and training programs at various sites for various target audiences. Perform scientific and health education research and provide consultation to professionals on health related topics. – Louisville Metro Health Department, Louisville, KY. Promoted to Epidemiologist.

8/98-7/02 - Environmental Health Specialist: Hazardous Materials Technician. Member of department's Health Emergency Action Team and county Multi-Agency Haz Mat Team. Responsible for responding to emergencies involving hazardous materials. Perform environmental inspections, assessments and investigations. Investigate environmental crimes and citizen complaints involving hazardous or toxic substances. Perform indoor air quality investigations. – Jefferson County Health Department, Louisville, KY. Accepted lateral transfer opportunity to Health Education Specialist II position.

1/94-8/98 - Paramedic: Provide advanced emergency medical care to sick and injured persons in the pre-hospital environment. Transport patients to hospital. – Yellow Emergency Medical Services and Operation Life Emergency Medical Services, Clark County, IN. Accepted Environmental Health Specialist position with Jefferson County Health Department.

8/92-1/94 - Emergency Medical Technician: Provide basic and intermediate level emergency medical care to sick and injured persons in the pre-hospital environment. Transport patients to hospital. – Putnam County Operation Life, Greencastle, IN. Accepted Paramedic position with Operation Life EMS in Clark County, IN.

7/88-7/92 - Gunner's Mate: Member of deployable law enforcement detachment (LEDET) from Coast Guard Group Baltimore. Served in Middle East Theater of Operations in Operation Desert Storm enforcing U.N. embargo of Iraq imposed by Security Council resolutions 661 and 665. Participated in maritime drug interdiction operations in the Caribbean Atlantic and East Pacific. Served on patrol boat CGC Point Hannon (WPB 82355). Participated in search and rescue and fisheries law enforcement missions in the North Atlantic. Held Secret security clearance. – United States Coast Guard.

### **HONORS/AWARDS/ACTIVITIES**

2001 – Phi Kappa Phi Academic Honors Society  
1994 – Valedictorian, Methodist Hospital Paramedic Training Program  
1991 – Navy Achievement Medal, Awarded for service in Operation Desert Storm  
1989 – Valedictorian, Maritime Law Enforcement Training, Boarding Officer Course  
1988 – Honor Graduate, US Coast Guard Recruit Training

### **PUBLICATIONS**

#### **Refereed Journals:**

Groenewold, MR. Comparison of two signal detection methods in a coroner-based system for near real time mortality surveillance. *Public Health Reports* 2007;122(4):521-30.

Groenewold, MR, Troutman, A, Nielsen, J. Association of environmental lead exposure with academic performance: Louisville, Kentucky, 2003. *Journal of the Kentucky Medical Association* 2007;105(3):97-109.

Groenewold, MR. Enhancing local health department disaster response capacity with rapid community needs assessments: validation of a computerized program for binary attribute cluster sampling. *Prehospital and Disaster Medicine* 2006;21(1):32-9.

Groenewold, MR. Analysis for space-time clustering of CJD cases, Kentucky, 1988-1997. *Journal of the Kentucky Medical Association* 2004;102(4):163-70.

#### **Other Journals:**

Groenewold, MR. Forensic epidemiology: Kentucky law enforcement officers train to investigate bioterrorism. *Kentucky Law Enforcement* 2007;6(3):30-34.

#### **National Meeting Presentations:**

Varaprasad, I, Aldrich TE, Kennedy, MA, Groenewold, MR. Border effects: Who's on first?? Poster presented at the 2006 PHIN Conference. September 2006; Atlanta, Georgia.

Groenewold, MR, Aldrich, TE. A coroner-based surveillance system in Louisville, KY. Poster presented at the CSTE 2004 Annual Conference. June 2004; Boise, Idaho.

Groenewold, MR. A model sentinel event surveillance system in Louisville, KY. Poster presented at the Environmental Public Health Tracking Conference. March 2004; Philadelphia, Pennsylvania.

Aldrich, TE, Groenewold, MR, Hicks, M, Lawson, P, Saad, F. Statistical assessment for clustering of rare events. Paper presented at the 2<sup>nd</sup> Annual Hawaii International Conference on Statistics and Related Fields. June 2003; Honolulu, Hawaii.

Groenewold, M, Lawson, P, Saad, F. & Aldrich, TE. Analyses for space-time clustering of rare events. Poster presented at the Clinical Research 2003 National Meeting. March 2003; Baltimore, Maryland.

Groenewold, MR, Aldrich, TE. Analysis for space-time clustering of CJD cases, Kentucky, 1988-1997. Poster presented at the 17<sup>th</sup> National Conference on Chronic Disease Prevention and Control. March 2003; St. Louis, Missouri.