LNCS 4506

Daniel Zeng et al. (Eds.)

Intelligence and Security Informatics: Biosurveillance

Second NSF Workshop, BioSurveillance 2007
New Brunswick, NJ, USA, May 2007
Proceedings

Springer

Daniel Zeng Ivan Gotham
Ken Komatsu Cecil Lynch
Mark Thurmond David Madigan
Bill Lober James Kvach
Hsinchun Chen (Eds.)

Intelligence and Security Informatics: Biosurveillance

Second NSF Workshop, BioSurveillance 2007
New Brunswick, NJ, USA, May 22, 2007
Proceedings

Springer

Volume Editors

Daniel Zeng
University of Arizona, Tucson, USA, E-mail: zeng@eller.arizona.edu

Ivan Gotham
New York State Department of Health, Albany, USA, E-mail: ijg01@health.state.ny.us

Ken Komatsu
Arizona Department of Health Services, Phoenix, USA, E-mail: komatsk@azdhs.gov

Cecil Lynch
University of California at Davis, Sacramento, USA, E-mail: colynch@ucdavis.edu

Mark Thurmond
University of California, Davis, USA, E-mail: mcthurmond@ucdavis.edu

David Madigan
Rutgers University, New Brunswick, NJ, USA, E-mail: dmadigan@rci.rutgers.edu

Bill Lober
University of Washington, Seattle, USA, E-mail: lober@u.washington.edu

James Kvach
Lexington, VA, USA, E-mail: jandk30@aol.com

Hsinchun Chen
University of Arizona, Tucson, USA, E-mail: hchen@eller.arizona.edu

Library of Congress Control Number: 2007926312

CR Subject Classification (1998): H.4, H.3, C.2, H.2, J.3, I.6, G.3, K.4.1

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-72607-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72607-4 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12066490 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4506

Preface

The 2007 NSF BioSurveillance Workshop (BioSurveillance 2007) was built on the success of the first NSF BioSurveillance Workshop, hosted by the University of Arizona's NSF BioPortal Center in March 2006. BioSurveillance 2007 brought together infectious disease informatics (IDI) researchers and practitioners to discuss selected topics directly relevant to data sharing and analysis for real-time animal and public health surveillance. These researchers and practitioners represented a wide range of backgrounds including but not limited to epidemiology, statistics, applied mathematics, information systems, computer science and machine learning/data mining.

BioSurveillance 2007 aimed to achieve the following objectives: (a) review and examine various real-time data sharing approaches for animal and public health surveillance from both technological and policy perspectives; (b) identify key technical challenges facing syndromic surveillance for both animal and human diseases, and discuss and compare related systems approaches and algorithms; and (c) provide a forum to bring together IDI researchers and practitioners to identify future research opportunities. We are pleased that we received many outstanding contributions from IDI research groups and practitioners from around the world. The one-day program included one invited presentation, 17 long papers, six short papers, and two posters.

BioSurveillance 2007 was jointly hosted by: the University of Arizona; University of California, Davis; Rutgers, The State University of New Jersey; and the University of Washington.

We wish to express our gratitude to all workshop Program Committee members and reviewers, who provided high-quality, valuable and constructive review comments. We would like to thank Ms. Catherine A. Larson and members of the Artificial Intelligence Laboratory and the Intelligent Systems and Decisions Laboratory at the University of Arizona for their excellent support. BioSurveillance 2007 was colocated with the 2007 IEEE International Conference on Intelligence and Security Informatics (ISI 2007). We wish to thank the ISI 2007 organizers and support staff for their cooperation and assistance. We also wish to acknowledge the Springer LNCS editorial and production staff for their professionalism and continued support for ISI and related events. Our sincere gratitude goes to all of the sponsors, especially the U.S. National Science Foundation as the main sponsor

May 2007

Daniel Zeng
Ivan Gotham
Ken Komatsu
Cecil Lynch
Mark Thurmond
David Madigan
Bill Lober
James Kvach
Hsinchun Chen

Organization

Organizing Committee

Conference Co-chairs

Hsinchun Chen	University of Arizona
James Kvach	(formerly with) Armed Forces Medical Intelligence Center, Department of Defense
Bill Lober	University of Washington
David Madigan	Rutgers University
Mark Thurmond	University of California, Davis

Program Co-chairs

Daniel Zeng	University of Arizona
Ivan Gotham	New York State Department of Health
Ken Komatsu	Arizona Department of Health Services
Cecil Lynch	University of California, Davis

Government Liaisons

Larry Brandt	National Science Foundation
Dale Nordenberg	Centers for Disease Control and Prevention
Henry Rolka	Centers for Disease Control and Prevention

Local Arrangements

Catherine A. Larson	University of Arizona

Program Committee

John Allegra	Emergency Medical Associates of New Jersey
Michael Ascher	Lawrence Livermore National Laboratory
John Berrezowski	Alberta Agriculture, Food and Rural Development, Canada
Judith Brillman	University of New Mexico
Ian Brooks	University of Illinois at Urbana-Champaign
David L. Buckeridge	McGill University
Howard Burkom	Johns Hopkins University
Kathleen Carley	Carnegie Mellon University

Wendy Chapman	University of Pittsburgh
Carlos Chavez-Castillo	Arizona State University
Dennis Cochrane	Emergency Medical Associates of New Jersey
Noshir Contractor	University of Illinois at Urbana-Champaign
Kevin Coonan	University of Utah and Health Data Security Inc.
Greg Cooper	University of Pittsburgh
Dan Desmond	SIMI Group
Millicent Eidson	New York State Department of Health
Daniel Ford	IBM Almaden Research
Carol Friedman	Columbia University
Sherrilynne Fuller	University of Washington
Colin Goodall	AT&T Labs
Richard Heffernan	NY City Department of Health and Mental Hygiene
Alan Hevner	University of South Florida
Paul Hu	University of Utah
Xiaohua (Tony) Hu	Drexel University
Jesse Huang	Peking Union Medical College, China
Jeffrey Johnson	San Diego County Health and Human Services Agency
James Kaufmann	IBM Almaden Research
Ken Kleinman	Harvard University
Eileen Koski	Quest Diagnostics
Yang Kuang	Arizona State University
Brian La Forgia	Bergen County Department of Health Services
Colleen Martin	Centers for Disease Control and Prevention
Marcello Pagano	Harvard University
Marc Paladini	New York City Department of Health and Mental Hygiene
Fred Roberts	Rutgers University
Debbie Travers	North Carolina Public Health Information Network
Jason T. L. Wang	New Jersey Institute of Technology
Xiaohui Zhang	Scientific Technologies Corporation

Table of Contents

Multiple Data Streams

Informatics Infrastructure and Case Studies

Part II: Short Papers

Part III: Extended Abstracts

Early Outbreak Detection Using an Automated Data Feed of Test Orders from a Veterinary Diagnostic Laboratory

Loren Shaffer[1], Julie Funk[2], Päivi Rajala-Schultz[1], Garrick Wallstrom[3], Thomas Wittum[1], Michael Wagner[3], and William Saville[1]

[1] Department of Veterinary Preventive Medicine, The Ohio State University, Columbus, Ohio 43210

[2] National Food Safety and Toxicology Center, Michigan State University, East Lansing, Michigan 44824

[3] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania 15219

shaffer.45@osu.edu, funkj@cvm.msu.edu, rajala-schultz.1@osu.edu, garrick@cbmi.pitt.edu, wittum.1@osu.edu, mmw@cbmi.pitt.edu, saville.4@osu.edu

Abstract. Disease surveillance in animals remains inadequate to detect outbreaks resulting from novel pathogens and potential bioweapons. Mostly relying on confirmed diagnoses, another shortcoming of these systems is their ability to detect outbreaks in a timely manner. We investigated the feasibility of using veterinary laboratory test orders in a prospective system to detect outbreaks of disease earlier compared to traditional reporting methods. IDEXX Laboratories, Inc. automatically transferred daily records of laboratory test orders submitted from veterinary providers in Ohio via a secure file transfer protocol. Test products were classified to appropriate syndromic category using their unique identifying number. Counts of each category by county were analyzed to identify unexpected increases using a cumulative sums method. The results indicated that disease events can be detected through the prospective analysis of laboratory test orders and may provide indications of similar disease events in humans before traditional disease reporting.

1 Introduction

Prompt detection of outbreaks might provide for earlier intervention efforts that result in minimizing their overall impact [11], [19], [28], [32]. Some animals are susceptible to infection from many of the same pathogens as humans, sometimes showing signs of disease earlier [1], [12]. Therefore, animals might be used as sentinels and provide for earlier recognition of disease outbreaks that could affect humans. As pet animals share much of the same environment as their human owners, they especially might prove to be valuable outbreak sentinels [2].

Most of the current disease surveillance systems used for animal populations are considered inadequate for detecting outbreaks of emerging disease, potential acts of bioterrorism, or outbreaks resulting from pathogens for which the system was not

D. Zeng et al. (Eds.): BioSurveillance 2007, LNCS 4506, pp. 1–10, 2007.

specifically designed for in a timely manner [16], [20], [21], [24]. Such functionality in animal-based systems has been considered important to our overall bioterrorism and disease outbreak preparedness capabilities [13], [14], [25], [28], [31], [32], [34]. Syndromic surveillance methods utilize population health indicators to warn of potential outbreaks earlier than reports of confirmed diagnoses. Although many sources of data have been investigated for syndromic surveillance in humans, there is paucity in the literature describing similar studies in animals [17].

Laboratories are recognized as important sources of data for disease surveillance in animals as well as humans [9]. Test orders for specimens submitted to commercial medical laboratories have been utilized as one of the data sources for syndromic surveillance in humans [5], [33]. Most of the private veterinary practitioners in the United States also submit specimens to commercial laboratories for diagnostic testing [15]. Through the utilization of data from these commercial laboratories, we might possibly achieve the benefit of the aggregation of many veterinary providers across a wide geographic area. Such centralized aggregation of data may be important in detecting certain outbreaks [11]. The results of a previous investigation conducted by us demonstrated the representation of companion animals in select veterinary diagnostic laboratory (VDL) data and indicated the potential for identifying clusters of cases through analysis of the aggregated orders [27].

Although laboratory analyses are not as frequently a part of the veterinary care of pet animals compared to the medical care of humans [31], we hypothesize that the consistency of test orders over time is such that increases in cases of disease will result in detectable increases in the number of test orders submitted by veterinarians that can be identified using prospective analysis.

2 Methods

We conducted a prospective study of laboratory orders submitted to IDEXX Laboratories, Inc. (Westbrook, Maine) for specimens originating from veterinary clinics in Ohio between September 1, 2006 and November 30, 2006. IDEXX transferred once daily to a server located at the Real-time Outbreak and Disease Surveillance (RODS) Laboratory (University of Pittsburgh, Pennsylvania), via secure file transfer protocol, an automatically generated text file containing records of laboratory orders for specimens received within the previous 24-hour period. Each record included the accession number assigned by IDEXX to the specimen, date and time that IDEXX received the specimen, 5-digit ZIP code of the clinic submitting the specimen, species of animal, and numerical code/s of the laboratory product/s ordered.

2.1 Mapping Laboratory Orders to Syndromic Category

We distributed a list of product descriptions ordered during a 2-week period to ten small and large animal veterinarians asking them to consider the diseases that they might use each product to confirm or rule out during the diagnostic process. The veterinarians then assigned each product to syndromic categories based on the expected presentation of these diseases. Eight categories were considered initially: respiratory, gastrointestinal, neurologic, behavioral, dermal, reproductive, non-specific, and sudden death. Seven of the ten surveyed veterinarians returned the categorized lists (Table 1). The behavioral and sudden death categories were subsequently removed based on zero responses from the surveyed veterinarians for these categories.

In addition to the surveyed veterinarians, two IDEXX laboratorians also reviewed the list of products. Based on their input and advice, five categories were added to further describe many of those products that had been classified into the non-specific category. These additional categories were endocrine, hepatic, infectious, febrile, and renal. Records were mapped to syndromic category based on the identifying number for the laboratory product ordered and appropriately classified as the server received them.

2.2 Statistical Analysis

We used frequency analysis to describe the representation of species groups and distribution of accessions by day of the week. The percentage of the total daily records included in the dataset for each 24-hour period was used to describe availability of records.

2.3 Detection Method

A cumulative sums (CuSum) method was used to analyze category counts, as records were received, for each Ohio County, as determined by the ZIP code. The value of the CuSum was calculated as

$$S_t = \max\{0, S_{t-1} + (X_t - (\mu_t + 0.5\sigma_t))/\sigma_t\}. \quad (1)$$

where X_t was the observed count at time t, μ_t the expected count (baseline), and σ_t the standard deviation of the counts used to determine the baseline. Daily analysis was performed automatically using the count from the current and previous six days for the observed value. A moving 7-day period was chosen to reduce the anticipated day-of-week effect in the data. The expected value was calculated by averaging the weekly counts for the previous 4-week period. We defined alerts as instances when the CuSum value equaled or exceeded five.

An alert period was defined as at least two consecutive days where the CuSum value exceeded the threshold. By using this two-in-a-row rule we were able to somewhat reduce the impact of single-day increases on weekly counts. Using this rule has been shown to increase the robustness of CuSum methods [22]. Alerts were considered for all syndromic categories except non-specific, which was mostly comprised of general screening tests such as blood chemistries. We investigated alerts by identifying the specific laboratory product or products involved and contacting select veterinarians located in the same area as the alert asking about their impressions of disease activity. Veterinarians may or may not have been IDEXX clients.

Table 1. Syndrome category descriptions distributed to veterinarian sample for grouping laboratory products

Example Diseases	Clinical Presentation	Syndrome Category
Glanders, Bordetella, Aspergillosis	• Coughing • Dyspnea • Nasal discharge	Respiratory
Salmonellsis, Clostridia-associated enterocolitis, Campylobacter	• Diarrhea • Vomiting • Colic	Gastrointestinal
Heartwater, plant poisoning, Botulism, Tetanus	• Convulsions • Paralysis • Staggering • Disturbed vision	Neurologic
Poxvirus, allergies, Foot and Mouth Disease	• Abscesses • Rash • Hair loss • Vesiculation	Dermal
Brucellosis, chronic Leptospirosis	• Retained placenta • Abortion • Orchitis	Reproductive
Plague, Tularemia, Anemia, early Leptospirosis	• Lethargy • Malaise • Weakness • Fever without defining associated sign	Non-specific
acute swine erysipelas, Anthrax, Red Water Disease	• Rapid onset of death without apparent cause • Death occurring after brief display of illness	Sudden Death
Rabies, Listeriosis	• Change in appetite without defining associated signs • Unexplained aggression • Disorientation	Behavioral

3 Results

3.1 Data Transfer

During the pilot, the daily transfer of data from IDEXX Laboratories was interrupted twice. The first interruption began on September 7 and continued through September 28. This interruption in data transfer occurred because the workflow involved in the transfer had been unscheduled and the job was mistakenly shut down. The second interruption occurred October 6 through October 9 for unknown reasons. The interruptions affected the transfer of 10,847 (22.6%) records. IDEXX forwarded records that were created during these times of interruption once the data feed was re-established providing for a complete time-series.

The pilot system relied upon transfer of data from IDEXX that was being queued in a test environment. The reliability of this environment was knowingly not as stable as a production environment would be. The interruptions experienced during this pilot would not be expected in a more stable production platform.

3.2 Descriptive Statistics

During the study period, IDEXX transferred records for 48,086 accessions. Specimens originated throughout Ohio and appeared to correlate with the population of each area. Accessions displayed an obvious and predictable day-of-week effect (Figure 1) with Sundays, Mondays, and days following holidays representing days with the lowest volume. Species represented by the accessions included canine (70.1%), feline (25.6%), and equine (2.1%).

An important consideration for the designers of any syndromic surveillance system is the timely availability of data [6], [30]. Earlier detection being the overall goal, the system must receive records, with the appropriate information for analysis, within a period that provides for improved timeliness of detection compared to traditional reporting systems. Excluding the accessions that occurred during the interruption periods (n=10,847), on average, 95% of daily records were received with the next day's dataset (Figure 2). Almost all (99.4%) records were received by the fourth 24-hour period.

3.3 Aberration Detection

The system identified nine alert periods during the study period using the CuSum detection method as previously described. All of the alerts involved canines and/or felines. The number of accessions generating the alerts ranged from eight to 43. No cause could be determined for three of the nine (33.3%) alert periods and two (22.2%) were possibly related to breeding operations that existed in the area (e.g. screening of litters for pathogens). Two (22.2%) others were potentially the result of provider interest. One veterinary practice located in an area where a gastrointestinal alert occurred reported being especially interested in educating clients about the risks from parasite ova. Another provider in an area where an endocrine alert occurred had recently been ordering an increased number of thyroid tests that were unrelated to increases in clinical disease. The remaining two (22.2%) alert periods were linked to verified disease activity in the pet population during the time of the alert.

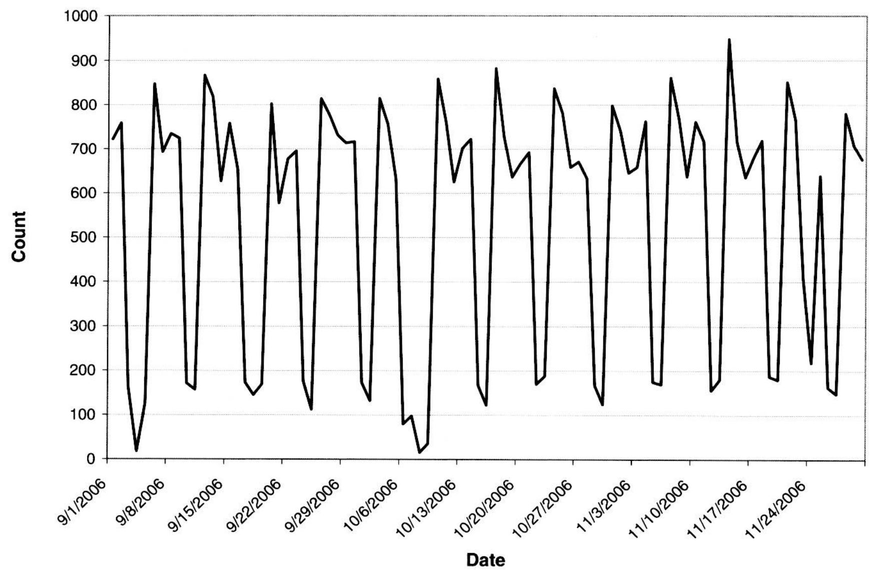

Fig. 1. Counts of specimens received by IDEXX from veterinary clinics in Ohio from September 1 through November 30, 2006

3.4 Case Reviews

On September 11, 2006, the system generated an alert in Preble County located in western Ohio. Cases (20 cats and 2 dogs) were equally distributed between two ZIP codes. Follow-up with area veterinarians confirmed that many small animal practices were treating an increased number of animals that lived or spent a significant amount time out-of-doors for unspecified gastrointestinal distress. Following consultation with the Ohio Department of Natural Resources, veterinarians suspected that the cases may have resulted from corona virus infections acquired from rodents (Melissa Howell, Preble County Health Department, personal communication). An increased number of rodents were noted in the area, coinciding with the harvesting of local grain fields. Veterinarians speculated that pets may have captured and consumed some of the rodents, resulting in the self-limiting intestinal condition. Although health authorities received no reports of human cases, the Real-time Outbreak and Disease Surveillance System used by the Ohio Department of Health indicated significant increases in both gastrointestinal-related chief complaints of emergency department patients and sales of anti-diarrheal medication in these areas during this time (L.S., unpublished data, 2006).

The pilot system generated a gastrointestinal alert for Lake County in northeastern Ohio on September 4, 2006. This alert included test orders for specimens originating from ten cats and three dogs submitted by clinics in two ZIP code areas. A local veterinarian from this county telephoned the State Public Health Veterinarian on

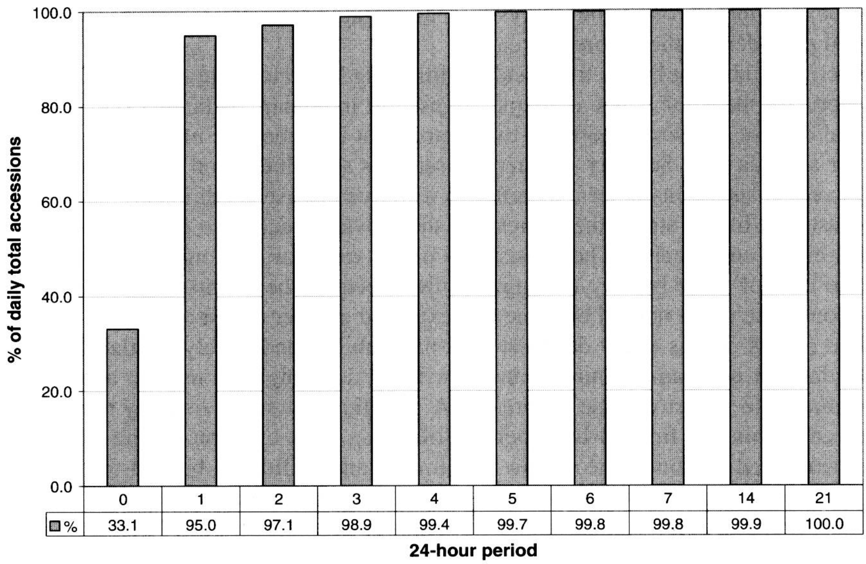

Fig. 2. Delay in receipt of daily records from IDEXX during prospective pilot

September 26, 2006 to inquire about a number of clients that had brought their pets presenting with vomiting and diarrhea (Nancy Niehaus, Lake County Health Department, personal communication). These clients had shared with the local veterinarian that they also were experiencing diarrhea. The Lake County Health Department reported on October 4, 2006 that they were investigating "a cluster of diarrheal illness in humans and their associated pet dogs."

4 Discussion

The primary purpose of this study was to explore the feasibility of using prediagnostic data from a VDL in a prospective manner to detect unexpected increases in the number of disease cases that might indicate an outbreak. We evaluated the feasibility by first determining the stability of electronic records and the success of automatically transferring them from the VDL for analysis, measured in terms of the percentage of complete records received in a timely manner. We then considered the representation of the records both by species of animal and geographic distribution. Finally, we investigated the alerts generated by the pilot system to validate if they might be associated with increases of disease.

While no single data source provides the capability to detect all outbreaks that may occur, veterinary providers may be desirable sources to include in surveillance activities for bettering our capabilities of detecting those outbreaks that result from emerging pathogens and potential bioweapon agents [10], [12], [16], [25], [32]. The change in the number of laboratory orders submitted by veterinary providers may be a valuable proxy to measure the number of individual cases they are treating. An increase in the number of these individual cases may result from an outbreak, detection of which