

Syndromic Surveillance System Using Emergency Department Patients Medical Information in Taiwan : The Automation Informatics System of Data Collection and Processing

Chan-Hsien Chiu^{a,i}, Tsung-Shu Wu^b, Jiunn-Shyan Julian Wu^{a,b}, Shiou-Wen Lu^a, Fuh-Yuan Shih^d, An-Jing Lung^e, Muh-Yung Yen^f, Kevin Chi-Ming Chang^a, Chao Hsiung^c, Jamii Wu^g, Chien-Hui Lu^h, Jr-How Chou^a, and Chwan-Chuen King^b

^aCenter for Disease Control Taiwan R.O.C., ^bInstitute of Epidemiology, National Taiwan University, ^cNational Health Institute of Research, ^dNational Taiwan University Hospital, ^eWan fang Municipal Hospital, Taipei City, ^fRen ai Municipal Hospital, Taipei City, ^gTatung System Technologies Inc., ^hINQGEN Technology Co., Ltd., ⁱGraduate Institute of Medical Informatics Taipei Medical University.

E-mail Address: zschou@seed.net.tw

Abstract

In order to establish an early emerging infectious disease and bioterrorism detection system, the emergency department (ED) visitors' medical information based syndromic surveillance was built up with automatically data collection and transmission in Taiwan, ROC. There were 189 hospitals took part in the syndromic surveillance system data supply since March, 2004. The remote program was used to send data to the center for diseases control (CDC) and connect real time with the hyper text transfer protocol over secure sockets layer (https). The data cleansing algorithm was set up by epidemiologists and clinical physicians using logical rules. Until August 12, there were 972,182 ED visit data storied in the data base with the error rate in 2.74%. Continuous data monitoring and system implementation research work are in progress.

Keywords: syndromic surveillance, emerging infectious disease, emergency department, Taiwan

Introduction

After global alert on detecting emerging infectious diseases (EID) in recent years and challenges of SARS epidemic in 2003, government health officials at the Center for Disease Control (CDC) in Taiwan realized the urgency to detect EID or bioterrorism, and had willingness to build up an early detecting automatic alert

system. Automation Syndromic Surveillance Planning Committee has been operated with team efforts since July, 2003.

Emergency department (ED)-based syndromic surveillance system has involved 189 hospitals' emergency department. The important data are hourly transferred to the CDC in Taiwan, in which the collected parameters information are downloaded and analyzed hourly. Signal-detection models had been built up, according to two Taipei municipal hospitals' historical data for detecting clusters or unusual cases with public health significance. The daily operation and data analysis are in progress every day.

Material and Methods

Data Collection and Transmission

At the beginning, two Taipei City Municipal Hospitals were our sites and then extended to 189 hospitals with emergency department that handle patients with emerging infectious diseases located all around in Taiwan. They were recruited to take part in the syndromic surveillance system by daily web-delivering ED Triage data to the CDC in Taiwan. The Triage station generates most information for syndromic surveillance, including admission time, date of birth, gender,

home-zip-code, body temperature, triage categories and chief complaints of all patients visited emergency department. In addition, the international classification of diseases, 9th revision, clinical modification (ICD-9-CM) immediately created by physicians or nurses based on the initial clinical judgment were also collected. Both health level seven protocols (HL-7) and the integrated electronic data files not only containing patients' clinical information but also hospital identification numbers are formatted to an extensible Markup Language (XML) file. The XML files are allowed to hourly automatic transfer information from previous enrolled hospitals (see table 1) to the CDC Taiwan. The system also accepts some hospital send the data with health level seven protocols (HL7) in XML files.

Table 1. Variables Collected from Hospital Emergency Departments*

*Only patient's medical information variables are presented.

<u>Field</u>	<u>XML Tag</u>	<u>HL7</u>
Patient Gender	Sex	PID.8
Patient's Date of Birth (Age)	Birthday	PID.7.2
Patient Address ZIP	ZipCode	PID.11.5
Hospital identification number	Hospital	PV1.3.4
Admission Date/Time	AdmitTime	PV1.44
ICD-9-CM	ICD9_1~ICD9_4	DG1.3
Chief complaints of patient	Subject	DG1.4
Major diagnostic category	Access Category	DG1.7
Body temperature	Temperature	OBX

All the sentinel hospitals have been recruited in this Automated Syndromic Surveillance for detecting emerging infectious diseases have an independent computer which installed with a MySQL server to save the data and a remote connecting program to respond from the collected data. Data files generated from different systems in 189 hospitals, including triage classification system, hospital information system (HIS) and clinical information system (CIS) were all hourly

transferred to the Microsoft SQL Sever 2000 with hyper text transfer protocol over secure sockets layer (HTTPS) to the Center for Diseases Control Taiwan. All communicating histories are recorded in a log file in the SQL server in Taiwan and daily monitored by informatics personnel. A written program within the system was set up to check each communication of data transfer and automatically sends a message to hospital to announce whether the transfer is successful or not.

Data Processing

Three different tables were designed to process the data. Entire information is initially transferred into the first table and generated an additional column with a serial number for each case. The system picks up the transferred data from the first table every six minutes and processes it into the second table. Data cleansing work is held in the second table by the system algorithm written by SQL commands. Cleaned data are then transferred to the third table to do further analysis (see fig. 1).

Data Cleansing and Standardization

Although only few variables are collected in each hospital, inconsistent data remain as major difficulty in doing analysis. Specific cleansing criteria, including logics, removing repeated data of the same person, were set for different variables. For the data deletion, only those cases whose chief complaint fields containing "test" and without ICD field or both fields were null were deleted. Hospital ID, date of birth, admission time, gender and home zip code are the key fields for double-check to remove those duplicated and repeated ones. The system has the capability for continuous checking the subject of hospital ID and time format of all timing fields then moved the error data to an "error table" and storage. Wrong format of ICD-9-CM data also moved to the "error database". All the process of deletion and removing procedure are recorded in the log file for monitoring. Additionally, fewer numbers of ED visits in those hospitals on certain specific days were

also reviewed for further improvements.

Results

The program on “Emerging Infectious Disease Surveillance Hospitals” started database linkage work from March 10, 2004 to May 24, 2004. Up to now, there are 189 hospitals linking to CDC syndromic surveillance system and transferring the daily data in a required format. The SQL server received 6,000 ED patients’ information in average per day. Overall, there are 972,182 ED visit data stored in CDC database by August 12, 2004 and the error rate was 2.74%. Two computer engineers check the log files and informed the hospitals on weekdays by telephone to correct the errors or to fill out the missing data. The daily counts of data transmission will be investigated and analyzed. The results will be present at this meeting.

Only few hospitals could not send ICD-9-CM data (5.04%), but almost half of them failed in sending chief complaints (47.4%). For example, there were 239,617 cleaned data received in July, when the system became more stabilized, and found that about 7.1% ICD information were lacking null and 54.82% of chief complaint data were missing. Since the quality of synchronous communication between hospitals and CDC in Taiwan was not stable, only twenty six percents hospitals sent data every day from the beginning up to Aug. 12, 2004 because certain hospitals sent the chief complaints in Chinese, increasing the difficulty for further analysis. The format of ICD-9-CM, which has 5 digits officially, also showed recognition problems by presenting 3 or 4 digits in ICD number creating confusion for the data analysis. Other errors, for instance, include duplicated data and missing data are automatically cleaned every 5 minutes by the system algorithm.

Discussions

Syndromic surveillance has been applied to many respiratory infections such as flu-like illness and measles-like illness. Continuous data monitoring system

must be operated immediately after the system sends the inspection messages to surveillance hospitals. At present, engineers have informed hospitals about data errors and the automatic error feedback system has been built up. Future efforts need more computer-science professionals and personnel of medical informatics at CDC in Taiwan to establish a system with the standard operating procedures for database maintenance and to provide more continuous on-job trainings.

Various types of chief complaints, ICD-9-CM coding, and home zip code have made the data analysis to become more complicate. Systematic approaches, such as using standardized format of selected syndrome and variables, to improve data quality and stability of data transmission need to be established and continuously evaluated. More research areas will be open up, including the Chinese styles of chief complaints, comparison between chief complaints and ICD-9 codes, different combinations of symptoms/signs and data linkage to establish the most sensitive, specific and timely syndromic surveillance system in Taiwan.

References

1. Bravata, D. M., K. M. McDonald, W. M. Smith, C. Rydzak, H. Szeto, D. L. Buckeridge, C. Haberland, and D. K. Owens. 2004. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. *Ann Intern Med* 140:910-22.
2. Buehler, J. W., R. L. Berkelman, D. M. Hartley, and C. J. Peters. 2003. Syndromic surveillance and bioterrorism-related epidemics. *Emerg Infect Dis* 9:1197-204.
3. Day, F. C., D. L. Schriger, and M. La. 2004. 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* 43:401-9.
4. Gesteland, P. H., R. M. Gardner, F. C. Tsui, J. U. Espino, R. T. Rolfs, B. C. James, W. W. Chapman, A. W. Moore, and M. M. Wagner. 2003. Automated syndromic surveillance for the 2002 Winter

- Olympics. *J Am Med Inform Assoc* 10:547-54.
5. Heffernan, R. 2004. Syndromic surveillance in public health practice, new york city. *Emerg Infect Dis* 10:858-64.
 6. Irvin, C. B., P. P. Nouhan, and K. Rice. 2003. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. *Ann Emerg Med* 41:447-52.
 7. Lazarus, R., K. Kleinman, I. Dashevsky, C. Adams, P. Kludt, A. DeMaria, Jr., and R. Platt. 2002. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerg Infect Dis* 8:753-60.
 8. Lewis, M. D., J. A. Pavlin, J. L. Mansfield, S. O'Brien, L. G. Boomsma, Y. Elbert, and P. W. Kelley. 2002. Disease outbreak detection system using syndromic data in the greater Washington DC area. *Am J Prev Med* 23:180-6.
 9. Lober, W. B., L. J. Trigg, B. T. Karras, D. Bliss, J. Ciliberti, L. Stewart, and J. S. Duchin. 2003. Syndromic surveillance using automated collection of computerized discharge diagnoses. *J Urban Health* 80:i97-106.
 10. Lombardo, J., H. Burkom, E. Elbert, S. Magruder, S. H. Lewis, W. Loschen, J. Sari, C. Sniegowski, R. Wojcik, and J. Pavlin. 2003. A systems overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). *J Urban Health* 80:i32-42.
 11. Moore, K. 2004. Real-time syndrome surveillance in Ontario, Canada: the potential use of emergency departments and Telehealth. *Eur J Emerg Med* 11:3-11.
 12. Platt, R., C. Bocchino, B. Caldwell, R. Harmon, K. Kleinman, R. Lazarus, A. F. Nelson, J. D. Nordin, and D. P. Ritzwoller. 2003. Syndromic surveillance using minimum transfer of identifiable data: the example of the National Bioterrorism Syndromic Surveillance Demonstration Program. *J Urban Health* 80:i25-31.
 13. Suyama, J., M. Sztajnkrzyer, C. Lindsell, E. J. Otten, J. M. Daniels, and A. B. Kressel. 2003. Surveillance of infectious disease occurrences in the community: an analysis of symptom presentation in the emergency department. *Acad Emerg Med* 10:753-63.
 14. Tsui, F. C., J. U. Espino, V. M. Dato, P. H. Gesteland, J. Hutman, and M. M. Wagner. 2003. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc* 10:399-408.

Acknowledgements

This work is supported partly by CDC Grant No. DOH92-DC-SA03. We would also like to thank Ms. HueiChe at NHRI.

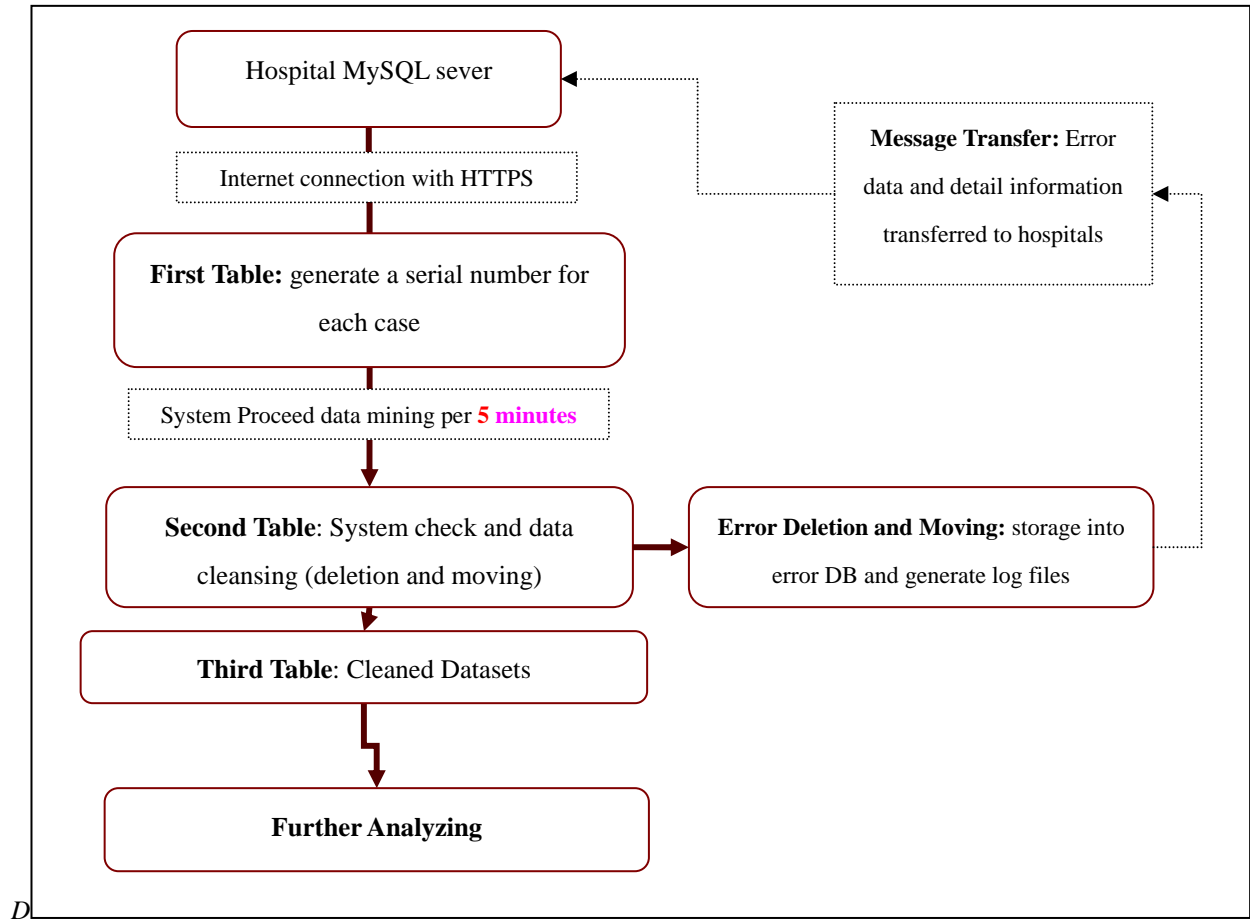


Figure 1. Flow-chart of Syndromic Surveillance System Data Processing

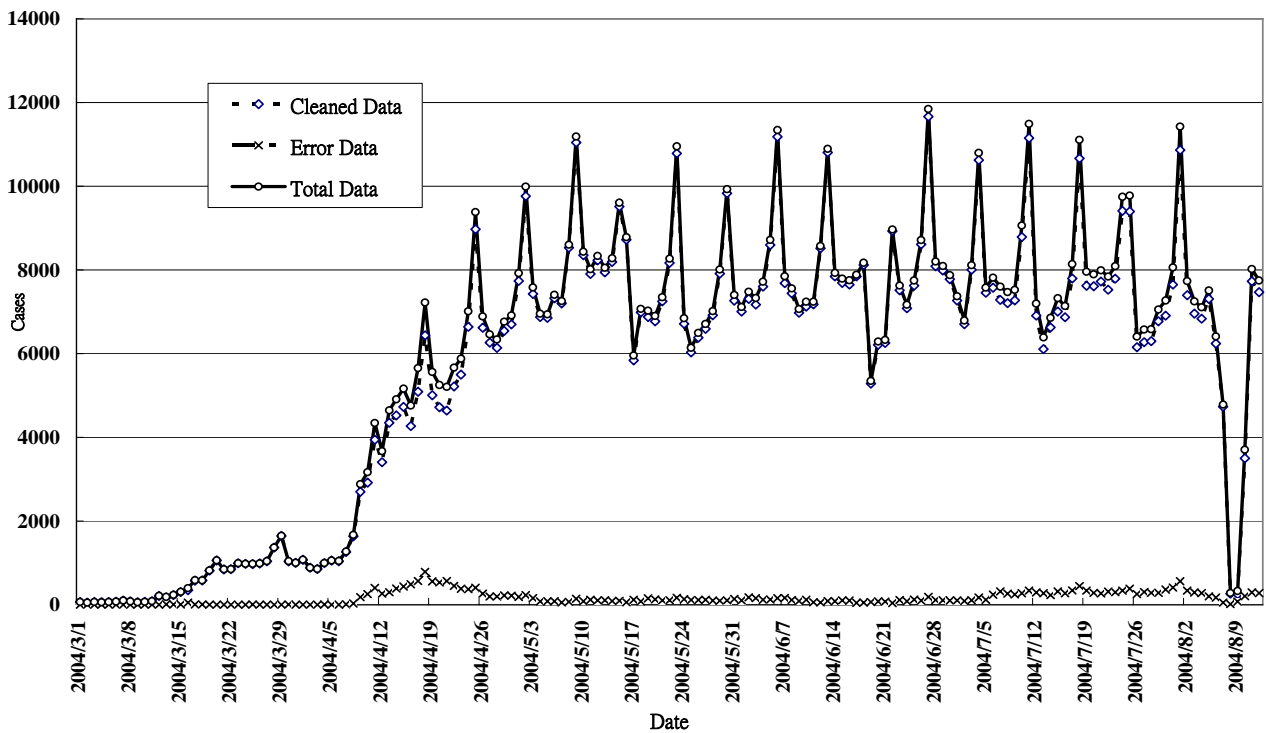


Figure 2. Daily Data of Syndromic Surveillance System*

*The duplication data were not shown in the figure.

