

# A nationwide web-based automated system for outbreak early detection and rapid response in China

Weizhong Yang,<sup>a</sup> Zhongjie Li,<sup>a</sup> Yajia Lan,<sup>b</sup> Jinfeng Wang,<sup>c</sup> Jiaqi Ma,<sup>a</sup> Lianmei Jin,<sup>a</sup> Qiao Sun,<sup>d</sup> Wei Lv,<sup>e</sup> Shengjie Lai,<sup>a</sup> Yilan Liao<sup>c</sup> and Wenbiao Hu<sup>f</sup>

Correspondence to Weizhong Yang (e-mail: yangwz@chinacdc.cn) and Zhongjie Li (e-mail: lizj@chinacdc.cn)

Timely reporting, effective analyses and rapid distribution of surveillance data can assist in detecting the aberration of disease occurrence and further facilitate a timely response. In China, a new nationwide web-based automated system for outbreak detection and rapid response was developed in 2008. The China Infectious Disease Automated-alert and Response System (CIDARS) was developed by the Chinese Center for Disease Control and Prevention based on the surveillance data from the existing electronic National Notifiable Infectious Diseases Reporting Information System (NIDRIS) started in 2004. NIDRIS greatly improved the timeliness and completeness of data reporting with real-time reporting information via the Internet. CIDARS further facilitates the data analysis, aberration detection, signal dissemination, signal response and information communication needed by public health departments across the country. In CIDARS, three aberration detection methods are used to detect the unusual occurrence of 28 notifiable infectious diseases at the county level and transmit information either in real time or on a daily basis. The Internet, computers and mobile phones are used to accomplish rapid signal generation and dissemination, timely reporting and reviewing of the signal response results. CIDARS has been used nationwide since 2008; all Centers for Disease Control and Prevention (CDC) in China at the county, prefecture, provincial and national levels are involved in the system. It assists with early outbreak detection at the local level and prompts reporting of unusual disease occurrences or potential outbreaks to CDCs throughout the country.

**A** aberration of disease occurrence means the occurrence of cases is in excess of normal expectancy in a certain region. Early detection of the aberration of infectious disease occurrence and rapid control actions are prerequisites for preventing the spread of outbreaks and reducing the morbidity and death caused by diseases.

After China had an outbreak of severe acute respiratory syndrome (SARS) in 2003, the government took efforts to enhance the capacity of infectious disease surveillance and successfully built the innovative web-based Nationwide Notifiable Infectious Diseases Reporting Information System (NIDRIS) in 2004. It enabled all the health care institutes across the country to report in real time individual case information of notifiable infectious diseases by Internet.<sup>1</sup> This system shortened the interval between case diagnosis and case reporting to within one day on average.<sup>2</sup>

However, enhancing the timeliness of data reporting is only the first step for outbreak monitoring and response. Effectively analysing and interpreting the large volume of reported data and rapidly distributing the results to the responders are also key components. Therefore, a tool was conceived to conduct automated and timely analyses and detection of aberration of infectious disease occurrence to facilitate a rapid response to outbreaks and to effectively communicate the outbreak information among Centers for Disease Control and Prevention (CDCs) in China. The universal availability of modern communication tools (such as computers, the Internet and mobile phones) in China also helped this idea to be realized.

In 2005, the China CDC, cooperating with the World Health Organization, initiated a national project to develop the China Infectious Disease Automated-alert and Response System (CIDARS). The system

<sup>a</sup> Chinese Center for Disease Control and Prevention (China CDC), Beijing, 100050, China.

<sup>b</sup> West China School of Public Health, Sichuan University, Chengdu, China.

<sup>c</sup> Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.

<sup>d</sup> Shanghai Pudong New Area Center for Disease Control and Prevention, Shanghai, China.

<sup>e</sup> Guangxi Center for Disease Control and Prevention, Nanning, China.

<sup>f</sup> School of Population Health, The University of Queensland, Brisbane, Australia.

Submission date: 31 October 2010; Publication date: 8 March 2011

doi: 10.5365/wpsar.2010.1.1.009

Table 1. Type of aberration detection method for different infectious diseases

Aberration detection methods	List of infectious diseases
1. Fixed-threshold detection method (FDM)	Type 1 diseases: plague, cholera, SARS, human avian influenza, poliomyelitis, pulmonary anthrax, diphtheria, filariasis, unexplained pneumonia
2. Temporal detection method (TDM)	Type 2 diseases: hepatitis A, hepatitis C, hepatitis E, measles, epidemic haemorrhagic fever, epidemic encephalitis B, dengue fever, bacillary and amoebic dysentery, typhoid and paratyphoid, epidemic cerebrospinal meningitis, scarlet fever, leptospirosis, malaria, influenza, epidemic mumps, rubella, acute haemorrhagic conjunctivitis, epidemic and endemic typhus, infectious diarrhoea (excluding cholera, dysentery, typhoid and paratyphoid)
3. Spatial detection method (SDM)	Type 2 diseases; same as TDM

was successfully implemented and began to operate nationwide in 2008. This paper introduces the design and development of CIDARS and reports the preliminary evaluation of the system's performance.

## OVERVIEW OF NATIONAL NOTIFIABLE INFECTIOUS DISEASE REPORTING SYSTEM

According to the Law of Prevention and Control of Infectious Disease in China, 39 infectious diseases are regulated as notifiable diseases. All cases of notifiable infectious diseases are diagnosed by clinicians using the uniform case definition issued by the Chinese Ministry of Health. A standard report form is used to collect patient's information, including name, gender, age, identification number, residential address, date of onset, date of diagnosis and diagnosis results. Since the implementation of NIDRIS in 2004, all notifiable infectious disease cases have been reported in real time directly from hospitals to the national infectious diseases surveillance database, located at the China CDC, Beijing, China.<sup>1</sup> NIDRIS covers all health care institutions across the country, including general hospitals, specialized hospitals, township and village clinics and private clinics.

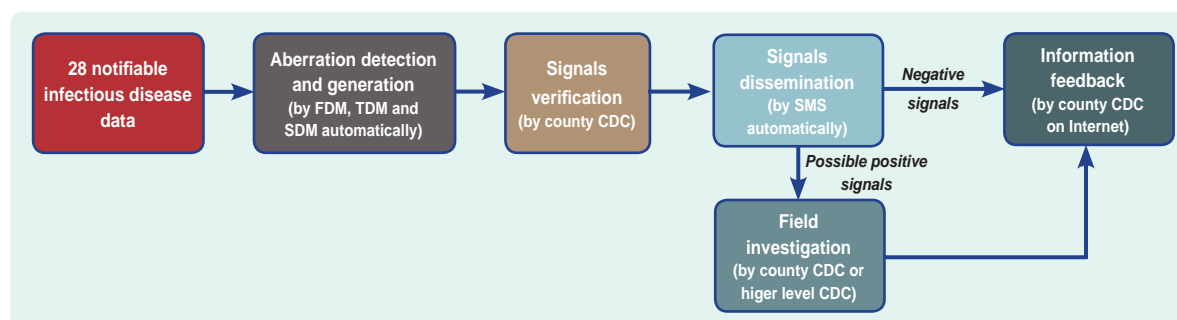
According to the annual report on disease surveillance in 2008, approximately 67 000 health institutions reported case information to NIDRIS and about 5 million infectious diseases cases were reported annually.<sup>2</sup>

## DESIGN AND IMPLEMENTATION OF CIDARS

### System description

CIDARS was developed based on the existing data from NIDRIS on 28 diseases (Table 1) that are outbreak-prone and require prompt action are included in the system. By integrating multiple aberration detection methods, CIDARS conducts real-time and daily analysis on the data and sends the abnormal signals to CDCs at the county level by short message service (SMS) using mobile phones. CDCs at national, provincial and city levels can also monitor the response process of each signal and provide timely technical guidance and support, if necessary. The system consists of four interconnected components: aberration detection, signal generation, signal dissemination and signal response information feedback (Figure 1). The unifying operational protocol of CIDARS on the workflow of these components was developed for the system users.

Figure 1. Flow diagram of China Infectious Diseases Automated-alert and Response System (CIDARS)



FDM - fixed-value detection method; TDM - temporal detection method; SDM - spatial detection method; SMS - short message service, possible positive signals - denoting a possible outbreak judged by county CDC staff after conducting signal verification; negative signals - not denoting a possible outbreak judged by county CDC staff after conducting signal verification.

### Aberration detection

The three aberration detection methods were developed and applied in CIDARS in two stages. At the first stage, two aberration detection methods, the fixed-threshold detection method (FDM) and the temporal detection method (TDM), were developed in 2006. One year later, the third method, the spatial detection method (SDM), was added and integrated with the first two methods. The 28 diseases were classified into two types according to severity, incidence rate and importance. They were analysed with one of the three different aberration detection methods (Table 1). The three methods are briefly described as follows:

#### (1) Fixed-threshold detection method

Type 1 diseases, includes nine infectious diseases characterized with higher severity but lower incidence, and are analysed using FDM with the threshold of one fixed value.<sup>3</sup>

#### (2) Temporal detection method

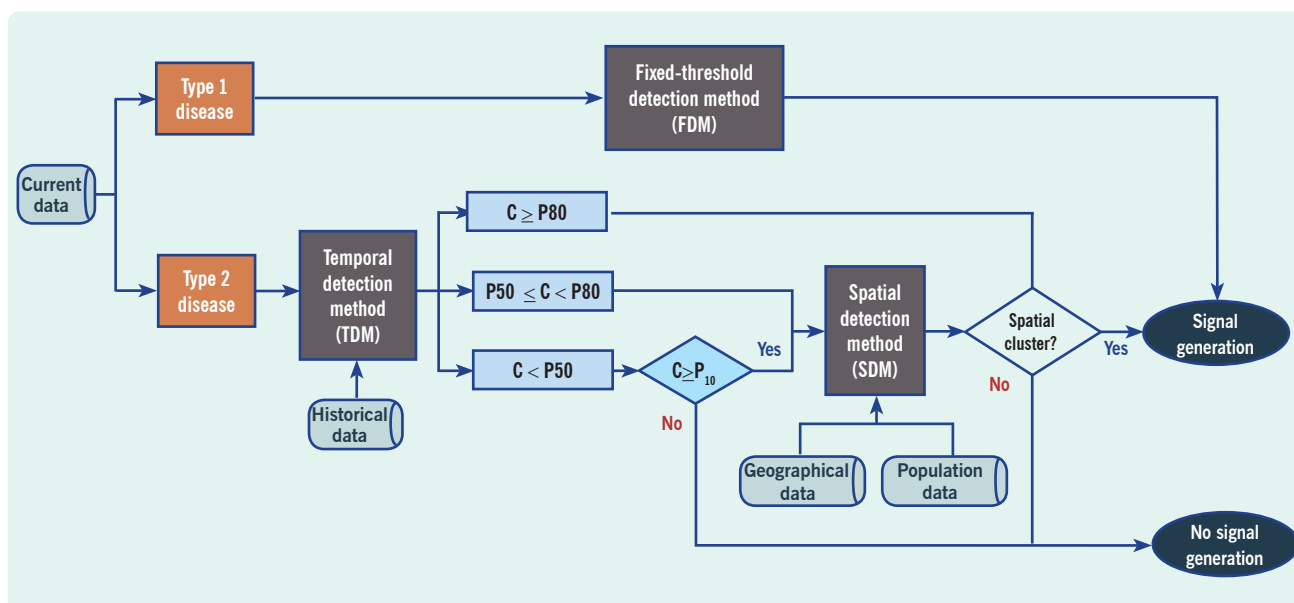
For type 2 diseases (more common infectious diseases), the moving percentile method is used to detect aberration of disease occurrence by comparing the reported cases in the current observation period to that of the corresponding historical period at the county level. To account for the day-of-week effect and the stability

of data, the most recent seven-day period is used as the current observation period and the previous three years as the historical period.<sup>4,5</sup> The number of cases in the current observation period is the sum of reported cases within the recent seven days. The corresponding historical period included, for each of the previous three years, the same seven-day period, the two preceding seven-day periods and the two following seven-day periods that resulted in 15 historical seven-day data blocks covering 105 days. We set the percentile of the 15 blocks of historical data as the indicator of potential aberration. The current observation period and historical data block are dynamically moved forward day by day.

#### (3) Spatial detection method

One SDM, the SaTScan method<sup>4</sup>, is used to search for spatial clusters of the incidence of type 2 diseases. SaTScan is a freely available spatial, temporal and space-time data analysis platform.<sup>6,7</sup> This model is applied to the data at the township level. The population data required by SaTScan were obtained from the Chinese Bureau of Statistics, and the geographic data were from the Chinese Institute of Geographic Sciences and Natural Resources Research. When the incidence of disease in certain geographic areas (one town or more than one town) is significantly higher than that of other areas in the county, this area is categorized as spatial clustering.

Figure 2. The aberration detection and signal generation technology road map of CIDARS



C: the sum of the reported cases during the current seven-day period; P: the percentile of historical data. Type 1 diseases include plague, cholera, SARS, human avian influenza, poliomyelitis, pulmonary anthrax, diphtheria, filariasis, and unexplained pneumonia. Type 2 diseases include hepatitis A, hepatitis C, hepatitis E, measles, epidemic haemorrhagic fever, epidemic encephalitis B, dengue fever, bacillary and amoebic dysentery, typhoid and paratyphoid, epidemic cerebrospinal meningitis, scarlet fever, leptospirosis, malaria, influenza, epidemic mumps, rubella, acute haemorrhagic conjunctivitis, epidemic and endemic typhus, infectious diarrhoea (excluding cholera, dysentery, typhoid and paratyphoid).

## Signal generation

Whether or not to generate a signal depends on the calculated results of these three aberration detection methods. The rules of signal generation are (Figure 2):

1. For type 1 diseases, the signal is immediately generated once one case is reported to NIDRIS.

2. For type 2 diseases, the generation of a signal is decided by the calculated results of both TDM and SDM, both of which are operated with certain logic sequence (Figure 2) and are conducted once a day at 24:00. The signal is finally generated when any one of the following requirements are met after the calculation process of TDM and SDM where  $C$  is the sum of cases during the current seven-day period and  $P$  is the percentile of the historical data:

- TDM:  $C \geq P80$ ;
- TDM:  $C \geq P50$  and  $C < P80$ , and SDM showing spatial clustering;
- TDM:  $C < P50$  and  $C \geq P10$ , and SDM showing spatial clustering.

## Signal dissemination

At least two epidemiologists in every CDC are designated to automatically receive the signals on their mobile phones by the SMS system located at the China CDC, Beijing, China. For type 1 diseases, the signal is distributed in real time, and for type 2 diseases the signal is released at 08:00 once a day.

## Signal response and information feedback

The signal response process includes two steps: signal verification and field investigation. The initial verification is conducted by epidemiologists in local CDCs by reviewing the reported cases in NIDRIS, completing a general assessment of information from other surveillance sources or directly contacting the reporting agencies. If the signal denoted one suspected outbreak after the initial verification, this signal would be determined as a possible positive signal, otherwise this signal would be determined as a negative signal. It is estimated that the verification of one negative signal may take about 10 minutes for one professional epidemiologist. Once a possible positive signal is determined, field investigation is conducted to confirm whether an outbreak is occurring.

The information on the signal verification and field investigation is fed back into CIDARS by local epidemiologists, so that the epidemiologists at the CDCs can actively monitor the outcome of signal verification and the involvement of the outbreak.

## ROLES OF SYSTEM USERS

China CDC took responsibility for the system design, development and maintenance as well as monitoring severe outbreaks. CDCs at provincial and prefecture levels took charge of the system's user management within their administrative areas, daily reviewing and following up on the signals response process. All CDCs at the county level are responsible for receiving and responding to the signals, and promptly feeding the response results into CIDARS.

## PRELIMINARY RESULTS

During the period 1 July 2008 to 30 June 2010, 221 counties from 10 provinces were selected to conduct the initial evaluation on CIDARS. For type 1 diseases, 308 signals were generated, involving nine diseases, 69 (22.4%) of which were identified as possible positive signals that triggered further field investigation, with nine cholera outbreaks confirmed. For type 2 diseases, 100 629 signals were triggered, including 19 infectious diseases, with about 4.4 signals per county per week on average. Among these, 1371 signals (1.36%) were verified as possible positive signals, and 167 outbreaks were finally confirmed by conducting field investigation. Generally, the percentage of possible positive signals to all signals of the respiratory diseases group (2.78%) was higher than that of zoonoses and vectorborne diseases group (1.95%) and food and waterborne diseases group (0.24%).

## DISCUSSION

The development and application of CIDARS was one significant activity to enhance the capacity of early outbreak detection and rapid response in China. It has been integrated into the routine work of outbreak monitoring and response for all of China's CDCs.

Compared to the manual analysis of surveillance data and reporting unusual information level by level, as done in the past in China, CIDARS greatly shortens the frequency of surveillance data analysis and that of outbreak communication among different CDCs. It also

lessens the workload of data collating and analysing for epidemiologists to a great extent. The web-based system was developed and is maintained by the national CDC. The local CDCs only need to use their existing mobile phones, a computer and the Internet to receive and review the signals and transmit information. No new equipment was needed, which reduced the cost for local users.

Many outbreak early warning systems disseminate the signal by e-mail which may make it hard to confirm that the information is received successfully and in a timely manner.<sup>3,8,9</sup> CIDARS uses an SMS platform and designates the specific mobile phones to receive the signal by short text message; the system automatically gets a confirming message which ensures accurate and timely dissemination. As opposed to some systems using only one-sided generation and distribution of the information, CIDARS has a good feedback function for processing signal responses and results to facilitate outbreak response cooperation and assistance, if necessary.

From the initial evaluation of the system, we found that CIDARS can quickly generate abnormal signals and effectively assist in the early detection and confirmation of some disease outbreaks, including both type 1 and type 2 diseases. However, the percentage of possible positive signals of all signals in CIDARS seems to be a little low. As we know, a low percentage of positive signals is a common deficiency facing many similar outbreak early warning systems.<sup>3,10–12</sup> The percentage of possible positive signals varied among the respiratory, zoonotic and vectorborne, and food and waterborne disease groups, which demonstrated that different algorithms need to be considered based on the epidemiological characteristics of the disease.

Although CIDARS is a powerful and sophisticated system, one challenge is to maintain normal operations of the system. Advanced computers with high-powered data calculation ability, the stability of Internet access as well as a professional system maintenance team are necessary. There are currently more than 6000 system users which raises the challenge of user management and training as staff turnover occurs.

One limitation of CIDARS is that it is hard to detect the outbreaks before the cases are diagnosed and reported by clinicians because the system is based on the

notifiable infectious disease surveillance data. Therefore, CIDARS sometimes may be less timely and sensitive than some other outbreak detection systems using data on pre-diagnosis of cases in hospitals, media reports or school absenteeism. In addition, many negative signals are currently generated by CIDARS, causing unnecessary signal response for local staff.

Some improvements to CIDARS should be considered in the future. More flexible and reasonable algorithms and parameters for aberration detection should be developed and calibrated for the different characteristics of particular diseases and various needs of different areas in order to improve the performance of outbreak detection. New diseases could be added into the system by local users to address priorities in a particular jurisdiction. Finally, more systematic evaluations of the performance of the system should be conducted, especially on the feedback from users.

### *Competing interests*

None declared.

### *Funding*

This study was supported by grants from the Ministry of Science and Technology of China (2002DIA40020, 2003DIA6N009, 2006BAK01A13, 2008BAI56B02, 2009ZX10004-201), and the China-WHO regular budget cooperation project (WPCHN0801617, WPCHN1002405).

### *Acknowledgements*

We thank Dr Chin-kei Lee (WHO Country Office in China) and Dr Archie Clements (University of Queensland, Australia) for giving comments and suggestions for improving this manuscript.

### *References:*

1. Wang L et al. Emergence and control of infectious diseases in China. *Lancet*, 2008, 372:1598–1605. doi:10.1016/S0140-6736(08)61365-3 pmid:18930534
2. *Annual Report on Morbidity and Mortality of Notifiable Infectious Disease in China in 2008*. Beijing, Chinese Center for Disease Control and Prevention, 2009.
3. Widdowson MA et al. Automated, laboratory-based system using the Internet for disease outbreak detection, the Netherlands. *Emerging Infectious Diseases*, 2003, 9:1046–1052. doi:10.3201/eid0909.020450 pmid:14519238
4. Centers for Disease Control and Prevention. Notifiable diseases/deaths in selected cities weekly information. *Morbidity and Mortality Weekly Report*, 2009, 58(38):1076–1087.

5. Hutwagner L et al. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *Journal of Urban Health*, 2003, 80 Supplement 1:i89–96. pmid:12791783
6. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine*, 1995, 14:799–810. doi:10.1002/sim.4780140809 pmid:7644860
7. Kulldorff M. A spatial scan statistic. *Communication in Statistics: Theory and Methods*, 1997, 26(6):1481–1496. doi:10.1080/03610929708831995
8. Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clinical Infectious Diseases*, 2004, 39:227–232. doi:10.1086/422003 pmid:15307032
9. Cakici B et al. CASE: a framework for computer supported outbreak detection. *BMC Medical Informatics and Decision Making*, 2010, 10:14. doi:10.1186/1472-6947-10-14 pmid:20226035
10. Chen JH et al. Use of Medicaid prescription data for syndromic surveillance—New York. *Morbidity and Mortality Weekly Report*, 2005, 54(Supplemental):31–34.
11. Hope K et al. Syndromic surveillance: is it a useful tool for local outbreak detection? *Journal of Epidemiology and Community Health*, 2006, 60:374–375. doi:10.1136/jech.2005.035337 pmid:16680907
12. Galit Shmueli HB. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 2010, 52:39–51. doi:10.1198/TECH.2010.06134