Development of a Web-Based Integrated Birth Defects Surveillance System in New York State

Ying Wang, Zhen Tao, Philip K. Cross, Linh H. Le, Patricia M. Steen, Gwen D. Babcock, Charlotte M. Druschel, and Syni-An Hwang

Background: Over the past decade, the Internet has become a powerful and effective tool for public health surveillance. The objectives of this project were to develop secure Web-based applications for Birth Defects Surveillance and to integrate them into routine surveillance activities of the New York State (NYS) Congenital Malformations Registry (CMR). Methods: The Web-based applications were developed on infrastructure of New York State Health Provider Network using JAVA programming language. In addition, SAS/IntrNet software (SAS Institute Inc, Cary, NC) was also used to leverage the data analysis and processing capabilities of SAS for generating real-time reports and performing statistical and spatial analyses. Results: Congenital Malformations Registry staff have developed and implemented a Web-based integrated birth defect surveillance system, which enables staff to routinely perform surveillance activities including monitoring the quality, timeliness, and completeness of case reporting by hospitals; matching the CMR cases to the vital records; conducting trends analysis on birth defect prevalence and mortality with data query and visualization capabilities; and performing temporal and spatial analysis. Conclusions: The CMR's Web-based integrated birth defects surveillance system empowers authorized users to perform routine surveillance activities using only a PC and a Web browser. This system will help NYS public health professionals and epidemiologists perform trend analyses and identify possible clusters of birth defects in space and time that may be related to environmental toxins.

KEY WORDS: birth defects surveillance, congenital malformations registry, GIS, New York State, visualization, Web-based

The Internet has become a powerful and effective tool for disease surveillance, information retrieval, information exchange, and communication. Over the past decade, Web-based disease surveillance systems have been developed and implemented by public health officials and researchers in the academic and medical environment around the world to facilitate the efficient systematic collection, analysis, and interpretation of health data from multiple, geographically distributed sources. This project was supported in part by the Environmental Public Health Tracking grant no. 15032301, funded by the US Centers for Disease Control and Prevention.

Corresponding Author: Ying Wang, MPH, PhD, Congenital Malformations Registry, Bureau of Environmental & Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Flanigan Sq, 547 River St, Troy, NY 12180 (wxy01@health.state.ny.us).

Ying Wang, MPH, PhD, is Research Scientist, Congenital Malformations Registry, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Troy; and Assistant Professor with School of Public Health, New York State University, Albany. Zhen Tao, MD, PhD, is Associate Computer programmer, Congenital Malformations Registry, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Troy. Philip K. Cross, BS, is Director, Congenital Malformations Registry, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Troy. Linh H. Le, MD, PhD, MPH, is Research Scientist, Bureau of HEALTHCOM Network Systems Management, Information Systems & Health Statistics Group, New York State Department of Health. Patricia M. Steen, BS, is Research Scientist, Congenital Malformations Registry, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Troy. Gwen D. Babcock, MS, is Research Scientist, Environmental Health Surveillance Section, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Troy; and Assistant Professor, School of Public Health, New York State University, Albany. Charlotte M. Druschel, MD, is Medical Director, Congenital Malformations Registry, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health, Troy; and Assistant Professor, School of Public Health, New York State University, Albany. Syni-An Hwang, PhD, is Director, Bureau of Environmental and Occupational Epidemiology, Center for Environmental Health, New York State Department of Health; Assistant Professor with School of Public Health, New York State University, Albany.
dispersed organizations. Since 2001, the Centers for Disease Control and Prevention (CDC) has been working with states and local health departments to develop and implement the National Electronic Disease Surveillance System (NEDSS). The NEDSS would facilitate the electronic transfer of information needed for public health from clinical information systems in the healthcare industry, better manage and enhance the large number of current surveillance systems, and allow the public health community to respond more quickly to public health threats. As of April 2005, a total of 27 state health departments and two municipal health departments (New York City and Los Angeles) had implemented the secure, Internet-based disease reporting system, which is the key component of the NEDSS.

New York State Department of Health (NYSDOH) has been an early adopter of electronic disease surveillance concepts and developed a statewide, secure, Web-based Health Provider Network (HPN) system for public health surveillance and information exchange with its federal, state, and local partners. Over the past several years, the NYS Congenital Malformations Registry (CMR) has developed a Web-based case reporting, data management, and communication system for the statewide birth defects registry using the NYSDOH’s HPN. By January 2006, the CMR had converted all 163 reporting hospitals statewide from a manual, paper-based reporting system to an electronic, Web-based case reporting, data management and communication system. This innovative system provides a platform-independent environment for data submission, retrieval, analysis, and communication. This system offers a cost-effective solution for participating hospitals and requires minimal technical assistance from CMR staff. The implementation of this Web-based reporting and communication system has resulted in the more timely submission of cases to the CMR and promoted effective communication between the CMR and reporting hospitals.

In the past several years, the CDC has developed the National Environmental Public Health Tracking (EPHT) program and allocated resources to funded states, local and academic efforts to develop EPHT capacity and science; improve surveillance of environmental hazards, human exposures, and health effects; and test the feasibility and utility of linking the data obtained from the Tracking Network. As part of the EPHT Network, NYSDOH received a grant from the CDC to conduct EPHT projects including improving and enhancing the statewide population-based birth defects surveillance system. The objectives of this project were to develop Web-based applications for birth defects surveillance based on existing infrastructure and emerging technologies and to integrate the use of these tools into the routine surveillance activities of the NYS CMR, using resources available from EPHT funds.

Methods

Congenital malformations registry

The CMR, which was established and began operations in late 1982, is one of the largest statewide, population-based birth defects registries in the nation. Using the passive method of reporting with an active follow-up report auditing system, the CMR annually receives birth defects reports for more than 10,000 children of New York State residents, which comprise approximately 4 percent of all live births.

System architecture

The system architecture is a multilayered application architecture utilizing shared resources including security, data processing and analysis services, and database. The security layer authenticates user identity and, based on the user’s permission, directs requests from the user’s browsers to the Web application layer that contains CMR Web applications. The application layer provides a robust environment to respond to the user’s request and perform common application services for the CMR. It also issues requests to the data processing and analysis layer and the database layer, which enables the users to store and retrieve CMR data.

Database

The birth defects data reported to the CMR are stored in a relational database system. In order to provide a fast and dynamic response to on-line data query and analysis, a cache table scheme was developed so that the data tables and aggregated counts are prepared and refreshed hourly.

Web applications

The CMR’s Web applications with user-friendly interfaces were developed using Web development tools such as Java. Structured Query Language was used in the Web application programs for users to view, update, or query case information dynamically from the CMR’s database using a Web browser. SAS/IntrNet, the newly developed software, functions as a Web gateway from the Web browser to the SAS data-processing layer. This gateway, written by using the Common Gateway Interface, provided access to data in combination with a powerful array of analysis and presentation procedures and was used for users to search and retrieve hospitals’ data submitted on-line, generate
real-time reports, and perform statistical and spatial analysis using the CMR’s databases.

The Web-based applications developed for the integrated birth defects surveillance system included CMR data quality assurance and quality control, completeness of case reporting, record linkage, prevalence and mortality analysis, and temporal and spatial analysis. Programs were developed to run against CMR’s relational databases to search and correct invalid data. The completeness of case reporting by hospitals was measured by the percent of reports for a selected reporting period using the average number of reports from previous reporting years as a baseline. Live birth data from the Vital Records Bureau of the NYSDOH were used as the denominators in calculating birth defect prevalence. Infant mortality rates of children with birth defects were calculated using CMR’s data (children who are alive) as the denominators.

A complex SAS program was developed for matching CMR cases to live birth data from the Vital Records Bureau of the NYSDOH. The deterministic method was used in developing matching strategies and assigning matching scores. The methodology of data linkage will be described in detail in a separate manuscript.

The cumulative sum (CUSUM) methods were used in developing applications for temporal analysis. As described previously in detail, the CUSUM methods sum the differences between the observed and expected number of birth defects and signal a potential change in prevalence when the sum has exceeded the selected decision limit and estimates the time of the change. The SaTScan software package for the Unix system, which uses spatial scan statistics, was used to develop Web-based geographic information system (GIS) tools for identifying clusters of birth defects in space and time. A detailed description of this application will be presented in a separate manuscript.

System security

The security system consists of both the infrastructure, policies, and protocols in place today, as well as novel functionality and approaches to support privacy and security protections including firewalls at strategic points that enforce access control policies between networks, and authentication and identity proofing procedures ensuring that only authorized users can access the application and data for which they are authorized. Rigorous steps have been taken to ensure that the security infrastructure of the NYSDOH’s HPN meets requirements of not only the department security policies and procedures but also the NYS Information Security Policy and Health Insurance Portability and Accountability Act of 1996 (HIPAA), and conforms to the ISO/IEC 17799 standard. Information submitted by the reporting hospitals cannot be accessed by unauthorized persons, is protected from unauthorized modification, and is available for use by the authorized users.

Results

During the past 2 years (2005–2006), a Web-based integrated birth defects surveillance system for the statewide birth defects registry in New York State has been developed and implemented using the NYSDOH’s HPN. The overview of the system is shown in Figure 1. Using a secure access procedure to the firewall protected system and any personal computer equipped with just an Internet browser, an authorized user is able to perform certain surveillance tasks depending on the level of the access to the system, using the developed functions listed on the system’s main menu as shown on the right section in Figure 1. The functionalities of the surveillance system, except for the Web-base case reporting, database management and communication tools which were described previously, are described below in detail.

CMR data quality assurance and quality control

Web-based applications were developed to allow CMR staff to routinely perform data quality assurance and quality control (QA/QC) activities. For instance, the link on the main menu, Check CMR Sybase tables (right section in Figure 1), leads to applications that enable CMR staff to check invalid data and track referential integrity of links among CMR’s relational data tables to make sure that no “orphan” records exist in the data tables.

Web applications were also developed for CMR staff to search and edit records with invalid or redundant British Pediatric Association (BPA) codes, which are used for coding the birth defects. For example, when a child with only a cleft palate condition is reported to the CMR, staff will code this case by using the BPA coding system as 749.0xx. The last 2 digits (“xx”) provide more detailed information about the cleft palate condition. Several months later, a new report may be submitted for the same child with the defect reported to be cleft palate with cleft lip. A new BPA code, 749.2xx, will then be added to this case. Because the BPA code, 749.2xx includes the condition code of 749.0xx, there is no need to keep both BPA codes. Thus, an application was developed to list the cases that have redundant BPA codes and remove those unnecessary BPA codes from the table.

Completeness of case reporting

The function, Completeness of Case Reporting (Figure 1), leads to a submenu with a list of applications...
FIGURE 1 ● System architecture.

(Figure 2A) that were developed to monitor the completeness of case reporting, such as searching and retrieving hospital submitted cases, generating real-time reports, and performing simple statistical analysis using the CMR’s database. For instance, in order to check the completeness of case reporting using the previous birth years’ reporting numbers as a baseline (function # 7 in Figure 2A), authorized users can select a reporting hospital and the birth years of interest using a Web query form (Figure 2B) and generate a real-time report table or chart (Figure 2C). As shown in Figure 2C, in comparison to the baseline average reports (birth years 1995–2002) the completeness of case reporting for a selected hospital (the actual hospital name is not shown for reasons of confidentiality) for the birth years 2004–2006 was 99.7 percent, 97.6 percent, and 91.1 percent, respectively. By reviewing this report, CMR staff will be able to see whether the hospital has been submitting an appropriate number of cases routinely and to identify hospitals with low or inconsistent case reporting rates.

Record linkage

Figure 3 gives an overview of the powerful on-line tools (accessed via the Records linkage link in Figure 1) developed for the record linkage of CMR cases to vital records such as birth certificate files. This user-friendly application provides three functions (Figure 3A) including (1) automatically matching the CMR cases to the birth certificate data for a selected birth year and updating the Sybase tables with selected information from the birth certificate files for the matched cases (Figure 3B); (2) generating a list of records with relatively low matching scores and pairing them with detailed case information from both CMR and birth certificate data for staff interactive matching (Figure 3C); (3) checking and removing duplicate records in the CMR data identified from record linkage. It was found that sometimes two or more records in the CMR’s case table were matched to one record in the birth certificate data for a selected birth year due to the existence of duplicate records (<1%) in the CMR database. The third function of the record linkage application enables the CMR staff to check and remove these duplicate records.

Prevalence and mortality analysis

The functions, Prevalence and Trends and Mortality and Trends (Figure 1), include Web-based tools developed for authorized users to perform trends analysis on birth defect prevalence and infant mortality with query capabilities for selecting disease-related information. As shown in Figure 4A, the birth defect prevalence can be calculated from the CMR data tables and examined by birth year, birth defects categories, or geographic location (county of residence at birth or health service area). Using the built-in on-line query form (Figure 4B), birth defect prevalence can be presented graphically (Figure 4C) by major organ systems for selected birth years and geographic locations. Applications were also developed using the GIS component of SAS/IntrNet.
FIGURE 2  ●  The Web-based birth defects surveillance system of the New York Congenital malformations Registry—monitoring the completeness of case reporting: (A) the submenu; (B) the data query form; (C) the graphic presentation of the completeness of case reporting for the selected birth years and a reporting hospital.
to map the calculated birth defect prevalence and infant mortality rates among children with birth defects, for selected geographic units such as residence-at-birth county or health service area (data not shown).

**Temporal and spatial analysis**

Web-based GIS tools including CUSUM plots and spatial scan statistics were developed for users with access privileges to query the data for temporal and spatial analysis of specific birth defects in particular geographic areas. Figure 5 shows an example of a temporal analysis using the CUSUM method. As shown in Figure 5A, a user-friendly query form was built for authorized users to select variables of interest. Figure 5B shows an example of a binomial CUSUM plot against days, the smallest time unit of analysis available, for a selected birth defect in New York State during the birth years 2000–2005. Similarly, the Web application for the spatial analysis enables authorized users to query factors of interest and then perform scan statistics to identify birth defects clusters in space and time (data not shown).
The Web-based birth defects surveillance system of the New York Congenital Malformations Registry—prevalence and treatment analysis: (A) the submenu; (B) the data query form; (C) the graphic presentation of the birth defect prevalence by organ system.

Prevalence and Trends Analysis by

1. Birth Year
2. Birth Defects Category
3. Geographic Location: County or Health Service Area (HSA)

2. Calculate prevalence by Birth Defects Category: Please make the following selections

Year begin Year end*
2000 2003
2001 2004
2002 2006
2003 2006
2004 2007
2005 2008
2006 2009
2007 2010

Residence at birth County (--- OR ---) Health Service Area (HSA)

All counties
Upstate New York
New York City (5 counties)
01 ALBANY
02 ALLEGANY
03 BROOME
04 CATTARAUGUS
05 CAYUGA

* - The live birth data for 2006 are not available

Select a Birth defects category:

45 Malformations
11 Organ Systems

Select an output format: (PDF HTML RTF)

Submit Reset

New York State Congenital Malformations Registry
Total prevalence by birth defects category (unique children in each category)
Birth year = 2000–2005
County = all counties HSA = no selection
Prepared on: October 15, 2007

Cardiovascular
Central nervous system
Chromosomal
Digestive
Ear
Eye
Genitourinary
Integument
Musculoskeletal
Oral clefts
Respiratory

Prevalence (per 10,000 live births)
**FIGURE 5** The web-based birth defects surveillance system of the New York Congenital Malformations Registry—temporal analysis: (A) the data query form; (B) an example of binomial cumulative sum plot for selected birth years and birth defects.

---

**Discussion**

The significant advancement of information technology and informatics has facilitated development of Web-based applications to improve the timeliness of reporting, quality of data, interpretation of analyses, and dissemination results. As reported previously, the implementation of a secure, flexible, and user friendly Web-based reporting, data management, and communication system in CMR has resulted in better compliance and more timely submission of birth defects cases by reporting hospitals and increased data quality of the
submitted reports.\textsuperscript{18} There was a nearly 50 percent reduction in median days used for reporting by the electronic, Web-based reporting system when compared to the manual, paper-based reporting system. Moreover, the percentage of unspecified diagnoses for three selected birth defect categories examined was significantly decreased for cases submitted through the Web-based reporting system. Developing additional Web-based applications for birth defects surveillance and integrating the use of these tools into the routine surveillance activities of the CMR has greatly extended and enhanced the birth defects surveillance system in New York State.

Recent studies,\textsuperscript{27,28} which reviewed US states’ approaches to implement a Web-based data query systems (WDQSs), showed that at least 27 US states currently have WDQSs. Many of the systems have been developed by the individual states themselves. Although individual states have made different decisions regarding the WDQS, none of the states used a commercial off-the-shelf system to implement a WDQS. The Web-based integrated birth defects surveillance system in New York State developed by the CMR staff is the first system in the nation among the birth defects surveillance programs. This innovative system enables surveillance staff and researchers to perform a variety of tasks in birth defects surveillance and research, through a dynamic interface pertaining to birth defects data from a population-based registry on the World Wide Web. These Web-based applications allow authorized users to retrieve and query data, generate reports, and perform simple statistical analysis with little or no knowledge of SAS programming or Web-page design skills. In addition, the conversion of PC-based report monitoring, data quality control, and data matching programs to Web-based applications has resulted in the automation of a majority of the routine tasks of the CMR and reduced information technology staff’s workload so that they will have more time to design and develop new applications to improve and expand the system.

Geographic information system has been widely used in public health as stand-alone tools deployed on personal computers equipped with sophisticated graphical display software. Since the beginning of the 21st century, Web-based GIS systems have been developed with the rapid development of World Wide Web and Internet technology.\textsuperscript{11,28,33} Web-based GIS systems that perform most of the processes on the server do not require users to install and maintain GIS software on personal computers and require less expertise in GIS for users. The GIS components of our integrated Web-based birth defects surveillance system enable authorized users to map birth defect prevalence and infant mortality rates among children with birth defects for a selected geographic unit, monitor potential change of birth defect prevalence in time, and identify birth defect clusters in space and time.

To further enhance the Web-based integrated birth defects surveillance system, CMR staff are currently developing new applications such as on-line tools that conduct statistical testing for trends analysis applications, perform street-level geocoding of CMR cases, and generate hospital report cards that can be used by CMR staff to routinely evaluate reporting status and alert reporting hospitals with fewer than expected birth defects reports. Moreover, improvements can also be made to the Web-based applications. For instance, the built-in query system can be modified so that users will be able to retrieve data using their own search criteria.

In conclusion, a Web-based integrated birth defects surveillance system has been successfully developed and implemented using NYSDOHs HPN. The system includes previously developed on-line data reporting, database management, and communication facilities.\textsuperscript{16} These user-friendly Web-based tools enable CMR staff to control and ensure the quality of the CMR data, monitor the timeliness and completeness of case reporting by hospitals, and match CMR cases to the vital records. It also allows authorized users to conduct trends analysis with query capabilities on birth defect prevalence and infant mortality of the children with birth defects and perform temporal and spatial analysis over time and space to identify geographic clusters of birth defects that could be related to exposure to teratogens in the environment. This innovative system provides a platform-independent environment for authorized users to perform routine surveillance activities without the need for additional special hardware, software, or programming and has greatly enriched and enhanced the birth defects surveillance in New York State.

REFERENCES


