

Developing a Database Management System to Support Birth Defects Surveillance in Florida

Jason L. Salemi, MPH^{a,b}; Kimberlea W. Hauser, MBA^a; Jean Paul Tanner, MPH^{a,b};
Diana Sampat, MPH^{a,b}; Jane A. Correia, BS^c; Sharon M. Watkins, PhD^c; Russell S. Kirby, PhD^{a,b}

Abstract: The value of any public health surveillance program is derived from the ways in which data are managed and used to improve the public's health. Although birth defects surveillance programs vary in their case volume, budgets, staff, and objectives, the capacity to operate efficiently and maximize resources remains critical to long-term survival. The development of a fully-integrated relational database management system (DBMS) can enrich a surveillance program's data and improve efficiency. To build upon the Florida Birth Defects Registry—a statewide registry relying solely on linkage of administrative datasets and unconfirmed diagnosis codes—the Florida Department of Health provided funding to the University of South Florida to develop and pilot an enhanced surveillance system in targeted areas with a more comprehensive approach to case identification and diagnosis confirmation. To manage operational and administrative complexities, a DBMS was developed, capable of managing transmission of project data from multiple sources, tracking abstractor time during record reviews, offering tools for defect coding and case classification, and providing reports to DBMS users. Since its inception, the DBMS has been used as part of our surveillance projects to guide the receipt of over 200 case lists and review of 12,924 fetuses and infants (with associated maternal records) suspected of having selected birth defects in over 90 birthing and transfer facilities in Florida. The DBMS has provided both anticipated and unexpected benefits. Automation of the processes for managing incoming case lists has reduced clerical workload considerably, while improving accuracy of working lists for field abstraction. Data quality has improved through more effective use of internal edits and comparisons with values for other data elements, while simultaneously increasing abstractor efficiency in completion of case abstraction. We anticipate continual enhancement to the DBMS in the future. While we have focused on enhancing the capacity of our DBMS for birth defects surveillance, many of the tools and approaches we have developed translate directly to other public health and clinical registries.

Key words: birth defects, congenital malformations, database, Florida, registry management, surveillance

Introduction

The value of any public health surveillance program is derived from the ways in which data are managed and used to improve the health of the population it serves.¹ Birth defects surveillance programs are no exception. The ability to meet program goals is driven by the accuracy, completeness, and timeliness of data. Although birth defects surveillance programs vary in their case volume, budgets, staff, and objectives, each program's capacity to operate efficiently and maximize resources remains critical to its long-term survival. The development of a fully-integrated relational database management system (DBMS) can enrich a surveillance program's data with built-in quality assurance (QA) and quality control (QC) measures, while improving efficiency by decreasing procedural complexities associated with administrative tasks.

In 1998, the Florida Department of Health (FDOH) established the Florida Birth Defects Registry (FBDR), a statewide, population-based, passive surveillance system.

Since its inception, the FBDR has monitored structural, functional, and biochemical abnormalities identified within the first year of life in live-born children of women who are Florida residents at the time of delivery. Cases are identified by collecting information from secondary data sources including the Florida Office of Vital Statistics birth records, the Agency for Health Care Administration hospital discharge data, the Children's Medical Services (CMS) Regional Perinatal Intensive Care Centers data, CMS Early Steps Program data, and CMS service-related data sets. Birth defect diagnoses are recorded using the International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) codes. These data sets are merged to develop an unduplicated inventory of infants with birth defects in Florida. The published birth defects prevalence data currently submitted for Florida to the National Birth Defects Prevention Network (NBDPN) are derived from the FBDR.²

The FDOH has readily acknowledged the limitations of the FBDR, primarily, the lack of a mechanism to confirm

"Developing a Database Management System to Support Birth Defects Surveillance in Florida"

^aBirth Defects Surveillance Program, Department of Pediatrics, College of Medicine, University of South Florida, Tampa, FL. ^bDepartment of Community and Family Health, College of Public Health, University of South Florida, Tampa, FL. ^cBureau of Environmental Public Health Medicine, Division of Environmental Health, Florida Department of Health, Tallahassee, FL.

Address correspondence to Jason L. Salemi, MPH; Birth Defects Surveillance Program, University of South Florida, 13201 Bruce B. Downs Blvd., MDC56, Tampa, FL, 33612-3805. Telephone: (813) 396-9242, email: jsalemi@health.usf.edu.

diagnoses derived solely from ICD-9-CM codes. Over the years, the FDOH provided funding to the University of South Florida Birth Defects Surveillance Program (BDSP) to develop, pilot, and operate an enhanced surveillance system in targeted areas with a more comprehensive approach to case identification and diagnosis confirmation than the FBDR. An essential first step of the project was to develop a database capable of managing and storing demographic and clinical data abstracted from the medical records of infants born with birth defects. However, changes in funding sources, project objectives, and manpower required that the database be transformed from an instrument that merely manages the data collection process into one that can guide the mechanical and administrative aspects of the surveillance system. The objective of this article is to describe the process by which our surveillance system's database evolved into a DBMS capable of managing transmission of project data from multiple sources, tracking abstractor time during record reviews, offering tools for defect coding and case classification, and providing reports to users.

Methods

Using the NBDPN Guidelines for Conducting Birth Defects Surveillance,¹ BDSP staff consulted with abstractors, clinicians, information technology specialists, and program coordinators regarding data capture procedures, transmission of data, data file management, data confidentiality, and data security. Each surveillance project posed unique operational challenges. The catchment area for our current project covers 14 counties, including over 80 birthing facilities and more than 70% of the approximately 240,000 annual resident live births in Florida. Despite requests to birthing facilities for transmission of case lists electronically and in a standard format, data structure and medium of submission varied significantly according to facility resources and information management systems. The sheer volume of correspondence with hospitals, requesting case lists and medical records and scheduling multiple site visits per year, was considerable. This required that the DBMS be flexible in managing administrative as well as case data. Also, since wireless Internet connectivity was unavailable in many health information management (HIM) departments, the DBMS had to operate as a stand-alone utility, housing a subset of all potential and confirmed case records for each abstractor in the field. Microsoft® Office Access ("Access") was chosen as the most parsimonious platform for the DBMS, with sufficient functionality to support nearly all necessary and desired activities for the BDSP. Access supports a variety of data field types and is capable of importing or linking directly to data stored in a variety of other formats. It also contains query, form, and report utilities and interfaces that were used to automate time-consuming administrative functions.

Access contains built-in features that improve data quality, such as referential integrity and record-locking in a multi-user environment. However, sophisticated data quality checks and advanced automation require an additional tool. Access databases are supported by Visual Basic for Applications (VBA), a functionally rich and flexible programming language that enables the manipulation and

control of data validation rules and user interface features including skip patterns and custom error messages.

The management of the enhanced birth defects surveillance system required extensive data linkage efforts, both to prevent record duplication and to link case records to external data sources. Although Access with VBA programming can accomplish data linkage, two external software packages, SAS 9.1³ and FRIL (fine-grained record linkage),^{4,5} were used due to their availability, programmatic flexibility, relative ease-of-use, and ability to extend traditional record linkage tools with a richer set of parameters.

Security features were put in place to protect against data loss and maintain data confidentiality. The password-protected DBMS resides in a restricted access folder on a server behind a University of South Florida firewall that incorporates authentication and identity proofing measures, as well as nightly data backup. While in the field, laptops containing the DBMS use fingerprint identification for access, are locked in the vehicle trunk when driving, and are taken into the hotel room with the abstractor for overnight stays.

Results

Description of the DBMS

The DBMS used for the BDSP's enhanced surveillance projects is driven by a front-end menu system that guides the user to several sections within the database and is organized primarily according to the user's role. In the abstractors' section, computerized case records are created, queried, and updated, incoming records are screened for record duplication, and transmission of data (ie, case lists, request for medical records) to and from facilities are tracked. In another section, medical geneticists review malformation and other diagnostic information. A decision tree is used for case classification and final diagnosis. Data managers run quality control assessment reports and manage data linkage in another domain. The last section contains reports that summarize administrative and case data. This secure, cost-effective application offers functional features targeted at the maintenance of data in an organized environment.

Utilization of the DBMS

Since 2004, the FDOH has partnered with the BDSP to design, implement, and manage two enhanced birth defect surveillance projects. Both projects monitored select subsets of birth defects in targeted geographical areas. There was some overlap in the selected birth defects and catchment areas covered between these projects since they differed in funding sources, project intent, and non-overlapping time periods of surveillance.

The first project was funded by a cooperative agreement from the Centers for Disease Control and Prevention's (CDC) National Center on Birth Defects and Developmental Disabilities. It focused on exploring strategies for conducting active surveillance for selected conditions, identifying opportunities to assist with referral for services, and promoting folic acid and other birth defects prevention and education activities. As part of this project, enhanced surveillance was conducted in 56 hospitals across nine counties in Florida

Table 1. Summary of 5421 Confirmed Cases Abstracted as Part of Two Enhanced Birth Defects Surveillance Projects Managed by the Florida Birth Defects Surveillance Program, by Defect Type, 2003–2008

Defect Type	Project 1 (2003–2006)	Project 2 (2007+)
<i>Central Nervous System</i>		
Anencephaly	134	34
Spina bifida	170	80
Other CNS defects	1192	^b
<i>Orofacial Clefts</i>		
Cleft lip with or without cleft palate	534	202
Cleft palate alone	348	130
<i>Abdominal Wall</i>		
Gastroschisis	311	144
Omphalocele	109	^b
<i>Sensory</i>		
Ear defects (anotia, microtia)	46	^b
<i>Chromosomal</i>		
Trisomy 13	68	^b
Trisomy 18	147	^b
Trisomy 21	788	297
<i>Cardiovascular</i>		
Tetralogy of Fallot	^a	107
Transposition of the great vessels	^a	167
Hypoplastic left heart syndrome	^a	79
<i>Genitourinary</i>		
Hypospadias	^a	663
<i>Musculoskeletal</i>		
Reduction deformity, upper limbs	^a	49
Reduction deformity, lower limbs	^a	31
Total Cases^c	3461	1960

^aDefect not included in the case definition for Project 1

^bDefect not included in the case definition for Project 2

^cThe total number of cases is less than the sum of all defect types since each case may have more than one defect.

covering pregnancy outcomes from January 1, 2003 to December 31, 2006. The surveillance goal was to identify and collect information on those pregnancy outcomes affected by selected defects. Birth defects and counts determined in the Project 1 enhanced surveillance efforts are presented in Table 1. The second project is part of Florida's Environmental Public Health Tracking (EPHT) initiative, funded by CDC's National Center on Environmental Health. EPHT is the ongoing collection, integration, analysis, interpretation, and dissemination of data from environmental hazard monitoring, human exposure, and health effects (including birth defects) surveillance. Since 2007, the BDSP has conducted enhanced surveillance of 12 specific defects

in a 14-county catchment area. Birth defects and counts determined in the Project 2 enhanced surveillance efforts are also presented in Table 1. Over the course of both projects, collaboration with HIM departments was integral to our understanding of facility reporting capabilities, which guided initial database development and periodic revisions. Over the last 2 years, the database has evolved into an invaluable system that has vastly improved the quality and security of our data, the efficiency with which we manage our data, and the capacity we have to report on our data. The DBMS has also been easily modifiable and flexible with regard to system enhancements. The addition, removal, and editing of variables collected, data entry/processing forms, QA and QC processes, and reporting mechanisms has required relatively minimal time and expertise.

Computerized collection and transmission of data. The BDSP has utilized multiple sources of case identification, including case reports from catchment facilities and electronic records ascertained by querying vital statistics and hospital discharge data. The DBMS has enhanced the BDSP's capacity to receive computerized data from reporting facilities, eliminating the need to manually enter data that is already captured electronically. The format and structure of all incoming electronic data is reviewed and, if necessary, converted into a standard format before importing into the DBMS. All new electronic records are screened for duplication using linkage software (see section on ability to link surveillance data), and paper-based submissions undergo manual de-duplication using a record query tool within the DBMS. During automated importation or manual entry, data are passed through a series of validation rules to identify data compliance issues and communicate desired modifications to reporting facilities.

Automation and time-saving features. One of the most time-consuming aspects of the enhanced surveillance projects has involved correspondence with each hospital which includes preparing medical record request lists and tracking requested records over time. Useful tools were designed and developed for the DBMS using the query and report utilities within Access, supplemented with VBA programming, to automate some processes such as drafting letters and generating case lists. Once case lists have been prepared, every potential case identified requires a record review to evaluate case criteria. Although the primary unit of interest exists at the infant level, hospital admission-level information (often with multiple admissions per infant) drives record review. The DBMS has facilitated this process by automatically generating request lists and allowing the abstractor to modify lists with an admission-level toggle box. In addition, contact information for participating facilities has been managed by the DBMS, enabling automated generation of customized request letters for HIM departments. The DBMS has also automatically tracked the number of times a medical record is requested from a reporting facility, allowing rapid identification and follow-up of records that cannot be located. Following the record review, cases are removed from the current facility's request list but may still appear on other facilities' lists should further records be eligible for review.

Figure 1. Using the database management system to code birth defects: within each case record (top); using a defect-level coding tool (bottom)

General Info | Admissions | ICD-9 Codes | Birth Details | Diagnostic Tests I | Diagnostic Tests II | Malformations | Malf-Specific Info | Referrals | Comments 09212

Enter child malformations

Name/description:
Down syndrome

Important comments/details:
47, XX, +21 (karyotype)

Method of diagnosis: karyotype

Timing of diagnosis: Postnatal

Malformation confirmed?

Texas 6-digit code: 758.000

Add Another Malformation

Record: 14 | 2 of 7 | No Filter | Search

View existing child malformations entered

Defect	Code
Atrial septal defect	745.590
Down syndrome	758.000
Down syndrome facies	744.910
Nuchal fold	744.500
Patent ductus arteriosus	747.000
Patent foramen ovale	745.590
Wide spaced toes	755.600

Record: 14 | No Filter | Search

CLOSE Child Form

STOP RECORD REVIEW

MALF_CD	MALF	MALF_C	STUDY
746.700	Hypoplastic left heart syndrome	With aortic atresia and mitral stenosis	06023
746.700	Hypoplastic left heart syndrome	With severe left ventricle hypoplasia, mitral valve hypoplasia and aortic atresia	07409
746.700	Hypoplastic left heart syndrome	With probable mitral valve atresia	07548
746.█	Hypoplastic left heart syndrome	With intact atrial septum, severe mitral stenosis, vsd, severe aortic arch hypoplasia and large pda	07556
	cleft palate	No cleft lip; full and soft palate cleft, redundant tissue at midline	07537
	Cleft palate	Isolated bilateral; record notes defect was not diagnosed prenatally	09942
	Cleft lip and palate	Bilateral complete	07551
	Cleft lip	No cleft palate; left incomplete with alveolar notch; repaired at Shands at 4 months of age	07426

The NBDPN deems proper coding of diagnostic information an essential aspect of birth defects surveillance⁶. Following medical record review and diagnosis confirmation, the BDSP codes every confirmed birth defect using a specific 6-digit British Paediatric Association (BPA) code modified by the birth defects branch of the CDC. Abstractors initially coded defects within each case record (Figure 1, top); however, the DBMS incorporated a coding tool allowing abstractors to code identical or similar defects across multiple

cases at the same time (Figure 1, bottom). The result was a significant gain in coding efficiency and standardization.

Quality assurance and quality control. The DBMS was designed to implement real-time QA and QC methods to prevent problems at the source of data collection. VBA programming was used to develop range checks and sophisticated multi-variable logic tests that are evaluated immediately. Message boxes to alert the user appear instantly; impossible values must be corrected, while

borderline or implausible values require confirmation before exiting a particular case record. Calculations and conversions are made automatically whenever possible (ie, birth weight, weight gain during pregnancy) and drop-down boxes are used to facilitate data standardization and ease-of-data entry. The DBMS also incorporates transaction logs to document the “who”, “when”, and “what” regarding record additions, deletions, or modifications. Total abstraction time for each record is tracked to monitor timeliness and to estimate future time in the field for a given case load.

Ability to link surveillance data. SAS 9.1³ or FRIL^{4,5} has been used to link all incoming electronic data against existing case data in an effort to identify duplicate records prior to importation or manual entry into the DBMS. As perinatal data from many sources are linked to vital records, all enhanced surveillance records reviewed by BDSP staff, whether case or non-case, are linked to birth or fetal death certificates. Certificate numbers from validated links are imported back into the DBMS. This capability is essential to the conduct of epidemiologic studies that necessitate linking cases to data sources housing information on study subjects.

Case-classification tool. The BDSP captures detailed clinical information on each case infant and mother, including relevant perinatal data, diagnostic test results (eg, genetic

and blood tests, ultrasounds, MRIs, x-rays, etc), exposures and family history, verbatim and coded descriptions of every major and minor birth defect, and methods of defect diagnoses. Case data are reviewed by a clinical geneticist for diagnosis confirmation and classification. The goal of case classification is to apply knowledge of embryologic and pathogenetic mechanisms to place cases in groups that are more comparable for analysis.⁷⁻⁹ To facilitate geneticists' case-classification, the DBMS uses an interactive decision tree (Figure 2), modeled after a classification algorithm developed for the National Birth Defects Prevention Study.⁶ After the geneticist progresses through a series of decision points, the DBMS classifies the case into isolated, multiple, and syndromic categories and the geneticist provides a final diagnosis with relevant details.

DBMS case review and utilization experience. Case review and abstraction began in April 2004 and is ongoing for Project 2 (EPHT). Since its inception, the DBMS has been used as part of our surveillance projects to guide the receipt of over 200 case lists and review of 12,924 (Project 1 and 2 total) fetuses and infants (with associated maternal records) suspected of having selected birth defects in over 90 birthing and transfer facilities. Following abstractor review, 7503 (58.1%) cases failed to meet the case criteria leaving

Figure 2. Interactive decision tree used to facilitate geneticists' classification of cases within the database management system

Unique STUDY ID: 09212 **Facility:** MEMORIAL HOSPITAL MIRAMAR
Case Type: Child **Pregnancy Outcome:** Live Birth

Classification: Syndrome of Known Etiology
Primary diagnosis:
Notes:

Documented child malformations:

Defect	Code
Atrial septal defect	745 590
Down syndrome	758 000
Down syndrome facies	744 910
Nuchal fold	744 500
Patent ductus arteriosus	747 000
Patent foramen ovale	745 590
Wide spaced toes	755 600

Record: 1 of 7

CLASSIFY CASE **OPEN Maternal Form** **OPEN Child Form** **Exit Form**

5421 confirmed cases across the two projects. Table 1 shows all confirmed cases abstracted as part of both projects, tabulated by type of defect. There are currently over 2000 records pending record review for Project 2. We began tracking the time it takes to review case and non-case records in 2007, and now have data on over 3700 record reviews. We estimate that, on average, it takes our staff 7 minutes to evaluate case criteria and another 30 minutes to locate and enter infant and maternal data into the DBMS, although there is significant variation depending on the complexity of the case and organization of the medical record.

As the enhanced DBMS operationalized administrative functions such as preparing letters, cover sheets, and medical record request lists for each hospital, and tracking individual record requests, we estimate that it results in saving approximately one hour per case list, which translates into 2–3 full-time weeks for an abstractor over the course of a year.

Discussion

Our experience with the evolving DBMS for birth defects surveillance in Florida has provided both anticipated and unexpected benefits. Automation of the processes for managing incoming case lists has reduced clerical workload considerably, while improving accuracy of working lists for field abstraction. Data quality has improved through more effective use of internal edits and comparisons with values for other data elements, while simultaneously increasing abstractor efficiency in completion of case abstraction. Enhancements to the DBMS have also improved the processes for de-duplication of case lists, coding of birth defects, and classification of cases. Automation of internal and external reports has improved programmatic operations. Theoretically, use of the enhanced capabilities of the DBMS has also decreased the generation of paper records; however, we have been unable to verify this outcome in practice.

To address current limitations and to meet evolving program goals, we anticipate continual enhancement to the DBMS. Currently, the DBMS only supports the enhanced surveillance projects previously described and does not support the FBDR since that system is based solely on data linkage and not medical record review at facilities. The desired expansion of the FBDR into a full-fledged statewide program based on active case-finding methods will require a DBMS with simultaneous multi-user functionality. As health care facilities become more comfortable with the use and operation of Web-based data acquisition modalities, we will explore the feasibility of front—and back-end database technologies utilizing Web-based interfaces. We also plan to explore transitioning the database repository to a network

server using Microsoft® SQL Server or a similar platform. This will facilitate integration of geographic information systems (eg, ArcGIS, ESRI Redlands, CA) and statistical analysis programs (eg, SAS³) as tools for data management, exploratory data analysis, and reporting.

While we have focused on enhancing the capacity of our DBMS for birth defects surveillance, many of the tools and approaches we have developed translate directly to other public health and clinical registries. Our methods can also enhance population-based surveillance for developmental disabilities, sequelae of adverse perinatal outcomes, specific pediatric disorders including cystic fibrosis, fragile X, muscular dystrophy, and fetal alcohol-related disorders, or conditions affecting adults including amyotrophic lateral sclerosis, multiple sclerosis, and other conditions of interest for environmental public health or the health of the aging population.

Acknowledgements

The FDOH and the University of South Florida BDSP would like to acknowledge the following organizations and individuals for contributions to this project: the CDC National Center for Birth Defects and Developmental Disabilities (Cooperative Agreement U50/CCU423304-02); the CDC National Center for Environmental Health (Cooperative Agreement U38 EH000177-01); Greg Kearney, PhD for his guidance and expertise with the EPHT initiative; and Suzanne Sage, RN, MEd for her proficiency in data abstraction and insight into DBMS modifications.

References

1. National Birth Defects Prevention Network (NBDPN). *Guidelines for Conducting Birth Defects Surveillance*. Sever, LE, ed. Atlanta, GA: National Birth Defects Prevention Network, Inc.; June 2004.
2. Population-based Birth Defects Surveillance data from selected states, 2001–2005. *Birth Defects Res A Clin Mol Teratol*. Dec 2008;82(12):831–961.
3. SAS Institute Inc. *SAS 9.1*. Cary, NC: SAS Institute; 2004.
4. Jurczyk P, Lu JJ, Xiong L, Cragan JD, Correa A. FRIL: A Tool for Comparative Record Linkage. *AMIA Annu Symp Proc*. 2008:440–444.
5. Jurczyk P, Lu JJ, Xiong L, Cragan JD, Correa A. Fine-grained record integration and linkage tool. *Birth Defects Res A Clin Mol Teratol*. Nov 2008;82(11):822–829.
6. Rasmussen SA, Olney RS, Holmes LB, Lin AE, Keppler-Noreuil KM, Moore CA. Guidelines for case classification for the National Birth Defects Prevention Study. *Birth Defects Res A Clin Mol Teratol*. Mar 2003;67(3):193–201.
7. Khoury MJ, Botto L, Mastroiacovo P, Skjaerven R, Castilla E, Erickson JD. Monitoring for multiple congenital anomalies: an international perspective. *Epidemiol Rev*. 1994;16(2):335–350.
8. Martinez-Frias ML, Frias JL, Rodriguez-Pinilla E, et al. Value of clinical analysis in epidemiological research: the Spanish registry experience. *Am J Med Genet*. Nov 1 1991;41(2):192–195.
9. Martinez-Frias ML, Frias JL, Salvador J. Clinical/epidemiological analysis of malformations. *Am J Med Genet*. Jan 1990;35(1):121–125.