ISSN 1945-6123

National Cancer Registrars Association

Journal of Registry Management

Spring 2013 • Volume 40 • Number 1

Published by the National Cancer Registrars Association • Founded in 1975 as *The Abstract*

NCRA Launches the New Center for Cancer Registry Education!

www.CancerRegistryEducation.org

The Center for Cancer Registry Education (CCRE) provides easy access to high-quality educational programming to support both seasoned professionals and those new to the cancer registry field. The website offers a variety of products and services, allowing you to tailor your training and manage CE credits.

Use your NCRA Website member log-in and password!

Questions? Call 703-299-6640 Ext. 317 or e-mail ccre@ncra-usa.org

The CCRE Website is partially funded by the Centers for Disease Control and Prevention, Cooperative Agreement U58DP003842.

Journal of Registry Management

is published quarterly by the
National Cancer Registrars Association
1340 Braddock Place, Suite 203
Alexandria, VA 22314
(703) 299-6640
(703) 299-6620 FAX

Address change of address and subscription correspondence to:
National Cancer Registrars Association
1340 Braddock Place, Suite 203
Alexandria, VA 22314

Address all editorial correspondence to:
Vicki Nelson, MPH, RHIT, CTR, Editor
CDC/NCCDPHP/DCPC/CSB
4770 Buford Drive, MS K-53
Atlanta, GA 30341-3717
Email: izo4@cdc.gov

Letters to the Editor
Letters to the Editor must be signed and include address and telephone number; none will be published anonymously. Letters subject to editing.

Editorial Policy
Opinions expressed in articles appearing in *Journal of Registry Management* are those of the authors and not necessarily those of the National Cancer Registrars Association, or the editor of *Journal of Registry Management*.

Subscription to *Journal of Registry Management*
Subscription is a benefit of membership in the National Cancer Registrars Association. Individual subscriptions may be purchased for $40/year U.S. dollars. Single copies may be purchased for $12 U.S., $15 International, prepaid only.

Advertising Sales
Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the *Journal* does not indicate or imply endorsement by the *Journal* or NCRA. Advertising sales should be directed to Paula Spitler, NCRA, 1340 Braddock Place, Suite 203, Alexandria, VA 22314, (703) 299-6640, (703) 299-6620 FAX.

National Cancer Registrars Association

Journal of Registry Management

Spring 2013 • Volume 40 • Number 1

Contents

Production and Printing
The Goetz Printing Company

Indexing

The *Journal of Registry Management* is indexed in the National Library of Medicine's MEDLINE database. Citations from the articles indexed, the indexing terms (key words), and the English abstract printed in *JRM* are included and searchable using PubMed.

For your convenience, the *Journal of Registry Management* is indexed in the 4th issue of each year and on the Web (under "Resources" at **http://www.ncra-usa.org/jrm**). The 4th issue indexes all articles for that particular year. The Web index is a cumulative index of all *JRM* articles ever published.

Guest Editors' Introduction

Enhancing Methods for Population-Based Birth Defects Surveillance Programs

Russell S. Kirby, PhD, MS, FACE[a]; Wendy N. Nembhard, PhD, MPH[b]

In this issue of the Journal of Registry Management, we have continued our collaboration with the National Birth Defects Prevention Network (NBDPN) to promote research aimed at improving and enhancing birth defects surveillance methods. NBDPN is committed to the primary prevention of birth defects and to the improvement of outcomes for children and families living with birth defects through the use of birth defects surveillance data for research, program planning, and program evaluation. Members of NBDPN include staff from population-based birth defects surveillance programs across the United States, as well as clinicians, public health professionals, and researchers involved with birth defects epidemiology, primary and secondary prevention activities, program planning, and evaluation.

The birth defects articles included in this Spring 2013 issue of the Journal of Registry Management were selected from those submitted in response to a call for manuscripts distributed to all state birth defects surveillance programs, NBDPN members, and the birth defects surveillance list serv, as well as posted on the NBDPN Web site (**http://www.nbdpn.org**). The papers included in this issue had both editorial and formal peer review. It is our hope that the methods and findings from these papers will contribute to the continual improvement of the science and practice of birth defects surveillance in the United States and around the world. We are pleased to include 2 international contributions: one describes the methods and strategies used to improve registry quality in a birth defects registry in Argentina and the other examines the issues involved in obtaining statutory notification for a registry of birth defects and cerebral palsy in Western Australia. In this issue, we also include a study presenting results of a statewide survey of hospitals in the state of New York regarding their plans for conversion from ICD-9-CM to ICD-10-CM disease and procedure coding. Another birth defects paper describes a record linkage framework for integrating population-based public health databases into a multi-layered maternal and child health database. The approach developed in the paper is generalizable to almost all birth defects surveillance programs, and should greatly enhance the utility of these data to address questions concerning health services utilization, costs/charges, rehospitalization, treatments and procedures, and co-occurring conditions.

Many dedicated individuals contributed their time and effort to assist with the publication of this issue. These include the authors of all the submitted manuscripts and the following peer reviewers: Suzanne Block, Amy Case, Derek Chapman, Glenn Copeland, Jan Cragan, Dan Driggers Charlotte Druschel, Marcia Feldkamp, Deborah Fox, Debra Kane, Peter Langlois, Melanie Lockhart, Rachel Richesson, Lowell Sever, Csaba Siffel, Carol Stanton, Shihfen Tu, and Dante Verme. We also thank the members of the NBDPN Publications and Communications Committee, and Vicki Nelson for her help and assistance with the submission and publication of these manuscripts. We also recognize Cara Mai for assisting in coordinating this special issue and her tremendous support of birth defects surveillance at the state and national levels. We also thank the Division of Birth Defects and Developmental Disabilities at the National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, for its support of the NBDPN.

[a]Department of Community and Family Health, College of Public Health, University of South Florida, Tampa, FL. [b]Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL.

Original Article

A Survey on Readiness and Needs Regarding the Transition from ICD-9-CM to ICD-10-CM

Ying Wang, PhD[a]; Zhen Tao, PhD[a]; Deborah Fox, MPH[a]; Patricia M. Steen, BS[a]; Charlotte M. Druschel, MD[a]

Abstract: The New York State Congenital Malformations Registry (CMR) conducted a Web-based survey to assess reporting hospitals' readiness and needs with regards to the transition from the ICD-9-CM to ICD-10-CM for diagnoses and ICD-10-PCS for procedure coding system (PCS). The survey contains 8 questions focusing on the transition to collect information about case reporting methods, anticipated plan and date for the transition, and the needs from the CMR for the process. In September 2012, a link to the Web-based survey was sent to all 158 CMR reporting hospitals requesting completion of the online survey. By October 31, 2012, 91 (60%) out of 158 reporting hospitals completed the survey. For the question "*When will your facility be ready to report to the CMR using the ICD-10 coding system?*", a majority (71%) of the respondents answered October 1, 2014. With regard to the method they plan to use for converting from ICD-9-CM to ICD-10-CM/PCS, 51% will rely on a crosswalk provided by vendors and 7% will use the general equivalence mapping method. Nearly half (45%) of the respondents were interested in implementing a dual reporting system (accepting both ICD-9-CM and ICD-10-CM/PCS by the CMR). For the question, "*What specific information would you like the Congenital Malformations Registry to provide in regards to reporting to this registry using ICD-10?*", 30% of the respondents requested a list of reportable ICD-10-CM/PCS codes and related descriptions, 10% requested an ICD-9-CM to ICD-10-CM/PCS crosswalk, and 19% requested keeping them updated with information about the transition and implementation of ICD-10-CM/PCS. Among the respondents who provided comments at the end of the survey, more than half (55%) stated that they are in the process of transition and 27% expressed thanks and appreciation for CMR's leading effort on the transition project. This online survey enabled the CMR staff to assess readiness and identify the needs of the hospitals regarding the transition. This information will help the CMR with appropriate planning for our own transition and enable us to meet the needs of hospital reporters.

Key words: *survey, ICD-9-CM, ICD-10-CM, coding system, online reporting, transition, congenital malformations registry, New York State hospitals*

Introduction

The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), based on the World Health Organization's 9th Revision, International Classification of Diseases (ICD-9),[1] has been used in the United States for more than 30 years as the standardized health-care coding and reimbursement system for health management and clinical purposes. ICD-9-CM enables health-care providers and researchers/epidemiologists to assign codes describing and classifying diagnoses and procedures associated with heath care encounters including visits to physicians' offices and emergency rooms, hospital admissions and ambulatory surgeries. In 1990, the World Health Organization created and released the 10th revision of the ICD codes (ICD-10),[2] which have been expanded to more than 70,000 codes (compared to fewer than 15,000 codes in ICD-9) by primarily adding detail, laterality (right-left location on the body), and first versus later episode codes. For the past 20 years, over 100 industrialized countries have already adopted ICD-10 to classify diseases and related health problems because of the many benefits it provides, such as enhanced accuracy of coding and payment for reimbursement for services rendered, improved quality of care and documentation, increased specificity to identify diagnoses and procedures, improved disease reporting, and outbreak data and increased data mining capabilities for analysis of diagnosis.

The United States developed a Clinical Modification to ICD-10 and created the ICD-10-CM coding system for diagnoses and ICD-10-PCS for procedure coding system (PCS) in 1998 as the replacement for ICD-9-CM. The National Center for Health Statistics started using ICD-10 to code and classify mortality data from death certificates in 1999.[1] Health care providers and payers in the United States were initially scheduled to adopt ICD-10-CM/PCS in 2008 for morbidity, diagnosis and procedure coding, but the Centers for Medicare and Medicaid Services pushed back the deadline due to substantial costs and potential challenges in the implementation of ICD-10-CM/PCS such as planning and documentation; education and training of coders, physicians and users of the coded data; information technology (IT) changes (new hardware and software necessary for the updated system); and financial and administrative transaction coding under the Health Insurance Portability and Accountability Act (HIPAA).[3-6] The implementation of ICD-10-CM/PCS was further delayed to October 1, 2014, the compliance deadline for nationwide conversion to ICD-10-CM/PCS codes, by final rule issued in 2012 by the Centers for Medicare and Medicaid Services.

The Congenital Malformations Registry (CMR) of the

[a]New York State Department of Health, Albany, New York.

Address correspondence to Ying Wang, PhD, Congenital Malformations Registry, Center for Environmental Health, New York State Department of Health, Empire State Plaza, Corning Tower, Room 1203, Albany, NY 12237. Telephone: (518) 402-7990. Fax: (518) 402-7969. Email: wxy01@health.state.ny.us.

New York State Department of Health (NYSDOH), one of the largest statewide, population-based birth defects registries in the nation, receives birth defect reports on more than 11,000 children annually among nearly 250,000 live births in New York State (CMR, 2009). Hospitals and physicians are required to report children under 2 years of age, who were born or reside in New York State with a congenital malformation, chromosomal anomaly, or persistent metabolic defect, through CMR's electronic, Web-based reporting system.[7-8] The submitted electronic records to the CMR include demographics of the patients and diagnostic information including ICD-9-CM codes and narratives describing birth defect conditions. Thus, the CMR will need to modify its current reporting and database management systems to accommodate the changes relating to the nationwide implementation of ICD-10-CM/PCS across all health-care settings. In preparation for that, this article presents the findings from the survey conducted by the CMR among all reporting hospitals to assessing hospitals' readiness and needs regarding ICD-9-CM to ICD-10-CM/PCS transition.

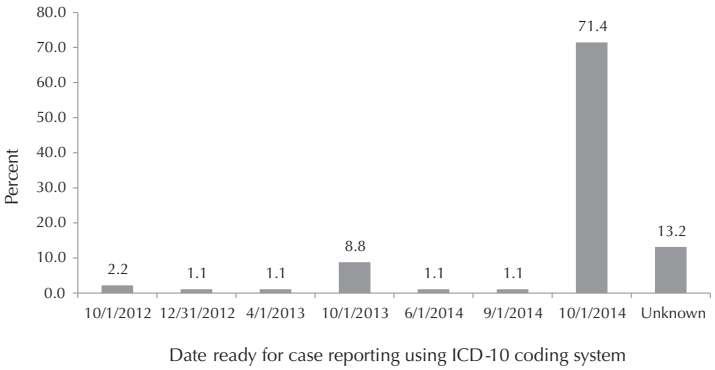
Methods

Survey Questionnaire

The survey questionnaire consisted of 8 questions with the focus on the transition from the ICD-9-CM to ICD-10-CM/PCS coding system:

1. **When will your facility be ready to report to the CMR using the ICD-10 coding system?** Open-ended. Example: October 2012. *Note: we omitted the term "CM" in the survey questionnaire.*
2. **How does your facility plan to convert ICD-9 to ICD-10?** Select one: Use general equivalence mappings / Rely on crosswalk provided by a vendor / Not sure / Other—specify.
3. **Initially, will your facility be interested in taking advantage of a dual reporting system that allows you to send in both ICD-9 and ICD-10 codes for the same malformation?** Select one: Yes / No / Not sure.
4. **Which of the following tools is your facility currently using for reporting cases to the CMR?** Select one: File upload / Data entry / Both.
5. **If your facility uses the file upload method, approximately what date will your IT department complete the modification of the file upload programs for reporting in ICD-10?** Open-ended. Example: October 1, 2013.
6. **Is your facility currently using HANYS (Healthcare Association of New York State) to report CMR cases?** Select one: Yes / No / Not sure.
7. **Is there a specific person in your department who is overseeing the ICD-9 to ICD-10 conversion?** Select one: Yes / No / Not Sure. **If yes, please provide the contact information.** Name and phone number.
8. **What specific information would you like the Congenital Malformations Registry to provide your department in regards to reporting to this registry using ICD-10?** Open-ended.

Figure 1. Responses to Question 1, *"When will your facility be ready to report to the CMR using the ICD-10 coding system (example: October 2012)?"* (n=91). New York State Congenital Malformations Registry's Survey on Transition from ICD-9-CM to ICD-10-CM/PCS among Reporting Hospitals, September 2012



In order to ensure a reasonably high response rate, identifying information such as the names of the responders and their institutions was not collected in the survey. However, the Internet protocol address (a numerical label assigned to each desktop or laptop computer) of the responders was recorded by the online survey tool and allowed us to identify and remove duplicate responders.

Web-based Survey

The survey questionnaire was uploaded to the Internet using Web-based survey software. In September 2012, an email containing a cover letter explaining the purpose of the survey and a Web link to the Web-based survey was sent to the directors of the Medical Records Department of all 158 CMR reporting hospitals across New York State to invite them to complete the online survey. Two weeks after the first mailing, a subsequent follow-up email was sent to all the hospitals to thank the ones who responded to the survey and invite non-respondents to participate.

Data Analysis

Responses to the survey were collected, downloaded, and converted into a dataset for analysis. Summary statistics, simple and stratified, were generated using SAS (SAS Institute, Cary, North Carolina). Facilities were not required to answer all survey questions, as such percentages were calculated using the number of hospitals that responded to each individual question.

Results

By October 31, 2012, 91 out of 158 reporting hospitals completed the survey. The response rate was 60%. Figure 1 presents responses to open-ended question 1, "*When will your facility be ready to report to the CMR using the ICD-10 coding system?*"(n=91). A majority (71%) of the respondents answered October 1, 2014; 9% answered October 1, 2013; and 13% were not sure. For question 2 regarding the method for converting ICD-9-CM to ICD-10-CM/PCS, about 51% of the 91 respondents answered that they will rely on a

Table 1. Responses to Questions 3-7 about Case Reporting and ICD-9 to ICD-10 Transition in the Survey (n=91), New York State Congenital Malformations Registry's Survey on Transition from ICD-9-CM to ICD-10-CM/PCS among Reporting Hospitals, September 2012

Questions	*Responses*		
	Yes	No	Not sure
	n (%)	n (%)	n (%)
Will your facility be interested in taking advantage of a dual reporting system that allows you to send in both ICD-9-CM and ICD-10-CM codes for the same malformation? (n=91)	41 (45.1)	10 (11.0)	40 (44.0)
Which of the following methods does your facility currently use for reporting cases to the CMR? (n=91)			
Online data entry	48 (52.7)	–	–
Online file upload	28 (30.8)	–	–
Both data entry and file upload	13 (14.3)	–	–
Not sure/unknown	–	–	2 (2.2)
If your facility uses the file upload method, approximately at what date will your IT department complete the modification of the file upload programs for reporting in ICD-10-CM (example? (n=41)			
October 1, 2013	4 (10.0)	–	–
October 1, 2014	16 (39.0)	–	–
Not sure/unknown	–	–	21 (51.0)
Is your facility currently using HANYS* (Healthcare Association of New York State) to cases reporting? (n=91)	22 (24.2)	44 (48.4)	25 (27.5)
Is there a specific person in your department who is overseeing the ICD-9-CM to ICD-10-CM conversion? (n=91)	54 (59.3)	19 (20.9)	18 (19.8)

*HANYS is a third party vender that helps hospitals to meet their state reporting requirements.

crosswalk provided by a vendor, 7% will use the general equivalence mapping method (which serves as 2-way translation dictionaries for diagnosis and procedure codes from which crosswalks can be made), 10% use other methods (ICD-9-CM and ICD-10-CM/PCS coding books), and 33% were not sure (data not shown in tables/figures).

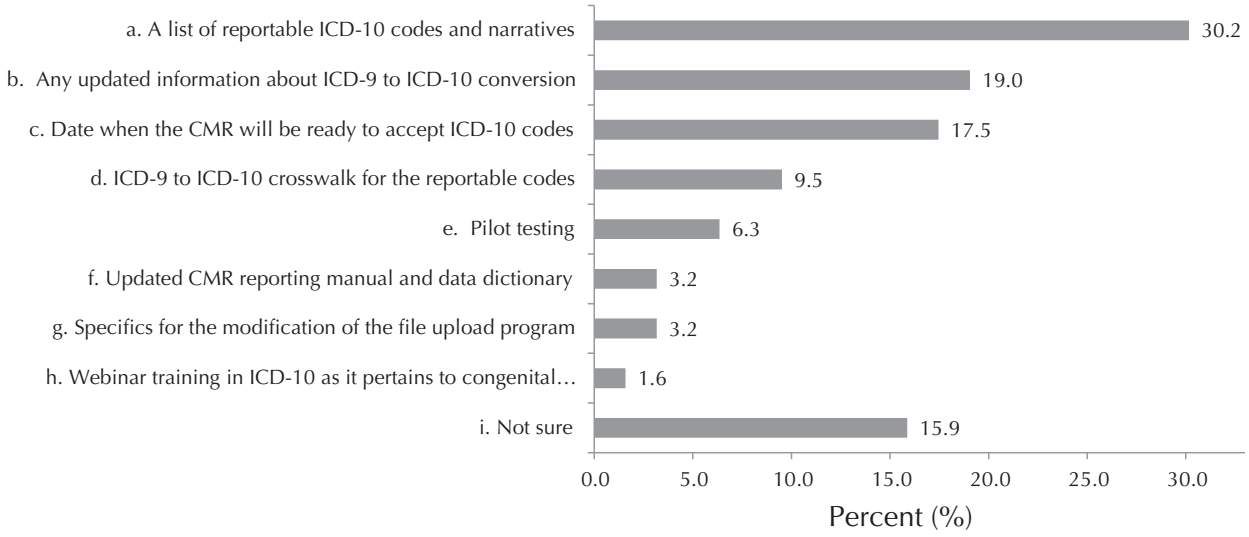
Table 1 summarizes the responses to the survey questions 3-7 about case reporting and ICD-9-CM to ICD-10-CM/PCS transition. Of the 91 responding hospitals, 45% showed interest in taking advantage of a dual reporting system that allows facilities to send in both ICD-9-CM and ICD-10-CM/PCS codes to the CMR for the same malformation; 24% are currently using HANYS (Healthcare Association of New York State) to report cases to the CMR. HANYS is a third party vendor that helps hospitals to meet their state reporting requirements including IT software updates. For the case-reporting methods, 53% of the respondents are currently using CMR's online data entry system; 31%, the file-upload method; and 14%, both methods. When the responding hospitals who are currently using online file-upload method for case reporting (n=41) were asked "*What date will your Information Technology Department complete the modifications of the file upload-programs for reporting in ICD-10?*", about 10% answered October 1, 2013; 39% answered October 1, 2014; and 51% were not sure. When asking the 91 respondents if there is a specific person in their facility who oversees the ICD-9-CM to ICD-10-CM/PCS transition, 59% provided that information.

Figure 3 summarizes the responses (n=70) to the open-ended question 8, "*What specific information would you like the Congenital Malformations Registry to provide your department in regards to reporting to this registry using ICD-10?*". One third (30%) of the respondents requested a list of reportable ICD-10-CM/PCS codes and related descriptions, 19% would like the CMR to keep them updated with information about CMR's transition plan and implementation, 18% wanted to be informed about the date when the CMR is ready to accept ICD-10-CM/PCS codes, 10% requested an ICD-9-CM to ICD-10-CM/PCS crosswalk for the reportable ICD-10-CM/PCS codes, and 6% were willing to participate the pilot testing for the new reporting system. Interestingly, a few responding hospitals (2%) requested the CMR to provide webinar trainings in ICD-10-CM/PCS as it pertains to congenital malformations.

Discussion

ICD-10-CM/PCS is the biggest change in standard health-care coding systems in decades. Implementation of ICD-10-CM/PCS in the United States will impact every system, process, and operation that uses diagnostic codes at all health-care (providers and payers) surveillance and research settings. To ensure that systems and processes on both sides, the CMR (receiving birth defect cases) and the reporting hospitals (submitting the case reports) are

Figure 2. Responses to Question 8, *"What specific information would you like the Congenital Malformations Registry to provide for your department in regards to reporting using ICD-10)?"* (n=70). New York State Congenital Malformations Registry's Survey on Transition from ICD-9-CM to ICD-10-CM/PCS among Reporting Hospitals, September 2012



updated to be ICD-10-CM/PCS compliant by the mandated implementation date of October 1, 2014, a strategic plan is necessary for a smooth and timely transition. The current survey enables the CMR staff to assess the readiness and the needs for the transition and help in developing the transition plan.

It is encouraging to find that a majority (87%) of the respondents stated that they will be ready to report to the CMR using the ICD-10-CM/PCS coding system by October 1, 2014 (Figure 1). The current survey found that 24% of the reporting hospitals who responded to the survey currently use a vendor for system updating and fulfilling their state reporting requirements (Table 1). The CMR is responsible for contacting the vendors in regards to their services and plans for ICD-10-CM/PCS implementation and provide technical assistance in modifying database structure and file format for online case reporting to the CMR.

The CMR's online reporting system provides 2 reporting methods: 1) online data entry that allows submitting one report at a time by a coder or staff, and 2) online file upload that provides a file transfer tool for a hospital's medical records department to submit a batch of records at one time.[8] Hospitals using CMR's online file-upload tool will need assistance by their IT staff to modify their current file format with regard to the changes the CMR will make for accepting ICD-10-CM/PCS codes. Our survey found that only about half (49%) of the responding hospitals who use CMR's online file-upload tool stated that their IT staff will complete the modification for reporting using ICD-10-CM/PCS codes by October 1, 2014. CMR staff will contact all file-upload hospitals to make them aware of the October 1, 2014 deadline and work with their IT staff to ensure timely implementation of the changes. Although only about 59% of the survey respondents provided the contact person who oversees the transition project, CMR staff will identify the contact people of all reporting hospitals and work closely with them on this transition project.

Although there is no grace period for using ICD-10-CM/PCS coding system after the deadline of October 1, 2014, the CMR plans to develop a dual system to accept both ICD-9-CM and ICD-10-CM/PCS codes. The CMR routinely conducts audits on reporting hospitals to search for unreported cases using hospital discharge files recorded in the previous years (2-3 years prior to the current reporting year). Thus, the CMR's reporting system and database system should be able to accept not only ICD-10-CM/PCS codes for children born and diagnosed on and after October 1, 2014, but also ICD-9-CM codes for children born, diagnosed and recorded/coded before October 1, 2014. Our survey found that nearly half (45%) of the respondents are interested in a dual reporting system that allows the hospitals to send in both ICD-9-CM and ICD-10-CM/PCS codes for each record. This finding supports our plan of developing such a dual reporting system. The dual coding system will enable CMR staff to analyze comparative data between the 2 code sets. It is possible that some hospitals intend to be fully converted and plan to use the dual system before the deadline arrives. The CMR plans to complete the transition process and the implementation of the dual reporting system a few months before the deadline and conduct pilot testing with those hospitals who are ready for using the dual reporting system.

About one third of the survey respondents are not sure about the methods they will use for converting from the ICD-9-CM to ICD-10-CM/PCS. This suggests the need for educating and training health-care providers and office staff so that they become familiar with the new coding system and the crosswalk between the 2 coding systems, in order to ensure accurate clinical documentation, case reporting and database management. CMR staff have been trained and prepared for the new coding systems, including the

new diagnosis and procedure codes specific to congenital malformations. In addition, CMR staff organized a webinar training course about ICD-10-CM/PCS in 2012 for coders from CMR reporting hospitals.

Eleven respondents provided comments at the end of the survey and nearly one third of them expressed thanks and appreciation for CMR's leading effort on the ICD-10-CM/PCS transition project (data not shown). According to the findings from our survey, CMR staff will develop a detailed ICD-10-CD/PCS implementation plan with staff responsibilities including modification of the online reporting and database management systems, preparation of a list of reportable ICD-10-CM/PCS codes and narratives, development of a crosswalk for conversion, preparation of an updated CMR reporting manual and data directory, providing additional training webinars, development of strategies for pilot testing and final launch of the new system. As requested by the survey respondents, we will communicate, in a timely fashion, with the reporting hospitals to keep them updated with the project progress, provide them with newly developed documents, and inform them about pilot testing and the final date when the CMR is ready to accept ICD-10-CM/PCS codes.

One of the limitations of the current survey is the relatively low participation rate (60%). A higher response rate may have been achieved if telephone follow-ups of the non-respondents were conducted, but due to limited staff time and resources, this was not done. Furthermore, the survey has a limited number of questions and thus did not collect other useful information such as the concerns and barriers that the hospitals may have in implementing the ICD-10-CM/PCS. However, additional questions would have placed additional demands on respondents that may have resulted in a further reduction in participation rates. For the similar concern, this survey did not collect identifying information such as the name and institution of the respondents. Without the names of the responding hospitals, we were not able to conduct comparative analysis between the responding and non-responding hospitals and thus, could not estimate the impact of the selection bias, if any, on the survey results.

In conclusion, the current statewide survey among all CMR reporting hospitals assessed hospitals' readiness and needs regarding the transition from ICD-9-CM to ICD-10-CM/PCS. This information will help the CMR with appropriate planning for our own transition and enable us to meet the needs of hospital reporters.

References

1. Classification of Diseases, Functioning, and Disability. An online resource provided by the Centers for Disease Control and Prevention. Available at: **http://www.cdc.gov/nchs/icd.htm**.
2. International Classification of Diseases. An online resource provided by World Health Organization. Available at: **http://www.who.int/classifications/icd/en/**.
3. Chute CG, Huff SM, Ferguson JA, Walker JM, Halamka JD. There are important reasons for delaying implementation of the new ICD-10 coding. *Health Aff*. 2012;31:836-842.
4. Hazlewood A. *ICD-9-CM to ICD-10-CM: Implementation Issues and Challenges*. AHIMA's 75th Anniversary National Convention and Exhibit Proceedings. October 2003. Available at: **http://library.ahima.org/xpedio/groups/public/documents/ahima/bok3_005426.hcsp?dDocName=bok3_005426**.
5. Sanders TB, Bowens FM, Pierce W, Stasher-Booker B, Thompson EQ, Jones WA. The road to ICD-10-CM/PCS implementation: forecasting the transition for providers, payers, and other healthcare organizations. *Perspect Health Inf Manag*. 2012;9:1f.
6. Steindel SJ. International classification of diseases, 10th edition, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. *J Am Med Inform Assoc*. 2010;17:274-282.
7. Steen PM, Wang Y, Tao Z, Cross PK, Druschel CM. Implementing a web-based case reporting and communication system among hospitals reporting to the birth defects registry in New York State. *J Public Health Manag Pract*. 2008;14:E11-E16.
8. Wang Y, Cross PK, Steen PM, et al. Development of a Web-based Reporting, Management, and Communication System for the Statewide Birth Defects Registry in New York. *J Registry Manage*. 2007;34:45-52.

Original Article

Collaborating with Consumers: the Key to Achieving Statutory Notification for Birth Defects and Cerebral Palsy in Western Australia

Carol Bower, MBBS, MSc, PhD, FAFPHM, DLSHTM, FPHAA[a,b]; Anne McKenzie[b]; Linda Watson[a]; Adrian Charles, MA, MB, BChir, MD, MRCP (UK), MRCPCH, FRCPath, FRCPA[a,c]

Abstract: Introduction: The Western Australian Birth Defects Registry and the Western Australian Cerebral Palsy Register used multiple sources of voluntary notification without consent and have a high level of case ascertainment, but there were concerns over privacy and a call for statutory notification. Objective: To seek consumer consensus on whether notification to the registers should be statutory or only with consent. Methods: Two facilitated workshops for consumer and community members of groups representing people with birth defects, cerebral palsy and disability, and the Western Australian Health Consumers' Council. Results: Parent groups and the Health Consumers' Council were unanimous in their support for statutory notification, with 3 conditions: that comprehensive and open information be provided to consumer groups and community; that consumers have input into the development of statutory notification; and that an opt-out clause be included. A Consumer Reference Group was established. They decided on a name for the new register (Western Australian Register of Developmental Anomalies), developed an opt-out clause and reviewed drafts of the regulations for statutory notification. The regulations came into effect in January 2011. Conclusions: Consumers were key to achieving statutory notification. We encourage others to engage with their consumers and community in equal partnership for mutual benefit.

Key words: *consumers, birth defects, cerebral palsy, surveillance, register*

Introduction

The Western Australian Cerebral Palsy Register was established in 1977 and the Western Australian Birth Defects Registry in 1980 to obtain high-quality, complete, and population-based information on cerebral palsy and birth defects respectively in Western Australia and to use this information to: establish how often these conditions occur; conduct research into their causes and prevention; investigate changes in their frequency; evaluate screening, treatment and prevention interventions; assist in planning health care facilities; provide information to health professionals; and increase knowledge generally about them.[1,2]

Since their inception, notifications to the registers have come from several vital statistical sources (the Midwives' Notification of Birth forms, death registrations and the hospital morbidity system) and a large number of voluntary sources, including private practitioners, diagnostic, and treatment services. While some notifiers either informed parents or sought parental consent before notification, most cases were ascertained without consent. Validation studies have shown that case ascertainment by the registers is high[3-5] and the data from both registers have been used extensively for monitoring, research, and health service evaluation (for example[6-9]).

While notification of birth defects to registers in Victoria and New South Wales[10,11] and both birth defects and cerebral palsy in South Australia[12] is covered by legislation, this was not the case in Western Australia (WA) until 2011.

Other valuable data collections in Western Australia are statutory (for example, cancer) and on several occasions since establishment of the WA cerebral palsy and birth defects registers, statutory notification had been considered but rejected by the Department of Health, mainly because the registers were working well with a system of non-statutory notification. However, with growing community and medical practitioner concern, changes in attitudes towards consent and national privacy legislation, this situation was no longer deemed to be tenable. The National Privacy Principles (1988) in relation to use and disclosure require an organization not to disclose personal information about an individual for a secondary purpose (such as a notification to a register) unless the individual has consented to disclosure or, if the information is health information for research or public health, it is approved by the Privacy Commissioner, or it is required by law (**http://www.privacy.gov.au/materials/types/infosheets/view/6583**). Several medical practitioners had indicated that they would not continue to notify to the registers in the absence of statute and nor were they able to obtain individual consent, in some instances because they had no direct contact with the parent (for example, laboratory pathologists).

Thus, in 2004, a further request was made to the

[a]Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, Women and Newborn Health Service. [b]Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia. [c]Pathwest Laboratory Medicine WA.

Address correspondence to Professor Carol Bower, Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, Women and Newborn Health Service, PO Box 134, SUBIACO Western Australia 6904. Telephone: 08 9340 2721. Fax: 08 9340 2636.
Email: caroline.bower@health.wa.gov.au.

Minister for Health for notification to become statutory in WA. The Minister was supportive of the request, but wanted consumer consultation before proceeding.

Written advice was sought from the Health Consumers' Council in WA and from several community support groups for parents and a presentation was made to the Board of the Health Consumers' Council. While the parent groups were in favor of statutory notification, the Health Consumers' Council (HCC), as a rights-based organization, believed the registers should require consent.[13] When these findings were provided to the Minister, he replied that until consensus was reached, statutory notification would not proceed.

Methods and Results

In order to address this issue, we took 2 complementary approaches. A computer-assisted telephone (CATI) survey of a random sample of the adult population of WA was undertaken and has been reported separately.[14]

Secondly, we consulted with the Consumer Advocate (AM), who manages a joint consumer and community involvement program for researchers at the Telethon Institute for Child Health Research and the University of Western Australia School of Population Health.[15] On her advice, we conducted 2 workshops with consumer representatives of families with birth defects and cerebral palsy and the Health Consumers' Council, to seek consumer consensus on the preferred model for notification. We report here the conduct and outcomes of the 2 workshops and the steps taken following the workshops to achieve statutory notification.

Workshop 1

Support groups representing people with birth defects, cerebral palsy and disability, the WA Health Consumers' Council, the Genetic Support Council of WA, the Ministerial Advisory Council on Disability, the advisory committees of the 2 registers, register staff and Department of Health staff were invited to send representatives to participate in the workshop. Nineteen people attended. One of us (AM) facilitated the workshop and another (CB) presented information about the nature, functions and sources of notification of the registers and experience of other consent-based registers.[16-20] The results of the CATI survey[14] were also presented, which indicated limited knowledge of birth defects and the Western Australian Birth Defects Registry but general acceptance of statutory notification. Three possible models were described: the current model (non-statutory notification); the consent model (notification only with consent); and the statutory model (statutory notification without consent).

Following a question and answer session, the register and Department of Health staff left the room while the consumers discussed the options further. The consumers decided that they needed more information, wider consultation, and another workshop. They concluded that the current model should continue in the meantime and that, whatever the final model, parents and community must be better informed about the registers.

Workshop 2

For the second workshop, we engaged an external facilitator. The aim was to explore the issues further and reach a consensus for either statutory notification or notification with consent, using a method aligned with "dynamic facilitation" (**http://www.iaf-methods.org/methods**). Dynamic Facilitation is a form of facilitating where people address difficult issues creatively and collaboratively, through a process of talking and thinking that builds mutual respect, trust, and the sense of community. The dynamic facilitator helps foster shifts of heart and mind by following the natural flow of conversation and supporting group spontaneity, with the aim of reaching consensus.

Invitations to attend were extended to all support groups invited to the first workshop and to others nominated by them. There were 10 attendees, and several potential participants who were unable to attend provided out-of-session input. No register or Department of Health staff attended, apart from the medical officer of the Birth Defects Registry (CB). Neither the medical officer nor the consumer advocate contributed to any of the discussions except to clarify matters of fact. It was made clear to the participants that, should they reach a consensus decision, that consensus decision would be the model under which the registers would function in the future.

The facilitator asked participants to list the issues they wanted to cover in the workshop. These included: access to the data collected; how to do the best for parents of children with birth defects or cerebral palsy; how to inform parents and the community that the registers exist; the best model to maximize research outcomes; how to ensure the community embraces the need for a register; how much information is retained and how individuals can opt out; how to reach everyone to obtain accurate information; the vulnerability of new mothers when gathering information; who is responsible for notification; and issues of consent.

Following this, 2 presentations were made. The first, by the Birth Defects Registry medical officer, was a repeat of that given at the first workshop. The second, by the executive director of the Health Consumers' Council was about the council's role as an independent community-based organization, representing the consumers' "voice" in health policy, planning, research and service delivery. Community members with whom the council had spoken were surprised that the Department of Health holds such a large amount of health data without their knowledge or consent. The council was of the view that consent should be sought and given before inclusion of data on individuals on the registers.

Participants then considered the best and worst aspects of the current register model. The best was thought to be the high quality of the data collected and its availability for planning and improving facilities, public health, research and prevention, and the worst was the low community awareness of the registers.

In their final consideration, participants reached a unanimous decision in favor of statutory notification, with 3 conditions: that comprehensive and open information

Figure 1. Timeline from commencement of WA Cerebral Palsy and Birth Defects Registers to Enactment of Statutory Notification

be provided to consumer groups and community; that consumers have input into the development of statutory notification; and that an opt-out clause be included. All groups involved expressed appreciation of being able to contribute. The executive director of the Health Consumers' Council noted that the consultation process was an example of good practice and was the first time that the community had been invited to give input into the reporting process of a register in Western Australia.

Next Steps

The Minister for Health was informed of the consumer and community consensus for statutory notification, following which he consented to statutory notification proceeding and the Legal and Legislative Services in the Department of Health began drafting regulations to combine cerebral palsy and birth defects notification in 1 statutory register.

Several attendees at the workshops volunteered to be members of a consumer reference group (CRG) for the new register, to ensure that the conditions under which statutory notification had been agreed to by the consumers were met.

The CRG devised an opt-out clause for the regulations, allowing parents/guardians (or affected adult individuals) to have identifying information (names, address) removed from the register 6 years or more after first notification,

while retaining the diagnostic information. The CRG also requested that the new register be called the Western Australian Register of Developmental Anomalies (WARDA), to better reflect the 2 groups of conditions to be included (birth defects and cerebral palsy). Both the name change and the opt-out clause were included in the regulations. The consumer reference group reviewed and revised the draft regulations on several occasions and the regulations were made law in January 2011.

To increase the visibility of the WARDA, the CRG assisted with the development of a Web site[21] and a brochure for the new register. The CRG constructed a list of 59 consumer groups to whom the information about the WARDA and a supply of brochures were sent. A copy of the regulations is available on the Web site.[21]

The CRG continues to support the WARDA, with updates for the Web site and presenting at community forums about the WARDA. They also meet with prospective WARDA researchers to advise on matters such as approaches to parents, questions important to parents and how to ask them. The CRG requests that all researchers using WARDA data provide a lay summary of their proposal, which will be made publicly available on the WARDA Web site.

The staff of the WARDA attended a workshop on involving consumers in research, through the Consumer Participation Program at the Telethon Institute for Child Health Research[15] and the process undertaken by the register has been included as a good practice example of consumer and community involvement in a resource manual for researchers.[22]

Discussion

Through a process of consultation and discussion, consumer and community participants reached consensus for statutory notification of cerebral palsy and birth defects in Western Australia.

This conclusion is in agreement with the results of the CATI survey of a sample of 600 randomly selected WA adults conducted in WA during 2006. Respondents were asked their views about the statutory collection of identifiable data by the WA Birth Defects Registry and the extent to which the use of the data was perceived to be an invasion of privacy.[14] In that survey, 96% felt the data collected was useful information for Western Australia and 79% supported a new law for mandatory notification.

Since their inception, both the birth defects and cerebral palsy registers have had links with support groups through presentations, collaboration in research[23] and health promotion (eg, promoting folic acid supplements for the prevention of neural tube defects), by providing register data to lay support groups to assist them in seeking services for children with birth defects and cerebral palsy and, like many registers, including consumer representatives on the registers' advisory committees. However, there had not previously been such an engagement of several groups simultaneously, as reported here, to consider how the registers operate.

How to involve consumers and the effects of such involvement in developing healthcare policy and research, remain largely unevaluated and, in the absence of evaluated strategies, Nilsen et al suggest relying on advice based on practical experience and common sense.[24] This is what we did, seeking the advice of the consumer advocate, the experience of the Health Consumers' Council and through the engagement of a professional facilitator.

There is now a considerable body of evidence documenting problems with consent-based registers or surveillance systems, the major ones being incomplete and biased registration.[16-20] Incomplete registration for birth defects registers is clearly illustrated in a survey of European congenital anomaly registers.[20] Of 35 EUROCAT registries surveyed, 29 responded, 8 of them requiring opt-in informed consent. The experience of these 8 registers showed that informed consent is a serious threat to a high level of case ascertainment, largely because of logistical problems for busy clinicians to seek consent and multiple approaches to parents. Participants in our workshop also raised concerns about the appropriate time to seek consent. They felt that a parent who is emotionally distressed (as many will be when their child is first diagnosed) could be spared having to make a decision about consent if it was not a requirement.

We made it clear to the consumers and community members that, given the registers were for the benefit of Western Australian families, the decision about how the registers are run should be made by the people most concerned about children with birth defects and cerebral palsy—their parents. Ultimately, we believe this is what convinced the Health Consumers' Council and others to reach consensus in favor of a statutory register.

Although consumers in our workshops acknowledged the high case ascertainment, accuracy and independence of the registers in their current state and their value as a public health tool, they placed a high priority on the community in general and parents in particular needing to be better informed about the registers. Supporting this call from consumers are the results of earlier studies in Western Australia that have shown that the majority of participants in 2 community surveys thought birth defects only affected 1 in 200 births or less.[25,26] Furthermore, in the CATI survey, only 6% of respondents were aware that there was a birth defects register in Western Australia.[14]

Of course, consumer involvement is not limited to registers—it is also important in health service delivery and research. Furthermore, the growth of genomic analysis that will likely affect not just individuals but families and communities, makes such involvement essential to ensure that the consumer and community voice is heard and heeded.

Conclusion

The consumer and community members were key to achieving statutory notification for the WARDA. They understood clearly that they were empowered to make the decision between a consent-based and a statutory register. We trusted that they would make the right decision for Western Australia and they trusted that we would honor that decision. Consumers understand research and its

benefits, and researchers need to understand the value of involving consumers in their research. We encourage others to engage with their consumers and community in equal partnership for mutual benefit.

Acknowledgements

We thank the participants in both workshops, the consumer groups who provided additional comment and Jillian Mercer, Mercer Management.

We are grateful to the contributions of the staff of WARDA (Edwina Rudy, Ann Callaghan, Jennifer Quick, and Alison Rowley) and the members of the WARDA Consumer Reference Group (Rachel Skoss [Chair], Belinda Frank, Michèle Kosky, Wendy Langford, Anne McKenzie, Charlie Rook, Beth Stein).

References

1. Bower C, Ryan A, Rudy E, Cosgrove P. *Report of the Birth Defects Registry of Western Australia 1980-2003*. Report Number 11. Subiaco: King Edward Memorial Hospital, Centre for Women's Health; 2004.
2. Watson L, Blair E, Stanley FJ. *Report of the Western Australian Cerebral Palsy Register to Birth Year 1999*. Perth: Telethon Institute for Child Health Research; 2006.
3. Bower C, et al. Ascertainment of pregnancies terminated because of birth defects: effect on completeness of adding a new source of data [erratum]. *Teratology*. 2001;63:164.
4. Bower C, Forbes R, Ryan A, Rudy E. Validation studies from the Western Australian Congenital Malformations Registry. *Community Health Stud.* 1990;14:274-277.
5. Bower C, Ryan A, Rudy E. Ascertainment of pregnancies terminated because of birth defects: effect on completeness of adding a new source of data. *Teratology*. 2001;63:23-5 [erratum appears in *Teratology* 2001;63:164].
6. Blair EM, Stanley FJ. Intrapartum asphyxia a rare cause of cerebral palsy. *J Paediatr.* 1998;112:515-518.
7. Blair EM, Watson L, Stanley FJ. Life expectancy among people with cerebral palsy in Western Australia. *Develop Med Child Neurol.* 2001;43:508-515.
8. O'Leary P, Breheny N, Dickinson J, et al. First trimester combined screening for Down syndrome and other fetal anomalies in Western Australia. *Obstet Gynecol.* 2006;107:869-876.
9. Hansen M, Kurinczuk JJ, Bower C, Webb S. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *New Engl J Med.* 2002;346:725-730.
10. Centre for Epidemiology and Research, NSW Department of Health. New South Wales Mothers and Babies, 2003. *NSW Public Health Bull.* 2004;15(s-5):79.
11. Riley M, Halliday J. *Birth Defects in Victoria 2003-2004*. Melbourne: Victorian Perinatal Data Collection Unit, Victorian Department of Human Services; 2006.
12. South Australian Birth Defects Register. *2003 Annual Report of the South Australian Birth Defects Register.* Adelaide: Women's and Children's Hospital; 2005.
13. Health Consumers' Council (Inc). Proposed changes to arrangement for the Birth Defects Registry. *Health Matters* [Newsletter of Health Consumers' Council WA (Inc)]. 2005;60:17-18.
14. Molster C, Bower C, O'Leary P. Community attitudes to the collection and use of identifiable data for health research—is it an invasion of privacy? *Aust NZ J Public Health*. 2007;31:313-317.
15. Community Participation Program, Telethon Institute for Child Health Research. Available at: **http://www.childhealthresearch.org.au/community-engagement/community-participation-program.aspx**. Accessed January 24, 2013.
16. Tu JV, Willison DJ, Silver FL, et al. Impracticability of Informed Consent in the Registry of the Canadian Stroke Network. *N Engl J Med.* 2004;350:1414-1421.
17. Verity C, Nicoll A. Consent, confidentiality, and the threat to public health surveillance. *BMJ*. 2002;324:1210-1213.
18. Al-Shahi R, Vousden C, Warlow C. Bias from requiring explicit consent from all participants in observational research: prospective, population based study. *BMJ*. 2005;331:942.
19. Williamson OD, Cameron PA, McNeil JJ. Medical registry governance and patient privacy. *Med J Aust.* 2004;181(3):125-126.
20. Busby A, Ritvanen A, Dolk H, et al. Survey of informed consent for registration of congenital anomalies in Europe. *Br Med J.* 2005;331:140-141.
21. Western Australian Register of Developmental Anomalies. Available at: **http://kemh.health.wa.gov.au/services/register_developmental_anomalies**. Accessed January 24, 2013.
22. McKenzie A., Hanley R. *Consumer and Community Participation in Health and Medical Research: A practical guide for health and medical research organizations*. 2nd printing. 2007.
23. Bower C, Stanley F, Croft M, De Klerk N. Absorption pteroylpolyglutamates in mothers of infants with neural tube defects. *Br J Nutr.* 1993;69:827-834.
24. Nilsen ES, Myrhaug HT, Johansen M, Oliver C, Oxman AD. *Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material (Review)*. The Cochrane Library. 2007(2).
25. Bower C. What do we know about birth defects? In: Durston B, Ghisalberti A, Moore D, eds. *Genetics: the Facts, Fiction, Issues and Ethics—Proceedings of the 1998 Lecture Series*. Perth: Human Genetics Society of Australasia, Western Australian Branch; 1999:21-28.
26. Bower C, Blum L, Grace L. What do women of child-bearing age know about birth defects and folate? *Health Promot J Aust.* 1992;2:53-54.

Original Article

Creation and Evaluation of a Multi-layered Maternal and Child Health Database for Comparative Effectiveness Research

Jason L. Salemi, MPH[a]; Jean Paul Tanner, MPH[b]; Marie Bailey, MA, MSW[c]; Alfred K. Mbah, PhD[a]; Hamisu M. Salihu, MD, PhD[a]

Abstract: Background: As high-speed computers and sophisticated software packages for data linkage become increasingly available, investigators from nearly every arena are creating massive databases for epidemiologic and comparative effectiveness research (CER). Decisions made during database construction have a major impact on the accuracy and completeness of the data. Considering their potential use in informing health-care decisions, it is vital that we increase transparency of these data, including a thorough understanding of the record linkage strategy implemented and an evaluation of linked and unlinked records so that potential biases can be addressed. Methods: Our target population included infants born to Florida-resident women from January 1, 1998 through December 31, 2009 with a valid birth certificate record. We used a stepwise deterministic record linkage strategy to link to any and all inpatient, ambulatory, and emergency department hospital visits from birth through December 31, 2010, and to identify deaths that occurred within the first year of life. Thus, each infant was followed up for at least 1 year after birth or until death, up to a maximum of 13 years. We investigated linkage rates and associations between *linked status* (linked vs unlinked) and a host of maternal and infant demographic and reproductive characteristics, all extracted from the birth certificate files. Bivariate county-level maps were created to describe the impact of both maternal race/ethnicity and maternal nativity on the geographic variation in linkage rates. Results: During the 13-year study period, there were 2,549,738 birth certificate records for infants born alive to Florida-resident women, and with no indication of an adoption. We were able to link 2,347,738 (92.1%) birth certificate records to an infant birth hospitalization record. The highest crude unlinked rates were seen among infants who died during their first year of life (35.9%), births in which the documented principal source of payment was "self-pay" (28.1%), and infants born to mothers with less than a ninth-grade education (26.0%), who were foreign-born (12.9%), and who self-identified as Hispanic (12.8%). After adjusting for other related and potentially confounding variables, several of these infant and maternal characteristics were associated with increased odds of failure to link infant birth records. Conclusion: Using a stepwise deterministic linkage approach, we achieved a high linkage rate of several data sources, and produced a reliable, multipurpose database that can be used for observational, comparative effectiveness, and health services research in maternal and child health (MCH) populations. Our findings underscore the importance of evaluating routinely collected health data and increasing clarity regarding the strengths and limitations of linked electronic data sources. The resultant database will be of immense utility to researchers, health planners, and policy makers as well as other stakeholders interested in MCH outcome studies.

Key words: *record linkage, vital statistics, hospital discharge data, unlinked records, maternal and child health, comparative effectiveness research*

Introduction

Despite the United States' immense wealth and its per capita spending on health care, Americans die and suffer from illness and injury at rates that far exceed those in other developed countries.[1] In need of national health reform, the US government allocated $1.1 billion for comparative effectiveness research (CER) through the American Recovery and Reinvestment Act (ARRA). CER is a rigorous scientific approach that seeks to inform decisions about health policy and clinical care through the acquisition, utilization, and synthesis of data and evidence from a spectrum of epidemiologic study designs.[2] Although randomized clinical trials (RCTs) are considered the gold standard, methodologically-speaking, RCTs are often expensive and limited in practical applicability and generalizability. Thus, CER based on "real-world" observational data is rapidly becoming a tool of choice, particularly with the increased existence and availability of rich administrative and clinical

[a]The MCH Comparative Effectiveness Research Group, Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL. [b]Birth Defects Surveillance Program, Department of Community and Family Health, College of Public Health, University of South Florida, Tampa, FL. [c]Florida Department of Health, Office of Health Statistics and Assessment, Bureau of Community Health Assessment, Tallahassee, FL.

Address correspondence to Jason L. Salemi, MPH, Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, 13201 Bruce B Downs, MDC56, Tampa, FL 33612-3805. Telephone: (813) 974-7716. Email: jsalemi@health.usf.edu.

This project was supported by grant number R01HS019997 from the Agency for Healthcare Research and Quality (AHRQ). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality, or the University of South Florida.

databases.[3,4] Through record linkage, detailed, individual-level sociodemographic, health, and geospatial information from large samples can be combined over many years to establish population-based disease registries and conduct longitudinal analyses that would otherwise not have been possible.[3,5-8]

As high-speed computers and sophisticated software packages for data linkage become increasingly available, investigators from nearly every arena are creating massive databases for research.[5] Decisions made during database construction, including the use of deterministic vs probabilistic linkage algorithms, the linking variables selected, and the threshold levels chosen that define a link/non-link, have a major impact on the accuracy and completeness of the data. Considering their potential use in informing health-care decisions, it is vital that we increase transparency of these data, including a thorough understanding of the record linkage strategy implemented and an evaluation of linked and unlinked records so that potential biases can be addressed.[5,9]

In Florida, the state department of health and university partners have collaborated for many years to link vital records and hospital discharge data together for various purposes, including the establishment of the population-based Florida Birth Defects Registry (FBDR).[10] However, funding restrictions limited the sophistication of the linkage algorithms, breadth of data linked, and ability to evaluate linkage results. In September 2010, the University of South Florida (USF) was awarded ARRA funding through the Agency for Healthcare Research and Quality (AHRQ) to improve its statewide, hospital-based encounter-level data for the purpose of producing an accurate and reliable evidence base for CER in maternal and child health.[11] By adding new data sources and incorporating a refined approach to record linkage, the 3-year project will also help to improve existing public health surveillance systems in Florida. The primary aims of this paper are to: 1) describe, in detail, the creation of a clinically enhanced multipurpose maternal and child health administrative dataset; 2) evaluate the dataset by comparing maternal and infant characteristics among linked and unlinked records; and 3) discuss the potential biases that need to be considered if using the database for CER or other public health surveillance/research purposes.

Methods

Data Sources

Vital statistics data. Using standard forms recommended by the US Public Health Conference on Records and Statistics, the Florida Office of Vital Statistics (VS) compiles records and reports of live births, deaths, and fetal deaths for "resident" events (occurs in a Florida resident regardless of place of occurrence) and "recorded" events (occurs in Florida even if Florida is not the usual place of residence).[12] Vital records are completed/filed by physicians, midwives, or funeral directors, then submitted to local registrars who forward them to VS for incorporation into a statewide database.[12] The birth certificate data contain pregnancy and delivery information on infants (eg, birth weight, gestational age, Apgar scores, abnormal conditions, congenital malformations) and their birth mothers (eg, pregnancy history, obstetric procedures, morbidities, tobacco and alcohol use during pregnancy). Death and fetal death data include information on the occurrence, timing, and location of death, as well as the underlying and contributing causes of death (up to 20) identified using International Classification of Disease, Clinical Modification, 9th or 10th Edition (ICD-9-CM, ICD-10-CM) codes. All VS data contain sociodemographic characteristics and detailed personal identifiers for the infant, mother, and father. In response to the National Center for Health Statistics' 2003 revisions of the birth, death, and fetal death standard certificates, VS modified the birth (in 2004), death (in 2005), and fetal death (in 2006) certificates, primarily to elicit more specific responses (eg, can select multiple races), and collect new data elements (eg, maternal pre-pregnancy height and weight, maternal infections, principal source of payment for delivery).[13] Each VS database is person-level and uniquely identified by an administrative "state file number" (SFN) that is specific to that database (ie, if an infant has both a birth and death record, the SFN would be different in each file).

Hospital discharge data. In Florida, the Agency for Health Care Administration (AHCA), the chief health policy and planning entity for the state, collects discharge data from a wide range of health-care facilities in accordance with Florida statutes and the Florida Administrative Code. State law requires that all facilities (excluding military, state, and federally-operated hospitals) submit data on all civilian hospital discharge records to AHCA on a quarterly basis. Hospital inpatient (HIP) data, which include all acute, intensive care, and psychiatric live discharges including newborn live discharges and deaths, are collected from acute care hospitals, short and long-term psychiatric facilities, and comprehensive rehabilitation facilities.[14,15] AHCA also collects ambulatory data from freestanding ambulatory surgical centers, radiation therapy centers, lithotripsy centers, cardiac catheterization labs, and short-term acute care hospitals. These data include events that are primarily surgical in nature or that involve specific invasive diagnostic procedures.[14] Lastly, beginning in 2005, AHCA has collected emergency department (ED) data, which captures visits in which there is an ED registration but the patient is not admitted for inpatient care.[14] Collectively, the hospital discharge data contain patient demographics, facility and payer information, specific clinical diagnoses (up to 34) and medical procedures performed (represented primarily by ICD-9-CM codes), and detailed hospital charges for specific revenue code groups. Although no names are available, personal identifiers including dates of birth (DOB) and Social Security numbers (SSNs) are present. In addition, HIP records for patients under 2 years of age contain an extra field for their birth mother's SSN, which proves critical for maternal and child health data linkages. These data are collected at the discharge level, with the potential for multiple discharges per person. A system-generated ID number serves as an individual discharge record's unique

identifier, and multiple records for the same person will not possess the same ID number.

Target Population

Our target population included infants born to Florida-resident women from January 1, 1998 through December 31, 2009 with a valid birth certificate record. We used record linkage to compile any and all HIP, ambulatory, and ED hospital visits through December 31, 2010, and to identify deaths that occurred within the first year of life. Thus, each infant would be followed up for at least 1 year after birth or until death, up to a maximum of 13 years (for an infant born in 1998). We excluded infants whose birth certificate record indicated an adoption (0.3%). Since all VS records are legal, administrative documents, a court order can mandate changes to a record at any time. For adoptions, sociodemographic information on the biological parents is replaced with that of the adoptive parents, and confidentiality safeguards preclude the inclusion of these records.

Maternal hospital discharge data and maternal death records from January 1, 1998 through December 31, 2010 were also linked to each infant included in the study. To allow for examination of events that occurred before conception, during pregnancy, and after birth, no restrictions were made as to the timing of events that were compiled and linked. For an infant born in 2005, we would still attempt to link to any and all hospitalization records for the infant's mother during the entire 13-year study period.

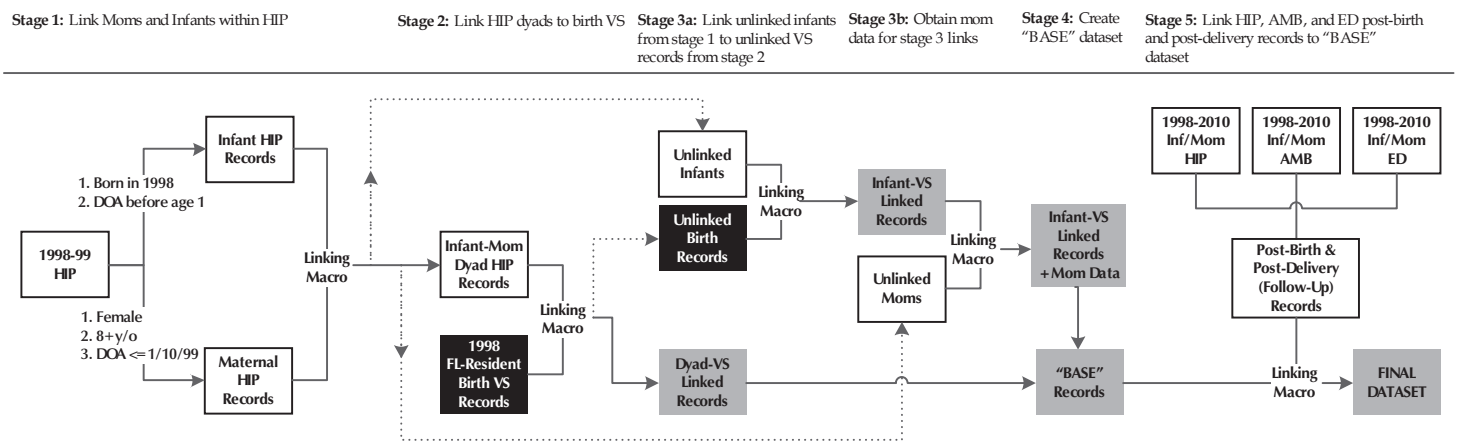
Record Linkage

Overarching approach to linkage. The primary goal of this project was to link birth certificate records for our target population to their HIP birth record, or if this could not be established, a HIP record that was closest to birth. This is what we refer to as the "base linkage," since it establishes the pool of linked infants that will comprise the clinically-enhanced dataset, and to whom we will link all subsequent infant and maternal data. The creation of the base dataset occurs in 4 stages that are depicted in Figure 1.

We do not attempt to link birth certificate records directly to HIP records using infant data alone. The variables common to these 2 datasets that could be used in the linkage are: 1) infant DOB, 2) infant SSN, 3) facility of care/birth, 4) ZIP code of residence, 5) county of residence, 6) plurality, and 7) INFANTLINK. For patients under 2 years of age, AHCA includes the INFANTLINK variable in the HIP data, which is intended to represent the birth mother's SSN. However, it is important to note that despite the availability of some personal identifiers, the infant's SSN is missing in ~93% of HIP records in an infant's first year of life. Moreover, approximately 15% of infant HIP records are missing the INFANTLINK variable, and others suffer from typos, transpositions, and misinformation that reduce the usefulness of the variable. Thus, our first step in data linkage is to link infant birth and maternal delivery HIP records together, creating a maternal-infant dyad. Doing so allows us access to a variable from the maternal HIP record that is critical for subsequent linkage to birth certificate records—the maternal DOB. When the maternal DOB is used in combination with the infant's DOB and facility of birth, it forms a powerful group of identifiers that can be used in record linkage, even in the absence of valid SSNs.

In stage 2, maternal-infant dyads from stage 1 are linked to infant birth certificate records. Other infants from stage 1 could not be matched to their birth mothers within the HIP database; however, in stage 3a we attempt to link them directly to their birth certificate records, without the added benefit of having mother's DOB available. Since a successful link in stage 3a adds maternal information from the birth certificate to information from the infant's

Figure 1. Overarching Approach to Linking Infant Birth Certificate Records to Infant and Maternal Hospital Discharge Records*



HIP=hospital inpatient database. AMD=ambulatory database. ED=emergency department database. VS=vital statistics. FL=Florida. Inf=Infant.
Solid line=path of processed or linked records. Dotted line=path of unlinked records.
White=dataset including only hospital discharge records. Black=dataset including only birth vital records. Gray=dataset including linked records.
*This diagram illustrates an attempt to link the 1998 birth cohort as an example. This process would have been completed for the 1998-2009 birth cohorts.

HIP birth record, we make another attempt to link to the maternal delivery HIP record (stage 3b). In stage 4, we combine valid links from stages 2 and 3, which constitutes the "base" dataset for a given birth cohort. The resultant dataset is comprised of mostly birth certificate records linked to both infant birth hospitalizations and maternal delivery hospitalizations (see Results section), which is ideal for research that seeks to establish the birth/delivery baseline. However, we do not exclude the small proportion of records in which the linked infant HIP record is one that occurs shortly after birth, or in which a maternal delivery HIP record could not be linked. Their inclusion allows the final dataset to be multipurpose in nature, since surveillance activities (eg, birth defects registry) are interested in capturing cases over a specified time frame, regardless of whether the condition was present and diagnosed at birth.

The secondary goal of this project was to expand the base file into a longitudinal dataset that would permit tracking of maternal and clinical events over time. Thus, in stage 5, the base dataset from stage 4 is linked to hospital HIP, ambulatory, and ED records throughout the study period. This process (stages 1-5) is repeated for each birth cohort and compiled to create the final 1998-2009 clinically-enhanced maternal-infant database.

Stepwise deterministic data linkage strategy. For each of the stages described in the previous section, we linked data using a *modified* stepwise deterministic record linkage strategy. Despite the growing popularity of probabilistic record linkage,[16-21] we wanted the ability to incorporate our years of experience and detailed knowledge regarding the quality of linking variables in component datasets. Such information tends to be automatically generated by a probabilistic strategy.[21] Moreover, considering the intended use of the final database in research and surveillance activities, our stakeholder group sought to minimize false positive links, something more easily controlled using a predominantly deterministic approach.[21,22]

A general overview of stepwise deterministic record linkage has been described elsewhere.[22,23] Using knowledge about the quality and availability of linking variables, a *hierarchical* series of linking steps is constructed, with each step consisting of a group of identifiers upon which records will be linked. In order to be considered a link, 2 records must exhibit exact agreement on all identifiers listed in a particular step. After each step, linked records are removed and the unlinked records from both datasets proceed to the next step, in which a different set of identifiers are specified. However, in some cases it is not prudent to remove linked records from both datasets prior to a subsequent step. For example, when linking maternal and infant HIP records, we expect that a single maternal delivery record may link to multiple infant birth records (eg, in the case of a twin birth). Furthermore, due to differing availability and quality of personal identifiers, both infants' records may not link to the maternal record within the same step. In this case, removing linked records from both datasets would decrease the sensitivity of the linking algorithm. Thus, we specify 1 dataset in which records are not removed after each step, and 1 in which records are removed.

Our approach is hierarchical because the ordering of steps is extremely important, with higher confidence steps (eg, exact match on maternal SSN and DOB, infant DOB, infant sex, facility of birth) preceding lesser confidence steps (eg, infant DOB, facility of birth, ZIP code). This specification requires expertise in working with the datasets and variables being linked.

Although deterministic linkage is often considered "all-or-none" because records must match on the entire set of identifiers to be linked during any given step, *partial* and *crossover* agreement is easily incorporated into the strategy. Examples of partial agreement in our algorithm include various date field combinations (eg, dates are within 2 days, or match on month and year only), and 8/9- or 7/9-digit agreement on SSNs in order to accommodate simple typos and transpositions. Crossover agreement is very important to the sensitivity of our linkage strategy, and was a key factor in our choice of a flexible software program for conducting linkages. A crossover link is one in which a match is desired not only on variables designed to capture the same information in 2 datasets, but also on variables originally intended to capture different information. As mentioned previously, HIP records for patients 2 years of age and under have an important variable that is meant to be the birth mother's SSN. However, it is not uncommon for the father's SSN to be documented in the record. Similarly, in hospital discharge data and VS data, SSNs for the infant, mother, and father may be switched. Thus, in specifying a series of linking steps, it is prudent to incorporate crossovers (eg, mother's SSN in AHCA links to father's SSN in the birth certificate record) into the algorithm. By default, missing values in 2 records for a particular variable will be treated as agreements. Since this greatly decreases the specificity of the link, missing values for variables in both datasets are recoded to values unique to each dataset to prevent linking on missingness. It is important that the selected replacement values not occur as real values in each dataset. For example, all missing character data in the birth certificate record was recoded to a caret (^), whereas missing character data in HIP records was recoded to a hashtag (#). A similar process was implemented for numeric variables and date fields.

Following a complete sequence of linking steps, a set of matched records is created. One-to-one matches (a single record from dataset A links to only 1 record from dataset B, and vice versa) are accepted, whereas one-to-many (a single record from dataset A links to multiple records from dataset B) are further scrutinized. In some instances, both records are validated as positive matches (eg, twins linking to the same maternal record). In other cases (eg, 2 different infant records link to the same maternal record), a series of decision rules is designed to select the "best" link.

Linking multiple births and records without a valid SSN. Our previous attempt to link these datasets focused on singletons with a valid infant or maternal SSN, which precluded research on multiple births and disproportionately eliminated populations more likely to be missing SSNs (eg, infants of foreign-born, underemployed, and undereducated parents). As part of our current strategy, we developed an algorithm for automating the process of

linking multiples births. We used a combination of variables capable of differentiating between multiples: 1) infant's SSN, 2) infant sex, and 3) specific ICD-9-CM codes (764, 765.0, and 765.1) whose fifth digit can classify birth weight into 250-gram categories. Birth order is not available in the AHCA data. The algorithm creates a score (0 to 100) that reflects our confidence that we have a link between a birth certificate and HIP record belonging to the same multiple. The algorithm awards 40 points for infant SSN agreement, 30 for agreement on sex, and up to 30 points for agreement on birth-weight category with more points awarded for higher concordance. If a birth certificate record pairs with 1 HIP record with a higher score than other HIP records for a set of multiples, the link is accepted and deemed the "best match." However, if VS and AHCA data for a set of multiples lack infant SSNs, shares the same sex, and has either missing or identical birth-weight categories, we do not want to lose the ability to link these records. Instead, each HIP birth record is matched to a birth certificate record based on the ordering of administrative identifiers in each dataset. For example, for a set of same-sex, similar birth weight twins, the first-ordered birth certificate number would get linked to the first-ordered record number in the birth hospitalization data. The assumption is that, more often than not, the multiple that is born first will be assigned the first birth certificate number and be documented first in the hospital discharge data. Although it is not a documented practice, data experts on our team believe this approach to be more accurate than pure random assignment of multiples' HIP records to birth certificate records.

For non SSN-based linkages, we relied on a strategy that requires a "one-to-one" linkage of first moms and babies within hospital-based records, and subsequently these dyads to birth vital records. Maternal delivery and infant birth HIP records were linked on facility of birth, county and ZIP code of residence, self-identified race, and proximity of infant's date of birth with mother's date of admission. Since one-to-many or many-to-one linkages might be indicative of uncertainty in which mother "belonged" to each infant, we accepted cases in which only 1 maternal record linked with only 1 infant record. We did allow for a maternal record to link to more than 1 infant record if both records indicated a multiple plurality. Despite the one-to-one linkage with the HIP data, the lack of a unique identifier conferred a level of uncertainty on the dyad pair. However, the creation of this dyad placed maternal and infant information together. Thus, the second stage of the non SSN-based linkage was to link the dyad HIP records to a birth certificate record using link steps that included *both* maternal and infant information (primarily infant's and mother's DOB), in addition to other variables.

Software and efficiency. After evaluating a myriad of software tools available to conduct deterministic and/or probabilistic record linkage, we opted to design our own SAS macro to implement the stepwise deterministic algorithm, and additional SAS code to guide post-linkage processing rules. Albeit more labor intensive to code, SAS readily handles extremely large datasets (our linkage includes millions of records in each dataset), and confers a high degree of customizability and control. The code was written to minimize, if not eliminate the need for manual review, which was important considering the size of the linked dataset. Thus, minimal programming changes must be introduced to accommodate new data from AHCA or VS.

Comparing Linked and Unlinked Records

Variables. The primary outcome in this analysis was the *linked status* of each infant's birth certificate record, which was dichotomized into *linked* and *unlinked* categories. To be classified as *linked*, the birth certificate record must have been matched to either a birth hospitalization record or another 1 of the infant's HIP records in the first year of life (although it is important note that, for over 99% of records, the established link was to a birth record).

We investigated bivariate associations between *linked status* and a host of maternal and infant demographic and reproductive characteristics, all extracted from the birth certificate files. We determined maternal race/ethnicity based on maternal self-report and first grouped women by ethnicity (Hispanic or non-Hispanic [NH]), with the NH group further subdivided by race (NH-white, NH-black, or other). Maternal nativity was dichotomized as US-born or foreign-born (ie, born outside the 50 United States). We categorized maternal age in years as <20, 20-29, 30-39, and ≥40, marital status as married or unmarried, and maternal education as <9 years, 9-11 years, 12 years, or >12 years. Method of delivery was either vaginal or C-section, parity was grouped into nulliparous, primiparous, and multiparous categories, and adequacy of prenatal care was determined using the revised graduated index algorithm,[24] which considers several factors: the trimester prenatal care began, the total number of prenatal visits reported, and the infant's gestational age. Women were classified as receiving intensive, adequate, intermediate, inadequate, or no levels of prenatal care. Maternal diabetes, hypertension, and tobacco and alcohol use during pregnancy were dichotomized as "yes/no" and identified using check boxes on the birth certificate. We also considered a few variables new to the birth certificate in 2004. Pre-pregnancy weight and height were collected, used to calculate a pre-pregnancy body mass index (BMI), and categorized into underweight (BMI<18.5), normal (18.5≤BMI<25), overweight (25≤BMI<30), and obese classes I (30≤BMI<35), II (35≤BMI<40), and III (BMI≥40). Principal source of payment was listed as private insurance, Medicaid, self-pay, and other. We also documented as "yes/no" whether mother or infant were documented as being transferred during the birth admission, and whether the pregnancy resulted from infertility treatment(s).

We designated plurality as singleton, twin, or higher-order multiples. Gestational age was categorized in weeks as very preterm (20-32), moderately preterm (33-36), term (37-42), and post-term (>42) based on the clinical estimate (CE) of gestation. When the CE was missing, we substituted gestational age calculated from the mother's date of last menstrual period. We grouped birth weight in grams as very low (125-1,500), low (1,500-2,499), and normal (>2,500). Lastly, we used linked death certificate information to document the occurrence and timing of infant death: no death,

same day as birth, other early neonatal (1-6 days), late neonatal (7-27 days), and postneonatal (28-364 days).

Statistical analyses. Descriptive statistics including frequencies and percentages were used to describe the distribution of linked and unlinked records by maternal and infant demographic and perinatal characteristics, using Wald chi-square tests of independence to test for differences. Before comparing linkage rates or constructing statistical models, we removed birth certificate records that we would not have expected to have a birth hospitalization record in AHCA, based on current data collection protocols and covered facilities. Thus, we excluded infants whose birth certificate indicated a birth taking place outside Florida, in a military, state, or federal hospital, or outside of the hospital (eg, birthing center, at home). Although linkage to AHCA data was still attempted, the reporting parameters for the HIP data gave these records very little chance to be linked, as reflected by their 9.7% linkage rate. After these exclusions, we reported the proportion of records that were missed (ie, the unlinked rate) by maternal and infant characteristics.

Unconditional logistic regression was used to estimate the adjusted odds ratios (AORs) and 95% confidence intervals (CIs) for factors significantly associated with our ability to link an infant's birth certificate record to its birth hospitalization record. In all models, the dependent variable was the *failure* to link the birth record. In addition to an unadjusted model, we constructed 2 multivariable models. In the first model, we included birth certificate records from the entire study period 1998-2009, considering only variables present in both versions of the live birth certificate. The second model incorporated additional variables introduced in the 2004 version of the birth certificate, but necessarily restricted to birth records in which the new certificate was used. Covariates considered for model inclusion were identified primarily by a literature review, expert opinion, and empirical univariate and bivariate analyses. Final models were computed using backward elimination of covariates with a significance level of 0.05. Both forward and stepwise selection procedures resulted in the same model specification. All statistical tests were 2-sided and declared significant at $p<0.05$. We performed all statistical modeling using SAS version 9.2 (SAS Institute, Cary, NC).

Bivariate county-level maps were created to describe the impact of both maternal race/ethnicity and maternal nativity on the geographic variation in linkage rates. The unlinked rate was broken down into 4 levels (<2%, 2-<5%, 5-<10%, and ≥10%) and represented by a choropleth map with a different shaded color for each level. We then overlaid circular symbols, whose size represented the proportion of live births in each county that came from foreign-born Hispanic mothers (<5%, 5-<10%, 10-<25%, and ≥25%). All maps were created using ArcMap 10.1 (ESRI, Redlands, CA).

Results

During the 13-year study period, there were 2,549,738 birth certificate records for infants born alive to Florida-resident women, and with no indication of an adoption. We were able to link 2,347,738 (92.1%) of these records to an infant HIP record in the first year of life. Over 99.4% of these "base" linkages had an indication of being the infant's birth (as opposed to a post-birth) record. Additionally, we were able to match nearly every (>99.7%) linked infant to a maternal delivery hospitalization, leaving 2,328,897 infants for whom we have both linked infant birth and maternal delivery hospitalizations. Among the 1,598,739 unique women in the linked dataset, most linked to only 1 (64.3%) or 2 (27.1%) live-born children, 8.1% linked to 3-4 children, and only 0.5% linked to 5 or more infants. A comparison of sociodemographic and perinatal characteristics among linked and unlinked infants is presented in Table 1. Multiples, and records linked without requiring an exact or partial SSN match, which were all excluded from previous attempts to link these data, now comprise 4% and 7% of the linked dataset, respectively.

Over the course of follow-up, we linked to over 2.3 million post-birth infant/child discharge records from the various AHCA databases (54.6% ED, 26.0% HIP, and 19.4% ambulatory). The number of post-birth records linked, per infant, ranged from 1 to 133, with 33.2% occurring in the first year of life, 25.5% in the second year, and a steady decline thereafter. We also linked to over 8.5 million hospital discharges among the mothers of these infants, ranging from 1 to over 700 per woman, with a similar breakdown by type of care received (50.9% ED, 26.9% HIP, and 22.2% ambulatory).

Prior to analyzing characteristics associated with lower linkage rates, we excluded 66,168 (2.6%) infants whose birth certificate indicated a birth taking place outside Florida, in a military, state, or federal hospital, or outside of the hospital (explained further in Methods section). Table 2 focuses on infant and maternal demographic/perinatal characteristics associated with reduced likelihood of establishing a record link. The highest crude unlinked rates were seen among infants who died during their first year of life (35.9%), births in which the documented principal source of payment was "self-pay" (28.1%), and infants born to mothers with less than a ninth-grade education (26.0%), who were foreign born (12.9%), and who self-identified as Hispanic (12.8%). After adjusting for other related and potentially confounding variables, several infant and maternal characteristics were associated with increased odds of failure to link infant birth records (Table 2). The strongest infant predictor was the occurrence and timing of death. When compared to infants who survived the first year of life, infants who died on the same day as birth, or subsequently in the early neonatal period (days 1-6), were 11 (AOR: 11.34; 95% CI: 10.53, 12.22) and 2 (AOR: 2.16; 95% CI: 1.90, 2.45) times more likely to be missed, respectively. Although triplets and higher order multiples were 2 times more likely to be missed than singletons, our algorithm for linking multiples resulted in a slightly (~10%) higher likelihood of successful linkage among twins. Those maternal characteristics most strongly associated with failure to link birth records were nativity, race/ethnicity, education, age, marital status, and adequacy of prenatal care. Infants of foreign-born women were over 4 times more likely to be missed than those born to US-born women (AOR: 4.05; 95%

Table 1. Distribution of Linked and Unlinked Birth Records, by Maternal and Infant Demographic and Perinatal Characteristics, Florida, 1998-2009

	Linked (n=2,347,738)		*Unlinked (n=202,000)*	
Characteristic	**n**	**%**	**n**	**%**
Maternal age (years)				
<20 years	255,753	10.9	30,073	14.9
20-29 years	1,219,762	52.0	109,986	54.4
30-39 years	808,302	34.4	57,049	28.2
40+ years	63,871	2.7	4,667	2.3
Missing	50	0.0	225	0.1
Maternal race/ethnicity				
Non-Hispanic white	1,154,113	49.2	67,438	33.4
Non-Hispanic black	520,140	22.2	34,801	17.2
Hispanic	578,235	24.6	91,616	45.4
Asian/Pacific Islander	61,268	2.6	4,445	2.2
Native American	5,691	0.2	1,063	0.5
Other	20,882	0.9	1,938	1.0
Missing	7,409	0.3	699	0.3
Maternal education				
8th grade or less	98,014	4.2	35,362	17.5
9th-11th, no diploma	343,985	14.7	34,103	16.9
High school diploma/GED	770,409	32.8	59,538	29.5
College	1,121,705	47.8	70,977	35.1
missing	13,625	0.6	2,020	1.0
Maternal country of birth				
US-born	1,679,135	71.5	91,647	45.4
Foreign-born	656,645	28.0	107,462	53.2
Missing	11,958	0.5	2,891	1.4
Marital status				
Married	1,389,775	59.2	108,284	53.6
Unmarried	957,746	40.8	92,725	45.9
Missing	217	0.0	991	0.5
Adequacy of prenatal care				
Intensive	130,586	5.6	6,297	3.1
Adequate	1,041,502	44.4	66,396	32.9
Intermediate	831,467	35.4	87,330	43.2
Inadequate	149,090	6.4	20,556	10.2
No Care	25,682	1.1	4,628	2.3
Missing	169,411	7.2	16,793	8.3
Parity				
Nulliparous	976,946	41.6	87,389	43.3
Primiparous	764,130	32.5	61,037	30.2
Multiparous	602,351	25.7	53,178	26.3
Missing	4,311	0.2	396	0.2

	Linked (n=2,347,738)		*Unlinked (n=202,000)*	
Characteristic	**n**	**%**	**n**	**%**
Method of delivery				
Cesarean section	758,639	32.3	52,244	25.9
Vaginal	1,589,088	67.7	148,722	73.6
Missing	11	0.0	1,034	0.5
Gestational age (weeks)				
Very preterm (20-32)	41,095	1.8	4,548	2.3
Preterm (32-36)	211,222	9.0	14,060	7.0
Term (37-42)	2,090,045	89.0	182,229	90.2
Post-term (43-44)	1,823	0.1	277	0.1
Missing	3,553	0.2	886	0.4
Birth weight (grams)				
VLBW (125-1500)	36,069	1.5	4,383	2.2
LBW (1500-2500)	161,934	6.9	10,597	5.2
Normal (2500-6000)	2,149,453	91.6	186,763	92.5
Missing	282	0.0	257	0.1
Plurality				
Singleton	2,273,300	96.8	195,651	96.9
Twins	70,907	3.0	3,889	1.9
Triplets or more	3,531	0.2	281	0.1
Missing	0	0.0	2,179	1.1
Infant sex				
Male	1,202,670	51.2	102,648	50.8
Female	1,145,043	48.8	99,328	49.2
Missing	25	0.0	24	0.0
Infant death				
No infant death	2,332,850	99.4	198,749	98.4
Infant death (days)				
Same day as birth	3,884	0.2	2,286	1.1
Early neonatal (1-6)	2,760	0.1	396	0.2
Late neonatal (7-27)	2,371	0.1	191	0.1
Postneonatal (28-364)	5,873	0.3	378	0.2
Maternal tobacco use				
Yes	189,990	8.1	9,372	4.6
Quit*	20,150	0.9	1,292	0.6
No	2,134,942	90.9	190,666	94.4
Missing	2,656	0.1	670	0.3
Maternal alcohol use				
Yes	8,535	0.4	568	0.3
No	2,337,257	99.6	195,646	96.9
Missing	1,946	0.1	5,786	2.9

Table 1, cont. Distribution of Linked and Unlinked Birth Records, by Maternal and Infant Demographic and Perinatal Characteristics, Florida, 1998-2009

	Linked (n=2,347,738)		*Unlinked (n=202,000)*	
Characteristic	**n**	**%**	**n**	**%**
Maternal diabetes				
Yes	89,615	3.8	6,208	3.1
No	2,254,188	96.0	194,218	96.1
Missing	3,935	0.2	1,574	0.8
Maternal hypertension				
Yes	129,757	5.5	8,863	4.4
No	2,217,981	94.5	193,137	95.6
Induction of labor				
Yes	508,449	21.7	27,686	13.7
No	1,835,743	78.2	172,977	85.6
Missing	3,546	0.2	1,337	0.7
Birth attendant				
Physician (MD, DO)	2,080,499	88.6	156,734	77.6
Midwife (CNM, LM)	256,384	10.9	41,918	20.8
Other	10,777	0.5	3,288	1.6
Missing	78	0.0	60	0.0
Infant year of birth				
1998	176,317	7.5	14,156	7.0
1999	177,987	7.6	14,209	7.0
2000	183,858	7.8	15,597	7.7
2001	186,499	7.9	14,986	7.4
2002	186,860	8.0	14,915	7.4
2003	193,907	8.3	13,699	6.8
2004	198,065	8.4	16,062	8.0
2005	205,016	8.7	17,629	8.7
2006	210,517	9.0	23,552	11.7
2007	213,149	9.1	23,442	11.6
2008	210,804	9.0	18,596	9.2
2009	204,759	8.7	15,157	7.5

	Linked (n=2,347,738)		*Unlinked (n=202,000)*	
Characteristic	**n**	**%**	**n**	**%**
Variables only available from 2004-2009				
Pre-pregnancy BMI				
Underweight	58,473	4.7	5,063	4.4
Normal	573,572	46.2	53,309	46.6
Overweight	272,583	21.9	26,163	22.9
Obese-I	132,191	10.6	10,270	9.0
Obese-II	57,830	4.7	3,518	3.1
Obese-III	37,560	3.0	1,741	1.5
Missing	110,101	8.9	14,374	12.6
Mother/infant transferred				
Yes	15,675	1.3	1,030	0.9
No	1,226,635	98.7	113,408	99.1
Pregnancy resulted from infertility treatment				
Yes	7,605	0.6	326	0.3
No	1,230,789	99.1	113,633	99.3
Missing	3,916	0.3	479	0.4
Principal source of payment				
Private Insurance	553,712	44.6	15,902	13.9
Medicaid	544,552	43.8	36,456	31.9
Self-pay	90,103	7.3	38,643	33.8
Other	18,403	1.5	16,701	14.6
Missing	35,540	2.9	6,736	5.9

*The "quit smoking" option was only available as a response from 2004-2009.

CI: 3.98, 4.11). Even after adjustment for nativity, Hispanic and Native-American race/ethnicity remained significant factors (Hispanics—AOR: 1.66; 95% CI: 1.62, 1.69; Native Americans—AOR: 1.70; 95% CI: 1.56, 1.85). Infants of young women (under 20 years of age), and unmarried women, were both moderately (30-40%) more likely to have been missed. We also experienced dose-response relationships for maternal education and adequacy of prenatal care with lower educational attainment and lower levels of prenatal care associated with an increased odds of failing to establish a link (p_{trend}<0.01). Lastly, among variables that were only present with the new version of the birth certificate (2004+), principal source of payment and transfer status were most strongly associated with the ability to link records (Table 2). Compared to births paid for primarily by private insurance, those that were self-pay, "other" (which includes charity cases), and those paid for by Medicaid were 5.2, 4.1, and 1.8 times more likely to be missed. Birth certificate records with an indication of infant and/or maternal transfer to another facility were less likely to be missed than records without an indication of transfer (AOR: 0.72; 95% CI: 0.67, 0.78).

We further investigated 2 characteristics that identify women and infants who have been disproportionately underrepresented in studies using previous versions of these linked data: maternal nativity and race/ethnicity. Figure 2 depicts the strong direct association between the annual

Table 2. Proportion of Birth Records that Could Not be Linked, by Maternal and Infant Demographic and Perinatal Characteristics, Florida, 1998-2009

Characteristic	*n**	*% not linked*	*Crude OR (95% CI)*	*Adjusted OR† (95% CI)*	*Adjusted OR‡ (95% CI)*
Overall§	2,483,570	5.7	N/A‖	N/A	N/A
Maternal age (years)					
<20 years	280,593	9.1	1.62 (1.59, 1.64)	1.30 (1.27, 1.32)	1.30 (1.27, 1.34)
20-29 years	1,291,770	5.8	reference	reference	reference
30-39 years	844,310	4.5	0.76 (0.75, 0.77)	0.88 (0.87, 0.89)	0.95 (0.93, 0.97)
40+ years	66,800	4.7	0.79 (0.76, 0.82)	0.83 (0.80, 0.87)	0.91 (0.87, 0.96)
Maternal race/ethnicity					
Non-Hispanic white	1,177,523	2.3	reference	reference	reference
Non-Hispanic black	544,839	4.9	2.22 (2.19, 2.26)	1.07 (1.05, 1.09)	1.07 (1.04, 1.10)
Hispanic	661,797	12.8	6.30 (6.22, 6.39)	1.66 (1.62, 1.69)	1.62 (1.57, 1.66)
Asian/Pacific Islander	63,468	3.7	1.65 (1.58, 1.72)	0.61 (0.58, 0.64)	0.68 (0.64, 0.72)
Native American	6,402	11.6	5.65 (5.23, 6.11)	1.70 (1.56, 1.85)	1.99 (1.77, 2.24)
Other	21,859	4.8	2.18 (2.05, 2.32)	0.93 (0.87, 1.00)	0.94 (0.88, 1.01)
Maternal education					
8th grade or less	132,027	26.0	6.40 (6.30, 6.50)	2.74 (2.69, 2.78)	2.33 (2.27, 2.38)
9th-11th, no diploma	373,389	8.2	1.62 (1.60, 1.65)	1.42 (1.40, 1.45)	1.38 (1.35, 1.41)
High school diploma/GED	810,339	5.2	reference	reference	reference
College	1,152,591	2.9	0.55 (0.54, 0.56)	0.70 (0.69, 0.71)	0.80 (0.78, 0.82)
Maternal country of birth					
US-born	1,716,919	2.5	reference	reference	reference
Foreign-born	752,920	12.9	5.80 (5.73, 5.87)	4.05 (3.98, 4.11)	3.30 (3.22, 3.38)
Marital status					
Married	1,446,366	4.2	reference	reference	reference
Unmarried	1,036,954	7.9	1.98 (1.96, 2.00)	1.39 (1.37, 1.41)	1.24 (1.21, 1.26)
Adequacy of prenatal care					
Intensive	133,750	2.5	reference	reference	reference
Adequate	1,082,355	4.0	1.59 (1.53, 1.65)	1.34 (1.29, 1.39)	1.18 (1.11, 1.26)
Intermediate	892,949	7.1	2.95 (2.85, 3.06)	1.75 (1.69, 1.82)	1.51 (1.42, 1.61)
Inadequate	166,100	10.6	4.57 (4.40, 4.74)	1.93 (1.85, 2.01)	1.54 (1.45, 1.65)
No Care	28,473	12.2	5.36 (5.10, 5.63)	2.19 (2.07, 2.31)	1.61 (1.49, 1.74)
Parity					
Nulliparous	1,037,716	6.0	1.02 (1.01, 1.04)	1.15 (1.13, 1.17)	1.14 (1.11, 1.16)
Primiparous	803,963	5.2	0.88 (0.86, 0.89)	1.03 (1.01, 1.04)	1.03 (1.01, 1.05)
Multiparous	637,349	5.9	reference	reference	reference
Method of delivery					
Cesarean section	801,258	5.4	0.91 (0.90, 0.92)	N/A	N/A
Vaginal	1,682,299	5.9	reference		

Table 2, cont. Proportion of Birth Records that Could Not be Linked, by Maternal and Infant Demographic and Perinatal Characteristics, Florida, 1998-2009

Characteristic	*n**	*% not linked*	*Crude OR (95% CI)*	*Adjusted OR† (95% CI)*	*Adjusted OR‡ (95% CI)*
Gestational age (weeks)					
Very preterm (20-32)	44,795	9.0	1.63 (1.57, 1.68)	N/A	1.16 (1.05, 1.29)
Preterm (32-36)	221,720	5.1	0.88 (0.87, 0.90)		1.05 (1.02, 1.09)
Term (37-42)	2,210,929	5.7	reference		reference
Post-term (43-44)	2,059	11.7	2.19 (1.92, 2.51)		1.51 (1.11, 2.04)
Birth weight (grams)					
VLBW (125-1500)	39,659	9.8	1.79 (1.73, 1.85)	1.09 (1.04, 1.15)	1.01 (0.90, 1.12)
LBW (1500-2500)	169,806	5.0	0.87 (0.86, 0.89)	0.95 (0.92, 0.97)	0.90 (0.87, 0.94)
Normal (2500-6000)	2,273,738	5.7	reference	reference	reference
Plurality					
Singleton	2,404,100	5.7	reference	reference	reference
Twins	73,764	4.1	0.70 (0.67, 0.73)	0.90 (0.87, 0.94)	0.78 (0.73, 0.82)
Triplets or more	3,785	6.7	1.19 (1.05, 1.36)	2.18 (1.90, 2.51)	2.00 (1.62, 2.46)
Infant sex					
Male	1,271,635	5.7	reference	N/A	N/A
Female	1,211,889	5.8	1.01 (1.00, 1.03)		
Infant death					
No infant death	2,465,824	5.6	reference	reference	reference
Infant death (days)					
Same day as birth	5,995	35.9	9.35 (8.87, 9.86)	11.34 (10.53, 12.22)	11.56 (10.40, 12.85)
Early neonatal (1-6)	3,086	11.8	2.24 (2.01, 2.50)	2.16 (1.90, 2.45)	2.63 (2.21, 3.13)
Late neonatal (7-27)	2,519	7.0	1.26 (1.08, 1.47)	1.14 (0.96, 1.35)	1.21 (0.97, 1.52)
Postneonatal (28-364)	6,146	5.3	0.94 (0.84, 1.05)	0.90 (0.80, 1.02)	0.99 (0.84, 1.16)
Maternal tobacco use					
Yes	195,107	3.0	0.49 (0.48, 0.51)	0.86 (0.83, 0.88)	0.78 (0.75, 0.81)
Quit	20,682	3.0	0.48 (0.44, 0.52)	0.78 (0.72, 0.85)	0.83 (0.76, 0.90)
No	2,264,857	6.0	reference	reference	reference
Maternal alcohol use					
Yes	8,743	3.5	0.59 (0.52, 0.66)	N/A	N/A
No	2,472,669	5.7	reference		
Maternal diabetes					
Yes	94,311	5.1	0.88 (0.86, 0.91)	N/A	N/A
No	2,385,101	5.7	reference		
Maternal hypertension					
Yes	136,037	4.8	0.81 (0.79, 0.84)	1.06 (1.03, 1.09)	1.13 (1.10, 1.18)
No	2,347,533	5.8	reference	reference	reference
Induction of labor					
Yes	528,881	3.9	0.61 (0.60, 0.62)	0.83 (0.81, 0.84)	0.79 (0.77, 0.81)
No	1,950,996	6.2	reference	reference	reference

Table 2, cont. Proportion of Birth Records that Could Not be Linked, by Maternal and Infant Demographic and Perinatal Characteristics, Florida, 1998-2009

Characteristic	*n**	*% not linked*	*Crude OR (95% CI)*	*Adjusted OR† (95% CI)*	*Adjusted OR‡ (95% CI)*
Birth attendant					
Physician (MD, DO)	2,200,058	5.5	reference	N/A	reference
Midwife (CNM, LM)	275,196	7.4	1.37 (1.35, 1.40)		0.93 (0.91, 0.95)
Other	8,256	7.1	1.32 (1.21, 1.43)		1.00 (0.88, 1.14)
Infant year of birth					
1998	184,544	4.8	reference	N/A	N/A
1999	186,656	4.9	1.03 (1.00, 1.07)		
2000	194,329	5.7	1.20 (1.17, 1.23)		
2001	196,153	5.2	1.09 (1.06, 1.12)		
2002	196,391	5.1	1.08 (1.05, 1.11)		
2003	202,677	4.6	0.97 (0.94, 1.00)		
2004	208,468	5.3	1.11 (1.08, 1.14)		
2005	217,764	6.1	1.30 (1.26, 1.33)		
2006	228,553	8.1	1.77 (1.73, 1.82)		
2007	230,612	7.8	1.70 (1.65, 1.74)		
2008	223,368	5.9	1.25 (1.21, 1.28)		
2009	214,055	4.6	0.97 (0.94, 0.99)		
Variables only available from 2004-2009					
Pre-pregnancy BMI					
Underweight	61,916	5.9	0.93 (0.90, 0.97)	N/A	0.95 (0.91, 0.98)
Normal	610,345	6.3	reference		reference
Overweight	292,015	6.9	1.11 (1.09, 1.13)		0.98 (0.96, 1.00)
Obese-I	139,969	5.8	0.92 (0.89, 0.94)		0.92 (0.89, 0.94)
Obese-II	60,427	4.5	0.70 (0.68, 0.73)		0.90 (0.86, 0.94)
Obese-III	38,870	3.5	0.54 (0.52, 0.58)		0.88 (0.83, 0.93)
Mother/infant transferred					
Yes	15,961	4.9	0.76 (0.70, 0.81)	N/A	0.72 (0.67, 0.78)
No	1,306,859	6.4	reference		reference
Pregnancy resulted from infertility treatment					
Yes	7,812	2.8	0.42 (0.37, 0.48)	N/A	N/A
No	1,310,869	6.4	reference		
Principal source of payment					
Private Insurance	563,901	1.9	reference	N/A	reference
Medicaid	575,147	5.6	2.98 (2.92, 3.05)		1.83 (1.79, 1.88)
Self-pay	124,824	28.1	19.82 (19.37, 20.27)		5.19 (5.05, 5.33)
Other	21,322	16.7	10.17 (9.76, 10.59)		4.10 (3.92, 4.29)

*Frequencies may not add to the total due to missing values

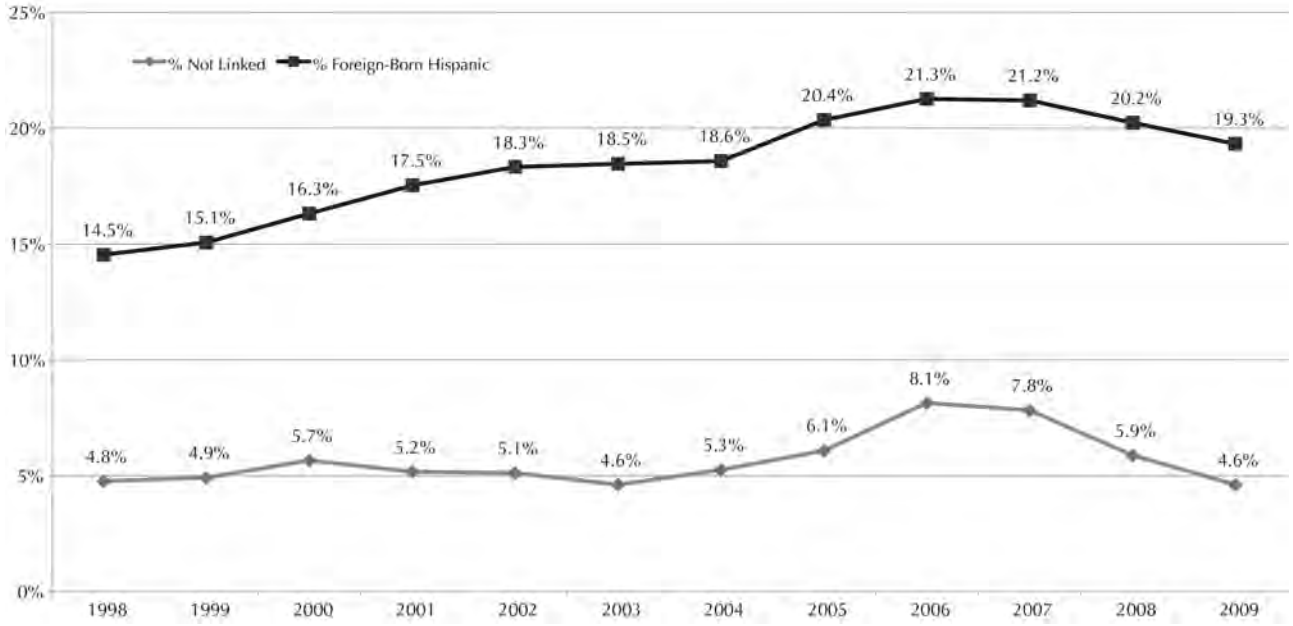
†Multivariable model included birth records and variables spanning both versions of the live birth certificate (1998-2009, n=2,483,570)

‡Multivariable model restricted to birth records and variables from the new version of the live birth certificate (2004-2009, n=1,322,820)

§Excludes births outside Florida, those in facilities not designated as hospitals, and those in military hospitals

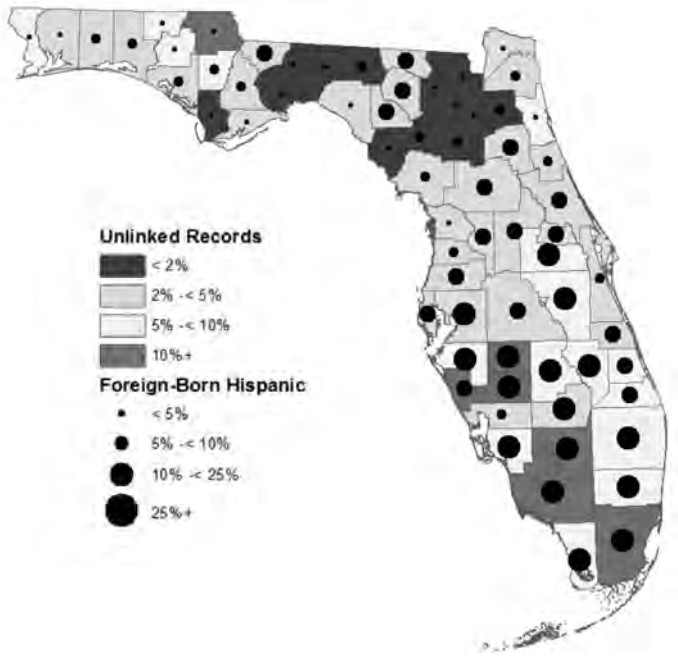
||N/A: variable was not selected for inclusion in the model. All other variables in the table were included in the model. Final models were computed using backward elimination of covariates with a significance level of 0.05.

Figure 2. Proportion of all Birth Certificate Records Not Linked and the Proportion of Mothers that were Foreign-born Hispanics, by Year, 1998-2009



trends in linkage rates and the proportion of birth records in which the biological mother of the baby self-identified as a foreign-born Hispanic woman, which was most notable from 2004 through 2009. These factors also explain a great deal of the geographic variation in linkage rates. Figure 3 highlights the strong correlation between a county's foreign-born, Hispanic makeup and the calculated linkage rate. In South Florida, large, densely populated counties including Miami-Dade, Broward, and Palm Beach, that also have large foreign-born Hispanic populations tended to have higher-than average unlinked rates. However, the highest unlinked rates were seen in other, smaller southwest counties in which agricultural harvesting employs many migrant workers (eg, Hardee, DeSoto, Collier). Table 3 better identifies the subpopulations most difficult to link, likely due underreporting or misreporting of important information (eg, SSN) used in our linkage algorithms. The majority of Asian (88.9%) and Hispanic (70.3%) mothers were not born in the United States, whereas most NH-blacks (74.9%) and nearly all (92.7%) of NH-whites were US-born. However, for each race/ethnic group, birth records for infants born to foreign-born mothers were significantly more difficult to link, with the nativity disparity the strongest for Native Americans and Hispanics. Even among Hispanics, there were large differences in linkage rates between US-born women and their foreign-born counterparts. For example, the proportion of records not linked is similar for Mexican, Puerto Rican, and Cuban women born in the United States. However, for Puerto Ricans, the foreign-born rate is nearly the same (2.7% vs 2.5%), for Cubans it is over doubled (5.8% vs 2.2%), and for Mexicans there is greater than a 10-fold difference (30.4% vs 2.9%).

Figure 3. Proportion of all Birth Certificate Records not Linked and Proportion of Mothers that were Foreign-born Hispanics, by Maternal County of Residence, 1998-2009



Discussion

In this study, we developed a stepwise, deterministic data linkage strategy and created a longitudinal database capable of examining clinical events that occur to mother and child from delivery/birth up to 13 years later. Despite not being able to incorporate name or street address information, we successfully linked vital records and hospital discharge data for over 92% of Florida-resident births,

Table 3. Proportion of Birth Records that Could Not be Linked, by Maternal Nativity and Race/Ethnicity (Including Hispanic Subgroup), Florida, 1998-2009

	US-born		*Foreign-born*		
Maternal race/ethnicity	**Total births**	**% not linked**	**Total Births**	**% foreign born**	**% not linked**
Non-Hispanic white	1,089,917	2.0	85,269	7.3	5.0
Non-Hispanic black	404,837	3.7	135,465	25.1	8.3
Asian/Pacific Islander	6,859	2.2	55,127	88.9	3.8
Native American	4,418	2.5	1,894	30.0	31.8
Other	9,891	2.6	11,928	54.7	6.7
Hispanic	195,230	2.6	461.646	70.3	17.0
Mexican	39,988	2.9	122,899	75.5	30.4
Puerto Rican	66,421	2.5	50,593	43.2	2.7
Cuban	48,854	2.2	76,793	61.1	5.8
Central/South American	27,741	3.3	195,937	87.6	17.1
Other known Hispanic	12,226	2.7	15,424	55.8	10.7

including nearly 96% of twin and higher order multiple births and 59% of birth records missing both maternal and child SSNs. Although probabilistic methods may have increased the sensitivity of our linkage strategy,[17,20,21,25] the algorithm designed allowed us to have stringent control over the occurrence of false positive matches while minimizing the need for extensive manual review. It also permitted us to better incorporate specific knowledge we had concerning each data source and variables used in the linkage.

As observational research makes up an increasing proportion of the evidence base for CER, it is imperative that reporting is thorough and transparent. Quality initiatives such as the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations aim to "ensure clear presentation of what was planned, done, and found in an observational study."[26] For surveillance and research based on the linkage of observational data sources, this includes an evaluation of the linked data that goes beyond a basic assessment of traditional linkage error (eg, missed matches and false positive matches). In this project, we evaluated our linked maternal and child health database by comparing linked and unlinked records to assess whether linkage errors occurred randomly (probability of being missed is the same for all mom/infant dyads) or non-randomly (probability of being missed is not the same for everyone and depends on certain characteristics or circumstances).[9] Consistent with other MCH data linkage projects,[25,27] we found that *linked status* (linked vs unlinked) was strongly associated with various sociodemographic and perinatal attributes. Unfortunately, we observed the lowest rates of linkage amongst some of the most vulnerable, high-risk populations in Florida.

Infants of unmarried Hispanic women born outside of the United States, with low education and poor access to prenatal care—characteristics often associated with adverse pregnancy outcomes—were least likely to be captured by our linkage algorithms. Birth and hospital discharge records for these moms and infants had higher rates of missing or erroneous SSN information, greatly reducing the likelihood for a record match. It was the distribution of these foreign-born, Hispanic women that explained much of the geographic variation in record linkage rates. Although many studies, regardless of the variables available for linkage or the linkage strategy, have consistently reported lower linkage rates for Hispanics,[28-32] it is important to recognize the substantial differences we observed by Hispanic subtype. Women born in Puerto Rico are legal US citizens and are more likely to possess SSNs and be linked, whereas those born in South and Central America (particularly Mexico) are more likely to be undocumented migrant and seasonal farmworkers. The latter are predominantly located in Southwest Florida, frequently claim Florida as their permanent residence, and often attain less than a seventh-grade education,[11] earn extremely low wages,[33] experience deplorable working conditions,[34,35] do not have access to regular, affordable health care,[33] and disproportionately lack or misreport identifying information.

In addition to sociodemographic characteristics, we also found that the probability of linking an infant's birth certificate to an HIP record varied according to factors associated with time spent in the health-care system. Whereas we successfully linked approximately 95% of all birth records, we failed to link over one third of infants who died within minutes to hours of delivery. Many of these infants may not have been formerly admitted to the hospital; therefore, the birth event would not have resulted in an HIP record being generated, precluding our ability to establish the desired link.[36] Conversely, we found that transfer of the mother and/or baby from the birth hospital to another facility was associated with a 28% higher probability of establishing a record link. In such cases, the medical reasons for the transfer are more likely to result in recurrent hospitalizations and longer lengths of stay, increasing the opportunity for more identifying information to be collected and thus for a link to be established.[37]

A detailed assessment of our data linkage process and quantification of linkage rates was necessary since a small amount of linkage error can result in substantially biased results, particularly when errors occur in a non-random fashion.[38] The impact of linkage errors will depend upon the way in which the database is being used (eg, surveillance vs research) and the nature of the analysis (eg, exposures, outcomes, cofactors examined). For example, Lariscy (2011) used a dataset, in which health surveys were linked to death certificates, to investigate disparities in survival by ethnicity and nativity.[32] He demonstrated that relaxing and tightening cut-points that defined a link/non-link (that is, varied the types of linkage errors made) resulted in substantial changes in the reported disparity comparing foreign-born Hispanics to US-born Whites. When criteria were strict (minimized false positives), a 22% risk *reduction* was reported for foreign-born Hispanics, compared to a 24% *increased* risk when criteria were relaxed (maximized sensitivity).[32] Even when differential linkage rates can not entirely account for an observed ethnic disparity, the accuracy of mortality risk estimates is reduced.[31] If our MCH database is to be used to establish a population-based public health surveillance system in which the entire birth cohort (including records that could not be linked to a HIP record) is included in the denominator, and cases are determined by ICD-9-CM codes in the HIP file (only available for linked records), the under-ascertainment for particular groups should be thoroughly described.[10] Alternatively, if the database is being used for epidemiologic research or CER, unlinked records are excluded, meaning that certain subgroups (eg, foreign-born Hispanics) are underrepresented in the dataset and the impact on internal and external validity should be discussed.

Interpretation of our evaluation should consider several limitations. A more complete evaluation and quantification of any data linkage project would ideally include a comparison of the linked database with a "gold-standard" and sensitivity analyses comparing multiple data linkage strategies. Unfortunately, an external dataset in which the true match status for all records is known (the "gold standard") does not exist for these data. Also, limited funds and personnel precluded our ability to develop, test, implement, and evaluate multiple linkage strategies. Second, the sociodemographic and perinatal characteristics investigated as part of our evaluation were taken from the birth certificate record, in which many factors are self-reported (eg, race, ethnicity, tobacco and alcohol use), unconfirmed (eg, maternal hypertension), and thus, subject to misclassification. However, birth certificates can also produce highly accurate data for select variables (eg, maternal age, method of delivery, birth weight). Lastly, although our data linkage algorithms incorporated successful strategies for linking birth records to birth hospitalization records for multiple births and records without a valid SSN, these 2 subgroups are extremely difficult to "follow" over time using data linkage. For example, without sufficient discriminating information, it is uncertain when 10 post-birth hospitalization records for a pair of twins all belong to 1 twin, are split equally amongst the twins, or some other variation. Thus, for CER studies that extend beyond events that occur at delivery/birth, one must acknowledge and attempt to quantify the bias associated with a reduced ability to link post-birth hospital discharge records among these subgroups.

In summary, using a stepwise deterministic linkage approach, we achieved a high linkage rate of several data sources, and produced a reliable, multipurpose database that can be used for observational, comparative effectiveness, and health services research in MCH populations. Our findings underscore the importance of evaluating routinely collected health data and increasing clarity regarding the strengths and limitations of linked electronic data sources. The resultant database will be of immense utility to researchers, health planners, and policy makers as well as other stakeholders interested in MCH outcome studies.

Acknowledgements

The authors acknowledge the following organizations and individuals for contributing to this project: the Agency for Healthcare Research and Quality for promoting the enhancement of statewide, hospital-based, encounter-level databases (grant number R01HS019997); Social and Scientific Systems, Inc. for their coordination and direction of enhanced state data grants; staff of the Florida Department of Health and the Agency for Health Care Administration for providing access to these data and ongoing support for this data linkage effort; the Florida Birth Defects Registry Consortium for their numerous consultations and feedback on linkage algorithms and products; and Bill Sappenfield, MD, MPH for the instrumental role he played in encouraging and sustaining interagency collaboration.

References

1. National Research Council and Institute of Medicine. U.S. Health in International Perspective: Shorter Lives, Poorer Health. Panel on Understanding Cross-National Health Differences Among High-Income Countries, Steven H. Woolf and Laudan Aron, Eds. Committee on Population, Division of Behavioral and Social Sciences and Education, and Board on Population Health and Public Health Practice, Institute of Medicine. Washington, DC; 2013.
2. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)*. 2010;29(10):1818-1825.
3. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011;32:91-108.
4. Shah BR, Drozda J, Peterson ED. Leveraging observational registries to inform comparative effectiveness research. *Am Heart J*. 2010;160(1):8-15.
5. Qayad MG, Zhang H. Accuracy of public health data linkages. *Matern Child Health J*. 2009;13(4):531-538.
6. Cameron CM, Purdie DM, Kliewer EV, McClure RJ, Wajda A. Population health and clinical data linkage: the importance of a population registry. *Aust N Z J Public Health*. 2007;31(5):459-463.
7. Herrchen B, Gould JB, Nesbitt TS. Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. *Comput Biomed Res*. 1997;30(4):290-305.
8. Howe GR. Use of computerized record linkage in cohort studies. *Epidemiol Rev*. 1998;20(1):112-121.
9. Adams MM, Kirby RS. Measuring the accuracy and completeness of linking certificates for deliveries to the same woman. *Paediatr Perinat Epidemiol*. 2007;21(suppl 1):58-62.
10. Salemi JL, Tanner JP, Block S, et al. The relative contribution of data sources to a birth defects registry utilizing passive multisource ascertainment methods: does a smaller birth defects case ascertainment net lead to overall or disproportionate loss? *J Registry Manage*. 2011;38(1):30-38.

11. Data Innovations. *Healthcare Cost and Utilization Project (HCUP)*. 2012. Available at: **http://www.hcup-us.ahrq.gov/datainnovations/grants.jsp**. Accessed November 13, 2012.
12. Florida Department of Health. *Florida Vital Statistics Annual Report*. 2011. Available at: **http://www.flpublichealth.com/VSBOOK/pdf/2011/vscomp.pdf**. Accessed November 13, 2012.
13. National Center for Health Statistics. Report to the Panel to Evaluate the U.S. Standard Certificates. 2001. Available at: **http://www.cdc.gov/nchs/data/dvs/panelreport_acc.pdf**. Accessed December 2, 2012.
14. Agency for Health Care Administration. *AHCA Patient Data Submission Guide, version 2B*. 2012. Available at: **http://ahca.myflorida.com/SCHS/DataCollection/docs/DataGuideVersion2Bupdate08072012.pdf**. Accessed November 29, 2012.
15. Agency for Health Care Administration. *Florida Center for Health Information and Policy Analysis: Data Catalog*. 2012. Available at: **https://floridahealthfinderstore.blob.core.windows.net/documents/researchers/OrderData/documents/Data_catalog%20Revised%20January%202012.pdf**. Accessed November 29, 2012.
16. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol*. 2002;31(6):1246-1252.
17. Clark DE, Hahn DR. Comparison of probabilistic and deterministic record linkage in the development of a statewide trauma registry. *Proc Annu Symp Comput Appl Med Care*. 1995:397-401.
18. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med*. 1995;14(5-7):491-498.
19. Mason CA, Tu S. Data linkage using probabilistic decision rules: a primer. *Birth Defects Res A Clin Mol Teratol*. 2008;82(11):812-821.
20. Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol*. 2007;60(9):883-891.
21. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 2011;64(5):565-572.
22. Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. *Stat Med*. 2002;21(10):1485-1496.
23. Fornari C, Madotto F, Demaria M, et al. *Record-linkage procedures in epidemiology: an Italian multicentre study* [in Italian]. *Epidemiol Prev*. 2008;32(suppl 3):79-88.
24. Alexander GR, Cornely DA. Prenatal care utilization—its measurement and relationship to pregnancy outcome. *Am J Prev Med*. 1987;3(5):243-253.
25. Bentley JP, Ford JB, Taylor LK, Irvine KA, Roberts CL. Investigating linkage rates among probabilistically linked birth and hospitalization records. *BMC Med Res Methodol*. 2012;12(1):149.
26. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4(10):e297.
27. Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. *Paediatr Perinat Epidemiol*. 2006;20(4):329-337.
28. Adams MM, Berg CJ, McDermott JM, et al. Evaluation of reproductive histories constructed by linking vital records. *Paediatr Perinat Epidemiol*. 1997;11(1):78-92.
29. Elo IT, Turra CM, Kestenbaum B, Ferguson BR. Mortality among elderly Hispanics in the United States: past evidence and new results. *Demography*. 2004;41(1):109-128.
30. Liao Y, Cooper RS, Cao G, et al. Mortality patterns among adult Hispanics: findings from the NHIS, 1986 to 1990. *Am J Public Health*. 1998;88(2):227-232.
31. Palloni A, Arias E. Paradox lost: explaining the Hispanic adult mortality advantage. *Demography*. 2004;41(3):385-415.
32. Lariscy JT. Differential record linkage by Hispanic ethnicity and age in linked mortality studies: implications for the epidemiologic paradox. *J Aging Health*. 2011;23(8):1263-1284.
33. Villarejo D. The health of U.S. hired farm workers. *Annu Rev Public Health*. 2003;24:175-193.
34. Alderete E, Vega WA, Kolody B, Aguilar-Gaxiola S. Lifetime prevalence of and risk factors for psychiatric disorders among Mexican migrant farmworkers in California. *Am J Public Health*. 2000;90(4):608-614.
35. Rivara FP. Fatal and nonfatal farm injuries to children and adolescents in the United States. *Pediatrics*. 1985;76(4):567-573.
36. Tanner JP, Salemi JL, Hauser KW, Correia JA, Watkins SM, Kirby RS. Birth defects surveillance in Florida: infant death certificates as a case ascertainment source. *Birth Defects Res A Clin Mol Teratol*. 2010;88(12):1017-1022.
37. Zingmond DS, Ye Z, Ettner SL, Liu H. Linking hospital discharge and death records—accuracy and sources of bias. *J Clin Epidemiol*. 2004;57(1):21-29.
38. Goldstein H, Harron K, Wade A. The analysis of record-linked data using multiple imputation with data value priors. *Stat Med*. 2012;31(28):3481-3493.

Original Article

Strategies to Achieve Sustainability and Quality in Birth Defects Registries: The Experience of the National Registry of Congenital Anomalies of Argentina

Boris Groisman[a]; María Paz Bidondo[a]; Juan Antonio Gili[b]; Pablo Barbero, PhD[a]; Rosa Liascovich, PhD[a]

Abstract: In many low- and middle-income countries, birth defects are not considered a public health priority and are perceived by the medical community as rare, unpreventable events. In this context, a registry of birth defects should address not only the collection, analysis, and dissemination of information but also contribute to local interventions like prevention, diagnosis, and treatment. We describe the National Registry of Congenital Anomalies of Argentina (RENAC) in terms of case definition, data collection, quality assurance, and data sending, coding, analysis, and information dissemination and we present the strategies used to ensure its sustainability. We emphasize strategies for motivating the people collecting data, such as training activities, participation in research projects, returning the processed data, making useful clinical information available, giving non-monetary rewards, and linking cases to genetic services.

Key words: *birth defects, public health surveillance, registry, infant mortality, Argentina*

Introduction

A major concern for people running a birth defects (BDs) registry is to achieve sustainability and good quality.[1-3] In many low- and middle-income countries, BDs are not considered a public health priority and are perceived as rare, unpreventable events. Therefore, it is essential to develop strategies to ensure the registry's continuity over time by having funding and political support, but also by developing a simple reporting system and motivating the people involved in it. The purpose of this article is to show, based on our experience in the National Registry of Congenital Anomalies of Argentina (RENAC), strategies to ensure sustainability and quality of a registry of BDs in a middle-income country.

Argentina has a population of 40,117,096 in an area of 2,780,400 square kilometers (1,727,660 square miles).[4] In 2011, there were 758.042 live births. Health care is provided by public and non-public sectors, the latter funded by unions and pre-private plans. The public sector provides free services under the coordination of national, provincial, or municipal authorities. The infant mortality rate (IMR) decreased from 33.2/1,000 live births in 1980 to 11.7/1,000 live births in 2011. BDs account for 25% of IMR, being the second leading cause after perinatal causes.[5]

After the "thalidomide tragedy," the Latin American Collaborative Study of Congenital Malformations (ECLAMC) was established in South America.[6] In the 1960s, BDs were not considered a public health issue in the region. ECLAMC was created as a voluntary, nongovernmental network of hospitals in a case-control study. It was the only data source, but coverage was low. In Argentina vital statistics records mortality by BDs, but birth certificates do not include information about them. The increase in the contribution of BDs to IMR and the lack of official information about birth prevalence were 2 reasons for the creation of the RENAC. There was a need to improve care of affected newborns and, because Argentina has an important agricultural activity and there is social concern about pesticides and other contaminants in relation to BDs, there was a need to monitor BDs prevalence.

Strategies

We wanted the registry to be an official health policy, but also to ensure its sustainability and quality by motivating those responsible for collecting the data. As members of the National Center of Medical Genetics (CNGM), we lobby the National Ministry of Health (NMoH) to include BDs in the agenda. We contacted ECLAMC for advice. The registry was established in 2009 with 2 objectives: 1) To generate epidemiological data about distribution and determinants of BDs; 2) To improve the care of affected newborns, since genetic services in Argentina are sparse and physicians are often unaware of how to manage BDs.

The registry had to be national and representative. In Argentina, 99% of births occur in hospitals. Therefore, the registry is hospital based and has a central coordination. RENAC Coordination consists of 4 professionals from the CNGM (authors BG, MPB, PB, and RL), with support of a part-time administrative assistant and a statistician (author JAG). It was necessary to prioritize hospitals with adequate resources (trained physicians, X-rays, ultrasound, etc) and with a high number of births to ensure enough cases, so the registration routine is not sporadic. These criteria are in-line with the concentration of births in hospitals of high complexity, which take place in Argentina.[7] Consequently, we decided that RENAC, in the first stage, had to include public hospitals with 1,000 births or more per year. We are at this stage working with 120 hospitals with 300,000 births

[a]National Center of Medical Genetics (CNGM), National Ministry of Health. [b]Latin American Collaborative Study of Congenital Malformations (ECLAMC).

Address correspondence to Boris Groisman, Medical Geneticist, Centro Nacional de Genética Médica, Av. Las Heras 2670, 3er piso, ZIP code: 1425. Email: bgroisman@gmail.com.

per year, covering 75% of births of the public sector and 40% of the country.[8] At a later stage, RENAC will include non-public hospitals by convincing the authorities of the advantages of belonging to a BDs registry, like getting support in clinical management of affected newborns.

Argentina has a decentralized health-care system and hospitals are under different jurisdictions, so it was necessary that the NMoH conduct agreements with local authorities in charge of the hospitals. Agreements include the appointment of neonatologists for data sending, the return of the data encoded by the RENAC coordination, provision of paper forms and manuals and funding for an annual meeting. Along with hospital authorities, 2 neonatologists are invited by the Coordination to integrate the RENAC team as advocates for the issue of BDs. Reporting is mandatory for hospitals included in the RENAC. Available resources determined the case definition and data collection, the quality assurance (QA) and data sending, coding, analysis, and information dissemination.

Case Definition and Data Collection

The inclusion criteria was restricted to live births or stillbirths weighing 500 grams or more, with major morphological BDs as defined by EUROCAT,[9] identified from birth until discharge, and detected by physical examination or complementary studies. Functional disorders were excluded because they often require lab diagnosis and follow-up. We focused on BDs easily identifiable by a trained neonatologist. Since in Argentina elective termination of pregnancy for fetal anomaly is illegal, there is no available data and they are therefore excluded.

Because hospitals have computers with an Internet connection, but not electronic medical records, we developed a system for data collection using a paper form and Excel file with data then sent through the Internet. The form is attached to every maternal clinical record and is completed by physicians attending births, who write whether the newborn presents BDs or not. If so, the physicians fill the form with a verbatim description of BDs[6,10] and complete additional variables available in the clinical record, without the need to ask the mother for further information. RENAC neonatologists supervise this task following standardized procedures in the RENAC manual. Variables included are: sex, birth status (live birth, stillbirth), twinning, condition when sending data (discharged alive, dead, not discharged, referred to another hospital), gravidity, gestational age, weight, maternal age, prenatal diagnosis and place of residence of the mother at birth.[11] Only reporting neonatologists and the Coordination team has access to this password-protected data.

Quality Assurance and Data Sending

Neonatologists load the total hospital births and the data from the forms into an Excel file. The file is sent monthly to the Coordination center through a restricted access Web site. Neonatologists have their profile in the Web site with name, photo, hospital and city. The Web site is a forum running vBulletin 4 software hosted by Amazon Web Services. It allows data sending, solving of operational issues and QA of data by the Coordination team. If BD descriptions are unclear, the Coordination team asks for more information. Neonatologists may enter the forum at any time and send messages with pictures of patients, X-rays and other studies allowing members to discuss clinical cases. Before sending this information, informed consent is asked to the parents of the newborns. The Coordination team suggests diagnosis, the finding of associated BDs, and referral to genetic services. Forum interaction increases social cohesion among participants who feel themselves as members of the same team.

Figure 1. Proposed Strategies to Achieve Sustainability and Quality in a BDs Registry of a Low- or Middle-income Country

- Contact organizations with experience in BDs monitoring for advice (ie, other registries, the International Clearinghouse for Birth Defects Surveillance and Research)
- Lobby for the registry to be an official health policy
- Be in-line with the raising social concerns (ie, the effect of pesticides) and the relative increase of BDs in the IMR
- Include the improving of health care of affected newborns as a main objective
- Start with few high complexity hospitals and then extend the registry from there.
- Choose as reporting members, people already attending births (neonatologists, pediatricians, midwives).
- Include BDs feasible to diagnose by the reporting members and define a core of data to collect for each case. Including follow-up can increase sensibility of the registry, but also make it more complex and less sustainable.
- Use a verbatim description of BDs instead of a checklist.
- Develop an easy-to-use system for data sending. If possible, use participatory information and communications technologies.
- Centralize coding and, if possible, use experts. Use a system that allows comparisons with other sources (ie, ICD-10).
- Consider dissemination of information as a main activity of the registry, adapting the message to different stakeholders (health authorities, reporting members, scientific community, mass media, patients' associations, etc).
- Motivate the people who report by helping them in diagnosis and appropriate referral of cases, as well as training activities, participation in research projects, giving them back the processed data, and non-monetary or monetary rewards.

Coding, Analysis, and Information Dissemination

To ensure quality and homogeneous criteria, each description is reviewed and coded by a medical geneticist of the Coordination following pre-established criteria. To allow comparisons with other sources, we use the ICD-10 with the BPA modification.[12]

We analyze prevalence of BDs selected by clinical relevance, frequency, and comparability with other registries,

using STATA software. In the first stage we use births from included hospitals as denominators when calculating rates. As we move forward to later stages with higher coverage, we will use population births as denominators.

Information dissemination is performed through a printed annual report, an annual meeting funded by the NMoH and the electronic sending of each hospital's report. Information is disseminated to stakeholders, including the participating neonatologists who feel empowered when using locally their own processed data.

We stimulate the neonatologists with training activities and participation in research projects, eg, studies with a case-control design in selected hospitals to research on environmental risk factors. We give them back the processed data and we make available useful clinical information, facilitating sample referral and links with geneticists. The annual meeting is an essential opportunity for live interaction of members. We show how RENAC worked in the previous year, train new members, and discuss clinical cases. Special mentions are given to outstanding neonatologists, according to the timely sending of data, the quality of descriptions, local use of the data, and the number of clinical cases presented for discussion in the Web site.

Figure 2. Disadvantages and Challenges of the Proposed Strategies

- Included BDs detected until discharge results in underreporting of BDs that can be found later, such as congenital heart disease
- The lack of controls restricts the possibility of performing routinely analytical studies of research for risk factors. For case-control studies we design special projects involving selected hospitals.
- The design of the system requires a coordination team familiar with information and communications technologies and that people who report have computers with Excel and Internet access.
- To protect patient privacy and confidentiality of the data, it is necessary to have a data-sending and storage system that is secure and password-protected.
- While this design allows the support of neonatologists in clinical issues, it takes time to achieve the high coverage needed to perform surveillance.
- Since the reports are done remotely, it is difficult to perform quality controls based on review of medical records of the affected newborns.

Summary

To achieve sustainability and quality, we consider it important to have funding and political support, but also to train and motivate those who report by helping them in diagnosis and appropriate referral of cases. The diligence with which the information is collected depends on the motivation of those responsible for collecting it[13] (Figure 1). In Figure 2, we summarize the disadvantages and challenges of this approach. Finally, we think that BDs registries in low- and middle-income should address not only the collection, analysis, and dissemination of information but also contribute to local interventions like prevention, diagnosis and treatment of BDs.

References

1. CDC. Guidelines for evaluating surveillance systems. *MMWR*. 1988;37(S-5).
2. Centers for Disease Control and Prevention. *CDC Updated Guidelines for Evaluating Public Health Surveillance Systems*. Recommendations from the Guidelines Working Group Centers for Disease Control and Prevention. 2001;50(RR13):1-35
3. National Birth Defects Prevention Network. *Guidelines for conducting birth defects surveillance*. June 2004. Available at: **http://www.nbdpn.org/docs/NBDPN_Guidelines2008.pdf**. Accessed January 20, 2013.
4. INDEC. National Institute for Statistics and Census. 2010. Available at: **http://www.indec.gov.ar**. Accessed January 20, 2013.
5. National Ministry of Health, Department of Health Statistics and Research, Vital Statistics. 2011. Available at: **http://www.deis.gov.ar/Publicaciones/Archivos/Serie5Nro55.pdf**. Accessed January 20, 2012.
6. Castilla EE, Orioli IM. *ECLAMC: the Latin-American collaborative study of congenital malformations*. Community Genet. 2004;7(2-3):76-94.
7. Speranza AM, Kurlat I. Regionalization of perinatal care: a strategy to reduce infant and maternal mortality [in Spanish]. *Argentinean J Pub Health*. 2011;2(7).
8. Groisman B, Menazzi S, Gili JA, Bidondo MP, Barbero P, Liascovich R. RENAC Annual Report 2012. Available from the corresponding author.
9. EUROCAT Guide 1.3 and reference documents. Instructions for the Registration and Surveillance of Congenital Anomalies. Available at: **http://www.eurocat-network.eu/content/EUROCAT-Guide-1.3.pdf**. Accessed March 8, 2013.
10. Luquetti DV, Koifman RJ. Surveillance of birth defects: Brazil and the US. *Cien Saude Colet*. 2011;16(suppl 1):777-785.
11. Bidondo MP, Flores R, Barbero P, Groisman B, Menazzi S, Liascovich R. RENAC 2012 operating manual. Available from the corresponding author.
12. Crawshaw P. The new BPA classification. *Arch Dis Child*. 1995;73(6):563-567.
13. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. *Bull World Health Organ*. 1994;72(2):285-304.

Original Article

The Impact of the Benign Brain Tumor Cancer Registries Amendment Act (Public Law 107-260) on Non-malignant Brain and Central Nervous System Tumor Incidence Trends

Bridget J. McCarthy, PhD[a]; Carol Kruchko[b]; Therese A. Dolecek, PhD[a]

Abstract: The study objective was to investigate patterns of reported non-malignant brain and CNS tumor incidence over a time period encompassing 1997-2008 during which time the Benign Brain Tumor Cancer Registries Amendment Act (PL 107-260) was passed and implemented. Analyses of 75,350 incident non-malignant brain and CNS tumors from eleven population-based central registries revealed that there were statistically significant increases in the age-adjusted incidence rate for non-malignant tumors for those diagnosed prior to 2002 and over the time period from 2002 until 2005. However, no significant change in the age-adjusted incidence rate for non-malignant tumors was observed over the time period 2005 to 2008 indicating that the incidence from this time period may quantify the "true" incidence of non-malignant brain and CNS tumors in the United States.

Key words: *brain, non-malignant, central nervous system, incidence, patterns*

Introduction

Brain and central nervous system (CNS) tumors are often devastating both in terms of morbidity and mortality and the importance of requiring the reporting of all primary brain tumors regardless of tumor behavior (malignant or non-malignant) has been recognized.[1,2] The Central Brain Tumor Registry of the United States (CBTRUS), in collaboration with participating state cancer registries, demonstrated in 1992 the feasibility of collecting data on all primary brain and CNS tumors in the United States[3] and has since promoted the collection of these data globally.[4] Passed in 2002, the Benign Brain Tumor Cancer Registries Amendment Act (Public Law 107-260; **ftp.resource.org/gpo.gov/laws/107/publ260.107.pdf**; accessed February 3, 2012) required central cancer registries supported by the National Program of Cancer Registries (NPCR) to expand data collection on primary brain and CNS cancer incidence to include tumors of non-malignant (benign and uncertain) behavior in addition to malignant behavior beginning with diagnosis year 2004. In keeping with the spirit which advocated for enactment of this law, other standard setters in surveillance including the Surveillance, Epidemiology and End Results (SEER) program and the North American Association of Central Cancer Registries (NAACCR) agreed to comply with the new statute. This united support enabled national public surveillance of the incidence and mortality of brain and CNS tumors.

Analyses of incidence data prior to diagnosis year 2004 in the United States have shown increasing trends over time for all primary and malignant primary brain tumors.[5-18] However, trends in the incidence of primary malignant brain tumors in more recent time periods have been flat or decreasing.[19,20] Significant changes in the coding, classification, and particularly, the ascertainment and reporting of brain tumors have occurred over the last 2 decades. Among the most significant of these changes was achieving consensus on the classification of the brain and CNS[21] and the efforts to reconcile the most recent coding and classification schemes, ICDO-3[22] and WHO 2000[21] which paved the way for a site definition to guide the collection of these tumors and a reporting scheme for comparing estimates of primary brain tumors across registries in 2000.[23] Although some cancer registries have routinely collected all primary brain and CNS tumors, the extent of collection and reporting of non-malignant tumors has not been consistent. These factors, along with implementation of Public Law 107-260, have undoubtedly influenced non-malignant primary brain and CNS tumor incidence patterns. Thus, the primary objective of this study was to evaluate patterns of reported incidence rates of non-malignant brain tumors diagnosed over a time period which spans the introduction and implementation of Public Law 107-260.

Methods

The Central Brain Tumor Registry of the United States (CBTRUS) has compiled population-based incidence data on all primary brain and CNS tumors, regardless of biologic behavior, since 1992. Data from 11 population-based state

[a]Division of Epidemiology and Biostatistics and Institute for Health Research and Policy, School of Public Health, University of Illinois at Chicago, Chicago, IL.
[b]Central Brain Tumor Registry of the United States (CBTRUS), Hinsdale, IL.

Address correspondence to Carol Kruchko, Central Brain Tumor Registry of the United States, 244 E. Ogden Ave., Ste. 116, Hinsdale, IL 60521. Telephone: (630) 655-4786. Fax: (630) 655-1756. Email: ckcbtrus@aol.com

All analyses were conducted under contract to the Central Brain Tumor Registry of the United States which receives funding in support of its analyses in 2012 from National Brain Tumor Society, Pediatric Brain Tumor Foundation and from the Cooperative Agreement 1U58DP003831 from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

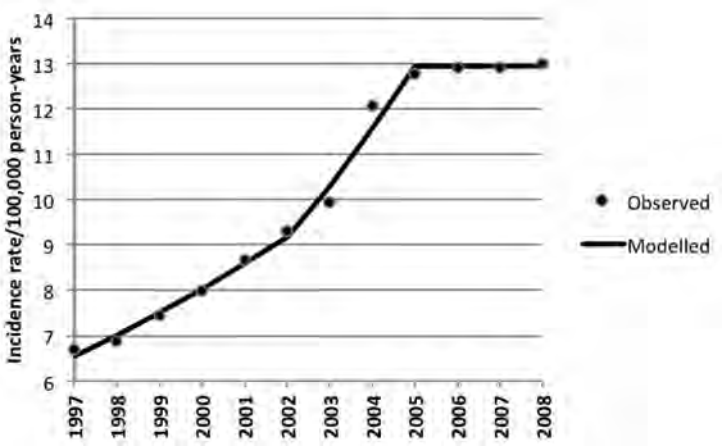
cancer registries (Arizona, Colorado, Delaware, Idaho, Maine, Massachusetts, Minnesota, Montana, North Carolina, New York, and Virginia) that collaborated with the CBTRUS and collected both malignant and non-malignant primary brain tumors diagnosed from 1997-2008 were analyzed. Representing close to 22% of the population in the United States, almost all of these central registries currently have achieved gold standard certification from NAACCR. Use of these data was approved by the University of Illinois at Chicago Institutional Review Board. Primary brain and CNS tumors were defined using the International Classification of Diseases for Oncology (ICD-O-3)[22] site codes of C70.0-C72.9, C75.1-C75.3 and C30.0 (histology codes 9522-9523). Non-malignant tumors were defined as those with ICD-O-3 behavior codes of "0" (benign) or "1" (uncertain).

Age-adjusted incidence rates and confidence intervals at the 95% level were calculated using SEER*Stat 7.0.9.[24] Population data available from the US Census Bureau were obtained from the National Cancer Institute Surveillance, Epidemiology and End Results (SEER) Program Web site (**seer.cancer.gov/popdata/**) to calculate incidence rates. Incidence rates per 100,000 were analyzed for each respective diagnosis year and were age-adjusted to the 2000 US Standard Population. To further investigate the potential for sharp changes in age-adjusted incidence rates over time, Joinpoint 3.5.2 (piece-wise regression) software was utilized.[25] Join points correspond to a point in time of a change in the trend where 2 different sloped lines come to a juncture, and the software fits the simplest join-point model that the trend data will allow. Using the grid search method, the permutation test model (model: ln[y]=xb) assessed changes in age-adjusted incidence rates with a minimum number of 3 observations from a join point to either end of the data and a minimum of 3 observations between 2 join points. The annual percent change (APC) with corresponding 2-sided 95% confidence intervals (CI) for each trend segment was calculated with Joinpoint 3.5.2 software using weighted least squares regression.

Results

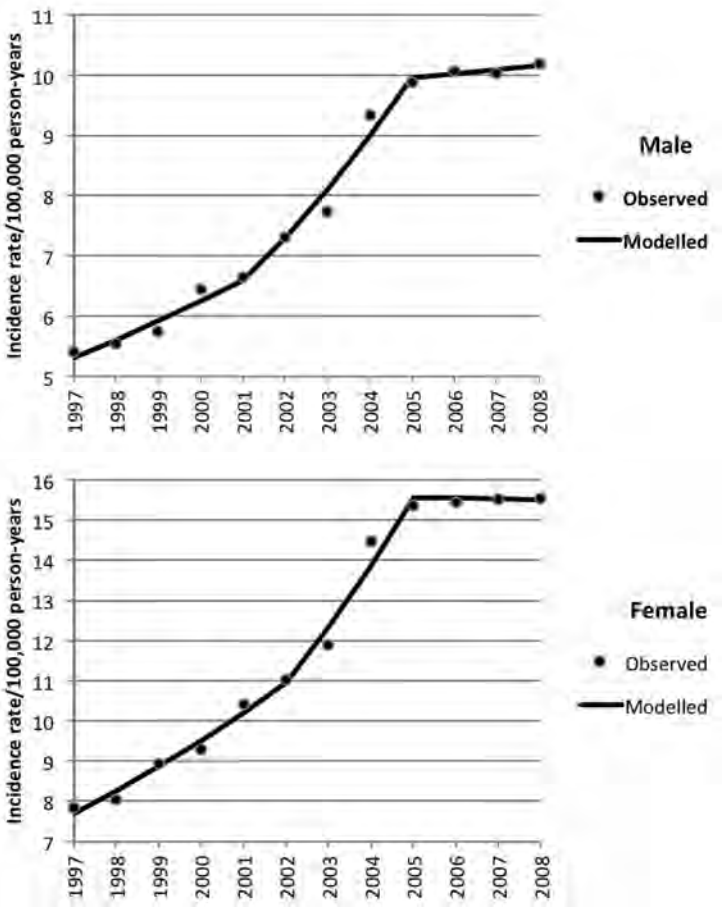
A total of 75,350 incident non-malignant brain and CNS tumors diagnosed from 1997-2008 were included in these analyses. A join-point analysis of the non-malignant brain and CNS tumor incidence over time revealed 2 junctures where the slope of the age-adjusted incidence rate trend line changed (Figure 1). Overall, a statistically significant increase in the age-adjusted incidence rate for non-malignant tumors diagnosed prior to 2002 was found (APC=7.0). During that time period, the age-adjusted incidence rate increased from 6.7 in 1997 to 9.3 per 100,000 person-years in 2002. A shift in the slope of the age-adjusted incidence rate trend was observed over the time period from 2002-2005, with a statistically significant increase in the non-malignant age-adjusted incidence rate (APC=12.2). The age-adjusted incidence rate during this time period increased more rapidly, from 9.3 in 2002 to 12.8 per 100,000 person-years in 2005. This shift in rates was primarily driven by the "jump" in age-adjusted incidence rates from diagnosis year 2003 (9.9 per 100,000 person-years) to diagnosis year 2004 (12.1

Figure 1. Trend in Age-Adjusted Incidence Rates for Non-Malignant Brain and CNS Tumors; CBTRUS 11 State Central Cancer Registries, 1997-2008



Incidence rates were per 100,000 and were age-adjusted to the 2000 US Standard Population. Analyses included data provided through the NPCR-CSS mechanism from state cancer registries in Arizona, Colorado, Delaware, Idaho, Maine, Massachusetts, Minnesota, Montana, North Carolina, New York, and Virginia.

Figure 2. Trends in Age-Adjusted Incidence Rates for Non-Malignant Brain and CNS Tumors by Gender, CBTRUS 11 State Central Cancer Registries, 1997-2008



Incidence rates were per 100,000 and were age-adjusted to the 2000 US Standard Population. Analyses included data provided through the NPCR_CSS mechanism from state cancer registries in Arizona, Colorado, Delaware, Idaho, Maine, Massachusetts, Minnesota, Montana, North Carolina, New York, and Virginia.

per 100,000 person-years). As previously noted, diagnosis year 2004 was the first year mandated for implementation of the law. Conversely, no significant changes for non-malignant age-adjusted incidence rates were observed over the time period 2005-2008 (APC=0.0), with the rates slightly increasing from 12.8 in 2005 to 13.0 per 100,000 person-years in 2008.

A similar pattern was found in both males and females when analyzed separately (Figure 2). Males demonstrated changes in the slope of the age-adjusted incidence rate trend in 2001 and 2005, with a significant increase from 1997-2001 (APC=5.1), a larger increasing incidence from 2001-2005 (APC=10.8), and a flattening out of the incidence from 2005-2008 (APC=0.7). From 1997-2001, the age-adjusted incidence rate increased from 5.4 to 6.6 per 100,000 person-years, while from 2005-2008, the age-adjusted incidence rate for non-malignant brain tumors slightly increased in males from 9.9 to 10.2 per 100,000 person-years. The slope of the age-adjusted incidence rate trend in females significantly increased from 1997-2002 (APC=7.4), increased at a faster rate from 2002-2005 (APC=12.4), and showed no change in the age-adjusted incidence rate from 2005-2008 (APC=-0.2). From 1997-2002, the age-adjusted incidence rate increased from 7.8 to 11.0 per 100,000 person-years, while from 2005-2008, the age-adjusted incidence rate changed very little (15.4 to 15.5 per 100,000 person-years, respectively).

Discussion

The Benign Brain Tumor Cancer Registries Amendment Act (Public Law 107-260) has had a profound impact on non-malignant brain and CNS tumor incidence patterns in the United States. The study findings indicated substantial changes in non-malignant-specific reporting across the time period 1997-2008, particularly for the time period surrounding implementation of the law in diagnosis year 2004. A significant increase in the age-adjusted incidence of all primary and malignant brain and CNS tumors in the United States before the early 2000s has been noted by others.[5-16,18] Studies which have included data after this time period have reported flat or downward trends in the age-adjusted incidence of malignant brain tumors.[19,20] Many of these previous studies only included data on malignant brain tumors and those studies that did include non-malignant tumors reported data prior to diagnosis year 2004 and, therefore, do not reflect the impact of Public Law 107-260. Much of the large increasing trend in incidence of non-malignant brain tumors prior to 2004 was likely attributable to factors associated with refinement of standards, variable reporting requirements, and legislative inconsistencies that influenced case ascertainment. As mentioned previously, coding and classification changes for brain and CNS tumors were implemented during this time. Alternatively, some of the increase in incidence may be related to environmental exposures, diet, or other factors that could not be assessed in this data analysis.

The increasing trend in brain and CNS tumor age-adjusted incidence between 2002-2005 seen in this study is reminiscent of the increase in brain tumor incidence reported after the introduction of CT scans and MRIs.[11,12,14,15,27-29] This increasing trend in reporting of non-malignant brain tumor incidence most likely reflects many dynamic factors and an enormous amount of activity in the cancer registry community preparing for and adapting to the new legislation targeted for implementation in diagnosis year 2004.

Although the collection of non-malignant brain and CNS tumors was voluntary prior to 2004, among all CBTRUS collaborating state cancer registries, some actively collected data on non-malignant tumors, while others passively collected data on these tumors[30,31] At least 1 state cancer registry collected data on non-malignant brain tumors but did not collect data on non-malignant spinal cord tumors.[30] In addition, tumors that were not histologically confirmed may not have been required to be reported to the state cancer registry.[30] As a large percentage of non-malignant brain and CNS tumors are not histologically confirmed, but rather diagnosed by radiography or other non-invasive means,[31] this resulted in an underreporting of non-malignant tumors. It is apparent that data collected prior to 2004 significantly underestimated the true incidence of non-malignant brain tumors. It is likely that some continued under-reporting in the years directly following enactment of the law (eg, diagnosis year 2004) occurred as the state cancer registries worked to ensure reporting from all sources.

Looking at its data from 2004-2007, the NAACCR Data Use and Research Committee Data Assessment Work Group involving benign/borderline brain and ONS tumors reported at the NAACCR Annual Meeting in 2011 that incomplete data for non-malignant brain tumors are likely to be found in NAACCR central registries especially for states with low rate ratios and low rates for non-malignant brain tumors.[31] The possible underreporting of cases detected radiographically without microscopic examination has also been noted in a study of intracranial meningiomas in Denmark, Finland, Norway, and Sweden diagnosed between 1968-1997.[32] More recently, an 18% increase in reporting of non-malignant brain tumors through the use of electronic capture of radiology reports was reported by a single institution.[33]

The relatively constant non-malignant brain and CNS tumor incidence rates during 2005-2008 suggest stabilization in reporting under the Act's governance. Current collection of non-malignant brain and CNS tumors in the United States as reflected in diagnosis years 2004-2008 has been guided by Uniform Data Standards and under 1 federal law. State cancer registries are now required to actively collect data on all brain and CNS tumors (ICD-O-3 codes C70.0-72.9 and C75.1-75.3) regardless of behavior and method of diagnostic confirmation. Quality control measures to ensure complete ascertainment of brain and CNS tumors, especially non-malignant tumors, will continue to be essential.

In summary, under mandatory collection with standardized reporting requirements, it is believed that the reported age-adjusted incidence of non-malignant brain and CNS tumors in the United States is more closely reflecting the "true" incidence. Given the findings of the study, it should also be emphasized that any evaluation of trends in non-malignant or total brain and CNS tumors must be made cautiously, and only if a registry can satisfy the high-quality

standards for diagnosis years prior to implementation of the law in 2004. Trends in malignant brain and CNS tumors may be evaluated from earlier years depending upon the completeness of case ascertainment of the respective data set.

Acknowledgements

The authors gratefully acknowledge the collaborators at state registries who provided data for this analysis through the NPCR-CSS mechanism: Georgia Yee, Arizona Cancer Registry; Randi Rycroft, Colorado Central Cancer Registry; Betsy Cromartie, Delaware Cancer Registry; Stacey Carson, Cancer Data Registry of Idaho; Molly Schwenn, Maine Cancer Registry; Susan Gershman, Massachusetts Cancer Registry; Sally Bushhouse, Minnesota Cancer Surveillance System; Debbi Lemons, Montana Central Tumor Registry; Maria Schymura, New York State Cancer Registry; Karen Knight, North Carolina Central Cancer Registry; Jim Martin, Virginia Cancer Registry; and to Reda Wilson, Program Epidemiologist, National Program of Cancer Registries.

References

1. Davis FG, Bruner JM, Surawicz TS. The rationale for standardized registration and reporting of brain and central nervous system tumors in population-based cancer registries. *Neuroepidemiology*. 1997;16(6):308-316.
2. Fulton J, Kruchko C. Requiring the Reporting of Non-malignant Tumors of the Brain and Central Nervous System. *J Registry Manage*. 1999;26(1):20-23.
3. Davis FG, Malinski N, Haenszel W, et al. Primary brain tumor incidence rates in four United States regions, 1985-1989: a pilot study. *Neuroepidemiology*. 1996;15(2):103-112.
4. McCarthy BJ, Schellinger KA, Propp JM, Kruchko C, Malmer B. A Case for the Worldwide Collection of Primary Benign Brain Tumors. *Neuroepidemiology*. 2009;33(3):268-275.
5. Jukich PJ, McCarthy BJ, Surawicz TS, Freels S, Davis FG. Trends in incidence of primary brain tumors in the United States, 1985-1994. *Neuro-Oncol*. 2001;3(3):141-151.
6. Hoffman S, Propp JM, McCarthy BJ. Temporal trends in incidence of primary brain tumors in the United States, 1985-1999. *Neuro-Oncol*. 2006;8(1):27-37.
7. Ahsan H, Neugut AI, Bruce JN. Trends in incidence of primary malignant brain tumors in USA, 1981-1990. *Int J Epidemiol*. 1995;24:1078-1085.
8. Blair V and Birch JM. Patterns and temporal trends in the incidence of malignant disease in children: II. Solid tumours of childhood. *Eur J Cancer*. 1994; 30A:1498-1511.
9. Bunin GR, Feuer EJ, Witman PA, Meadows AT. Increasing incidence of childhood cancer: Report of 20 years experience from the Greater Delaware Valley Pediatric Tumor Registry. *Paediatr Perinat Epidemiol*. 1996;10:319-338.
10. Gurney JG, Davis S, Severson RK, Fang JY, Ross JA, Robison LL. Trends in cancer incidence among children in the US. *Cancer*. 1996;78:532-541.
11. Desmeules M, Mikkelsen T, Mao Y. Increasing incidence of primary malignant brain tumors: Influence of diagnostic methods. *J Natl Cancer Inst*. 1992;84:442-445.
12. Legler JM, Gloeckler Ries LA, Smith MA, et al. Brain and other central nervous system cancers: Recent trends in incidence and mortality. [erratum in *J Natl Cancer Inst*. 1999;91:1693]. *J Natl Cancer Inst*. 1999;91:1382-1390.
13. Polednak AP. Time trends in incidence of brain and central nervous system cancers in Connecticut. *J Natl Cancer Inst*. 1991;83:1679-1681.
14. Radhakrishnan K, Mokri B, Parisi JE, O'Fallon WM, Sunku J, Kurland LT. The trends in incidence of primary brain tumors in the population of Rochester, Minnesota. *Ann Neurol*. 1995;37:67-73.
15. Smith MA, Freidlin B, Ries LA, Simon R. Trends in reported incidence of primary malignant brain tumors in children in the United States. *J Natl Cancer Inst*. 1998;90:1269-1277.
16. Werner MH, Phuphanich S, Lyman GH. The increasing incidence of malignant gliomas and primary central nervous system lymphoma in the elderly. *Cancer*. 1995;76:1634-1642.
17. Deorah S, Lynch CF, Sibenaller ZA, Ryken TC. Trends in brain cancer incidence and survival in the United States: Surveillance, Epidemiology, and End Results Program, 1973 to 2001. *Neurosurgical Focus*. 2006;20(4):E1.
18. Saika K, Hirabayashi Y. Comparison of time trends in brain central nervous system incidence (1973-97) in East Asia, Europe, and USA, from Cancer Incidence in Five Continents Vols IV-VIII. *Jpn J Clin Oncol*. 2008;38(9):650-652.
19. Inskip PD, Hoover RN, Devessa SS. Brain cancer incidence trends in relation to cellular telephone use in the United States. *Neuro-Oncol*. 2010;12(11):1147-1151.
20. Kohler BA, Ward E, McCarthy BJ, et al. Annual report to the nation on the status of cancer, 1975-2007, featuring tumors of the brain and other nervous system. *J Natl Cancer Inst*. 2011;103(9):714-736.
21. Kleihues P, Cavenee WK, eds. Pathology and Genetics of Tumours of the Nervous System, 2nd Edition. Lyon: IARC Press; 2000.
22. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Perkin DM, Whelan S (eds): International Classification of Diseases for Oncology, 3rd edition. Geneva: World Health Organization; 2000.
23. McCarthy BJ, Surawicz T, Bruner JM, Kruchko C, Davis F. Consensus conference on brain tumor definition for registration, *Neuro-Oncol*. 2002;4:134-145.
24. Surveillance Research Program, National Cancer Institute SEER*Stat software version 7.0.9.
25. Joinpoint Regression Program [computer program]. Version 3.5.2. Statistical Research and Applications Branch, National Cancer Institute; 2011.
26. Helseth A. The incidence of primary CNS neoplasms before and after computerized tomography availability. *J Neurosurg*. 1995;83:999-1003.
27. Steinberg EP. The status of MRI in 1986: Rates of adoption in the United States and elsewhere. *Am J Roentgenol*. 1986;147:435-455.
28. Steinberg EP, Sisk JE, Locke KE. X-ray CT and magnetic resonance imagers: Diffusion patterns and policy issues. *N Engl J Med*. 1985;313:859-864.
29. CBTRUS. 2007-2008 *Statistical Report: Primary Brain Tumors in the United States, 2000-2004*. Hinsdale, IL: Brain Tumor Registry of the United States; 2008.
30. CBTRUS (2012). CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2008 (March 23, 2012 Revision). Source: Central Brain Tumor Registry of the United States, Hinsdale, IL. Available at: **http://www.cbtrus.org**. Accessed May 15, 2012.
31. DURC Data Assessment Work Group, Huang, B. Benign/Borderline Brain and ONS Tumors in the NAACCR Data [Presentation]. NAACCR Annual Meeting; 2011.
32. Klaeboe L, Lonn S, Scheie D, et al. Incidence of intracranial meningiomas in Denmark, Finland, Norway and Sweden, 1968-1997. *Int J Cancer*. 2005;117:996-1001.
33. March S, Cernile G, West K, Borhani D, Fritz A, Brueckner P. Application of Automated Pathology Reporting Concepts to Radiology Reports. *J Registry Manage*. 2012;39(3):95-100.

Original Article

An Automated Algorithm for Consolidating Dates of Diagnosis from Multiple Sources

Xiuling Zhang, PhD[a]; Amy R. Kahn, MS, CTR[a]; Francis P. Boscoe, PhD[a]; Paul M. Buckley, CTR[a]

Abstract: Background: Multiple dates of diagnosis are often received from different reporting sources at a central cancer registry. Resolving these inconsistencies can be a labor-intensive task. To our knowledge, no algorithms for the consolidation of diagnosis dates have been published. We present such an algorithm here. Methods: The algorithm uses a "take the best" heuristic approach, incorporating the reported dates of diagnosis, class of case, service type (a New York-specific item similar to type of reporting source), and the date of first contact. The algorithm was evaluated by comparing results to those obtained with manual review by experienced certified tumor registrars (CTRs). Results: From a sample of 209,907 tumors with multiple diagnosis dates reported to the New York State Cancer Registry (NYSCR), the algorithm determined a single date for 94.7% of these, with the balance designated for manual review. Of a sample of 636 tumors that were manually reviewed to evaluate the algorithm, the algorithm obtained the same year as the CTRs for 621 tumors (97.6%), the same month and year for 572 tumors (89.9%) and the same month, year, and day for 518 tumors (81.4%). There was much lower agreement between the manually derived dates and the originally consolidated dates. Conclusion: The algorithm presented here is accurate, efficient, and reliable, and hopefully will help the cancer registry community move toward standard practices for record consolidation.

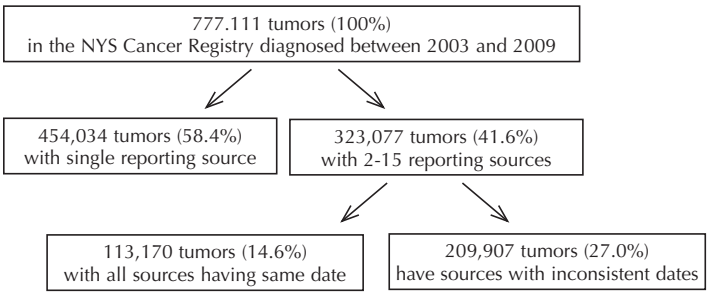
Key words: *date of diagnosis, record consolidation, algorithm*

Introduction

The date of diagnosis is a fundamental cancer registry data element. It is formally defined as the date of initial diagnosis for a tumor by a recognized medical practitioner, whether clinically or microscopically confirmed.[1] Each tumor should have a single valid date of diagnosis, but multiple dates are often reported owing to differing availability of diagnostic information and interpretations of coding rules by abstractors. Accurate and complete dates of diagnosis are essential for calculating survival, evaluating the comparative effectiveness of treatments, and assessing treatment delay.[2,3]

In the New York State Cancer Registry (NYSCR), 27% of the tumors diagnosed from 2003-2009 had more than 1 date of diagnosis reported (Figure 1). One way to resolve these differences is to review each case manually. This approach is common in central cancer registries but it is labor-intensive, time-consuming, and potentially subject to human error. A more efficient solution is to automate the process using an algorithm. This has the advantages of requiring minimal resources and being consistent and reproducible, though potentially at the expense of accuracy.

Figure 1. Characteristics of Reporting of Date of Diagnosis in the NYSCR



This paper reports on an automated consolidation algorithm developed by the NYSCR to select from among multiple reported dates of diagnosis. The algorithm was evaluated by comparing the algorithm-derived dates to existing consolidated dates and to dates manually derived by senior coding staff. The results suggest that the algorithm is sufficiently accurate to be employed on a routine basis.

Materials and Methods

A total of 777,111 tumors diagnosed from 2003-2009 were included in the analysis. Of these, 58% (454,034) had a single reporting source and 42% (323,077) had 2 or more sources. Of the tumors with multiple reporting sources, 35% (113,170) had identical dates of diagnosis for all of the sources and 65% (209,907) had non-identical dates. The tumors with non-identical dates accounted for 27% of the overall total. Further characteristics of these dates are given in Table 1.

Algorithm Logic

The algorithm consists of 6 basic steps (Table 2). It is an example of a "take the best" heuristic, wherein each step is applied in sequence for each tumor, and the process is halted as soon as a discrimination is made.[4,5] These steps were arrived upon following a lengthy process of trial and error and testing. The algorithm uses the following source-level data items: date of diagnosis (NAACCR data item

[a]New York State Cancer Registry, New York State Department of Health, Albany, NY.

Address correspondence to Xiuling Zhang, PhD, 150 Broadway, Suite 361, Albany, NY 12204-2719. Email: xxz04@health.state.ny.us.

Table 1. Characteristics of Non-identical Dates of Diagnosis Reported to the NYSCR

Completeness	
All dates incomplete	270 (0.1%)
Some dates incomplete	36,593 (17.4%)
All dates complete	173,044 (82.4%)
Time differences	
Different year	36,932 (17.6%)
Different month	101,118 (48.2%)
Different day	71,857 (34.2%)

390), class of case (item 610), date of first contact (item 580), and service type (a New York-specific item very similar to type of reporting source, item 500), along with the currently consolidated date of diagnosis.[1]

The first step removes unusable and redundant sources. Unusable sources include all those with unknown month and/or year, while redundant sources are those from the same facility with an identical class of case and date of diagnosis. For example, consider a tumor that has been reported to the registry 6 times: facility A reported date 1 once; facility B reported date 2 twice, using 2 different classes of case; and facility C reported date 1 twice and date 3 once, using the same class of case each time. Here 1 source would be removed, date 1 from facility C. The second step identifies some specific situations requiring manual review, while the third step identifies cases where the date ambiguity has been resolved by the prior steps.

The fourth and fifth steps prioritize certain class of case and type of reporting source combinations. In step 4, we prioritize certain non-analytic sources, namely those with class of case codes of 43 (pathology or other lab specimens only) or 30 (initial diagnosis and all first course treatment elsewhere and reporting facilities participated in diagnostic workup [for example, consult only, staging workup after initial diagnosis elsewhere]), or with missing class of case codes but with the service types of laboratory only, consult only or port catheter (the latter 2 are New York-specific items). The dates from these sources are usually complete dates since they are the specimen receipt dates. Step 5 categorizes the remaining sources into 1 of 3 tiers (Table 3). Those with rank 1 were assigned based on the likelihood that the cancer diagnosis probably took place in the reporting facility. Sources with rank 2 were assigned based on the likelihood that the treatment, but not the diagnosis, took place in the reporting facility. The remaining sources were assigned rank 3 since neither diagnosis nor treatment took place at the reporting facility. The final step simply consists of designating any remaining cases for manual review. The algorithm was coded using SAS version 9.2 (SAS Institute, Cary, NC, 2008); this code is available from the authors.

Algorithm Validation

Algorithm-derived dates of diagnosis were compared to the existing consolidated dates and categorized into those that agreed and those that disagreed, stratified by year, month, and day. In general, the existing consolidated dates reflected the date on the first report that was received for a tumor, though some had been consolidated to other dates. An experienced certified tumor registrar (CTR) was enlisted to manually consolidate a sample of 636 tumors, 225 where the algorithm-derived and originally consolidated dates agreed, and 411 where they disagreed. The sample was chosen to be roughly representative of cases resolved in steps 3 through 5 of the algorithm. A second CTR repeated the manual consolidation for a large subsample of the

Table 2. Steps in the Algorithm for Consolidating Date of Diagnosis

Step	*Description of consolidation procedure*	*% of tumors resolved*
1	Disregard all sources that have an unknown month and/or year of diagnosis. Discard all redundant sources (those from the same facility with the same class of case and date of diagnosis).	
2	Identify certain cases requiring manual review:	
	a. There are no remaining sources, because all were discarded in step 1 due to unknown month and/or year	a. 0.1%
	b. The consolidated date of diagnosis does not equal any of the reported dates.	b. 1.6%
	c. The difference between the earliest and latest reported date of diagnosis is 5 or more years.	c. 0.4%
3	If there is a single date of diagnosis across all remaining sources, use it as the consolidated value.	10.70%
4	If a source has class of case of 43, 30, or missing, and a service type of laboratory only, consult only, or port catheter, and it is the earliest date of diagnosis reported, then use it as the consolidated value.	10.50%
5	Assign source priority ranks using Table 3:	
	a. If the sources with rank 1 have a single date of diagnosis, use it as the consolidated date.	a. 48.6%
	b. If there are 2 sources with rank 1 with non-identical dates of diagnosis, use the earlier date as the consolidated value, even if the day is missing. Where the month and year agree, choose a complete date over an incomplete date.	b. 8.6%
	c. If there are 3 or more sources with rank 1 with non-identical dates of diagnosis, then require manual review.	c. 0.3%
	d. e. f. Repeat steps 5a-5c for rank 2.	d. 7.3% e. 9.0% f. 1.8%
6	Manually review all remaining cases.	1.00%

Table 3. Priority Ranks of Reporting Sources

Rank	*Class of case code*	*Service type (when class of case code is missing)*
1	00, 10-14, 34,35 (prior to version 12: 0, 1, 4)	Inpatient, non-NY case, private medical practitioner (office visit), laboratory followback
2	20-22, 36, 37, 40-42, 32* (prior to version 12: 2, 6, 3*)	Outpatient, clinic (within facility), ambulatory care center, radiation treatment only, DCO/followback
3	43, 30, 99, 38, 31, 33, 32† (prior to version 12: 7, 9, 5, 3†)	Laboratory—within facility, consult only, port catheter

*If the first contact date is within 60 days of the date of diagnosis

†If the first contact date is more than 60 days after the date of diagnosis

Table 4. Percent Agreement Between Algorithm-derived and Manually Derived Diagnosis Date

Step	*Number of tumors reviewed*	*Percent agreement by year*	*Percent agreement by year, month*	*Percent agreement by year, month, day*
3	70	100	100	97.1
4	45	95.6	88.9	77.8
5a	252	98	92.5	87.3
5b	103	98.1	92.2	77.7
5d	98	93.9	84.7	73.5
5e	68	100	75	63.2
Overall	636	97.6	89.9	81.4

tumors, and a third individual manually consolidated all tumors where the 2 CTRs disagreed. The percent agreements between the manually consolidated, algorithm-derived, and existing consolidated dates were calculated, with the manually consolidated values taken to be the gold standard.

Results

The percentage of tumors resolved by each step of the algorithm is given in Table 2. Overall, the algorithm provides a consolidated value for 94.7% of the tumors, with nearly half of all tumors resolved in step 5a.

For tumors consolidated by both CTRs, they agreed in 220 of 266 instances (83%). For the 46 cases in which they disagreed, the level of disagreement was by year for 5 tumors, by month for 20 tumors, and by day for 21 tumors. For these 46 tumors, the third reviewer was called in as a "tie-breaker". The third reviewer agreed with the first CTR 19 times, the second 22 times, and neither 5 times.

Taking the manually derived date as the gold standard, it agreed exactly with the algorithm-derived date for 518 of 636 tumors (81%), and with the existing consolidated date for 260 of 636 tumors (41%). Corresponding kappa statistics[6] are 0.72 for manual versus algorithm, signifying substantial agreement, and 0.09 for manual versus existing, signifying slight agreement. Results stratified by algorithm step and level of agreement are given in Table 4. The accuracy of the algorithm-derived results is seen to decline somewhat in steps 5d and 5e, where the rank 2 sources are considered, though the accuracy is still well above chance.

Discussion

Record consolidation lacks the same level of standardization as many other aspects of cancer registration. Instead, registries rely on their own historic practices and ad hoc rules. These range from logical ("choose a known value over an unknown value") to naïve ("always use the values from the first report received") to pragmatic ("manually resolve as many of the discrepancies as our resources permit"). What they do not do is systematically choose the best possible value for the greatest possible number of cases. A NAACCR Record Consolidation Committee has been active since the late 1990s, issuing occasional reports,[7-10] but their conclusions have not found their way into standard practice. In the first of these reports, the committee specifically reviewed date of diagnosis, summarizing the consolidation methods used by 4 participating registries. These included selecting the date with fewer unknown components, replacing unknown with known components, choosing the date associated with the diagnosing hospital, and following a class-of-case hierarchy; these loosely correspond with the steps in the NYSCR algorithm. More recently, staff at the Florida Cancer Data System (FCDS) developed a set of data consolidation algorithms encompassing many data items based on the practical experience of their CTRs.[11]

A very different approach to the problem of record consolidation can be found in the computer science literature. Here, the focus has been on the problem of identifying reliable sources of information from conflicting sources on the Internet, referred to as the "veracity problem" and the "truth-finding problem."[12-13] The mathematics involved tend to be complex, but the approach can be summarized as follows: "A Web site is trustworthy if it provides many pieces of true information, and a piece of information is likely to be true if it is provided by many trustworthy Web sites."[12] An example is given of the online search query, "What is the height of Mount Everest?", which returns 4 different answers among the top 20 results; which one should be considered the most trustworthy?[12] While the analogy to cancer registration is not exact—the number of conflicting sources under consideration in a cancer record is typically 2 or 3, not the far larger numbers characteristic of the Internet—the applicability of this line of research to disease registration certainly deserves further exploration.

The algorithm we present here is generally accurate, efficient, and reproducible, but its major drawback is that it is not scalable. That is, it would be prohibitive to attempt to repeat this for the hundreds of other items collected by cancer registries. Our approach could be viewed as the "most complex" scenario: our 6-step algorithm yielded good results, but most of the value came from just 3 steps, and most of that value came from just a single step

(acknowledging that a programmer turning our logic into computer code would not necessarily count the steps in the same way). This is consistent with work on algorithms and heuristics in the field of cognition,[14] and with an evaluation of the NAACCR Hispanic Identification Algorithm (NHIA), which found that nearly all of the accurate results can be obtained by ascertaining 2 pieces of information: 1) was the patient born in a Spanish-speaking country and 2) does the patient have a last name that is heavily Hispanic, according to the United States Census.[15] This suggests that the broad approach taken in Florida is likely to be effective, even though it has not yet been rigorously evaluated.

In summary, the algorithm we describe here resolved nearly 95% of the diagnosis date discrepancies and got the "correct" answer 81% of the time, acknowledging that our coders only agreed on what this correct answer was 83% of the time. In addition to being automated and generally accurate, the algorithm has the advantage of being completely reproducible, thereby entirely eliminating the problem of inter-coder disagreement (excepting the approximately 5% of cases we designate for manual review). In addition, the effects of any proposed modifications to the algorithm—deciding to alter the order of the steps, for example—can be rapidly quantified and assessed. We believe that requiring manual review of 5% of the caseload is both manageable and the appropriate way of handling the most ambiguous and unusual situations, but even applying the automated rules to some or all of these cases would not have had a major impact on the overall results. We hope that this algorithm will advance the discussion about record consolidation generally and will represent a step toward wider adoption of standard practices.

Acknowledgements

We acknowledge Beth Bernard and Laura Soloway of the NYSCR for their assistance in reviewing cases. This work was supported in part by the Centers for Disease Control and Prevention's Cooperative Agreement U58/DP000783, awarded to the New York State Department of Health through the National Program of Cancer Registries. It would not have been possible without the diligence of the dedicated certified tumor registrars of the hospitals of New York or the coding of the New York State Cancer Registry.

References

1. Thornton M, O'Connor L, eds. *Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Record Layout Version 12*. 14th ed. Springfield, IL: North American Association of Central Cancer Registries; 2009.
2. Woods LM, Rachet B, Ellis L, Coleman MP. Full dates (day, month, year) should be used in population-based cancer survival studies. *Int J Cancer.* 2012;131(7):E1120-E1124.
3. Tate AR, Martin AGR, Murray-Thomas T, Anderson SR, Cassell JA. Determining the date of diagnosis—is it a simple matter? The impact of different approaches to dating diagnosis on estimates of delayed care for ovarian cancer in UK primary care. *BMC Med Res Methodol.* 2009;9:42.
4. Gigerenzer G, Goldstein DG (1999) Betting on one good reason: the Take the Best heuristic. In: Gigerenzer G, Todd PM, the ABC Research Group, eds. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press; 75-95.
5. Lee MD, Zhang S. Evaluating the coherence of take-the-best in structured environments. *Judgment and Decision Making*. 2012;7(4):360-372.
6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
7. Ross F, et al. *Report of the Record Consolidation Committee*. Springfield, IL: North American Association of Central Cancer Registries; 1999.
8. Braun J, Ross F, et al. Report of the Record Consolidation Committee to the Board of Directors of NAACCR. Submitted March 1, 2000. Springfield, IL: NAACCR, 2000.
9. Braun JE, Ross FE, Capron SJ, eds. *Creation of a Record Consolidation Test File: Report to the NAACCR Board*. Springfield, IL: North American Association of Central Cancer Registries; 2003.
10. Ross FE and Capron SJ. *Record Consolidation Test with the 2007 Multiple Primary/Histology Rules: Report to the NAACCR Board.* Springfield, IL: North American Association of Central Cancer Registries; 2008.
11. Levin G, Scharber W, Herna M, Stearns P, Peace S. Automated Tumor Consolidation: The Florida Algorithm. Proceedings, North American Association of Central Cancer Registries Annual Conference, Portland; June 2012.
12. Yin X, Han J, Yu PS. Truth Discovery with Multiple Conflicting Information Providers on the Web. *IEEE Transactions on Knowledge and Data Engineering* 2008;20(6):796-808.
13. Zhao B, Rubinstein BIP, Gemmell J, Han J. A Bayesian Approach to Discovering Truth from Conflicting Sources for Data Integration. Proceedings of the VLDB Endowment. 2012;5(6):550-561.
14. Gigerenzer G, Todd PM, the ABC Research Group, eds. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press; 1999.
15. Boscoe FP, Schymura MJ, Zhang X, Kramer RA. Heuristic algorithms for assigning Hispanic ethnicity. *PLoS ONE*. In press.

Original Article

Cancer Registry Enrichment via Linkage with Hospital-Based Electronic Medical Records: A Pilot Investigation

Monique N. Hernandez, PhD[a]; Lydia Voti, DSc, MS[a]; Jason D. Feldman, MPH[b]; Stacey L. Tannenbaum, PhD[c]; Wendy Scharber, RHIT, CTR[d]; Jill A. MacKinnon, PhD[a]; David J. Lee, PhD[a,e]; Youjie X. Huang, MD, MPH, DrPH[b]

***Abstract*:** Background: Hospital electronic medical record (EMR) systems are becoming increasingly integrated for management of patient data, especially given recent policy changes issued by the Centers for Medicaid and Medicare Services. In addition to data management, these data provide evidence for patient-centered outcomes research for a range of diseases, including cancer. Integrating EMR patient data with existing disease registries strengthens all essential components for assuring optimal health outcomes. Objectives: To identify the mechanisms for extracting, linking, and processing hospital EMR data with the Florida Cancer Data System (FCDS); and to assess the completeness of existing registry treatment data as well as the potential for data enhancement. Methods: A partnership among the Florida Department of Health, FCDS, and a large Florida hospital system was established to develop methods for hospital EMR extraction and transmission. Records for admission years between 2007 and 2010 were extracted using ICD-9-CM codes as the trigger and were linked with the cancer registry for patients with invasive cancers of the breast. Results: A total of 11,506 unique patients were linked with a total of 12,804 unique breast tumors. Evaluation of existing registry treatment data against the hospital EMR produced a total of 5% of registry records with updated surgery information, 1% of records with updated radiation information, and 7% of records updated with chemotherapy information. Enhancement of registry treatment information was particularly affected by the availability of chemotherapy medications data. Conclusion: Hospital EMR linkages to cancer disease registries is feasible but challenged by lack of standards for data collection, coding and transmission, comprehensive description of available data, and the exclusion of certain hospital datasets. The FCDS standard treatment data variables are highly robust and complete but can be enhanced by the addition of detailed chemotherapy regimens that are commonly used in patient centered outcomes research.

Key words: *cancer registries, hospital electronic medical record, data linkage*

Introduction

The electronic medical record (EMR) was introduced in the late 1960s by Lawrence L. Weed to provide physicians with easily organized patient records and to enhance utilization of these records for improved patient care. Subsequently in 1972, the Regenstrief Institute developed a structured modular system of EMRs that is still in existence today, for the purpose of integrating information to clinics, laboratories, radiology departments, and pharmacies.[1] By 1998, a number of institutions, such as Columbia Presbyterian Medical Center in New York City and the Latter-Day Saints Hospital in Salt Lake City, had incorporated an EMR within their specific hospital system.[2] Incorporating EMRs inside institutions was expected to benefit 3 entities: 1) patients, through improved care, 2) physicians, through more efficient organization, and 3) hospitals, through enhanced quality of care and reduced cost.[2-4] The EMR systems were initially developed to improve clinical care and facilitate billing, rather than for the conduct of research. This summary mostly focuses on EMRs, which are created for patients of health facilities and reviewed by clinicians, but also includes electronic health record (EHR) systems that enable the sharing of patient health information between institutions.

The Institute of Medicine proposed the reorganization of the entire health-care information infrastructure with a technology-forward model to reduce medical errors, advance health-care delivery and quality, and progress medical and health provider research.[5] Moreover, the Patient Protection and Affordable Care Act of 2010 mandates patient-centered outcomes research requiring standard setting by encouraging uptake of EHR from all providers.[6] Between 2009 and 2011, non-federal acute care hospitals' adoption of at least a basic EMR more than doubled from 16% to 35%, and more than 85% of hospitals intend to attest to Meaningful Use of an EHR under the Centers for Medicaid and Medicare Services (CMS) Incentive Program by 2015.[7]

Recent studies incorporating EMRs in cancer research represent the emergence of a new field of investigation.

[a]Florida Cancer Data System, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine. [b]Florida Department of Health, Bureau of Epidemiology. [c]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine. [d]Registry Widgets, Inc. [e]Department of Epidemiology and Public Health, Miller School of Medicine, University of Miami.

Address correspondence to Monique N. Hernandez, PhD, Florida Cancer Data System, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, P.O. Box 016960 (D4-11), Miami, FL 33101. Telephone: (305) 243-9673. Fax: (305) 689-7057. Email: Mhernandez5@med.miami.edu.

Funded through the Agency for Healthcare Research and Quality Multiple Chronic Conditions Research Network, grant number R24 H2019658-01.

Denny, et al compared EMR data collection to manual collection in order to identify individuals in need of colorectal cancer (CRC) screening.[8] The EMR algorithm employed natural language processing which could: 1) detect specific terminology, 2) categorize decision-making for CRC screening, and 3) schedule testing more successfully than manually abstracted information.[8] EMRs at one Veteran's Hospital were used to determine if newly diagnosed CRC patients had been properly screened for the presence of Lynch syndrome, a genetic condition which increases the risk for CRC.[9] Results indicated that most possible cases were missed (only 7% meeting guidelines were referred for genetic evaluation) and that there was inadequate information in the medical record to make a more precise determination in approximately half of the cases. In another study, investigators were able to take advantage of the EMR system to identify at-risk patients in whom an elevated PSA level was recorded but who, otherwise, would not have been referred to a urologist.[10]

Use of the EMRs to link to other databases has also recently been undertaken by investigators. A 2012 study linked the General Practice Research Database, which contains computerized medical records maintained by general practitioners in the United Kingdom, to a Cancer Registry in England, and found no increased cancer risk for diabetic patients taking certain glucose-lowering medications.[11]

There have been only limited attempts at cancer data enrichment through use of the EMR. In one study, an oncology EMR identified the tumor stage for 4 cancers (breast, lung, colorectal, and prostate), which was then linked with medical and pharmacy claims data to develop algorithms for patient classification of metastatic cancer.[12] Unfortunately, there was poor predictability of metastatic cancer by this algorithm; this negative finding may have been due in part to misclassification of ICD-9-CM codes. Nevertheless, these investigators suggested that claims data could be a valid tool for future research if standardized language were utilized in the United States.[12]

Clinicians and researchers share a desire for enriched cancer registry information; however there are obstacles due to a lack of uniformity of standardized language, formatting, and transmission specifications. A 2012 survey in Alabama found that when cancer registrars were asked to rate the benefits of using EHR in the cancer registry (on a scale of 1 to 5 where 1 = strongly disagree and 5 = strongly agree), the mean score was 4.2 (standard deviation [SD], 0.8) for the benefit, "it would allow the registrar more time for retrieving and analyzing data for clinicians and researchers."[13] The mean score was 3.8 (SD, 0.9) for how important the impacts would be if using the EHR with the cancer registry in health-care facilities "for advancing clinical, epidemiological, and health services research."

The rationale for testing the utility of enhancing the cancer registry through EMR hospital linkages is abundant. For example, discovering cases that would otherwise be missed, (casefinding), adding enriched information to the record that is too expensive and time consuming to manually collect within the hospital cancer registry (eg, specific chemotherapy dosing), and possible validation of cancer abstracts. Other positive outcomes would be the improved quality of data collection and potential for rapid case ascertainment.[14] Enriched EMRs facilitate the evaluation of patient care, and clinical and organizational efficiency, including patient-focused treatment options salient to comparative effectiveness research.[15]

The long-term vision of an EMR-linked cancer registry is the direct abstraction of certain relevant registry data. This linkage has the potential to enhance data collection efficiency combined with the ability to collect more detailed treatment information, and patient comorbidity information for use by researchers and clinicians. A secondary goal of such linkage is to establish a systematic rapid transmission of patient health data beyond the current model where facilities report cases anywhere within a 6-month interval. The objective of this project was to identify the mechanisms for extracting, linking, and processing hospital EMR data with the Florida Cancer Data System (FCDS); and to assess the potentials of EMR for enhancement of cancer registry data as well as to provide a pilot dataset for breast cancer patient-centered outcomes research.

Methods

The Florida Department of Health, Bureau of Epidemiology was awarded funding from the Agency for Healthcare Research and Quality (AHRQ) to establish a real-time linkage between the FCDS, and a hospital EMR system to extract data and enhance the medical information available for the statewide annual cancer incidence records. Included in the aims of the project was the development of an invasive breast cancer dataset for a piloted patient-centered outcomes research project. This dataset combined standard data elements from the FCDS registry and medical information from the hospital EMR. The linked dataset was intended to advance research on complex patients within the AHRQ Multiple Chronic Conditions Research Network.[16] This project was approved by the Florida Department of Health Institutional Review Board.

The Florida Cancer Data System

The FCDS (**http://fcds.med.miami.edu/**) is Florida's statewide, population-based cancer registry and has been collecting incident cancer data since 1981. FCDS is wholly supported by the State of Florida Department of Health, the National Program of Cancer Registries of the Centers for Disease Control and Prevention and the Sylvester Comprehensive Cancer Center at the University of Miami Miller School of Medicine. Under existing Florida statutes, all licensed hospitals, radiation therapy centers, laboratories, and ambulatory surgical facilities are required to report annual cases of cancer to the FCDS. Reported information includes patients' demographics, methods and results of diagnostic tests, and first course of treatment. Based on the most recent Florida Annual Cancer Report, there were 103,075 new primary cancers diagnosed among Florida's 17.5 million residents, at an age-adjusted rate of 441.2 per 100,000 people.[17]

Table 1. Hospital EMR Data Elements Identified for Central Registry Linkage and Validation		
NAACCR Item No.*	*Cancer Registry Data Item*	*Hospital EMR—Code System*
390	Date of Diagnosis	Admissions date
400	Primary Site	ICD-9 Principal and Secondary Diagnosis
1200	RX Date—Surgery	ICD-9 Proc/CPT Date
1210	RX Date—Radiation	ICD-9 Proc/CPT Date
1220	RX Date—Chemo	ICD-9 Proc/CPT Date
1230	RX Date—Hormone	ICD-9 Proc/CPT Date
1240	RX Date—BRM	ICD-9 Proc/CPT Date
1250	RX Date—Other	ICD-9 Proc/CPT Date
1285	RX Summ—Treatment Status	ICD-9 Proc/CPT/REV
1290	RX Summ—Surg Prim Site	ICD-9 Proc/CPT/REV
1292	RX Summ—Scope Reg LN Sur	ICD 9 Proc
1294	RX Summ—Surg Oth Reg/Dis	ICD 9 Proc
1360	RX Summ—Radiation	ICD-9 Proc/CPT/Radiation
1380	RX Summ—Surg/Rad Seq	ICD-9 Proc/CPT Dates -- Derived
1390	RX Summ—Chemo	ICD-9 Proc/CPT/medications
1400	RX Summ—Hormone	ICD-9 Proc/CPT/medications
1410	RX Summ—BRM	ICD-9 Proc/CPT
1420	RX Summ—Other	ICD-9 Proc/CPT
1570	Rad—Regional RX Modality	ICD-9/CPT/Radiation/Charge Description
1639	RX Summ—Systemic/Sur Seq	ICD-9 Proc/CPT—Derived
340	Tobacco History	Patient reported use; none, current, within past year, greater than 1 year ago
Florida-specific	Height	Not specific
Florida-specific	Weight	Not specific
230	Age at DX	Number
3110-3164	Comorbid Complications	ICD-9 CM/Health History

*North American Association of Central Cancer Registries.

Description of the Hospital System EMR

The Florida Department of Health and FCDS partnered with an American College of Surgeon accredited, Florida hospital system for EMR data linkage to the central cancer registry. The hospital system consisted of multiple hospital facilities, and had recently undergone implementation of a system-wide EMR project. The hospital's EMR project was a 7-year, 3-phase initiative, which began in early 2006 with the implementation of the project's Clinical Foundation: pharmacy, imaging, surgery, and clinical data. The recently completed phase II portion of the project included implementation of order management, nursing documentation, and Emergency Department Computerized Physician Order Entry (ED CPOE). Phase III extends CPOE to inpatients, included evidence based practice, and integrated plans of care. At the time of the cancer registry linkage, the hospital system had 2 database systems, a legacy health information management system and a newly adopted system being phased in over multiple years. The legacy system maintained non-clinical billing information and assigned a unique corporate patient identifier (CPI) to each patient to capture their associated encounters, or episodes of care within the hospital system. Due to phased implementation, data linked to the central registry included tables with records for partial years, such as pathology, and medication order and administration records. Others were complete datasets for all requested admission years.

The medication orders collected for this linkage spanned from October 2007 through July 2012, and comprised more than a million entries for all drug records that clinicians added to a patient medical chart. A medication may be ordered for a particular patient and never be given to that patient for a variety of sound clinical reasons. We restricted our medication data collection to 15 record order actions that reflected patterns of presumptive drug administration from a total of 28 possible record types that included: complete, resume, refill, cancel, suspend, and future discontinue. Medication administration reconciliation (MAR) records were also incorporated in the fourth quarter of 2009 to detail

Table 2. Hospital EMR Data Elements for Enhancement	
Data Table	*Data Elements*
Health history	Medical history I and II, patient self-reported family history, OB/Gyn, surgical
Medications: ordered and administered	Name, order date, strength dose and unit, route of administration, drug form
Pharmacy drug orders	Source order ID, Dispense date, order Mnemonic Ingredient, route of administration, ingredient catalog Mnemonic
Radiation therapy	Date, daily dose, cumulative dose, prior dose, total dose, physician notes, treatment number
Discharge instructions	Discharge instructions and values, including clinicians' observational notes relevant to comorbidities and complications, date of instructions
Charge codes	Charge description, Charge SVC code, service dates, CPT codes, REV codes
Pathology	Study description, specimen, report type, report
Comorbid complications	ICD-9-CM codes for principal and secondary diagnosis (up to 30 conditions)

drug doses, start and end times for patient bedside medication treatment, and their site of administration, such as "intravenous tube insertion to patient's left anterior chest." MAR records for anti-cancer drugs and neo-adjuvant breast cancer agents were filtered to isolate chemotherapy patients. All medication records contained a source order ID which was used with CPI numbers to map out the scope of treatments patients received within various hospital settings in order to review characteristics of care units and number of admissions to facilities.

Study Design

We captured hospital EMR data that primarily consisted of a) patient demographics, b) patient diagnostic information, including patient health histories, comorbid conditions, height, and weight, and c) treatment information, including detailed radiation, surgical, and medication treatment modalities; discharge summaries; and pathology reports for the invasive breast cancer pilot study. Using data standards issued by the North American Association of Central Cancer Registries,[18] the FCDS and hospital staff identified the most relevant hospital EMR data elements that have the potential to be used to derive and validate standard diagnostic and treatment information in the central cancer registry (Table 1). Additional hospital data elements were identified for enhancement of registry records, primarily used for the breast cancer pilot study (Table 2). The objective of the registry enhancement study was to identify registry records where first course of treatment could be updated or enhanced by hospital diagnostic and procedure data in order to evaluate the potential impact of hospital EMR data on existing registry treatment data.

Case Definition

Case identification was triggered by hospital admission records with a principal or secondary ICD-9-CM diagnostic code that met criteria for invasive and reportable cancers for admission years between 2007 and 2010.[19] The rationale behind the inclusion criteria were 1) the timing of the hospital EMR implementation schedule, and 2) to capture any hospital encounter related to invasive cancers, either diagnosed, treated or reported as a comorbid condition. Hospital-generated patient and encounter identifiers for records that met criteria were then used to query across hospital EMR data for extraction of detailed patient data.

Linkage with Central Registry

Hospital EMR data tables were transmitted to the FCDS through a secure file transfer protocol. At the time of data extraction the hospital system did not have the capability to transmit data in the HL7 Clinical Document Architecture (CDA), an XML-based markup standard that specifies encoding, structure, and semantics of clinical documents for exchange.[20] Therefore, data tables were submitted as flat files. Table layouts and data types were provided by the hospital system. For the purposes of the breast cancer pilot study, a subset of all transmitted EMR records was created consisting of admission records with a principal and/or secondary ICD-9-CM diagnosis between 174.0 and 175.9. The subset dataset was then linked to the central cancer registry using patient Social Security number, first and last name, and date of birth. Using a deterministic matching algorithm, a total of 11,506 unique patients were matched to a patient in the FCDS database, resulting in 12,804 primary tumor records and 53,940 unique hospital admission records. While the hospital EMR defined the patient dataset, all registry records for that patient were included in the final breast cancer pilot database, regardless of the reporting hospital or the date of diagnosis. This was to ensure capture of the entire diagnostic and treatment profile for each breast cancer patient as it is common among Florida cancer patients to receive diagnosis and care from multiple facilities. There were 102 non-female breast cancer patients records contributed to this dataset.

Identification of Procedure and Diagnosis Codes for Analysis

We examined all the procedure fields of the linked cancer records for information related to cancer treatment and, in particular, all codes pertaining to surgery, radiation and chemotherapy for breast cancers diagnosed in 2010. The diagnoses were coded using ICD-9 diagnostic coding scheme and the procedures were coded using the ICD-9 (procedure) and Current Procedural Terminology (CPT) coding schemes.[21,22] In general, the ICD-9-CM and CPT procedure fields contained information related to the treatment of the patient, but some information related to chemotherapy may also be included in the ICD-9-CM diagnoses fields. In the hospital admissions table there were 31 diagnosis and 21 procedure fields available in the breast-related EMRs that were transmitted and linked. The appropriate codes relevant to each of the 3 treatment modalities were identified based on the literature.[23-25] In

Table 3. Registry Codes to Identify Missing or Uninformative Treatment Values

NAACCR data item no.*	*Codes used to identify missing or uninformative values*
1290 – RX Summ	00 – None
Surg Prim Site	90 – Surgery, NOS
	99 – Unknown
1360 RX Summ Radiation	0 – None
	8 – Recommended, unknown if administered
	9 – Unknown
1390 RX Summ Chemo	0 – None
	8 – Recommended, unknown if administered
	9 – Unknown

*North American Association of Central Cancer Registries.

Table 4. Procedure Codes used for Breast Cancer Treatment Identification

Treatment	*Type of code*	*Codes*
Surgery	ICD-9	85, 85.21, 85.23, 85.41, 85.42, 85.43, 85.48,
	CPT procedures	19120, 19160, 19162, 19180, 19220, 19240, 19301, 19302, 19303, 19304, 19307, 38740, 38745
Radiation	ICD-9	92.2, 92.29
	CPT procedures	19297, 77401-77499 or 77750-77799
Chemotherapy	ICD-9	19297, 77401-77499 or 77750-77799
	ICD-9 diagnoses	V58.1, V58.11, V58.12, V66.2, V67.2
	CPT procedures	36561, 36569, 36571,36590, 36598, 96400-96549, J9000-J9999, Q0083-Q0085

addition, a frequency of all the codes was run and reviewed by clinicians, who identified additional codes to be included in the search. A complete list of all codes searched for can be found in Table 4.

Results

Of the FCDS 12,804 breast cancer cases that were linked to EMR records (11,506 patients), 1,584 cases had a breast cancer diagnosed in 2010, according to the FCDS. We focused on the cases diagnosed in 2010 and 3 treatment modalities: surgical, radiation and chemotherapy. As almost all these cases were analytic (95%), that is, they had been reported by facilities where the patients were diagnosed and/or treated, few cases had missing treatment information. Our investigation was widened to include the records

Table 5. Records with a Breast Cancer Diagnosis in 2010 (n=1,584)

FCDS records	*EMR enhancements*		
Surgical treatment	**Count**	**Percent**	**Total updated**
Missing	7	0.40%	11
No surgery	230	14.50%	
Surgery, NOS	8	0.50%	
Total	**245**		
Radiation treatment			
Missing	8	0.50%	10
No radiation	1,068	67.40%	
Recommended, Unknown if administered	7	0.40%	
Total	**1083**		
Chemotherapy treatment			
Missing	7	0.40%	81
No chemotherapy	1,040	65.70%	
Planned but not started at time of most recent follow up	44	2.80%	
Total	**1,127**		

that had no treatment or the treatment was uninformative according to FCDS (Table 3). The linked EMR records were located and all the ICD-9 procedure and CPT diagnoses fields were scanned for the codes listed in Table 4.

Surgical Treatment

For the 245 records (221 patients) that had no surgery or missing or uninformative surgery information in the FCDS data, 593 records were found in the EMR system. These records were scanned for the ICD-9 and CPT codes listed in Table 4. We found 41 records that could be used to enhance the FCDS surgical treatment data. The 41 records were reviewed and we found surgery information in the EMR dataset for 11 records that had no surgery captured by FCDS. For the rest of the records, the EMR dataset could potentially contribute new information such as procedures that were performed in addition to, or after, the surgical treatment for the breast cancer (example: excision of axillary lymph nodes, procedures indicating breast reconstruction, etc).

Radiation Treatment

For the 1,083 cases (855 patients) diagnosed in 2010, that had no radiation or missing radiation information or the radiation was recommended but unknown if administered, we searched the EMR dataset for the procedures listed in Table 4. Of these records, 10 EMR records had a procedure related to radiation therapy treatment: "Placement of radiotherapy after loading expandable catheter into the breast for interstitial radioelement application following partial mastectomy."

Table 6. Sample Chemotherapy Drug Table

Patient ID	Encounter type	Order Action type	Chemotherapy agent	Dose amount	Dose unit	Infused volume	Infused volume unit	Route of admin.	Site of admininstration
11893	Inpatient	Order	ondansetron	4	mg	0		IV	"Arm, R upper"
11923	Inpatient	Complete	dexamethasone	4	mg	1	tab(s)	PO	
12450	Inpatient	Order	ondansetron	4	mg	2	mL	IV	"Arm, R upper"
20252	Recurring	Complete	trastuzumab	457	mg	0		IV	"Anterior chest, R"
23817	Inpatient	Order	cefepime	2	g	0		IVPB	Port

Chemotherapy Treatment

According to FCDS, 1,127 cases (891 patients) had no chemotherapy or missing chemotherapy information or the chemotherapy was recommended but unknown if administered. The corresponding patients were located in the EMR dataset. We found 2,894 records for these patients, and their records were searched for procedures listed in Table 2. Of the 2,894 records found in the EMR dataset for these patients, 2,737 corresponded to 2010 admissions and 81 of them contained codes that were either specific to chemotherapy drug administration or suggestive of planned courses of treatment such as the insertion of a portacath which can be used to potentially enhance the chemotherapy information. Chemotherapy agents filtered from MAR records yielded 11,051 cases for 516 patients with detailed dosage amounts. Of those, *complete* chemotherapy drug order action records (39%) displayed administration end times which verified that those agents were actually given. A sample of the chemotherapy drug table is provided in Table 6.

Discussion

The majority of cases in the FCDS record were complete with respect to the abstraction of surgical, radiation, and chemotherapy information. Nevertheless there were instances of missing information in the FCDS record that were found in the hospital EMR, demonstrating the potential for enrichment of registry data. Examples of surgery procedures and placement of radiation and chemotherapy treatment catheters, for instance, illustrate types of clinical data captured as English text and numeric values for cumulative doses administered. Additional information not collected by registries includes clinicians' notes highlighting comorbidities or the need for special care and follow-up treatments upon discharge.

The completion of this pilot linkage revealed several challenges for routine linkage of EMR and cancer registry records. First, data elements required by the cancer registry were incomplete. Several elements needed to be collected using proxy data. Completeness of various data fields that were available was sometimes sparse.

Second, EMR system documentation was not consistently available. In addition to minimal documentation for data definitions and value sets, documentation relating to the methods in which hospital staff and or computer systems add data into the EMR was incomplete. Investment of considerable staff time from both the FCDS registry and the hospital staff who manage the EMR system was required in order to interpret data fields and to understand the strengths and limitations of each system. Of note, establishing a common terminology, ie, standardized language for system functionalities and data elements, was needed in the early phase of the study. These tasks contributed to a data analyst's workload of up to 20 hours, weekly, during the busiest periods of the project. Conservative estimates of the registry's staff hours invested in EMR data processing, scrubbing and removing of sensitive elements, and quality assurance is half of a full-time data analyst's work schedule consistently spread over the duration of the project. Because the EMR system could not provide the data using 1 of the health-care standard record layout formats, significant FCDS staff resources were required to manipulate the EMR data into formats that could be linked to the registry database.

Third, the success of hospital data linkage partnerships depends on its prioritization at the leadership level of the hospital facility. Research is not typically a priority of most hospitals. Of note, our initial plan was to perform linkages with 2 Florida hospital systems. Of the 2 hospital systems that originally participated in the project, 1 had to pull out due to lack of EMR staff to devote to the project. Although a single experience, we believe it to be indicative of challenges for routine linkages of cancer registries with busy hospital systems.

Fourth, we discovered that EMR datasets within a single hospital system are not currently well integrated. At the start of this project, we anticipated enhancing the registry's chemotherapeutic information with more detailed drug and dosing information. However, the computerized dispensing information used by the pharmacy, which dispenses chemotherapeutic agents, was not integrated into this hospital's EMR system. Consequently, we were not successful in integrating this information into an enhanced registry record.

Patient medication treatment data reflect further lack of integration when considering that these records were phased in as MAR records in 2011 on a facility-by-facility basis across the hospital network. This created a lag in detailed treatment data available in some units. On a broader level, the collection and review of EMRs for this project elucidates that there is an uncertain proportion of treatment which likely occurs outside of inpatient settings, very possibly outside of this hospital network, such as in

the case of chemotherapy patients' scheduled follow-up screenings and recurring courses of medication at different care centers.

Problems and challenges instituting EMR systems in hospitals across the country also currently limit the capability of performing routine cancer registry linkage using this emerging technology. A recent report by the Office of the National Coordinator for Health Information Technology issued in 2011 stated that of the non-federal acute care hospitals surveyed, approximately 8.8% had a comprehensive EMR, while 26% had a basic EMR with or without clinician notes.[27] Although the Institute of Medicine created a complete list of prospective functionalities for hospital patient EMRs, there is a lack of agreement as to what functionalities define the crucial components that classify hospital-based EMRs.[27] As EMR technologies continue to evolve, individual facilities may prioritize system features before any 1 set of standardized functionalities are agreed upon.

While many challenges were identified through this project, it is of importance to note that utilizing EMR data with central cancer registry data has advantages. Collecting EMR data to enhance central cancer registries is less limited by case definition selection criteria given a nationally standardized set of reportable diagnostic codes as opposed to other chronic conditions that are less defined such as diabetes and asthma.[28] Additionally, given that cancer is reportable by state legislative mandates, patient tumors identified in the hospital EMR should already exist in the central cancer registry database, enabling linkages to data that are already standardized and assessed for quality. These data can serve as validation checks against the EMR to ensure data quality and consistency as promoted by Kahn et al.[29] The EMR would serve to enhance registry records with more detailed treatment data and patient comorbid complications often utilized in comparative effectiveness research, the benefit of which is the capture of routinely collected data from real world settings as opposed to through the more costly and resource intensive randomized control trials.[28] With the forthcoming implementation of Meaningful Use Stage II by CMS, central cancer registries will be receiving both hospital and physician reports directly from EMR vendors.[30] These reporters would only include entities that choose to participate in the stage II menu option for reporting to a central cancer registry, but as CMS reimbursements are increasingly tied to public health reporting, this is expected to expand current levels of non-hospital reporting significantly. Furthermore, these data transmissions require nationally standardized specifications and will include detailed treatment information useful for comparative effectiveness research.

There were limitations in this study. Although we had planned to collect records from 2 hospitals, only 1 was examined in this study. Thus generalizability of challenges, opportunities, and findings is limited. The hospital system we partnered with is an American College of Surgeons Commission on Cancer (CoC)-accredited member. CoC accreditation is designed to enhance patient care and treatment, and includes a strong cancer surveillance component. Assigning significant resources to this study may be a reflection of the hospital's commitment to its cancer program. CoC-accredited hospitals may, on average, have more complete records submitted to the cancer registry relative to non-CoC-accredited hospitals. It is therefore possible that EMR linkage with non-CoC-accredited facilities may result in a higher yield of missed and or incomplete treatment data. Finally, we only performed our linkage on 1 type of cancer. The level of data enrichment may vary for cancers with more or less complex treatment regimens.

The size of our linked dataset reflects a unique advantage for this pilot study. The patients' data incorporated is highly representative of the overall population of Florida women in some important respects, beginning with incidence of breast cancer by race and ethnicity. Patients in our study group were 90% white, 7% black, 6% Hispanic, and 2% Asian. The 2009 breast cancer statistics for the county where the reporting hospital is located reveal that 84% were white and 12% were black, compared to 85% and 10%, respectively, for all cases in Florida.[31] Socioeconomically, 13% of patients in our study group resided in areas where greater than 20% of the population lives in poverty, the lowest socioeconomic status category in our dataset, while 16.9% of all Florida women are estimated to live in poverty.[32]

Despite the current limitations, there is potential for finding additional treatment information for certain modalities such as chemotherapy. For example, the chemotherapy data contained information on each chemotherapeutic agent used, specifying the name of the drug, the dose, the form (tablet, injection), the route of administration, and the frequency of administration, taken from MAR-level medication data. In theory, if there is a steady flow of information to the registry from an EMR system, it is possible to obtain more granular data on chemotherapy, allowing more detailed analysis of chemotherapy treatment. Therefore the utility of the medications information of the EMR to the registry needs to be investigated further.

Finally, although not the focus of this investigation, real-time linkage of cancer registries with EMR systems has the long-term potential of streamlining rapid case ascertainment and monitoring of patient quality of care outcomes.[33,34] For example, computerized programs could be written which immediately flag those who may be eligible for enrollment into clinical trials and to ensure that diverse and representative record samples could be identified for quality of care studies. Automated programs could also be designed to systematically track prescription drug dispensing and prevent complications by alerting clinicians of patients' health histories or active treatments that may trigger adverse outcomes.

Conclusion

This study demonstrated that EMRs provide detailed clinical data valuable for patient-centered outcomes research. We demonstrated that linking hospital EMR records and cancer registry records is feasible when sufficient resources are invested throughout the entire study process. The study validated that existing breast cancer registry records contained the most complete information

on the patient, tumor and treatment data, and confirmed that hospital data can expand patient clinical profiles. EMR records can be used for collection of detailed treatment information used in special studies. There is also great future potential for the routine enhancement of existing important cancer registry data fields and the collection of new fields as EMR standards are developed (eg, body-mass index, comorbidities, treatment complications, etc). Other longer term opportunities include integrated rapid case ascertainment systems and the efficient monitoring of patient quality of care outcomes. The findings of the study, along with the lessons learned provide an excellent starting point for creating significant opportunities for future public-private-partnerships, electronic health information sharing, and various applications for research.

Acknowledgements

The authors wish to thank the hospital personnel who generously gave their time to assist with data acquisition and troubleshooting. Specifically, we would like to thank Ann Davis and Bob Fulenwider from Bay Care Health System who were instrumental partners in this project.

References

1. Murray MD, Smith FE, Fox J, et al. Structure, functions, and activities of a research support informatics section. *J Am Med Inform Assoc*. 2003;10(4):389-398.
2. Sujansky WV. The benefits and challenges of an electronic medical record: much more than a "word-processed" patient chart. *West J Med*. 1998;169(3):176-183.
3. McDonald CJ, Tierney WM. Computer-stored medical records. Their future role in medical practice. *JAMA*. 1988;259(23):3433-3440.
4. Shekelle PG, Morton SC, Keeler EB. *Costs and benefits of health information technology: Evidence report/technology assessment*. Costs and benefits of health information technology. Evidence report/technology assessment. 2006(132):1-71.
5. Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. *J Med Internet Res*. 2005;7(1):e3.
6. Hughes GB. Evidence-based medicine in health care reform. *Otolaryngol Head Neck Surg*. 2011;145(4):526-529.
7. Charles D, Furukawa M, Hufstader M. Electronic Health Record Systems and Intent to Attest to Meaningful Use among Non-federal Acute Care Hospitals in the United States: 2008-2011. The Office of the National Coordinator for Health Information Technology. ONC Data Brief. No. 1. February 2012. Available at: **http://www.healthit.gov/media/pdf/ONC_Data_Brief_AHA_2011.pdf**. Accessed December 12, 2012
8. Denny JC, Choma NN, Peterson JF, et al. Natural language processing improves identification of colorectal cancer testing in the electronic medical record. *Med Decis Making*. 2012;32(1):188-197.
9. Singh H, Schiesser R, Anand G, Richardson PA, El-Serag HB. Underdiagnosis of Lynch syndrome involves more than family history criteria. *Clin Gastroenterol Hepatol*. 2010;8(6):523-529.
10. Erickstad L, Reed G, Bhat D, Roehrborn CG, Lotan Y. Use of electronic medical records to identify patients at risk for prostate cancer in an academic institution. *Prostate Cancer Prostatic Dis*. 2011;14(1):85-89.
11. van Staa TP, Patel D, Gallagher AM, de Bruin ML. Glucose-lowering agents and the patterns of risk for cancer: a study with the General Practice Research Database and secondary care data. *Diabtologia*. 2012;55(3):654-665.
12. Nordstrom BL, Whyte JL, Stolar M, Mercaldi C, Kallich JD. Identification of metastatic cancer in claims data. *Pharmacoepidemiol Drug Safety*. 2012;21(suppl 2):21-28.
13. Houser SH, Colquitt S, Clements K, Hart-Hester S. The impact of electronic health record usage on cancer registry systems in Alabama. *Perspect Health Inf Manage*. Epub 2012;9:1f.
14. D'Amore JD, Sittig DF, Ness RB. How the continuity of care document can advance medical research and public health. *Am J Public Health*. 2012;102(5):e1-4.
15. AHRQ. "What is Comparative Effectiveness Research?" AHRQ Effective Health Care Program. Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. Available at: **http://www.effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/**. Accessed November 26, 2012.
16. AHRQ. Agency for Healthcare Research and Quality. AHRQ's Multiple Chronic Conditions Research Network. Available at: **http://www.ahrq.gov/research/mccrn.htm**. Accessed November 26, 2012.
17. Ren C, Lim S, Hylton T, et al. *Florida Annual Cancer Report: 2007 Incidence and Mortality*. Tallahassee, FL: The Bureau; 2011:v.
18. Thornton M, ed. *Standards for Cancer Registries Volume II: Data Standards and Data Dictionary,* Record Layout Version 12.1, 15th ed. Springfield, IL: North American Association of Central Cancer Registries; 2010.
19. Florida Cancer Data System. Data Acquisition Manual 2010. Florida Department of Health; 2010.
20. Dolin RH, Alschuler L, Beebe C, et al. The HL7 Clinical Document Architecture. *J Am Med Inform Assoc*. 2001;8(6):552-569.
21. Centers for Disease Control and Prevention. International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). 2011. Available at: **http://www.cdc.gov/nchs/icd/icd9cm.htm**. Accessed November 26, 2012.
22. American Medical Association. *CPT 2011 Professional Edition*. American Medical Association Press; 2010.
23. Du X, Freeman JL, Goodwin JS. Information on radiation treatment in patients with breast cancer: the advantages of the linked medicare and SEER data. Surveillance, Epidemiology and End Results. *J Clin Epidemiol*. 1999;52(5):463-470.
24. Du XL, Key CR, Dickie L, et al. Information on chemotherapy and hormone therapy from tumor registry had moderate agreement with chart reviews. *J Clin Epidemiol*. 2006;59(1):53-60.
25. Du XL, Key CR, Dickie L, Darling R, Geraci JM, Zhang D. External validation of medicare claims for breast cancer chemotherapy compared with medical chart reviews. *Med Care*. 2006;44(2):124-131.
26. Kazley AS, Ozcan YA. Do hospitals with electronic medical records (EMRs) provide higher quality care?: an examination of three clinical conditions. *Med Care Res Rev*. 2008;65(4):496-513.
27. Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med*. 2009;360(16):1628-1638.
28. Desai JR, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. *Med Care*. 2012;50(suppl):S30-35.
29. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care*. 2012;50(suppl):S21-29.
30. Centers for Medicare and Medicaid Services. *EHR Incentive Programs*. Available at: **http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentivePrograms/**. Accessed December 5, 2012.
31. Florida Cancer Data System. *Interactive Maps and Rates*. Available at: **http://fcds.med.miami.edu/inc/statistics.shtml**. Accessed September 1, 2012.
32. U.S. Census Bureau; American Community Survey, State and County QuickFacts 2011. Available at: **http://quickfacts.census.gov/qfd/states/12/12057.html**. Accessed September 1, 2012.
33. Pearson ML, Ganz PA, McGuigan K, Malin JR, Adams J, Kahn KL. The case identification challenge in measuring quality of cancer care. *J Clin Oncol*. 2002;20(21):4353-4360.
34. Ruth M. Rapid Case Ascertainment. Available at: **http://www.unclineberger.org/research/rapid-case-ascertainment**. Accessed December 5, 2012.

Special Feature

Raising the Bar: Busyness Does Not Equal Competence

Michele Webb, CTR

Do you ever wonder why you are so busy? Probably not because you're too busy to stop and even start thinking about it! But, let's hit the pause button and consider this common problem.

Several months ago I was chatting with a friend and heard myself say, "Busy is the new normal." If that were true, then feeling overwhelmed, exhausted, ineffective, nonproductive, frustrated, or discouraged because I may not see a dream become reality, would also have to be true, and that is just downright wrong.

If you are like me, you frequently hear people say they are busy. If we tell ourselves this long enough, we begin to believe it and live as though it is an acceptable norm. And slowly, our strength, motivation and resolve is stripped away by this faulty thinking. When we live in this mindset, we stop performing optimally, and slowly become less productive and increasingly overwhelmed. It is time to wake up, rattle some cages, and change the thought processes and work of cancer registrars across the country.

The need and desire to be busy comes from many roots—cultural, economical, philosophical, and spiritual to name a few. But let's look at the psychological perspective. There is a misconception that busy registrars are important, competent, successful, and all-around good registrars. If you are busy, you are to be admired or even envied. And, quite often, the need to "out busy" your peers becomes an insane game of one-upmanship. But what if we are *too* busy? What happens when the busyness backfires and our lives become a runaway train?

Perhaps you can relate to the famous scene from "I Love Lucy," when Lucy and Ethel are working on the assembly line wrapping candies. At first it seemed like an easy job, the work was going at an easy pace. Lucy and Ethel were feeling competent and comfortable. And then the conveyor belt sped up—faster and faster. The candies keep coming, one after another and too closely packed together. Lucy and Ethel can't keep up. They try, but the belt is moving too fast. In a panic, they pretend they are keeping up by eating the candy, stuffing it into their hats, dropping it down their blouses, and hiding it from the boss. It is hilarious to watch but, at a certain level, so painful!

We can relate to this scene because it evokes a familiar, intense feeling. But our pursuit of competence and importance has gotten out of hand. By trying to be a "super registrar" we actually start feeling more *in*competent than we did before. The warning signs are there and we just need to take a step back, draw a deep breath, and slow things down. How do we do this? If you are willing to change, the solution is simple. Stop multitasking and start focusing your energy in a different way.

Multitasking is not working hard. It has been scientifically proven that multitasking does not result in a job well done. It only means that you have fooled yourself into believing that you are so important or so good at what you do, that you only need to do it halfway. We may think we are getting more work done when in reality we are not. To do your best work, do it one step at a time. You will enjoy the process more, give it the attention it deserves, and feel less anxious and rushed.

Now, if you think that by not multitasking you will fall behind in your work, or get in trouble with your boss, think again. The key to jumping ahead is to change your belief about busyness and scheduling is your first step to recovery.

Block out chunks of time, in 90 minute bursts, for focused work. You can do this for three to four bursts a day, but make sure it is quality time. This is the secret of many celebrities, highly successful business people, and millionaires. By using this technique, you will get solid results, be more productive, less stressed out, and more efficient.

Learn to say "no." I'm not saying you shouldn't help out, take on new projects at work, volunteer for NCRA or your state association, spend time in your community or other activities. But if you fill up your time with everyone else's projects and neglect your own, that's a problem. And you don't want to fall into the trap of volunteering and making a commitment to your peer groups and then failing to deliver. Volunteerism is a sign of maturity and passion for your work, but to become a volunteer member of a group and then claim busyness as an excuse for not contributing only demonstrates a lack of focus and causes others to label you as unreliable.

By taking our work at a slower, more manageable pace, you actually are giving yourself 2 enormously valuable gifts. The first is the gift of feeling capable by doing a few things well, which has far more value than doing many things poorly. The second gift you give yourself is the space and permission to think and feel. If you can slow down the

conveyor belt and put more space between the candies, so to speak, you can be in touch with your thoughts and values and make better choices about your work. In countless studies we are shown that by focusing our work we actually become more productive. Live and work to enrich your life, not deplete it.

As cancer registrars we all have things we want and need to do. But we also feel compelled to do things we should not be doing. We tend to take on tasks and responsibilities because of a false belief that they will make us feel better about work or meet some inflated expectation of our role as a registrar. If you can be honest with yourself, you may find you are hooked on busyness to satisfy a false belief of value or importance. Or maybe you work yourself so hard because you need to fill every nook and cranny of time so that there is no space for you to feel anything at all. As we mature, we need to become more intentional, more deliberate, and more mindful of our choices. When we nurture this kind of balance, the conveyor belt slows down. Life becomes manageable and our true competence is revealed.

Let's take a vow to stop complaining about how busy we are, acknowledge we created our own mess and refocus our energy by working smarter, not harder or longer. By changing our thoughts, which are the command posts for our life, we can make something happen, feel better, be better registrars, and get closer to seeing our dreams come true. Are you willing to refocus the busyness in your life and become the cancer registrar you are meant to be?

Michele is a cancer registry speaker, educator, coach, and independent contractor living in Rancho Cucamonga, California. She is the founder of ***www.CancerRegistrar.com***, ***http://CancerRegistryAcademy.com***, *and* ***www.RegistryMindset.com*** *offering cancer registry leadership, mentoring and continuing education opportunities. Your comments are welcomed by email to michele@michelewebb.com.*

NCRA Marks its 40TH Anniversary in 2014

JOIN US IN NASHVILLE TO CELEBRATE!

NCRA 2014 Annual Educational Conference

May 15-18, 2014

Gaylord Opryland Resort & Convention Center
Nashville, TN

National Cancer Registrars Association

National Cancer Registrars Association
Education Foundation

The *JRM* Quiz and answers are now available through NCRA's Center for Cancer Registry Education (CCRE). For your convenience, the JRM article and quiz can be accessed online at **www.CancerRegistryEdcuation.org/jrm-quizzes**. Download the article, complete the quiz and claim CE credit all online.

Journal of Registry Management Continuing Education Quiz—SPRING 2013

CANCER REGISTRY ENRICHMENT VIA LINKAGE WITH HOSPITAL-BASED ELECTRONIC MEDICAL RECORDS: A PILOT INVESTIGATION

Quiz Instructions: The multiple choice or true/false quiz below is provided as an alternative method of earning CE credit hours. Refer to the article for the ONE best answer to each question. The questions are based solely on the content of the article. Answer the questions and send the original quiz answer sheet and fee to the NCRA Executive Office before the processing date listed on the answer sheet. Quizzes may not be retaken nor can NCRA staff respond to questions regarding answers. Allow 4–6 weeks for processing following the submission deadline to receive return notification of your completion of the CE process. The CE hour will be dated when it is submitted for grading; that date will determine the CE cycle year.

After reading this article and taking the quiz, the participants will be able to:

- Identify the mechanisms for extracting, linking, and processing hospital Electronic Medical Record (EMR) data with the Florida Cancer Data System
- Assess the completeness of existing registry treatment data as well as the potential for data enhancement
- Describe the challenges, advantages, and limitations of this study

1. Recent studies incorporating electronic medical records (EMRs) in cancer research represent the emergence of a new field of investigation, and include the:
   a) comparison of EMR data collection to manual collection in order to identify individuals in need of colorectal cancer (CRC) screening
   b) identification of at-risk patients in whom an elevated PSA level was recorded but who, otherwise, would not have been referred to a urologist
   c) linkage of computerized medical records to cancer registry records to determine if there was increased cancer risk for diabetic patients taking certain glucose lowering medications
   d) all of the above
2. The long-term vision of an EMR-linked cancer registry is _____; while a secondary goal of such linkage is _____.
   a) to establish a systematic rapid transmission of patient health data; the direct abstraction of certain relevant registry data
   b) the direct abstraction of certain relevant registry data; to establish a systematic rapid transmission of patient health data
   c) to identify the mechanisms for extracting, linking, and processing hospital EMR data with the Florida Cancer Data System (FCDS); to assess the potentials of EMR for enhancement of cancer registry data
   d) to assess the potentials of EMR for enhancement of cancer registry data; to identify the mechanisms for extracting, linking, and processing hospital EMR data with the Florida Cancer Data System (FCDS)
3. The objective of the registry enhancement study was to identify registry records where first course of treatment could be updated or enhanced by hospital diagnostic and procedure data.
   a) True
   b) False
4. The study focused on:
   a) colon cancer cases
   b) cases diagnosed in 2009 and 2010
   c) treatment with surgery, radiation, and chemotherapy
   d) non-analytic cases
5. According to Table 5, records with a breast cancer diagnosis in 2010 (n=1,584), linkage of Florida Cancer Data System (FCDS) breast cancer cases with the hospital EMR allowed for the greatest enhancement of records for which treatment modality?
   a) Surgery
   b) Radiation therapy
   c) Chemotherapy
   d) Hormone therapy
6. Challenges for routine linkage of EMR and cancer registry records may include:
   a) EMR datasets within a single hospital system are currently well integrated
   b) EMR system documentation is consistently available
   c) research is typically a priority of most hospitals
   d) data elements required by the cancer registry are incomplete
7. One advantage of utilizing EMR data with central cancer registry data is:
   a) collecting EMR data to enhance central cancer registries is more limited by case definition selection criteria
   b) patient tumors identified in the hospital EMR should already exist in the central cancer registry database
   c) the EMR results in registry records with less detailed treatment data
   d) it is more costly and resource-intensive to collect data from real world settings as opposed to conducting randomized control trials
8. Limitations in this study include:
   a) the level of data enrichment may vary for cancers with more or less complex treatment regimens
   b) only 2 hospitals were examined
   c) the hospital system they partnered with was not an American College of Surgeons Commission on Cancer (CoC) accredited facility
   d) the linkage was performed on several types of cancer
9. Real-time linkage of cancer registries with EMR systems has the long-term potential of streamlining rapid case ascertainment and monitoring patient quality of care outcomes.
   a) True
   b) False
10. According to the authors, this study demonstrates that:
    a) EMRs provide detailed clinical data valuable for patient-centered outcomes research
    b) linking hospital EMR records and cancer registry records is feasible when sufficient resources are invested
    c) existing breast cancer registry records contained the most complete information on the patient, tumor, and treatment data
    d) all of the above

Journal of Registry Management Continuing Education Quiz **Answer Sheet**

Available online at www.cancerregistryeducation.org/jrm-quizzes

Please print clearly in black ballpoint pen.

First Name | M.I. | Last Name

Address

Address

City | State/Province | ZIP Code/Postal Code

NCRA Membership Number (MUST complete if member fee submitted) | CTR Number

Instructions: Mark your answers clearly by filling in the correct answer, like this ➽ not like this ⋘. Passing score of 70% entitles one (1) CE clock hour per quiz.

Please use black ballpoint pen.

1	A	B	C	D
2	A	B	C	D
3	A	B		
4	A	B	C	D
5	A	B	C	D
6	A	B	C	D
7	A	B	C	D
8	A	B	C	D
9	A	B		
10	A	B	C	D

Submit the original quiz answer sheet only!
No photocopies will be accepted.

For Internal Use Only

Date Received:_____

Amount Received:_____

Notification Mailed: _____

This original quiz answer sheet will not be graded, no CE credit will be awarded, and the processing fee will be forfeited unless postmarked by:

June 30, 2014

Quiz Identification Number:

4001.01

JRM Quiz Article:

CANCER REGISTRY ENRICHMENT VIA LINKAGE WITH HOSPITAL-BASED ELECTRONIC MEDICAL RECORDS: A PILOT INVESTIGATION

⋙ Processing Fee: **Member $25** **Nonmember $35**

⋙ Payment is due with submission of answer sheet. Make check or money order payable to NCRA. US currency only. Do not send cash. No refund under any circumstances. Please allow 4–6 weeks following the submission deadline for processing.

Please check one:

⋙ Enclosed is check #_____ (payable to NCRA)

⋙ Charge to the following card:

⋙ MasterCard (16 digits) ⋙ Visa (13 or 16 digits) ⋙ American Express

Card Number_____ Exp. Date _____

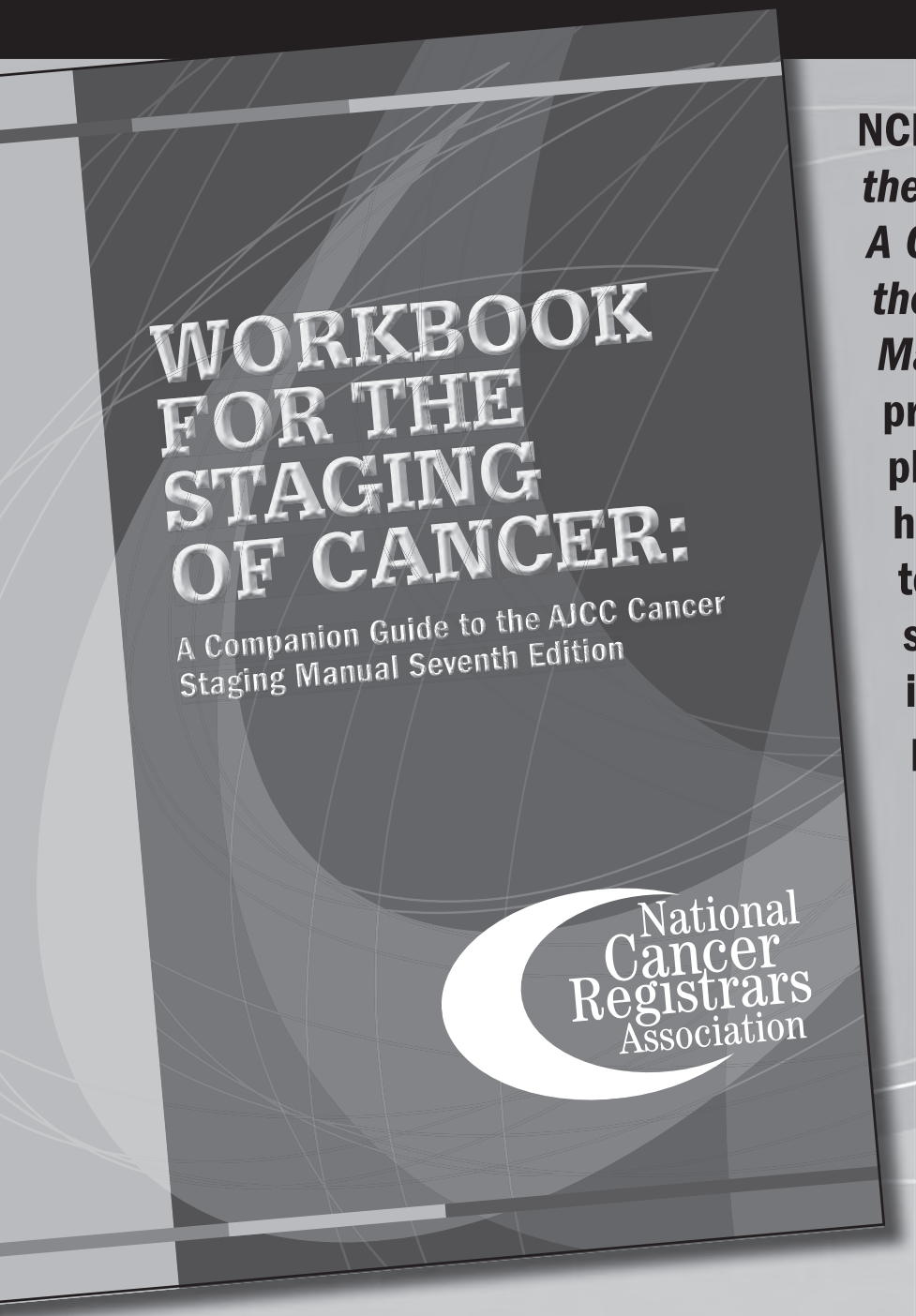
Signature _____

Print Cardholder's Name _____

Telephone # _____

Mail to: ***NCRA Executive Office***
JRM CE Quiz
1340 Braddock Place
Suite 203
Alexandria, VA 22314

NEW EDITION NOW AVAILABLE!



NCRA's *Workbook for the Staging of Cancer: A Companion Guide to the AJCC Cancer Staging Manual Seventh Edition* provides cancer registrars, physicians, and other health professionals a tool to understand cancer staging, including detailed information on its history, purpose, sources of information for staging, and various staging systems. Extensive exercises are included for ten primary cancers and the correct answers and corresponding rationales are provided.

Price:
NCRA member price: $109
Non-member price: $119

Order today at **www.ncra-usa.org**.

National Cancer Registrars Association

National Cancer Registrars Association
CALL FOR PAPERS

Topic:
1. Birth Defects Registries
2. Cancer Registries
Cancer Collaborative Stage
Cancer and Socioeconomic Status
History
3. Trauma Registries
4. Recruitment, Training, and Retention
5. Public Relations

The *Journal of Registry Management,* official journal of the National Cancer Registrars Association (NCRA), announces a call for original manuscripts on registry methodology or research findings related to the above 5 subjects, and related topics. Contributed manuscripts are peer-reviewed prior to publication.

Manuscripts of the following types may be submitted for publication:
1. **Methodology Articles** addressing topics of broad interest and appeal to the readership, including methodological aspects of registry organization and operation.
2. **Research articles** reporting findings of original, reviewed, data-based research.
3. **Primers** providing basic and comprehensive tutorials on relevant subjects.
4. **"How I Do It" Articles** describe tips, techniques, or procedures for an aspect of registry operations that the author does particularly well. The "*How I Do It*" feature in the *Journal* provides registrars with an informal forum for sharing strategies with colleagues in all types of registries.
5. **Opinion papers/editorials** including position papers, commentaries, essays, and interviews that analyze current or controversial issues and provide creative, reflective treatments of topics related to registry management.
6. **Bibliographies** which are specifically targeted and of significant interest will be considered.
7. **Letters to the Editor** are also invited.

Address all manuscripts to: Vicki G. Nelson, MPH, RHIT, CTR, Editor-in-Chief, *Journal of Registry Management,* (770) 488-6490, **vnelson@cdc.gov.**

Manuscript submission requirements are given in "Information for Authors" found on the inside back cover of each *Journal* and on the NCRA Web site at **http://www.ncra-usa.org/jrm.**

INFORMATION FOR AUTHORS

Journal of Registry Management (JRM), the official journal of the National Cancer Registrars Association, invites submission of original manuscripts on topics related to management of disease registries and the collection, management, and use of cancer, trauma, AIDS, and other disease registry data. Reprinting of previously published material will be considered for publication only when it is of special and immediate interest to the readership. *JRM* encourages authorship by Certified Tumor Registrars (CTRs); special value is placed on manuscripts with CTR collaboration and publication of articles or texts related to the registry profession. CTR continuing education (CE) credits are awarded; a published chapter or full textbook article equals 5 CE hours. Other published articles or documents equal CE hours. All correspondence and manuscripts should be addressed to the Editor-in-Chief, Vicki Nelson, MPH, RHIT, CTR at: izo4@cdc.gov, or at: CDC/NCCDPHP/DCPC/CSB, 4770 Buford Drive, MS K-53, Atlanta, GA 30341-3717, 770-488-6490 (office), 770-488-4759 (fax).

Manuscripts may be submitted for publication in the following categories: **Articles** addressing topics of broad interest and appeal to the readership, including **Methodology papers** about registry organization and operation; **Research papers** reporting findings of original, reviewed, data-based research; **Primers** providing tutorials on relevant subjects; and **"How I Do It"** papers are also solicited. **Opinion papers/editorials** including position papers, commentaries, and essays that analyze current or controversial issues and provide creative, reflective treatments of topics related to registry management; **Letters to the Editor**; and specifically-targeted **Bibliographies** of significant interest are invited.

The following guidelines are provided to assist prospective authors in preparing manuscripts for the *Journal*, and to facilitate technical processing of submissions. Failure to follow the guidelines may delay consideration of your manuscript. Authors who are unfamiliar with preparation and submission of manuscripts for publication are encouraged to contact the Editor for clarification or additional assistance.

Submission Requirements

Manuscripts. The terms manuscripts, articles, and papers are used synonymously herein. E-mail only submission of manuscripts is encouraged. If not feasible, submit the original manuscript and 4 copies to the Editor. Manuscripts should be double-spaced on white 8-1/2" x 11" paper, with margins of at least 1 inch. Use only letter-quality printers; poor quality copies will not be considered. Number the manuscript pages consecutively with the (first) title page as page one, followed by the abstract, text, references, and visuals. The accompanying cover letter should include the name, mailing address, e-mail address, and telephone number of the corresponding author. For electronic submission, files should be 3-1/2", IBM-compatible format in Corel WordPerfect™, Microsoft® Word for Windows®, or converted to ASCII code.

Manuscripts (Research Articles). Articles should follow the standard format for research reporting (Introduction, Methods, Results, Discussion, References), and the submission instructions outlined above. The introduction will normally include background information, and a rationale/justification as to why the subject matter is of interest. The discussion often includes a conclusion subsection. Comprehensive references are encouraged., as are an appropriate combination of tables and figures (graphs).

Manuscripts (Methodology/Process Papers). Methodology papers should follow the standard format for research reporting (Introduction, Methods, Results, Discussion), or for explanatory papers not reporting results (Introduction, Methods, Discussion), as well as the submission instructions outlined above.

Manuscripts ("How I Do It" articles). The "*How I Do It*" feature in the *Journal* provides registrars with a forum for sharing strategies with colleagues in all types of registries. These articles describe tips, techniques, or procedures for an aspect of registry operations that the author does particularly well. When shared, these innovations can help registry professionals improve their skills, enhance registry operations, or increase efficiency.

"*How I Do It*" articles should be 1,500 words or less (excepting references) and can contain up to 2 tables or figures. To the extent possible, the standard headings (Introduction, Methods, Results, Discussion) should be used. If results are not presented, that section may be omitted. Authors should describe the problem or issue, their solution, advantages (and disadvantages) to the suggested approach, and their conclusion. All submitted "*How I Do It*" articles will have the benefit of peer/editorial review.

Authors. Each author's name, degrees, certifications, title, professional affiliation, and email address must be noted on the title page exactly as it is to appear in publication. The corresponding author should be noted, with mailing address included. Joint authors should be listed in the order of their contribution to the work. Generally, a maximum of 6 authors for each article will be listed.

Title. Authors are urged to choose a title that accurately and concisely describes the content of the manuscript. Every effort will be made to use the title as submitted, however, *Journal of Registry Management* reserves the right to select a title that is consistent with editorial and production requirements.

Abstract. A brief abstract must accompany each article or research paper. The abstract should summarize the main point(s) and quickly give the reader an understanding of the manuscript's content. It should be placed on a page by itself, immediately following the title page.

Length. Authors are invited to contact the Editor regarding submission of markedly longer manuscripts.

Style. Prepare manuscripts using the *American Medical Association Manual of Style.* 9th ed. (1998)

Visuals. Use visuals selectively to supplement the text. Visual elements—charts, graphs, tables, diagrams, and figures—will be reproduced exactly as received. Copies must be clear and properly identified, and preferably e-mailed. Each visual must have a brief, self-explanatory title. Submit each visual on a separately numbered page at the end of the manuscript, following the references.

Attribution. Authors are to provide appropriate acknowledgment of products, activities, and support especially for those articles based on, or utilizing, registry data (including acknowledgment of hospital and central registrars). Appropriate attribution is also to be provided to acknowledge federal funding sources of registries from which the data are obtained.

References. References should be carefully selected, and relevant. References must be numbered in order of their appearance in the text. At the end of the manuscript, list the references as they are cited; do not list references alphabetically. Journal citations should include author, title, journal, year, volume, issue, and pages. Book citations should include author, title, city, publisher, year, and pages. *Authors are responsible for the accuracy of all references*. Examples:

1. LeMaster PL, Connell CM. Health education interventions among Native Americans: A review and analysis. *Health Education Quarterly.* 1995;21(4):521–38.

2. Hanks GE, Myers CE, Scardino PT. Cancer of the Prostate. In: DeVita VT, Hellman S, Rosenberg SA. *Cancer: Principles and Practice of Oncology*, 4th ed. Philadelphia, PA: J.B. Lippincott Co.; 1993:1,073–1,113.

Key words. Authors are requested to provide up to 5, alphabetized key words or phrases which will be used in compiling the Annual Subject Index.

Affirmations

Copyright. Authors submitting a manuscript do so on the understanding that if it is accepted for publication, copyright in the article, including the right to reproduce the article in all forms and media, shall be assigned exclusively to NCRA. NCRA will not refuse any reasonable requests by the author(s) for permission to reproduce any of his or her contributions to the *Journal.* Further, the manuscript's accompanying cover letter, signed by all authors, must include the following statement: "We, the undersigned, transfer to the National Cancer Registrars Association, the copyright for this manuscript in the event that it is published in *Journal of Registry Management.*" Failure to provide the statement will delay consideration of the manuscript. It is the author's responsibility to obtain necessary permission when using material (including graphs, charts, pictures, etc.) that has appeared in other published works.

Originality. Articles are reviewed for publication assuming that they have not been accepted or published previously and are not under simultaneous consideration for publication elsewhere. If the article has been previously published or significantly distributed, this should be noted in the submission for consideration.

Editing

Journal of Registry Management reserves the right to edit all contributions for clarity and length. Minor changes (punctuation, spelling, grammar, syntax) will be made at the discretion of the editorial staff. Substantive changes will be verified with the author(s) prior to publication.

Peer Review

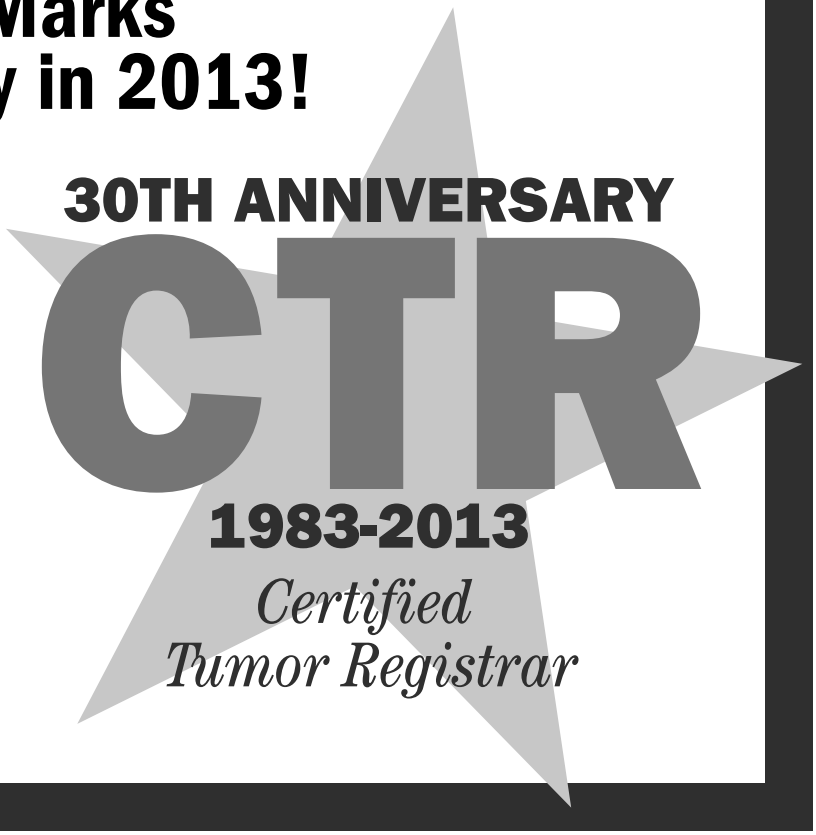
Contributed manuscripts are peer-reviewed prior to publication, generally by 3 reviewers. The *Journal* Editor makes the final decision regarding acceptance of manuscripts. Receipt of manuscripts will be acknowledged promptly, and corresponding authors will be advised of the status of their submission as soon as possible.

Reprints

Authors receive 5 complimentary copies of the *Journal* in which their manuscript appears. Additional copies of reprints may be purchased from the NCRA Executive Office.

CTR Credential Marks 30th Anniversary in 2013!

NCRA salutes all CTRs on their achievement in attaining this important credential.



National Cancer Registrars Association

Journal of Registry Management

NCRA Executive Office
1340 Braddock Place
Suite 203
Alexandria, VA 22314