NOW AVAILABLE!

Second Edition of NCRA’s Popular Cancer Case Studies Workbook

NCRA’s second edition includes 40 NEW cases to provide cancer registrars additional opportunities to practice assigning AJCC TNM Stage and coding SEER Summary Stage.

GO TO www.ncra-usa.org/casestudies TO ORDER!
Contents

Letter from the Editor
Vonetta L. Williams, PhD, MPH, CTR ................................................................. 3

Letter to the Editor
Notes on Brazilian Cancer Registries ............................................................... 4
Gil Patrus Pena, MD

Original Articles
Copy Number Variants and Congenital Anomalies Surveillance: A Suggested Coding Strategy Using the Royal College of Paediatrics and Child Health Version of ICD-10 .......................................................... 6
Tanya Bedard, MPH; R. Brian Lowry, MD, DSc, FRCPC; Barbara Sibbald, MSc;
Mary Ann Thomas, MD, CM, FRCPC, FCCMG; A. Michiel Innes, MD, FRCPC, FCCMG

Validating Procedures used to Identify Duplicate Reports in Haiti’s National HIV/AIDS Case Surveillance System ................................................................. 10
Chris Delcher, PhD; Nancy Puttkammer, PhD; Réginald Arnoux, MD;
Kesner Francois, MD; Mark Griswold, MSc; Irum Zaidi, MPH;
Yves Anthony Patrice Joseph, MD, MPH; Barbara J. Marston, MD

Linking the National Health Interview Survey with the Florida Cancer Data System: A Pilot Study ................................................................................ 16
Laura A. McClure, MSPH; Eric A. Miller, PhD; Stacey L. Tannenbaum, PhD;
Monique N. Hernandez, PhD; Jill A. MacKinnon, PhD; Yulei He, PhD;
William G. LeBlanc, PhD; David J. Lee, PhD

Collaborative Development and Results of a Nigerian Trauma Registry ........... 23
Laura D. Cassidy, MS, PhD; Oluwole Olaomi, MBBS, FWACS, FACS; Allison Ertl, MS;
Emmanuel A. Ameh, MBBS, FWACS, FACS

Using Data on Tumor Grade in Cancer Registries to Enhance Surveillance of Oropharyngeal Cancers in Relation to the Human Papillomavirus Epidemic .... 29
Anthony P. Polednak, PhD

Features and Other Journal Departments
Encouraging Health Information Management Graduates to Pursue Cancer Registry Careers ................................................................. 37
Jennifer Peterson, MS, RHIA, CTR

How I Do It
How to Build an Abstract .............................................................................. 42
Vickie Hauhee, Med, CTR

Raising the Bar
The Privilege of Being Invited In .................................................................. 44
Michele Webb, CTR

Spring 2016 Continuing Education Quiz ......................................................... 46
Deborah C. Roberson, MSM, CTR; Denise Harrison, BS, CTR

Call for Papers
Trauma Registry Call for Papers ................................................................. 48
Information for Authors .............................................................................. 49
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.
Greetings, Colleagues!

I learned something new about myself this week. My team and I completed an assessment of our strengths and, based on the assessment, my top 2 strengths are being a mental archiver of data, words, and information, and being self-motivated to accomplish tasks on a daily basis. In working with data, databases, reports, and research on a daily basis, I was confident that my primary strength would be analytics. I was wrong. I have a strength that I was not aware of. Do you know your strengths? We tend to focus on our weaknesses, but have you ever thought about sharpening or fine-tuning your strengths? I encourage each of you to step outside the box to learn and be mindful of your strengths. Share and communicate your strengths with your supervisor, manager, and colleagues in case they need to use your strengths in a particular area to accomplish a common team goal. In addition, share your strengths with your family and friends. This was a great learning experience for me and will allow me to leverage the strengths of my team members so that we can collectively accomplish our department strategic priorities. It also provided me with the opportunity to focus on projects that target my strengths.

In this issue of the Journal of Registry Management, we highlight a birth defects article from Tanya Bedard, MPH, and team on a coding strategy for congenital anomalies using ICD-10. In addition, we have original manuscripts from Chris Delcher, PhD, and colleagues detailing validation procedures used in Haiti’s HIV/AIDS surveillance system and Laura McClure, MSPH, and team who evaluated the linkage of data from the Florida Cancer Data System and the National Health Interview Survey. Finally, Anthony Polednak, PhD, evaluated the use of tumor grade to enhance the surveillance of oropharyngeal cancers and human papillomavirus (HPV).

Our special feature article is from Jennifer Peterson, MS, RHIA, CTR, who encourages health information management graduates to consider a career in the cancer registry field. In continuation of the Journal of Registry Management’s efforts to highlight information that is relevant to those who are abstracting, our How I Do It article is from Vicki Hawhee, MEd, CTR, and she provides a step-by-step guide to the process she uses to build an abstract. Lastly, Michele Webb, CTR, highlight the value of sharing your story and listening to someone else’s story in her article entitled “The Privilege of Being Invited In.”

In my winter 2015 letter, I encouraged each of you to write down things or people you are thankful for. Please don’t forget to do this. We thank you for your continued support of the Journal of Registry Management.

Respectfully,

Vonetta L. Williams, PhD, MPH, CTR
Editor-in-Chief, Journal of Registry Management
National Cancer Registrars Association
Notes on Brazilian Cancer Registries

Gil Patrus Pena, MD

The activities of cancer registries are complex and involve huge efforts in case identification, abstraction of information of the tumor and the patient, coding of variables (morphology, topography, metastasis site, and occupation), inclusion of the record in a database, and quality review.

In Brazil, the population-based cancer registries record incident cancer cases within a defined geographic region in a specified period. They provide the basis for cancer incidence estimates. Hospital-based cancer registries have the task of collecting all cancer cases attending that particular institution, with special interest in patient care and treatment results.

As with any other database, quality control is of paramount importance for information reliability. Databases from Brazilian population- and hospital-based registries can be accessed online at the website of the National Institute of Cancer of Brazil (Instituto Nacional do Câncer [INCA]). The use of this important tool for cancer epidemiological analysis should be encouraged, producing (as a secondary effect) quality enhancement of available data.

Using the table constructor accessible on the INCA website,1 we surveyed the data from hospital-based cancer registries in the state of Minas Gerais, Brazil, from 2009 to 2013 (all dates refer to calendar years). The cases were filtered by initial diagnosis in 2009 and primary site in the rectum (International Classification of Diseases for Oncology, third edition [ICD-O-3]2, C20.9). We evaluated the number of cases by morphology according to the hospital institution. The inquiry retrieved 891 cases from 35 hospitals. As expected, adenocarcinoma not otherwise specified (NOS) (8140/3) was the most frequent histological type, with 721 records (81.14%). Interestingly, there were 27 records (3.0%) of squamous cell carcinoma NOS (8070/3) of the rectum, of which 12 (44.4%) were reported from a single institution. In this particular institution, this histological type represented 66.7% (12/18) of all rectal cancers diagnosed in 2009.

We acknowledge the existence of cases of squamous cell carcinoma of the rectum, but they are rare. Two series of cases reported, respectively, 6 and 12 cases in periods of 15 and 22 years.3,4 The occurrence of 12 cases in a single year would merit further investigation. Is the registry in error? Are these cases from the anorectal junction? Is there a systematic error in morphology coding? Are these real cases? What would then distinguish the population attending this institution, regarding the risk of squamous cell carcinoma of the rectum?

More than drawing attention to possible flaws in the registry, we wish to emphasize the necessity of qualifying registry data. Besides the revision of individual cases, a global analysis of registered data is necessary, in the search for inconsistencies and peculiar aspects of cancer occurrence in the population attended in a particular institution or from a particular demographic group.

Another small exercise was conducted in the database from the population based cancer registries, also available on the INCA website.5 The aim of this survey was to evaluate the age distribution of bone tumors (osteosarcoma, chondrosarcoma, Ewing sarcoma). Taking data from all registries, we surveyed the years from 2005 to 2012. Cases were filtered by International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)6 C40 (bone and articular cartilage of limbs) or C41 (bone and articular cartilage of other and unspecified sites). The requested table displayed the number of cases by age interval, according to morphology. Cases of osteosarcoma (9180/3, 9181/3, 9183/3, 9185/3, 9186/3, 9192/3, 9193/3, 9195/3), chondrosarcoma (9220/3, 9221/3, 9230/3, 9231/3, 9240/3, 9242/3, 9243/3) and Ewing sarcoma/PNET (peripheral primitive neuroectodermal tumor) (9260/3, 9364/3, 9365/3) were grouped together. The results reflected the expected incidence for these tumors: osteosarcoma and Ewing sarcoma are diseases of young patients, while chondrosarcoma predominantly occurs later in life (Figure 1). However, there were epithelial

---

4Programa de Avaliação e Vigilância do Câncer e seus Fatores de Risco (PAV-MG) Secretaria de Estado da Saúde (SES/MG). Email: gilpena@gold.com.br.
tumor cases registered as primary bone tumors, including
in situ neoplasms (8010/2, 8010/3, 8020/3, 8041/3, 8050/3,
8051/3, 8070/2, 8070/3, 8075/3, 8140/3, 8211/3, 8240/3,
8260/3, 8310/3, 8480/3, 8490/3, 8510/3). They likely repre-
sent faulty records, possibly resulting from the notification
of metastases as primary bone cancer. The incidence esti-
mates calculated from these data may thus not be entirely
reliable. The inclusion of epithelial neoplasms in situ
(8010/2, 8070/2) indicates wrong codification of either
morphology or topography. Another possible error (of the
registry or of the histological classification of the neoplasm)
was the inclusion of 4 cases of myxoid chondrosarcoma as
primary bone tumor.

This brief analysis highlights, on the one hand, the
quality of information contained in the records, mirroring
the age distribution of primary tumors of bone in a consis-
tent manner. However, it also demonstrates imperfections
in the database that need to be better addressed.

We aimed at demonstrating the availability of the data-
bases of population- and hospital-based cancer registries
from Brazil, and at fostering their use, so that stored infor-
mation becomes a useful tool for investigators interested in
cancer epidemiology research. The widespread use of these
data, while revealing any flaws or inconsistencies, would
have the role to contribute to the continuous qualiﬁcation
of data.

The views expressed here are my own and do not
necessarily reﬂect those of the Programa de Avaliação e
Vigilância do Câncer e seus Fatores de Risco (PAV-MG).

References
1. Instituto Nacional de Cancer (Brasil) – Integrador RHC website. https://
December 4, 2015.
2. Organização Mundial da Saúde. CID-O – Classiﬁcação Internacional de
the rectum: report of six cases and review of the literature. Dis Colon
Rectum. 2002;45(11):1535-1540.
5. Instituto Nacional de Câncer (Brasil) – Ações e programas website.
http://www2.inca.gov.br/wps/wcm/connect/estatisticas/site/home/
6. Organização Mundial da Saúde. CID-10 – Classiﬁcação Estatística
Copy Number Variants and Congenital Anomalies Surveillance: A Suggested Coding Strategy Using the Royal College of Paediatrics and Child Health Version of ICD-10

Tanya Bedard, MPH; R. Brian Lowry, MD, DSc, FRCPChb,c; Barbara Sibbald, MS
c
Mary Ann Thomas, MD, CM, FRCP, FCCMChb,c; A. Micheil Innes, MD, FRCP, FCCMCb,c

Abstract: The use of array-based comparative genomic hybridization to assess DNA copy number is increasing in many jurisdictions. Such technology identifies more genetic causes of congenital anomalies; however, the clinical significance of some results may be challenging to interpret. A coding strategy to address cases with copy number variants has recently been implemented by the Alberta Congenital Anomalies Surveillance System and is described.

Key words: birth defect, congenital anomalies, copy number variant, surveillance, uncertain clinical significance.

Introduction

Congenital anomalies surveillance programs are familiar with the more commonly reported chromosomal abnormalities: trisomy 21 (Down syndrome), trisomy 18 (Edwards syndrome), and trisomy 13 (Patau syndrome). These conditions have been reported by cytogenetic laboratories using conventional cytogenetic techniques, such as banded chromosome analysis that typically has a resolution between 5 and 10 Mb. The corresponding phenotypes are well established, there are specific International Classification of Diseases (ICD)-9 and ICD-10 codes, and prevalence is well reported.

Fluorescence in situ hybridization (FISH) can detect changes of approximately 100 Kb or larger (eg, the submicroscopic deletion in the long arm of chromosome 22 [22q11.2] associated with different phenotypes and names, including DiGeorge syndrome and velocardiofacial syndrome). This chromosomal abnormality is relatively well defined, as are the associated phenotypes. Microdeletions identified by FISH may be captured by congenital anomalies surveillance systems with a code that represents an autosomal loss and microduplications with an autosomal gain.

More recently, chromosomal microarrays have been used to assess DNA copy number. This is an umbrella term used for array-based comparative genomic hybridization (CGH) and single nucleotide polymorphism arrays. Array CGH is the recommended first-tier investigation for patients with developmental delay, autism, multiple congenital anomalies, or dysmorphic features that are unexplained after a thorough history and examination. Results are either reported as normal (arr[1-22]x2,[XY]x1 or arr[1-22,X]x2) or having a copy number variant (CNV). A CNV is defined as a segment of DNA at least 1 Kb in size that differs in copy number compared with a representative reference genome. This term on its own does not imply clinical significance (benign or pathogenic) or relative dosage (deletion or duplication). While the relative dosage is straightforward, the clinical significance may be challenging to interpret. The recommended categories for clinical significance described by the American College of Medical Genetics and Genomics include benign, uncertain clinical significance, and pathogenic (Table 1).

Benign is defined as a CNV reported in multiple peer-reviewed publications or curated databases as a benign variant. Included in this category are common polymorphisms (documented in more than 1% of the population). Pathogenic CNVs are reported as clinically significant in multiple peer-reviewed publications, even if penetrance and expressivity are known to be variable. Penetrance is the percentage of individuals with a given genotype that exhibit the phenotype associated with the genotype; eg, if the penetrance is 80%, then 80% of people with the mutation will develop the disease and 20% will not. Expressivity is the extent to which the given genotype is expressed at the phenotypic level; eg, some family members with the same mutation will be significantly affected and others only mildly affected. These CNVs are usually larger (>3–5 Mb) and include syndromic loci.

The CNVs reported as uncertain clinical significance are subdivided into uncertain clinical significance likely pathogenic; uncertain clinical significance, likely benign; and uncertain clinical significance (Table 1). CNVs of uncertain clinical significance likely pathogenic are described in single case reports, and the genes within have gene functions that are likely relevant to the phenotypes. CNVs of uncertain clinical significance likely benign may not have genes or have genes that are known to be fairly redundant, or genes that

Original Article

Journal of Registry Management 2016 Volume 43 Number 1

6
have minimal documented information regarding function. They are described in a small number of cases in the general population but do not represent common polymorphisms.

CNVs of uncertain clinical significance contain genes within the CNV but it is not known whether the genes are dosage sensitive and clinical significance has not been established in the literature. Whether the CNV is inherited or de novo, as well as the size of the CNV, may also influence the interpretation of clinical significance.

The categories described by the American College of Medical Genetics and Genomics were used by the Alberta Congenital Anomalies Surveillance System (ACASS) to develop a coding strategy. The authors do not intend this article to assist with interpreting the clinical significance of CNVs. The aims were to promote discussion regarding the ascertainment of cases with CNVs and to describe a strategy for coding CNVs that is practical for the purpose of congenital anomalies surveillance.

### Methods

ACASS is primarily a passive system that depends on health professionals to report cases with congenital anomalies. Diagnoses can be clarified or verified by requesting additional information. ACASS uses multiple sources of ascertainment, including Alberta Vital Statistics (notices of birth and medical certificates of death or stillbirth), in-patient discharges, pathology, newborn screening programs, genetic clinics, and cytogenetic laboratories. Additional details regarding the methodology of ACASS have been previously described.6,7

## Table 1. Copy Number Variant Significance Descriptions

<table>
<thead>
<tr>
<th>Copy Number Variant Significance</th>
<th>Description</th>
</tr>
</thead>
</table>
| Benign                           | • Reported in multiple peer-reviewed publications or curated databases as a benign variant.  
• Common polymorphisms included |
| Likely benign                    | • May not have genes or have genes that are redundant or with minimal information regarding function  
• Described in a small number of cases |
| Uncertain Clinical Significance  | • Not known whether the genes are dosage sensitive  
• Clinical significance not established in literature |
| Likely pathogenic                | • Reported in single case reports  
• Genes have gene function likely relevant to phenotype |
| Pathogenic                       | • Reported as clinically significant in multiple peer-reviewed publications  
• Usually larger (>3–5 Mb)  
• Include syndromic loci |

In consultation with a local cytogeneticist and clinical geneticists, ACASS developed case and defect definitions along with a coding strategy to address cases with CNVs (Figure 1). Existing cases with CNVs previously reported to ACASS were discussed by reviewing genetic consultations and cytogenetic laboratory reports. An additional consideration was whether the CNV was of interest for later retrieval and evaluation. The codes used in this article reflect the Royal College of Paediatrics and Child Health (RCPCH) version of the ICD-10 coding system and the coding practices of ACASS. The authors acknowledge that different programs may have different coding practices and use different codes or versions of ICD that may not be as straightforward as the proposed strategy implies.

ACASS rejects benign CNVs and accepts the pathogenic ones, capturing the pathogenic CNVs with the codes representing a microdeletion (Q93.6) or microduplication (Q92.4) (Figure 1). Associated anomalies are captured with relevant codes and the specific microduplication or microdeletion is specified in a data field within the database for retrieval purposes.

If the CNV reported is described as uncertain clinical significance likely pathogenic, we would code these as we code the previously described pathogenic CNVs (Figure 1).

The CNV reported as uncertain clinical significance likely benign is not included as a case, particularly if it is inherited from a healthy parent and there are no congenital anomalies.

---

6. Likely benign includes genes that are dosage sensitive and clinical significance has not been established in the literature. Whether the CNV is inherited or de novo, as well as the size of the CNV, may also influence the interpretation of clinical significance.

7. Likely pathogenic includes genes with gene function likely relevant to phenotype. Additional details regarding the methodology of ACASS have been previously described.
reported (Figure 1). If congenital anomalies are reported the anomalies are captured with their corresponding ICD-10 RCPCH codes, however, the CNV would not be coded since it is likely unrelated to the phenotype.

The CNVs reported as uncertain clinical significance, where there are associated anomalies reported, are coded with the microduplication code followed by a V (Q92.4V) or the microdeletion code followed by a V (Q93.6V) (Figure 1). The "V" indicates that it is not known whether the CNV is truly pathogenic and plays a role in the phenotype, or if the CNV is a coincidental finding. ACASS would specify the CNV in a separate data field and the associated anomalies would be captured with their corresponding ICD-10 RCPCH code.

If there is a CNV of uncertain clinical significance reported without any associated congenital anomalies, ACASS would code Q99.7, which is a code currently not specified in the RCPCH modification of ICD-10 (Figure 1). There was a local consensus that this type of CNV may represent an abnormality and needs to be captured, although no anomalies or ineligible anomalies (eg, developmental delay) are reported and the exact significance has not been established. This is consistent with the decision to code other chromosomal abnormalities with no associated eligible anomalies reported. Trisomy 21 identified in a case with the characteristic facial appearance and mild to moderate intellectual disability would still be coded.

**Discussion**

The shift of cytogenetic testing from detecting changes in chromosomes of 5000 Kb or larger to the improved resolution offered by chromosomal microarrays detecting changes of 1 Kb is a significant genetic advancement. This improved technology, which is now commonly used in many centers, provides the opportunity for more children with multiple congenital anomalies to have a diagnosis and receive more targeted medical care, contributing to better health outcomes, and for families to be provided with more accurate recurrence risks. This relatively new genetic testing will ultimately impact reported prevalence and classification for etiological studies. The prevalence of certain conditions diagnosed with this type of testing will increase and cases with congenital anomalies that were once of unknown etiology will be more accurately classified with a known genetic cause.

The ability of surveillance systems to follow up cases to include new genetic diagnoses is suggested to maintain accuracy. As reported by Jackson et al, there may be challenges updating and reclassifying cases based on new genetic diagnoses and including cases that were once excluded because of uncertain clinical significance. The proposed coding strategy may alleviate some of these challenges by including cases with CNVs reported as uncertain clinical significance likely pathogenic and uncertain clinical significance with or without eligible congenital anomalies. The ICD-10 RCPCH codes with a “V” or Q99.7 could be changed as more is known about the CNV, either to drop the “V” if the CNV has become known to be pathogenic, or to not accept as an eligible case. Initially capturing cases with CNVs of uncertain clinical significance will reduce the possibility of missing eligible cases. By using the suggested ICD-10 RCPCH codes, these cases can be excluded from further analysis to limit artificial increases in prevalence (or they may be reported separately). Since this type of result is relatively new and significance is not established in the literature, ACASS includes these cases with the intention of future evaluation and the appreciation that some cases may be found to have benign CNVs.

Congenital anomalies programs need to be informed of new genetic testing and have the capability to ascertain and incorporate these results in a meaningful manner. The capacity to ascertain CNVs will vary by program, since the ascertainment of cases with a CNV optimally requires abstraction or reporting of the primary test result since a secondhand report may not be accurate. Program staff need sufficient education in reading complex laboratory reports and genetic consultations, as the interpretation of particular CNVs may not be specified or clear in medical records and may even change over time. However, including genetic clinics and cytogenetic laboratories as data sources, consulting with geneticists, keeping abreast of new genetic technologies, and participating in continuing-education opportunities will assist program staff. Although some birth defect programs may not routinely ascertain cases with CNVs due to a lack of resources, time, or expertise, when conducting etiological research using surveillance data, cases with CNVs (particularly those known to be pathogenic) need to be identified for proper classification.

While the goals of congenital anomalies surveillance programs are to contribute to the epidemiology of congenital anomalies, identify risk factors through research, and evaluate prevention strategies, there are also potential opportunities to contribute to establishing genotype-phenotype associations. Often, the reported CNVs are very rare. However, through comprehensive ascertainment using multiple data sources (including cytogenetic laboratories and clinical genetics), documentation, collaboration with programs, and the expertise of cytogeneticists and clinical geneticists, the CNVs that are today reported as uncertain clinical significance may in the future be determined as pathogenic or benign.

**Conclusions**

The proposed coding strategy is an initial attempt to address the complexity of CNVs and the uncertainty surrounding the interpretation of some reported results. The coding scheme presented reflects the current state of knowledge and will likely need modifying as new information arises. ACASS plans to assess the proposed strategy in the future by reviewing existing cases with CNVs of uncertain clinical significance in consultation with our geneticists and in the context of new knowledge. We acknowledge that congenital anomalies surveillance programs differ with regards to methodologies (eg, ascertainment, data sources, and follow-up) and resources. This paper is intended to promote discussion of how advanced genetic testing is impacting congenital anomalies surveillance.
References


Validating Procedures used to Identify Duplicate Reports in Haiti’s National HIV/AIDS Case Surveillance System

Chris Delcher, PhD; Nancy Puttkammer, PhD; Réginald Arnoux, MD; Kesner Francois, MD; Mark Griswold, MSc; Irum Zaidi, MPH; Yves Anthony Patrice Joseph, MD, MPH; Barbara J. Marston, MD

Abstract: Objectives: Valid deduplication of human immunodeficiency virus (HIV) case reports is critical to the utility of these data to inform HIV programs. The Haitian Ministry of Health (MSPP) and partners operate a case-based, national HIV/AIDS surveillance system (HASS), using deterministic and probabilistic procedures to identify duplicate records. These procedures are described and validated based on expert classifications. Methods: Two samples of HASS records identified as duplicates were selected: 100 pairs from deterministic and 100 pairs from probabilistic matching procedures (total: 200 pairs, 400 case reports). Clinical data from the national electronic medical record (iSanté) were reviewed and consensus gold-standard determinations on the status of duplications were made. False positive rates (FPR) were estimated by reviewing these records, while false negative rates were calculated (FNR) by using LinkPlus™ probabilistic linkage software. The effect of deduplication on total HIV case counts was demonstrated. Results: Review of deterministic matches yielded 99 true positives and 1 false positive (FPR, 1 per 100; 95% CI, 0.71–5.4). Review of probabilistic matches yielded a FPR of 6 per 100 (95% CI, 2.7–12.4). LinkPlus identified 1,491 probable matches among 68,393 records, representing a FNR of 2 per 100 (95% CI, 0.55–7.0). After adjustment, the estimated unique count of reported HIV patients in HASS was 211,885 (95% CI, 207,293–213,232) as of December 2013. Conclusions: Based on application of the established procedures, HASS conforms to the duplication performance standard recommended by the Centers for Disease Control and Prevention for HIV surveillance.

Key words: deduplication, Haiti, HIV surveillance

Introduction

The Joint United Nations Programme on HIV/AIDS (UNAIDS) estimates that 140,000 people are living with human immunodeficiency virus (HIV) in Haiti as of 2013. The Haitian Ministry of Health (MSPP) implemented a national HIV/AIDS case-based surveillance system (HASS) in 2008, which has produced estimates of new HIV diagnoses and other indicators for monitoring the status of the epidemic. In 2011 alone, HASS contained approximately 23,000 newly reported cases of HIV. HIV surveillance, from population counts to continuity of care, relies on accurate, unique identification of patients from case reports to reduce the possibility of duplication and/or improper data linkage.

In countries lacking unique identifiers, duplication rates can be high. Before Brazil implemented a national unique identifier in 2009, it was estimated that only 100 million people were represented in the 140 million records in one of the primary health information systems. Despite extensive discussion about implementing unique identification numbers in Haiti, progress has been delayed. Although many adults are assigned unique identification numbers for voting purposes, 53% of the population is ineligible to vote, and 19% of eligible voters are unregistered. Other proposed approaches have included biometric identifiers (eg, finger scans), health passports, and portable electronic medical records.

During HASS’s development and piloting phases, estimates showed the probability of duplication was high given the lack of a unique national identifier, clinic-to-clinic patient mobility in Haiti combined with limited inter-clinic communication, and stigma or incentives that would encourage the provision of false information. The MSPP partially addressed these concerns by supporting name-based HIV reporting to increase the likelihood of accurate epidemiologic counts, and applying deterministic and probabilistic patient-matching algorithms using names and other demographic variables.

HASS receives HIV case reports from all facility-based venues where HIV testing and counseling (HTC) services are provided, including HTC (formerly referred to as voluntary counseling and testing) sites, preventing mother-to-child transmission sites, and tuberculosis/HIV clinics, via both the MSPP’s Monitoring, Evaluation and Surveillance Interface (MESI) reporting system and from 3 centralized clinical systems. These systems are the iSanté electronic medical record (EMR) system, the Haitian Group for the Study of Kaposi’s Sarcoma and Opportunistic Infections (GHESKIO), and Partners in Health Zanmi Lasante. The iSanté EMR system is the MSPP’s primary EMR for the national HIV care and treatment program. iSanté was deployed in Haiti in 2005 and is currently in use in 98 urban and rural facilities located in all 10 administrative...
determine if the patient is the same person.11 The process associated with the human pattern recognition required to described as “probabilistic” due to the uncertainty asso-

values are not permitted to match. Records matching exactly in all of these fields automatically assigned to the same patient; missing the birthplace. Records matching exactly in all of these fields automatically generated for each record. The code consists of a pseudo-unique, 7-digit HIV reporting code (which is also automatically generated for each record. The code consists of the following 7 characters: first and surname initials, birth month and year, and the first initial of the mother’s first name (eg, XY0175Z).

Identifying duplicate records within HASS begins with a series of deterministic matches using first and surname, birth month and year, sex and mother’s first name, the first 4 letters of the patient’s first name, the reporting clinic, and the birthplace. Records matching exactly in all of these fields are automatically assigned to the same patient; missing values are not permitted to match.

The next step is human adjudication. This process is described as “probabilistic” due to the uncertainty associated with the human pattern recognition required to determine if the patient is the same person.11 The process starts by displaying records with the same pseudo-unique HIV case reporting code created in preprocessing, or same first and last name on the secure HASS website for visual inspection by epidemiologic staff in Haiti. Staff then decide if records represent the same or different persons based on variables such as the patient’s first and surname, mother’s maiden name, sex, birthdate, commune/department of residence, commune of birth, marital status, occupation, date of HIV diagnosis, reporting clinic, reporting system, and report date.

Methods

Validation of Deduplication Procedures

The national EMR system (iSanté) was selected for the validation exercise because of its programmatic and geographic representativeness. Due to limited resources and for ease of calculation, we selected 100 matched pairs of records generated from the deterministic routine and 100 identified from the probabilistic process to review and selected matches with exactly 2 possible records in HASS.

Analyses

Figure 1 provides an overview of the record review process. To evaluate the matches identified by HASS matching procedures, we used a 2-stage expert review process: (1) a central-level, administrative review and (2) a local-level, physician review for final adjudication. The administrative reviewer has worked with the iSanté EMR system since 2005. The physician reviewer is an internal medicine specialist with 7 years’ experience providing HIV care and 5 years as an HIV clinical trainer and mentor in Haiti.

The evaluated data fields and stepwise process used for administrative and physician review are shown in Table 1. When administrative review (Step 1) could not confirm a match using the fields shown, we asked for local physician review (Step 2). The physician had access to the complete longitudinal clinical record for each patient from the iSanté EMR.

Subjective interpretation was allowed. If differences between records were judged plausible given factors such as clinical measurement error, data entry errors, etc., they were accepted as matches. Examples are shown in Table 1.

We calculated false positive rates (FPRs) for the deterministic and probabilistic matched pairs using the final expert determination as the gold standard; 95% confidence intervals were calculated using the score interval method.12 For the calculations, matches unclassifiable by expert review were divided equally as matches and nonmatches.

To identify additional matches not identified by the HASS matching procedures, we used LinkPlus™, a probabilistic record-linkage software developed by the Centers for Disease Control and Prevention (CDC)’s Division of Cancer Prevention and Control.13 Designed to help cancer registries detect duplicate case reports, the expected inputs are data elements commonly found in disease registries (eg, first and surname, gender, race/ethnicity, US Social Security number), but the CDC indicates that LinkPlus can be used with “any type of data.”13 A recent study identified possible matches between patients attending different antenatal care clinics in Senegal using LinkPlus.14

LinkPlus (default settings) was used to identify possible matches among records that had not been detected
Table 1. iSanté Data Elements Used for Administrative and Clinical Review in this Study

<table>
<thead>
<tr>
<th>Expert Review</th>
<th>Data Elements Reviewed (X = Shared Element with HASS)</th>
<th>Data Presenting Ambiguity (Uncertain Cases)</th>
<th>Matching Logic</th>
</tr>
</thead>
</table>
| Administrative | • First, last name (X)  
• Mother’s first name (X)  
• National ID (X)  
• Clinic (X)  
• Gender (X)  
• Date of birth (X)  
• Total number of clinic visits (X)  
• First HIV test date (X)  
• Distinct height measures (with date)  
• Most recent weight measure (with date)  
• Earliest clinic visit date and encounter type  
• Most recent clinic visit date and encounter type  
• Most recent CD4 (with date)  
• Earliest ART regimen (with date)  
• Most recent ART regimen (with date)  
• Most recent gravida/para/living children (GPEV, with date)  
• Discontinuation (with reason and date) | • Different gender/  
Mother’s first name different  
• Date of birth with divergence in some elements of dd/mm/yy  
• Inconsistent information on visit history and date of death  
• Heights inconsistent  
• GPEV divergent | • Match  
• No Match  
• Date of birth with little or no divergence in dd/mm/yy  
• Similar mother’s first name  
• Similar height for adult or height trajectory for child  
• Visit histories compatible  
• Similar GPEV with plausible differences over time  

| Physician (same as administrative and:) | • Place of birth (free text)  
• Address (free text)  
• Emergency contact name (free text)  
• Spouse/partner name (free text)  
• Referral and transfer data (coded and free text)  
• Reason for discontinuation (free text)  
• Weight trajectory (free text)  
• CD4 trajectory (coded data)  
• Clinical encounter data (coded and free text) | • Place of birth or current place of residence distinct, but phonetically similar  
• Weight trajectory similar, but height different  
• Limited encounter and clinical information to make judgment | • Address or place of birth  
• Spouse/partner name  
• Transfer data consistent  
• Dates of death consistent  
• Consistent prior treatment history when transferring in  
• Consistent CD4 trajectory  
• Inconsistent visit history explained by data entry errors in visit dates/ “True” gender verifiable in the record by provider notes | • Adult vs pediatric patient  
• Place of birth is different  
• Name of emergency contact different |

ART, antiretroviral therapy; CD4, cluster of differentiation 4; GPEV, gravidity, parity, enfants vivants; HASS, HIV/AIDS surveillance system; HIV, human immunodeficiency virus; ID, identification.

As possible matches upon submission to HASS or been adjudicated as nonmatches by HASS surveillance staff. We matched on first, last, and middle name; birthdate; and sex. First and surname were used as blocking variables, increasing linkage efficiency in large data set, as suggested by the LinkPlus manual. The name-based matching used the Jaro–Winkler metric, comparing agreement between 2 strings accounting for random insertion, deletions, and transpositions. Birthdate matching accounted for the absence of in the month, day, and year elements. Sex had to match exactly.

Using a subset of iSanté records, we calculated a possible false negative rate (FNR) as the maximum number of pairs of duplicate records identified by LinkPlus divided by the total number of records considered to be unique following application of HASS deduplication procedures;
95% confidence intervals were calculated using the score interval method. Resource limitations prevented a detailed review of all matches identified by LinkPlus.

**Applying Deduplication Rates to HASS**

Our goal was to evaluate the effect (with confidence intervals) of deduplication on the total case counts in HASS. We compared the unadjusted number of cases reported to HASS (since December 2013) with adjusted counts using 4 different deduplication approaches: (1) deterministic matching using pseudo-unique HIV reporting codes manually entered from case reporting forms, (2) deterministic matching after basic data quality control to correct/generate missing reporting codes (eg, if the reporting code was missing birth month, we used birth month from the birthdate field), (3) deterministic and probabilistic matching results from HASS, and (4) the results from the latter adjusted for the estimated FPR and FNR.

**Results**

**False Positive Rate from Deterministic Matches**

Based on administrative review, 94 pairs identified by the deterministic procedures were considered true positive matches, 1 was a possible false positive match, and the validities of 5 were undetermined. After final adjudication, the 5 undetermined matches were identified as positive matches; thus, the totals were 99 true positives and 1 false positive, and the estimated false positive rate (FPR) was 1 case per 100 (95% CI, 0.71–5.4). The false positive result arose because fields in that patient’s EMR used for matching in HASS were updated after the case had been reported to HASS. After reviewing the updated record, it was clear the 2 patients were different, but currently no process for receiving retrospective updates in HASS exists.

**False Positive Rate from Matches Made by Human Adjudication**

After administrative review, 84 pairs identified by the probabilistic procedures were considered true positive matches, 2 were possible false positive matches, and the validities of 14 were undetermined. After final adjudication, the counts were 91 true positive matches, 3 false positives, and 6 undetermined matches. After applying the assumption that half undetermined matches were true matches, the estimated FPR was 6 cases per 100 (95% CI, 2.7–12.4).

**False Negative Rate from LinkPlus**

At the time of the validation exercise, HASS contained 68,393 records submitted from the iSanté EMR that did not match via the deterministic component or had been adjudicated during review as nonmatches. LinkPlus identified 1,491 probable matches in this group (FNR, 2.2 per 100 records [95% CI, 0.55–7.0]). As viewed on the LinkPlus user interface, the majority of these records appeared to be different patients that should not be matched.

**Applying Estimates to HASS**

By December 2013, HASS contained 302,718 HIV/AIDS case notification records from 4 reporting systems. Using current deterministic and probabilistic matches from HASS, the estimated unique patient count was 213,318. Of the 213,318 unique patients, there were 153,065 (72%) patients with 1 case notification record submitted; 41,059 (19%) with 2 records submitted; 12,129 (6%) with 3 records submitted; 4,177 (2%) with 4 records submitted; and 2,888 (1%) with 5 or more records submitted.

Among the 60,523 patients with 2 or more notification records, 39,741 and 20,512 patients were identified based on the deterministic and probabilistic matching, respectively. Application of the FPR from the deterministic validation component (0.01) to 39,741 patients suggests that 397 (95% CI, 278–2,146) were erroneously matched. Likewise, application of the FPR from the probabilistic validation component (0.06) suggests that 1,231 (95% CI, 551–2,543) were possibly erroneously matched. Based on application of the estimated FNR from the LinkPlus component (0.02) to the 153,065 patients presumed to be unique records, 3,061 (95% CI, 918–10,715) duplicate case reports may have gone undetected.

The case count would have been 167,954 had deduplication been based solely on the pseudo-unique reporting code. When data quality of the reporting code improved, the unique case count increased to 213,117, partially attributable to fewer missing reporting codes (and thus fewer instances of records assumed to be unique based on poor-quality codes).

Combining these adjustments, we obtained an estimate of 211,885 (95% CI, 207,293–213,232) total unique case counts since December 2013. Table 2 shows the impact of applying different adjustments based on our validation review components, and Figure 2 shows the impact on estimated total unique HIV case counts under each scenario.

<table>
<thead>
<tr>
<th>Match Type (Rate Type)</th>
<th>Error per 100 (95% CI)</th>
<th>Current Estimate (No.)</th>
<th>Adjustment (95% CI)</th>
<th>Adjusted Estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic (FP)</td>
<td>1 (0.7–5.4)</td>
<td>39,741</td>
<td>+397 (278–2,146)</td>
<td>40,138 (40,019–41,887)</td>
</tr>
<tr>
<td>Probabilistic (FP)</td>
<td>6 (2.7–12.4)</td>
<td>20,512</td>
<td>+1,231 (554–2,543)</td>
<td>21,743 (21,066–23,055)</td>
</tr>
<tr>
<td>Unmatched (FN)</td>
<td>2 (0.6–7.0)</td>
<td>153,065</td>
<td>-3,061 (-10,715 to -918)</td>
<td>150,004 (142,350–152,147)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>213,318</td>
<td>-1,433 (-6,025 to -86)</td>
<td>211,885 (207,293–213,232)</td>
</tr>
</tbody>
</table>

FP, false positive; FN, false negative.
Discussion

Case-based surveillance is one component of a country’s HIV surveillance activities that can provide critical information about HIV epidemics in many regions of the world. These ongoing data collection systems should be evaluated periodically to ensure they meet design objectives, including detection of duplicate case reporting. Reliable case counts may depend on the application of reasonable/validated approaches to identifying duplicate reports, particularly in settings where individuals are not assigned national unique identification numbers. Several deduplication procedures used by the national HIV/AIDS Surveillance System in Haiti were evaluated and validated. After deduplication, the system conforms to recommended CDC duplication performance standards. Without deduplication, the number of HIV case reports purported to represent a single person in the HASS system is markedly overestimated.

Figure 2. Comparison of Total Records Submitted to Haiti’s National HIV/AIDS Surveillance System (A) Versus Counts Using (B) the National Code “As Is”, (C) the National Code Cleaned, (D) Deterministic/Probabilistic (D/P) Methods Currently Used, and (E) D/P Methods after this Validation Review

The case count attained from identifying duplicates based on the pseudo-unique reporting code is similar to the case count following deterministic and probabilistic matching. However, now we have evidence that the deterministic/probabilistic matching is considerably more accurate at the patient level than the matching based only on the pseudo-unique code at the population and case level.

When deciding how best to deduplicate surveillance data, factors to consider include the data flow point where deduplication occurs, software environment, case volume, data quality, and staffing resources. Deterministic matching can be accomplished in many off-the-shelf or specialized database applications and packages, including Microsoft Access, SQL Server, and EpiInfo. However, extracting data from disparate data systems, preprocessing, and customizing deterministic algorithms require staff with computer programming skills. Human adjudication may identify matches missed by deterministic algorithms but may be inconsistent and burdensome. Thus, human review may be preferable when caseloads are manageable.

In countries like Haiti, where HIV case volumes are high and staffing resources are low, identifying ways to reduce the burden of human adjudication is important. For example, data fields were identified that may improve the ability to automatically discriminate between true and false matches, thereby reducing the pool of possible records requiring adjudication. These data fields include core variables already captured for HIV surveillance in Haiti (eg, patient full address, phone number, pediatric vs adult patient) and clinical variables in the iSanté EMR that could be incorporated in the future (eg, emergency contact name, most recent height, date and reason for discontinuation of treatment, evidence of transfer in or out). Other less resource-intensive approaches could include simple modifications of the manual review screen to increase the speed of pattern recognition (eg, color-coding mismatching fields between records).

Preliminary review of possible duplicates detected by LinkPlus indicated many were unlikely to be true duplicates. It is important to note that MSPP does not intend to adjust estimated surveillance totals based on the duplication identified by LinkPlus. The usefulness of LinkPlus will vary by country as the software was designed to work best with specific inputs (eg, names written in English, US Social Security numbers).

This review has important implications. First, we have shown that HASS provides an accurate, acceptable approach to patient record matching without a national identifier in Haiti. This finding should improve the confidence in the internal validity of future surveillance data. However, evaluation of other components of the surveillance system needs work. To examine external validity, one must make comparisons between findings from HASS and estimates from other sources. Second, we have identified a replicable set of algorithms and processes from health information systems in Haiti collecting the same set of identifiers. The methodology is being shared and evaluated within Haiti and other countries interested in case-based surveillance. Third, accurate patient matching allows us to understand and improve other aspects of longitudinal patient-level outcomes analyses such as transfer-adjusted analyses of patient retention. For example, Delcher et al (2012) reported that between 2006–2012, approximately a quarter of female patients originally diagnosed and reported from another clinic outside the GHESKIO network where an additional case report was generated. In a different study, GHESKIO researchers incorporated a HASS look-up step to understand patient transfer patterns. They found that, after accounting for transfers to facilities outside the GHESKIO network, estimated 24-month, lost-to-follow up rates changed from 52% to 43%. MSSP is establishing procedures for sharing deidentified information for patients in iSanté and other systems. Benefits of increased data sharing include consolidation of records for improved retention. Fourth, accurately matching patient records allows HASS to provide a clearer picture of HIV patient care from point of diagnosis through treatment by using the best information available across information systems. For example, initial data from counseling and testing case reports can provide...
robust risk-factor information, while EMR-based case data provides longitudinal data such as CD4 cell counts, antiretroviral treatment regimens, and other clinical variables. This assessment has several limitations. We selected our sample for convenience and only reviewed pairs with 2 possible matching records from the iSanté system. Thus, our findings may not be generalizable to the other EMR systems reporting to HASS. Further evaluations are needed to understand the validity of the matching algorithms for the full combined data set or other situations. Second, adjudication by local Haitian staff was used as the gold standard for this evaluation. This adjudication is subjective, and we have not validated the accuracy or reliability of their decisions. We are in the process of formally documenting the decision-making logic. Third, we used LinkPlus, a US system not developed to identify duplicate entries among previously deduplicated records in Haiti.

Our findings may not apply to systems with variable levels of data quality or different cultural practices (eg where similar names are more or less common or people are more or less likely to accurately report information such as birthdate). Additionally, the record-consolidation process used for surveillance and program evaluation purposes should be used with caution for patient management at the clinic-level.

Conclusion

In conclusion, the matching approach yielded an acceptable error rate for national-level HIV/AIDS surveillance purposes. These findings are being used to improve the accuracy of case reporting in Haiti. We recommend that countries develop strategies to prevent case duplication, especially when national identifiers are unavailable; iteratively test the strategy in coordination with local experts prior to scale-up; and periodically validate and modify the matching process.

References


Linking the National Health Interview Survey with the Florida Cancer Data System: A Pilot Study

Laura A. McClure, MSPH; Eric A. Miller, PhD; Stacey L. Tannenbaum, PhD; Monique N. Hernandez, PhD; Jill A. MacKinnon, PhD; Yulei He, PhD; William G. LeBlanc, PhD; David J. Lee, PhD

Abstract: Cancer registry data are vital for the surveillance of cancer trends, but registries are limited in the number of data items that can be collected for hypothesis-driven research. Linkage with other databases can add valuable information and is a potentially effective tool for increasing our understanding of and identifying the causes of cancer and health disparities along the cancer continuum. We conducted a pilot study to link data from the 1981–2010 Florida Cancer Data System (FCDS) with data from the 1986–2009 National Health Interview Survey (NHIS). The NHIS data contain important information on sociodemographics, screening behaviors, comorbidities, risk factors, health care access, and quality of life, which are not available from FCDS. The linkage resulted in a total of 6,281 linked cases. After removing cases with a first cancer diagnosis before 1981 (prior to FCDS creation) or missing date of diagnosis information, there were 1,908 cases diagnosed with cancer prior to their NHIS interview and 4,367 cases diagnosed after their NHIS interview. The enriched data set resulting from the linkage allows us to evaluate risk factors associated with developing cancer as well as conduct analyses on cancer survivorship issues and mortality. This pilot study demonstrates the feasibility and utility of a linkage between cancer registries and national health surveys, while also acknowledging the cost and challenges associated with such linkages.

Key words: cancer registry, linkage, National Health Interview Survey, population-based data

Introduction

Population-based cancer registries collect valuable information for the surveillance of cancer trends; however, registries are limited in the number of data items that can be collected for hypothesis-driven research. Currently, cancer surveillance research relies on data at the census tract level (for example, poverty level and education) or data obtained directly from a sample of cancer patients to determine relationships between sociodemographic cancer risk factors and survival. Linkage between cancer registries and nationally representative health databases can be an effective tool for increasing our understanding of individual-level risk factors and identifying the root causes of health disparities along the cancer continuum (eg, cancer development, stage at presentation, progression, and consequences of cancer, including treatment outcomes and survivorship). More specifically, the National Health Interview Survey (NHIS) is an official federal monitoring instrument for the objectives of the Healthy People 2010 and 2020 goals and contains valuable data on participant sociodemographics, health status, health behaviors, and health care utilization that are not available in cancer registry data. In addition, the National Cancer Institute has supported a series of periodic NHIS Cancer Control Supplements designed to monitor trends in cancer behavioral risk factors and cancer screening behaviors (most recently in 2000, 2005, 2010, and 2013). Another series of specific cancer screening questions was also administered in 2003 and 2008.

Because a cancer registry data linkage of this kind is so informative, the University of Miami and the Florida Cancer Data System (FCDS), in collaboration with the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC), conducted a pilot project linking the FCDS and the NHIS. The goal of this linkage was to determine the feasibility and administrative burden of conducting the linkage, to assess the utility of the resulting linked data, and to detail important lessons for potentially conducting a nationwide linkage. Here we describe the linkage process, the characteristics of the linked cancer cases, and the value of potential linkages of cancer registries nationwide.
Methods

Data Sources

National Health Interview Survey. The National Health Interview Survey (NHIS) (http://www.cdc.gov/nchs/nhis.htm) is an annual, cross-sectional household survey of the noninstitutionalized US civilian population that has been conducted continuously by the NCHS since 1957. The NHIS serves as the principal source of information on the health of the nation, collecting a wide range of health-related information, including population characteristics, health status and limitations, health care access and utilization, health insurance, and health behavior information on a representative sample of the US population. Data from the 1986–2009 NHIS were used for the linkage. Approval for the data linkage was obtained from the NCHS Ethics Review Board, the Florida Department of Health (FDOH) Human Subjects Committees Institutional Review Board, and the University of Miami Institutional Review Board.

NHIS participants eligible for the linkage included those with available personally identifiable information, who had provided their Social Security number (SSN), and who did not explicitly refuse to have their data linked (a question asked in the survey starting in 2007). NHIS respondents who asked to be withdrawn from the survey after the completion of the survey data collection period were not eligible for the linkage. Beginning with 2007 data, only primary respondents (those participants who shared their SSN directly, not via a household member) could be included in NCHS records linkage activities since subjects participating via proxy were not provided the opportunity to consent to record linkages. In addition, in 2007, the NHIS began only collecting the last 4 digits of respondents’ SSNs.

Florida Cancer Data System. Since 1981, the Florida Cancer Data System (FCDS) (http://fcds.med.miami.edu/) has served as the cancer registry for the state of Florida, operated under the FDOH via a contract with the University of Miami Miller School of Medicine and funded by the CDC as part of the National Program of Cancer Registries (NPCR). FCDS records more than 110,000 newly diagnosed cases of cancer per year, which represents approximately 6% of all US cancer cases. Additionally, the North American Association of Central Cancer Registries has awarded FCDS “Gold” status for timeliness and completeness of data for the past 13 years, and FCDS received a CDC/NPCR Registry of Distinction award in 2013. Data from the 1981–2010 FCDS were used for the linkage.

Linkage Methodology and Data Processing

To conduct the linkage, FCDS used LinkPlus software (CDC), which employs a probabilistic algorithm to match records in one file to another based on the variables that are available in both files. The record linkage procedure began by blocking a set of variables for which the variables had to match before additional variable comparison took place. These variables included 9- and 4-digit SSN, patient date of birth, and the patient’s first and last name using the NYSIIS (New York State Identification and Intelligence System) phonetic-matching algorithm. Blocking variables are variables common to the 2 files that are used to “block” (or partition) the 2 files. Only within these blocks are matching variables compared between the records. For files with millions of records, the total of all possible comparison pairs is too large for practical computation. Blocking is a way to reduce the computing cost by portioning files into mutually exclusive and exhaustive blocks and performing

<table>
<thead>
<tr>
<th>File 1—FCDS</th>
<th>File 2—NHIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Security number, 9 digits</td>
<td>Social Security number, 9 digits</td>
</tr>
<tr>
<td>Social Security number, 4 digits</td>
<td>Social Security number, 4 digits</td>
</tr>
<tr>
<td>Patient date of birth, YYYYMMDD</td>
<td>Date of birth, YYYYMMDD</td>
</tr>
<tr>
<td>Patient last name</td>
<td>Last name</td>
</tr>
<tr>
<td>Patient first name</td>
<td>First name</td>
</tr>
<tr>
<td>Patient sex</td>
<td>Sex</td>
</tr>
<tr>
<td>Patient middle initial</td>
<td>Middle name</td>
</tr>
<tr>
<td>Race</td>
<td>Race</td>
</tr>
<tr>
<td>City, 5 characters</td>
<td>City, 5 characters</td>
</tr>
<tr>
<td>ZIP code, 5 digits</td>
<td>ZIP code, 5 digits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field Name</th>
<th>NAACCR Item #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID Number</td>
<td>20</td>
</tr>
<tr>
<td>County at Diagnosis</td>
<td>90</td>
</tr>
<tr>
<td>Age at Diagnosis</td>
<td>230</td>
</tr>
<tr>
<td>Date of Diagnosis</td>
<td>390</td>
</tr>
<tr>
<td>Date of Diagnosis Flag</td>
<td>391</td>
</tr>
<tr>
<td>Primary Site</td>
<td>400</td>
</tr>
<tr>
<td>Laterality</td>
<td>410</td>
</tr>
<tr>
<td>Morphology ICDO3</td>
<td>522</td>
</tr>
<tr>
<td>Behavior ICOD3</td>
<td>523</td>
</tr>
<tr>
<td>Grade</td>
<td>440</td>
</tr>
<tr>
<td>Diagnostic Confirmation</td>
<td>490</td>
</tr>
<tr>
<td>Type of Reporting Source</td>
<td>500</td>
</tr>
<tr>
<td>SEER Summary Stage 2000</td>
<td>759</td>
</tr>
<tr>
<td>SEER Summary Stage 1977</td>
<td>760</td>
</tr>
<tr>
<td>ICDO3 Conversion Flag</td>
<td>2116</td>
</tr>
<tr>
<td>NCHS ID number</td>
<td>2220</td>
</tr>
<tr>
<td>Derived SS1977—Flag</td>
<td>3040</td>
</tr>
<tr>
<td>Derived SS2000—Flag</td>
<td>3050</td>
</tr>
<tr>
<td>Characteristic</td>
<td>All Linked Cancer Cases n (unweighted %)</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Total</td>
<td>8,110 (100.0)</td>
</tr>
<tr>
<td>Number of linked tumors</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6,667 (82.2)</td>
</tr>
<tr>
<td>&gt;1</td>
<td>1,443 (17.8)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4,074 (50.2)</td>
</tr>
<tr>
<td>Female</td>
<td>4,036 (49.8)</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
</tr>
<tr>
<td>18–39</td>
<td>821 (10.1)</td>
</tr>
<tr>
<td>40–64</td>
<td>3,784 (46.7)</td>
</tr>
<tr>
<td>≥65</td>
<td>3,505 (43.2)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1,066 (13.1)</td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>5,944 (73.3)</td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>978 (12.1)</td>
</tr>
<tr>
<td>Other</td>
<td>122 (1.5)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>1,826 (22.7)</td>
</tr>
<tr>
<td>High school graduate GED/some college</td>
<td>4,562 (56.8)</td>
</tr>
<tr>
<td>Post-high school degree</td>
<td>1,643 (20.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>79 (–)</td>
</tr>
<tr>
<td>Self-rated health</td>
<td></td>
</tr>
<tr>
<td>Excellent/Very Good/Good</td>
<td>6,356 (78.7)</td>
</tr>
<tr>
<td>Fair/Poor</td>
<td>1,724 (21.3)</td>
</tr>
<tr>
<td>Missing</td>
<td>30 (–)</td>
</tr>
<tr>
<td>Survey years</td>
<td></td>
</tr>
<tr>
<td>1986–1996</td>
<td>5,271 (65.0)</td>
</tr>
<tr>
<td>1997–2009</td>
<td>2,839 (35.0)</td>
</tr>
<tr>
<td>Cancer diagnosis</td>
<td></td>
</tr>
<tr>
<td>Before 1986</td>
<td>443 (5.5)</td>
</tr>
<tr>
<td>1986–1995</td>
<td>1,991 (24.6)</td>
</tr>
<tr>
<td>1996–2005</td>
<td>3,705 (45.7)</td>
</tr>
<tr>
<td>2006–2010</td>
<td>1,971 (24.3)</td>
</tr>
<tr>
<td>Years between survey and first cancer diagnosis</td>
<td></td>
</tr>
<tr>
<td>0–5</td>
<td>3,320 (41.5)</td>
</tr>
<tr>
<td>6–10</td>
<td>2,164 (27.1)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>2,513 (31.4)</td>
</tr>
</tbody>
</table>
comparisons only on records within each block. After these initial variables, additional variables were used for matching: sex, race, city name, and 5-digit ZIP code (Table 1). There was no order in the personal identifiers other than blocking the variables. The cutoff for true matches was a score of 30 or above, and the cutoff for false matches was a score of 16.9 or below. Questionable matches (n = 3,772) included records where the full 9-digit SSN did not match, which could be due to number transposition and/or using a spouse’s SSN, or cases in which the day and/or month of patient birth date does not match but the year of birth matches. Upon clerical review of these questionable matches, 35% were deemed to be true matches.

Missing data can be a concern when conducting data linkages. However, because FCDS is gold certified by the North American Association of Central Cancer Registries (NAACCR), we can be assured that only 2% of case reports are missing meaningful data on age, sex, and county, and less than 3% are missing information on race. Additionally, the number of data fields used for the linkage has not changed over the period of the linkage. For the NHIS, aside from SSN which was declined for 40% of cases, those eligible for the linkage had less than 2% missing data on the other linkage variables. We feel that missing data, therefore, did not present a significant challenge to the accuracy of this linkage.

The resulting deidentified linked file with the cancer registry variables, presented in Table 2, was returned to NCHS for processing. NCHS added public-use identifiers to enable the addition of NHIS survey data. The data set was also limited to adults aged 18 years and older because of the small number of children that were matched. Data were weighted using modified NHIS sample weights created to represent the civilian noninstitutionalized Florida population for each year of the survey. These weights were adjusted for linkage eligibility to account for potential differences in those who provided SSN and those who did not. The final linked data set was deposited at the NCHS Research Data Center (RDC) (http://www.cdc.gov/rdc/), secured and only available to approved researchers.

**Analysis of the Linked Data**

We conducted descriptive analyses at the RDC to characterize the linked cancer cases based on sociodemographic information, NHIS survey year, diagnosis year, and sequence of cancer diagnosis (before or after NHIS interview). Only the first cancer diagnosis, if multiple were present, was used in these analyses. Analyses were conducted using SAS version 9.3 software (SAS Institute) and SUDAAN (RTI International) to account for the complex survey design of the NHIS. Analyses describing the sociodemographic and cancer diagnosis factors of linked cancer cases by state of NHIS participation (Table 3) were not weighted. Analyses describing the number of linked cancer cases by site (Table 4) and those describing the age and diagnosis factors of linked cancer cases by time of diagnosis (ie, before or after and NHIS survey) (Table 5) were weighted to the modified NHIS sample weights as described above.

Because lessons learned from this pilot linkage were intended to assess the practicality of conducting a linkage with NCHS surveys and cancer registries nationwide, we estimated the number of NHIS cancer records we would expect in a nationwide linkage. The estimates were obtained by dividing the sample size obtained in this pilot linkage by 6%, the percentage of nationwide cancer cases represented in FCDS.

**Results**

Figure 1 depicts the linkage process and resulting sample sizes. There were 1,638,946 eligible records from the 1986–2009 NHIS linked to 2,520,333 eligible records from the 1981–2010 FCDS, resulting in 8,210 matches after processing by the NCHS. Of those, 59 were excluded due to missing cancer diagnosis date and 41 were excluded due to a single cancer diagnosis prior to 1981 (the FCDS does include some records obtained prior to official FCDS creation). The final linked data set, therefore, contained 8,110 matched cancer cases. This sample represents 6,281 cancer cases diagnosed

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>Cancer Diagnosed Prior to NHIS Interview (Prevalent Cases)</th>
<th>Cancer Diagnosed Subsequent to NHIS Interview (Incident Cases)</th>
<th>Cancer Diagnosed Prior to NHIS Interview (Prevalent Cases)</th>
<th>Cancer Diagnosed Subsequent to NHIS Interview (Incident Cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Breast</td>
<td>401</td>
<td>724</td>
<td>6,683</td>
<td>12,067</td>
</tr>
<tr>
<td>Prostate</td>
<td>403</td>
<td>666</td>
<td>6,717</td>
<td>11,100</td>
</tr>
<tr>
<td>Lung</td>
<td>89</td>
<td>613</td>
<td>1,483</td>
<td>10,217</td>
</tr>
<tr>
<td>Colorectal</td>
<td>233</td>
<td>446</td>
<td>3,883</td>
<td>7,433</td>
</tr>
<tr>
<td>Bladder</td>
<td>120</td>
<td>200</td>
<td>2,000</td>
<td>3,333</td>
</tr>
<tr>
<td>All cancers</td>
<td>1,908</td>
<td>4,367</td>
<td>31,800</td>
<td>72,750</td>
</tr>
</tbody>
</table>

*The number of cancer cases was estimated by dividing the sample size obtained in this pilot linkage by 6%, which represents the proportion of Florida cancer cases in the United States.*
### Table 5. Number of Linked Cancer Cases (Florida Residents Only) in the 1986–2009 National Health Interview Survey (NHIS) and 1981–2010 Florida Cancer Data System (FCDS) Linkage, Age, and Cancer Diagnosis Factors by Time of Diagnosis and Survey

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cancer Diagnosed Prior to NHIS Interview</th>
<th>Cancer Diagnosed Subsequent to NHIS Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (weighted %)</td>
<td>n (weighted %)</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–39</td>
<td>86 (4.6)</td>
<td>521 (11.7)</td>
</tr>
<tr>
<td>40–64</td>
<td>561 (28.6)</td>
<td>2,165 (49.5)</td>
</tr>
<tr>
<td>≥65</td>
<td>1,261 (66.8)</td>
<td>1,681 (38.9)</td>
</tr>
<tr>
<td>Survey years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986–1996</td>
<td>765 (32.5)</td>
<td>3,013 (63.6)</td>
</tr>
<tr>
<td>1997–2009</td>
<td>1,143 (67.5)</td>
<td>1,354 (36.4)</td>
</tr>
<tr>
<td>Cancer diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before 1986</td>
<td>349 (16.1)</td>
<td>–</td>
</tr>
<tr>
<td>1986–1995</td>
<td>941 (43.2)</td>
<td>753 (15.7)</td>
</tr>
<tr>
<td>1996–2005</td>
<td>576 (36.7)</td>
<td>2,235 (49.5)</td>
</tr>
<tr>
<td>2006–2010</td>
<td>42 (4.0)</td>
<td>1,379 (34.8)</td>
</tr>
<tr>
<td>Years between survey &amp; cancer diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–5</td>
<td>1,082 (53.5)</td>
<td>1,790 (44.7)</td>
</tr>
<tr>
<td>6–10</td>
<td>481 (25.8)</td>
<td>1,220 (26.7)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>345 (20.7)</td>
<td>1,357 (28.7)</td>
</tr>
</tbody>
</table>

Table 5 provides the estimated number of linked cancer cases we would expect to identify if a linkage were conducted with cancer registries nationwide. Based on the sample size obtained in this pilot linkage and because the FCDS represents approximately 6% of all cancer diagnoses in the United States, we estimate that a nationwide linkage would yield a total of 104,550 linked cancer cases and would include approximately 31,800 prevalent and 72,750 incident cases. Of note, this projection provides a crude estimate of the scale of the nationwide linkage program. The actual numbers will be influenced by the number years of registry operation in each state as well as variations in the number of state-specific NHIS participants given the NHIS sampling frame.

Among the Florida residents diagnosed with cancer prior to their NHIS interview (Table 5), approximately two-thirds of cases were in patients aged 65 years and older. Sixty-eight percent had their NHIS interview in 1997–2009, and 80% were diagnosed with cancer in 1986–2005. Over half of cancer cases in this group were diagnosed within 5 years between their survey and cancer diagnosis.
years of their NHIS interview. Among those diagnosed after their NHIS interview, 39% were aged 65 years and older, 64% had their interview in 1986–1997, 84% were diagnosed in 1996–2010, and 44% were diagnosed within 5 years of their interview.

Discussion

This linkage results in an enhanced data set of persons diagnosed with cancer before and after their participation in the nationally representative NHIS, creating distinctive cohorts of both cancer survivors and future cancer cases. Linking data from the NHIS and FCDS also provides a unique resource to evaluate clinical cancer registry variables in conjunction with socioeconomic status, risk factors, screening behaviors, and health care utilization variables that are typically not available in cancer registries. As we have shown in descriptive analyses, important variables in cancer surveillance research such as education and self-rated health can be obtained through linkage with the NHIS. This combination of variables available through NHIS linkage provides greater opportunities at the population-level to investigate the influence of a multitude of factors on the development of cancer. As an example, what is the attributable risk of developing lung or breast cancer for smokers and obese individuals compared to nonsmokers or those of healthy weight? In terms of cancer survivorship, this data resource could be used to characterize the quality of life among cancer survivors. For example, adjusting for time since diagnosis and cancer stage, does a cancer survivor’s sociodemographic characteristics or health behaviors influence their quality of life and how do these associations differ from cancer nonsurvivors? Results from analysis of this enriched database could be extremely useful in informing interventions to help prevent cancer and support cancer survivors.

Not only does this linkage provide a distinct resource of variables not otherwise available, it also removes the need to rely on area-based measures, which can be inaccurate or out-of-date, or small patient samples, which can be costly to obtain and possibly unrepresentative of the general cancer population. In addition, linking multiple years of cancer registry data to the NHIS provides a longitudinal component to the cross-sectional survey data and a retrospective look at cancer patients’ risk history.

Since conducting the linkage, our study team has undertaken analyses to demonstrate the utility of this resource for hypothesis-driven research. Specifically, we have examined the health status of the cancer survivors in this linked sample. Theses analyses reveal that, compared to those without a cancer history, cancer survivors have an increased likelihood of activity limitations and fair or poor self-rated health. In addition, those diagnosed with late-stage disease bear an even greater burden of activity limitations than those with early stage disease. We have additional analyses underway to describe the health status of cancer survivors by occupational status, the prevalence of health behaviors and outcomes among cancer survivors, and the factors associated with mortality among these survivors.

Having demonstrated the feasibility of a linkage between cancer registry data and the NHIS, further opportunity lies in the potential linkage of national registry data and other national health surveys. Since 1986, more than 1.8 million people have participated in health surveys conducted by the NCHS. Along with the NHIS, these national health surveys include the National Health and Nutrition Examination Surveys (NHANES) and the National Health Care Surveys, among others. Collectively, these surveys contain substantial information on demographics, socioeconomic status, medical expenditures, health status, and cancer risk factors (including the periodic cancer supplements), as well as extensive clinical health data and biological measures (NHANES only). Additionally, these national databases have been linked to the National Death Index to obtain mortality information, and other data sets, such as Medicare claims, to obtain health care utilization data.

It is important to note potential obstacles that were encountered and may impact a linkage between NCHS surveys and cancer registries nationally. Many of the challenges faced in this linkage were a result of linking a nationally representative sample with a single state registry. For example, our first obstacle was the task of accounting for the complex survey design of the NHIS in the analyses. Survey participants have different selection probabilities and not accounting for this in an analysis can result in biased estimates and exaggerated differences between groups. Second, there was the challenge of creating the most appropriate weights for these linked cancer cases, which are needed to represent the Florida population overall based on each year of the NHIS. This created a complication with survey participants from other states who were diagnosed with cancer in Florida since participants from other states receive a weight of zero. As a result, these “movers” are not included in weighted analyses, which reduced the sample size for analyses. We are also unable to identify those people who took the NHIS in Florida but were subsequently
diagnosed with cancer in a state other than Florida. Both of these limitations are a result of linking nationwide data (NHIS) with state-level data (FCDS) and could presumably be averted with a linkage of registries nationwide.

The administrative and financial costs and effort associated with performing this linkage reflect another potential challenge toward the move to a national linkage. The 4 major steps toward creating a linked file included: (1) the administrative and legal process of initiating a data exchange agreement between NCHS and FCDS, (2) NCHS preparation and delivery of the file to be linked by FCDS, (3) linkage of this file at the FCDS and secure return of linked cancer data to the NCHS, and (4) quality control checks and preparation of the data for placement in the RDC. Person-hours were not carefully monitored during this process but the effort was not trivial.

Conclusion

Cancer registry linkage to other data sources has tremendous potential for enriching the value of national health surveillance systems which provides important information on sociodemographic characteristics, screening behaviors, comorbidities, risk factors, health care access, and quality of life. Our study demonstrated that registry linkage to the NHIS is a feasible endeavor, and it results in a unique data set for hypothesis-driven research, despite the challenges encountered. Because of the cost and effort required, it will be important to explore strategies to streamline linkage efforts with each state registry while maintaining the highest level of data security and integrity. Ultimately, the opportunities provided by linking survey data to cancer registries will need to be weighed against the cost and effort required before expansion of linkage to all state registries is implemented.

Acknowledgements

We would like to thank additional colleagues at the NCHS who collaborated with us on this project and were responsible for the processing of the linked data and creation of the final data set: Jennifer Parker, Donna Miller, Dean Judson, Keith Zevallos, Peter Meyer, and Hannah Day.

References

Collaborative Development and Results of a Nigerian Trauma Registry

Laura D. Cassidy, MS, PhD; Oluwole Olaomi, MBBS, FWACS, FACSc; Allison Ertl, MS; Emmanuel A. Ameh, MBBS, FWACS, FACSc

Abstract: Background: More than 90% of injury-related deaths and disabilities occur in low- and middle-income countries. The development of the Nigerian Trauma Registry (NTR) and the first descriptive data analysis of the patient characteristics, mechanisms of injury, injury severity, and treatments are reported. Methods: Existing data collection tools were modified to capture a minimum data set of variables reflective of the trauma experience in Nigeria. Data are collected using the secure, Web-based application, REDCap (Research Electronic Data Capture). Results: Two hospitals entered 564 patients into the registry. Motor vehicle accidents were the most frequently reported trauma (69.2%). Of the 51 fall injuries, 82% were from buildings. There were 229 mass casualties, including bus accidents (41.5%), bombings or blasts (28.8%), multiple vehicle accidents (23.6%), fires (3.1%), and civil conflicts or riots (3.1%). External soft tissue was the most commonly reported injury region followed by extremities, head and neck, face, abdomen, and thorax/spine. Only 18.1% of patients arrived by ambulance. There were 19 recorded in-hospital deaths and 79.0% of these were due to motor vehicle accidents. Conclusions: This is the largest report of injury surveillance in this country. These data are essential to inform policy makers about the increasing trauma burden and provide a strong advocacy tool, prevention opportunities, provisions for unmet capacity needs, and better allocation of limited health care resources. The NTR has demonstrated that development and implementation of an electronic trauma registry is feasible in low- and middle-income countries. The NTR evolved through international collaborations that included a partnership with an American epidemiologist and 2 Nigerian hospitals that contributed their individual and institutional capabilities. Local champions are required to drive the initiation and implementation of registries.

Key words: low and middle income countries, trauma, trauma registries, unfunded project

Introduction
Injury remains a leading global public health burden, resulting in 5.8 million deaths each year and leaving many more disabled.1-3 Over 10% of the global burden of disease is due to injury and at least 9 people die from injuries every minute, making injury the leading cause of death for men and women under the age of 45.1-3 Importantly, the burden of injury is disproportionately placed on low- and middle-income countries (LMICs), where more than 90% of injury-related deaths and disabilities occur.1-3

In 2004, The World Health Organization (WHO) published the Guidelines for Essential Trauma Care as an attempt to decrease injuries and mitigate injury disparities between high-income countries (HICs) and LMICs through the implementation of trauma care systems.1-3 Trauma care systems address all aspects of care through seamless transitions from the prehospital setting to long-term care.2,3,6-8 By integrating existing resources and improving the organization of trauma care, trauma systems in HICs have repeatedly demonstrated significant reductions in trauma mortality rates.3,6-8 Trauma registries are an integral component of trauma systems, because access to trauma-specific data increases the efficiency of trauma quality improvement.2,3 A trauma registry has been defined as “a disease-specific collection composed of a file of uniform data elements that describe the injury event, demographics, prehospital information, diagnosis, care, outcomes, and costs of treatment for injured patients.”9 Trauma registries can facilitate injury prevention through descriptive epidemiology; development of population-specific injury severity scales and stringent evaluation of these scales for reliability and validity; data to promote research for disaster preparedness; and evaluation of quality of care and quality improvement activities at individual trauma centers and across centers, including trends in care. A major barrier to implementation of quality improvements in the systems of trauma care in LMICs is the incomplete characterization of the basic epidemiology of trauma. However; the implementation of organized trauma systems and trauma registries has been minimal in LMICs.2 LMICs have the lowest trauma registry activity, with 83% of 571 publications originating from HICs and only 4% from LMICs.3

Lack of a trauma registry in most LMICs means that high-quality data are not available to obtain information about injury epidemiology, care, and patient outcomes.10-13 Where available, such data sources are frequently incomplete and are unreliable for injury surveillance, planning, prevention, and control. In one report from Malawi,12 a capture-recapture model using primary data from police- and hospital-based registers was used over a 1-year period
to estimate mortality from trauma. This was one method to overcome some of the limitations of incomplete data sources.

Where some form of registry exists in LMICs, they are often entirely paper-based, making data entry and retrieval cumbersome and time-consuming. Such registries may be incomplete and the efforts face significant barriers including lack of funding and unfavorable government health policies.14

Determining the volume and types of trauma patients treated is important for improving patient care and making data-driven resource allocation decisions; therefore, a Nigerian National Trauma Registry (NTR) was implemented for hospital performance improvement and quality improvement purposes. The study reported here describes the development of the NTR and the first analysis of NTR data to describe patient characteristics, mechanisms of injury, injury severity, and treatments.

Methods

Responding to the surgeons’ need for standardized trauma data in Nigeria, a collaboration was formed that included 2 Nigerian hospitals, surgeons, and an American epidemiologist. This collaboration tested the feasibility of collecting a minimum data set at 2 different institutions. The data collection tools developed in Haiti15 and Uganda16,17 were reviewed and modified to initially capture a minimum data set consisting of variables reflective of the trauma experience in Nigeria. The NTR records patient demographics, injury setting, transportation to hospital, vital signs, anatomic sites involved, treatment, mode of leaving the emergency department, and injury and arrival timings. This minimum data set is the first step toward creating an injury surveillance system. It provides basic information on patient volume, demographics, resource utilization for treatment, mortality status, and injury severity and type. It was not yet sophisticated enough to inform treatment decisions but provided valuable initial surveillance information.

The trauma registry was built using REDCap (Research Electronic Data Capture),18 which is an open-source, secure, Web-based application designed to support data capture. REDCap provides an intuitive interface for validated data entry; audit trails for tracking data manipulation and export procedures; automated export procedures for seamless data downloads to common statistical packages; and procedures for importing data from external sources. The system is password protected and approved users can enter and view data from their institutions only. However, de-identified data may be downloaded by an approved user across institutions. It includes ad hoc reporting tools, branching logic, file uploading, and calculated fields. It also reduces errors from version control in that the most current data are always available and multiple versions of the data are not being accessed simultaneously. Researchers interested in using REDCap can view the website that provides a list of institutions that are running REDCap.

REDCap can be installed in a variety of environments for compliance with standards such as HIPAA (Health Insurance Portability and Accountability Act), 21 CFR Part 11 (US Food and Drug Administration Code of Federal Regulations), FISMA (Federal Information Security Management Act), and international standards, and is fully personalized to meet your security policies and user needs. REDCap can run on different operating systems such as Linux, Unix, Windows, Mac. For the NTR, it was hosted by the Medical College of Wisconsin and accessed via the Internet at the 2 hospitals. It can be accessed from any computer or handheld device with Internet access. A new feature is that users can collect data using a mobile app on an iPhone, iPad, Android phone, or tablet. This enables the user to collect data offline if there is poor Internet connectivity and upload the data later. This feature is especially important in LMICs and was not available during the NTR study period.

In 2010, the Ahmadu Bello University Teaching Hospital, Zaria, Nigeria and the National Hospital, Abuja, Nigeria established the registry. The surgeons at each Nigerian partner hospital are responsible for data collection. It was important to modify the data collection process to accommodate the individual workflow and resources at each of the respective hospitals.

National Hospital, Abuja

A prototype data form was printed for the trauma physicians to study and they were trained to enter the data into REDCap. The data are obtained from patients at the trauma unit and are entered by the attending doctor after resuscitation and stabilization in the emergency department. The registrar at discharge entered the patient data forms into REDCap. Once the hospital experienced the importance of the registry, a dedicated medical record staff previously trained in cancer registry data collection was appointed to ensure more complete case ascertainment and data collection. Periodic cross-checks are conducted between the hospital records by the head of the trauma unit and the data is entered into REDCap.

Ahmadu Bello University and Ahmadu Bello University Teaching Hospital, Zaria

At this institution, data are entered for admitted trauma patients by accident and emergency department physicians and pediatric surgery trainees. Data entry is supervised by the consultant in charge of the emergency department and the co-investigator/chief of pediatric surgery. Initial data are collected in the emergency department, and in the wards for patients who are transferred to the wards.

Whenever possible, data are collected on admission, but this is often not achievable due to fact that the same doctors entering data are involved in actual patient care. Therefore, data are often collected when the emergency department is quiet or the next day, but on the same day for deceased patients. All the data variables for the registry were entered into a hardcover book and then were entered and uploaded in batches as work pressure and Internet access allowed. These hard copies are stored in locked offices and locked filing cabinets when not in use.

Because no ICD-9 (International Classification of Diseases, ninth revision) data are collected at hospitals in
Nigeria, the clinicians recorded the Abbreviated Injury Scale (AIS) for each of the body regions. The AIS is an anatomical scoring system that provides a reasonably accurate way of ranking the severity of injury for 6 body regions (head, face, chest, abdomen, extremities [including pelvis], and external). Injuries are ranked on a scale of 1 to 6, with 1 being minor, 5 being severe, and 6 being an unsurvivable injury.\textsuperscript{19,20} The Injury Severity Score (ISS) provides a measure of severity of multiple injuries and ranges from 1 to 75, with 75 being unsurvivable. The ISS is calculated as the sum of the square of the 3 highest AIS values.\textsuperscript{20} For analysis purposes, ISS is often categorized. An ISS of less than 15 is generally considered a less severely injured patient, 16 through 25 indicate moderate injury, and greater than 25 indicates a more severely injured patient. Data on intentional and unintentional injuries were collected using the WHO definition. Unintentional injuries include most road traffic, poisoning, falls, fire, and burn injuries. Intentional injuries include interpersonal violence. Data from the registry from April 2010 through June 2013 were analyzed using the statistical analysis package SAS 9.3\textsuperscript{21} and are reported here.

### Results

Overall, 564 patients were entered into the registry. Patient characteristics are shown in Table 1. Most variables had some missing data resulting in different denominators. The mean patient age was 30.4 ± 14.5 (range, 1 month to 72 years) and 77.6% were male. Blunt injury was the most commonly reported injury (66.0%) followed by burns (4.6%) and penetrating injury (29.4%); however, the majority of

#### Table 1. Patient Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>432</td>
<td>77.6</td>
</tr>
<tr>
<td>Female</td>
<td>125</td>
<td>22.4</td>
</tr>
<tr>
<td>Total</td>
<td>557</td>
<td>100.0</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–9</td>
<td>44</td>
<td>9.0</td>
</tr>
<tr>
<td>10–19</td>
<td>51</td>
<td>10.4</td>
</tr>
<tr>
<td>20–29</td>
<td>143</td>
<td>29.2</td>
</tr>
<tr>
<td>30–39</td>
<td>133</td>
<td>27.2</td>
</tr>
<tr>
<td>40–49</td>
<td>63</td>
<td>12.9</td>
</tr>
<tr>
<td>50–59</td>
<td>38</td>
<td>7.8</td>
</tr>
<tr>
<td>60–69</td>
<td>13</td>
<td>2.7</td>
</tr>
<tr>
<td>70–79</td>
<td>4</td>
<td>0.8</td>
</tr>
<tr>
<td>Total</td>
<td>489</td>
<td>100.0</td>
</tr>
<tr>
<td>Injury type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blunt</td>
<td>101</td>
<td>66.0</td>
</tr>
<tr>
<td>Burn</td>
<td>7</td>
<td>4.6</td>
</tr>
<tr>
<td>Penetrating</td>
<td>45</td>
<td>29.4</td>
</tr>
<tr>
<td>Total</td>
<td>153</td>
<td>100.0</td>
</tr>
<tr>
<td>Intent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intentional</td>
<td>108</td>
<td>19.7</td>
</tr>
<tr>
<td>Unintentional</td>
<td>440</td>
<td>80.3</td>
</tr>
</tbody>
</table>

Data were missing for this field. Most injuries were unintentional (80.3%).

As shown in Table 2, motor vehicle accidents were the most commonly reported mechanism of injury (69.2%), including cars (63%), minibuses (15%), motorcycles (10%), and cars hitting a pedestrian (9%). Highways and roads were the most common place of injury (71.8%). Of the 51 fall injuries, 82% were from buildings. There were 229 mass casualties, including bus accidents (41.5%), bombings or blasts (28.8%), multiple vehicle accidents (23.6%), fires (3.1%), and civil conflicts or riots (3.1%) (data not shown).

Patients suffered multiple injuries, with 1,106 individual injuries recorded for 564 patients. External soft tissue was the most common injury region (n = 288) followed by extremities (n = 267), head and neck (n = 241), face (n = 193), abdomen (n = 107), and thorax/spine (n = 106). As shown in Figure 1, based on AIS, the face and external soft tissue had the highest proportion of minor/moderate injuries, extremities had the highest proportion of serious/severe injuries, and head and neck had the highest proportion of critical/maximal injuries. AIS was recorded for 464 patients and ISS was calculated for these. Of these, 297 (65.1%) had an ISS below 15, indicating minor severity; 89 (19.5%) had an ISS greater than 15 and less than 26, indicating moderate severity; and 70 (15.4%) had an ISS greater than 25, indicating severe injuries.

Table 3 shows the percentage of procedures by injury severity. Patients received multiple procedures; therefore,

#### Table 2. Distribution of Reported Mechanism of Injury

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle</td>
<td>368</td>
<td>69.2</td>
</tr>
<tr>
<td>Fall</td>
<td>51</td>
<td>9.6</td>
</tr>
<tr>
<td>Stab/cut</td>
<td>51</td>
<td>9.6</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
<td>4.5</td>
</tr>
<tr>
<td>Gunshot</td>
<td>22</td>
<td>4.1</td>
</tr>
<tr>
<td>Burn</td>
<td>12</td>
<td>2.3</td>
</tr>
<tr>
<td>Animal bite/horn/hoof</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td>Sports</td>
<td>2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

![Figure 1. Injuries by Severity](image-url)
the percentages do not sum to 100%. The most common procedure for all injuries was laboratory procedures and radiographs. The most severely injured patients had the highest percentage of computed tomography (CT) scans, with 23 of the 70 severely injured patients receiving a CT scan (32.9%). They also had the highest percentage of airway management procedures (27.1%), neurosurgery (7.1%), chest surgery (8.6%), and abdominal surgery (8.6%). The patients with moderate injuries had the highest percentage of extremity surgeries (22.5%, compared to 11.8% in the lowest severity group and 20.0% in the highest severity group).

Of the 501 patients with mode of arrival recorded, only 18.1% arrived by ambulance and 70.1% arrived by other vehicle (Table 4). Overall, 28% were transferred to another hospital. There were 19 recorded deaths and 15 of these (79.0%) were due to motor vehicle accidents. Of the 196 patients with an injury severity and discharge status recorded, 18 (9.2%) died. Of the patients with an ISS below 15, 3 (2.3%) did not survive. Three patients (8.8%) with an ISS between 16 and 25 did not survive, and 12 patients (36.4%) with an ISS above 25 died.

### Table 3. Percentage of Individual Procedures Performed by Injury Severity Score (ISS) Group

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Mild Injury ISS &lt;15</th>
<th>Moderate Injury ISS 16–25</th>
<th>Severe Injury ISS 26–75</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiograph</td>
<td>216 72.7</td>
<td>79 88.8</td>
<td>49 70.0</td>
<td>344 75.4</td>
</tr>
<tr>
<td>Computed tomography scan</td>
<td>38 12.8</td>
<td>27 30.3</td>
<td>23 32.9</td>
<td>88 19.3</td>
</tr>
<tr>
<td>Transfusion</td>
<td>9 3.0</td>
<td>8 9.0</td>
<td>7 10.0</td>
<td>24 5.3</td>
</tr>
<tr>
<td>Airway management</td>
<td>16 5.4</td>
<td>11 12.4</td>
<td>19 27.1</td>
<td>46 10.1</td>
</tr>
<tr>
<td>Laboratory</td>
<td>238 80.1</td>
<td>76 85.4</td>
<td>59 84.3</td>
<td>373 81.8</td>
</tr>
<tr>
<td>Neurosurgery</td>
<td>4 1.4</td>
<td>3 3.4</td>
<td>5 7.1</td>
<td>12 2.6</td>
</tr>
<tr>
<td>Chest surgery</td>
<td>3 1.0</td>
<td>4 4.5</td>
<td>6 8.6</td>
<td>13 2.8</td>
</tr>
<tr>
<td>Abdominal surgery</td>
<td>6 2.0</td>
<td>2 2.2</td>
<td>6 8.6</td>
<td>14 3.1</td>
</tr>
<tr>
<td>Head surgery</td>
<td>12 4.0</td>
<td>6 6.7</td>
<td>7 10.0</td>
<td>25 5.5</td>
</tr>
<tr>
<td>Extremity surgery</td>
<td>35 11.8</td>
<td>20 22.5</td>
<td>14 20.0</td>
<td>69 15.1</td>
</tr>
<tr>
<td>Total</td>
<td>297 65.1</td>
<td>89 19.5</td>
<td>70 15.4</td>
<td>456 100.0</td>
</tr>
</tbody>
</table>

* Patients had multiple procedures so percentages do not total 100%.
* The denominator for the percentages is the number of mildly injured patients (n = 297).
* The denominator for the percentages is the number of moderately injured patients (n = 89).
* The denominator for the percentages is the number of severely injured patients (n = 70).

### Table 4. Mode of Arrival

<table>
<thead>
<tr>
<th>Arrival Method</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambulance</td>
<td>97</td>
<td>18.1</td>
</tr>
<tr>
<td>By foot</td>
<td>9</td>
<td>1.7</td>
</tr>
<tr>
<td>Other vehicle</td>
<td>375</td>
<td>70.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>16</td>
<td>3.0</td>
</tr>
<tr>
<td>Other</td>
<td>38</td>
<td>7.1</td>
</tr>
<tr>
<td>Total</td>
<td>535</td>
<td>100.0</td>
</tr>
</tbody>
</table>

### Discussion

Trauma registries are an integral part of comprehensive trauma systems. Injury surveillance is facilitated by generating relevant data in a timely manner that are critical for a coordinated approach to trauma care, including planning and developing prevention and control measures. It also provides important information on care and outcomes. Hospital-based registries are particularly helpful in evaluating morbidity, mortality, and outcome of treatment, which are important for maintaining a continuing quality improvement in trauma care. Analysis of these data informs the trauma system for training needs and planning resource allocation.

The results of the descriptive analysis should be interpreted with caution due to the preliminary nature of the data. These data have not been verified using quality control methods and there were some variables with a significant amount of missing data. However, it is the largest trauma data set in the country, and the results are consistent with what is known about these regions. Motor vehicle accidents were the most common mechanism of injury in the NTR. This is consistent with the limited reports from Nigeria. One previous report from Nigeria used a 2-stage stratified cluster sampling survey of households. The results from that survey estimated that the burden of road traffic injuries alone is 41 per 1,000 people, with a mortality rate of 1.6 per 1,000 people.22 These data can be used to target interventions to increase road traffic safety such as seat-belt use, installation of traffic signals, and enforcement of speed limits. A relatively small percentage of severely injured patients received surgery, indicating an opportunity to educate health care providers about treatment guidelines.
and triage. It may also reflect the need for more surgeons or operating rooms. Only 18% of the patients arrived by ambulance, with the majority arriving by other vehicle. This is especially problematic with mass-casualty events. There is an opportunity to improve triage and prehospital care; however, this also requires an increase in resources.

Although the burden of injuries is thought to be increasing in LMICs, the magnitude is difficult to ascertain due to a lack of accurate and reliable data and the fact that population-based data are not readily available in Sub-Saharan Africa. In the 1990s, efforts began to develop a trauma registry in 2 hospitals in Uganda. It later expanded across the country and remains active. The Kampala Trauma Score was developed as a direct result of the registry.\textsuperscript{16,17} In South Africa, the Cape Town Trauma Registry pilot study\textsuperscript{18} was able to determine injury pattern and spatial distribution of injuries in the area covered, but the injury setting was captured only 50% of the time. A hospital-based trauma registry was developed in Haiti that was tailored to local needs.\textsuperscript{15} It compared a coordinator-based trauma registry (using a trained trauma data collector or coordinator who is not involved in patient care) and a provider-based trauma registry involving inputs from relevant stakeholders in the hospital. The provider-based trauma registry was viewed as more appropriate and more conducive to expansion to outpatient trauma and was better suited to LMICs.

The success of the Kampala trauma registry and a small number of trauma registries in other LMIC settings\textsuperscript{16,17} have demonstrated the feasibility and utility of trauma registries in LMICs. However, most of these registries were paper-based and the use of an electronic trauma registry in LMICs remains limited and underutilized. The NTR was designed from the outset as a Web-based registry. This facilitates data capture and retrieval and makes data analysis and injury surveillance more robust and timely and less cumbersome.

The NTR has demonstrated that development, deployment, and implementation of electronic trauma registry is feasible and achievable in this setting. To date, this is the largest report of injury surveillance in this country. It is important to stress that the NTR evolved through international collaborations and partnerships that contributed their individual and institutional capabilities in trauma epidemiology, registry development, and software (REDCap) and Web-hosting capabilities. These capabilities are presently limited in many LMICs. The success of the NTR was also largely due to the efforts and passion of local individual surgeons for trauma care, injury research, and epidemiology. Such local champions are required to drive the initiation and implementation of such registries.

A number of shortcomings and limitations have been identified with the NTR. Incomplete data sets, particularly records of patient discharge and outcome, remain a challenge. This is primarily due to the added burden on surgical trainees and emergency department staff to enter data into the registry. In HICs, trauma registrars undergo standardized training for data collection and are dedicated to populating the trauma registry. However, in these 2 LMIC hospitals, there were no trained registrars. In addition, there was no funding to hire and train registrars; therefore, for this pilot, the clinical staff was responsible for data collection. Spatial distribution and mapping of trauma/injury-prone locations was not possible as such information was not captured in the registry. This information would be helpful for targeted prevention and control strategies, and it is planned to include this in the registry in future. There are also challenges to a Web-based registry in West Africa due to unreliable Internet access. Although Internet access is improving in most areas, an offline application which can be used on mobile devices is still necessary. Recorded data can then be periodically uploaded to the Web registry when Internet access is available.

There are many opportunities for process improvement and lessons to be learned from this study. The data captured during admission was the most complete; however, to improve data collection during hospitalization, dedicated, trained registrars are essential. Patients often leave against medical advice and their data and discharge status are not captured. Other variables had a high percentage of missing data, such as injury type, and the reasons for this omission are unclear and warrant further investigation. There are limited resources to perform quality checks of the data and often there are no paper charts to gather data retrospectively. Enhancements to REDCap include the ability to collect data on mobile devices which could improve the process. At this time, an approach to prospective, succinct, standardized data collection is the most feasible approach. It requires behavior modification and, as data are accumulated, analyzed and disseminated, the culture of data collection will grow and improve.

It is planned to progressively recruit other institutions and centers across Nigeria with the hope of achieving a country-wide NTR over the next few years. Given the enthusiasm and interest from individual local surgeons and the international partners, the prospect of achieving this is promising.

The data from the NTR, even at this early stage, may be used for advocacy, particularly to capture the attention and support of policy makers and other trauma stakeholders in Nigeria and Sub-Saharan Africa. This is particularly important as the campaign by the international surgical community gathers momentum towards the inclusion of surgical care in primary care in the Sustainable Development Goals (SDG). More than 190 world leaders have committed to 17 SDGs to help end extreme poverty, fight inequality and injustice, and fix climate change, with good health being one of the 17 goals. Trauma care is also an important component of universal health coverage.

The NTR offers a unique opportunity for trauma and injury surveillance in this setting. However, sustainability of the registry will depend on the recruitment of dedicated registry staff in participating centers to help with data capture and maintenance. Key registry personnel for trauma research and epidemiology are needed. Adequate funding is needed to achieve this goal.

In LMICs, it is often difficult to convince policy makers about the increasing burden of trauma, trauma deaths, and associated long-term disabilities without accurate and reliable data. Trauma registry data would be a strong advocacy...
tool, and should help in planning control measures, making provisions for unmet capacity needs as well as appropriate allocation of already limited health care funding and resources. The current model can be implemented in other LMICs as well.

Collaboration between the existing efforts and implementing lessons learned from HICs can leverage existing resources and expertise to strive toward a minimum standardized data set in LMICs. These data are essential to convince policy makers about the increasing burden of trauma, mortality, and associated long-term disabilities. These data would provide a strong advocacy tool, and help in planning control measures, making provisions for unmet capacity needs as well as appropriate allocation of already limited health care funding and resources.

Acknowledgements

The authors would like to acknowledge the support of the hospital administration at the National Hospital Abuja and Ahmadu Bello University and Ahmadu Bello University Teaching Hospital, Zaria as well as the physicians and staff who dedicate their time to data collection. The Research Methods Workshop facilitated by the Association for Academic Surgery provided the opportunity for the authors to meet, generate the idea and design of the registry, and collaborate on its implementation.

References

Using Data on Tumor Grade in Cancer Registries to Enhance Surveillance of Oropharyngeal Cancers in Relation to the Human Papillomavirus Epidemic

Anthony P. Polednak, PhD

Abstract: Background: Incidence rates have been increasing in US whites for squamous cell carcinoma (SCC) at anatomic sites involving the oropharynx (OP) and classified as potentially associated with human papillomavirus (HPV). Registries have not routinely collected data on HPV status of tumors. High tumor grade (poorly differentiated or undifferentiated), however, has been associated with HPV positivity in clinical studies. This study explored the potential value of adding data on tumor grade to registry-based surveillance efforts. Methods: Data were obtained on tumor grade for 39,907 OP SCCs diagnosed in 2000–2012 at HPV-associated OP sites in a research database for Surveillance, Epidemiology, and End Results (SEER) Program registries. Grade was compared by anatomic site. Annual percent change in the age-standardized incidence rate was estimated by joinpoint regression. Results: HPV-associated OP SCC sites were predominantly (87%) base of tongue and tonsils (BTT). High-grade comprised 40% for BTT vs 20%–30% for other HPV-associated sites. Temporal increases in total rates for BTT were evident for whites, who had statistically significant increases for high grade but persistently low rates for low grade. Rates increased for unknown grade, reflecting a decline in cancer-directed surgery. Conclusions: Findings support the use of tumor grade for surveillance of OP SCC in relation to the HPV epidemic. Future studies should include non-SEER registries, and also examine HPV status for unknown grade and consistency in grading among pathologists.

Key words: cancer registries, cancer surveillance, human papillomavirus, oropharynx cancer, tonsillar cancer

Introduction

Temporal increases in age-standardized incidence rates for squamous cell carcinomas (SCCs) of certain anatomic sites in the oropharynx (OP) in recent decades have been reported for US whites, using data from cancer registries in the Centers for Disease Control and Prevention (CDC) National Program of Cancer Registries (NPCR) and the National Cancer Institute (NCI) Surveillance, Epidemiology and End Results (SEER) Program. These increases have been interpreted as reflecting the epidemic of human papillomavirus (HPV) infection. Surveillance has involved selecting “potentially HPV-associated” or “HPV-related” sites in the OP, recognizing that not all cases actually involve HPV.

US registries have not routinely collected data on HPV status of all incident cancers. Only samples of cases have been studied for HPV status of tumor tissues. Data from the CDC Cancer Registry Sentinel Surveillance System, designed to assess specific types of HPV in HPV-associated cancers prior to the introduction of HPV vaccine in 2006, have been combined with data from the SEER Residual Tissue Repository. In tissues available for selected OP anatomic sites, from the 7 US population-based cancer registries involved, the proportion HPV-positive increased slightly over time but the sample sizes were limited (36 diagnosed in 1994–1999, 41 in 2000–2003, and 511 in 2004–2005). US registries, however, routinely collect data on histopathologic tumor grade, which describes the degree of differentiation or resemblance to normal (keratinized) squamous epithelium. In clinical studies, HPV-positive OP SCCs have been reported as predominantly poorly differentiated or nonkeratinized. In a clinical series of 60 patients at Johns Hopkins Hospital, high grade (poorly differentiated or undifferentiated) was statistically significantly more frequent in HPV-positive vs HPV-negative OP SCCs (59% vs 31%, respectively); 20 (71%) of all 28 poorly differentiated cases were HPV positive. Therefore, the HPV epidemic may have resulted in larger temporal increases in incidence rates for higher vs lower grade OP SCC.

The purpose of this study was to assess the completeness of tumor grade data in SEER, and their potential utility for surveillance of incidence of HPV-associated OP SCC using a research database for SEER registries. The focus was on SCC of base of tongue and tonsil (BTT), for which the evidence for the role of HPV is strongest. For example, in a systematic review/meta-analysis of studies worldwide, HPV DNA detection was 60.4% for OP SCC in North American studies, and highest for BTT. In a US registry-based study of tissues for a sample of patients with cancers at various anatomic sites, HPV DNA detection was 82.0% for tonsil and 70.0% for base of tongue, vs only 42.9% for other OP sites selected. Presence of viral E6/E7 oncogene mRNA, regarded as more indicative of a causal role for HPV, was not tested, but this marker has been found in large proportions of OP SCC with HPV detected. In contrast, HPV may be involved in only a small proportion of oral cavity SCCs.

Address correspondence to Anthony P. Polednak. Email: appoled7@yahoo.com.
Methods

The Database Used, and Selection of Site (Topography) and Morphology Codes

The de-identified SEER research database\(^8\) included diagnoses in 2000–2012 for all 18 SEER registries, including 4 registries that together cover all of California (Los Angeles county, San Francisco–Oakland metropolitan area, San Jose-Monterey area, and Greater California); 3 that together cover all of Georgia (Atlanta metropolitan area, rural Georgia, and Greater Georgia); 8 statewide registries (Connecticut, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, and Utah); 2 other metropolitan areas (Detroit, Michigan, and Seattle–Puget Sound, Washington); and 1 registry covering the Alaska Native population of Alaska. The 18 registries together cover about 28% of the US population (http://www.seer.cancer.gov).

International Classification of Diseases for Oncology Version 3 (ICD-O-3) codes selected for invasive cancers included SCC (morphology codes 8050–8084)\(^1-4\) for base of tongue (ICD-O-3 topography code C019) (n = 15,854), lingual tonsil (C024) (n = 321), tonsil (C090–C099, including tonsil not otherwise specified or NOS, code C099) (n = 18,361), and Waldeyer’s tonsillar ring (C142) (n = 7), or a total of 34,543 BTT SCCs. BT SCCs comprised 86.6% of all 39,907 SCCs at OP sites classified as HPV-associated in Table 1.\(^2-4\) The remainder were overlapping lesion of tongue (C02.8); other OP and OP NOS (C102, C108, C109); pharynx NOS (C140); and overlapping lesion of lip, OP, and oral cavity (C148).\(^2-4\) Aside from BTT (C019 and C090–C099), however, the other specific OP sites selected as potentially HPV-related have differed in various reports.\(^1,5-12\)

For comparison with OP, the distribution of tumor grade was also examined for SCC of oral tongue (C020-023) and tongue NOS (C029), which are sites not classified as potentially HPV-associated.\(^3\)

Variables Used in the Analyses and Statistical Methods

Tumor grade was coded in the database as “1” (well differentiated), “2” (moderately differentiated), “3” (poorly differentiated), “4” (undifferentiated or anaplastic), and “9” unknown.\(^8\) Other variables included sex, race, Hispanic–Latino vs non-Hispanic origin (ethnicity), reporting source (eg, hospital or radiology facility vs death certificate only or autopsy only), method of confirmation (eg, microscopic vs other), and receipt of cancer-directed surgery.\(^8\)

| Table 1. Distribution of Tumor Grade for Incident Invasive Squamous Cell Carcinomas at Anatomic Sites\(^a\) Classified as HPV-Associated: Using Data from SEER Program Registries, 2000–2012 |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|
| **Tumor Grade\(^b\)** | **1** | **2** | **3, 4** | **Unknown** | **Total** |
| **Site (ICD-O-3 code)** | **n** | **%** | **n** | **%** | **n** | **%** | **n** | **%** |
| 1. Oropharynx (OP): Base of tongue and tonsils only | | | | | | | | |
| Base of tongue (C019) | 811 | 5.1 | 5,309 | 33.5 | 5,933 | 37.4 | 3,801 | 24.0 |
| Lingual tonsil (C024) | 13 | 4.0 | 107 | 33.3 | 123 | 38.3 | 78 | 24.3 |
| Tonsil (C090–099), Waldeyer's ring (C142) | 740 | 4.0 | 6,410 | 34.9 | 7,652 | 41.7 | 3,566 | 19.4 |
| **Subtotal, BTT** | (1,564) | **4.5** | (11,826) | **34.2** | (13,708) | **39.7** | (7,445) | **21.6** |
| 2. Other OP, OP NOS (C102,108,109) | 157 | 6.0 | 892 | 34.3 | 787 | 30.2 | 767 | 29.5 |
| Pharynx NOS (C140) | 73 | 5.1 | 543 | 38.1 | 390 | 27.3 | 420 | 29.5 |
| Overlapping lesion of lip, OP and oral cavity (C148) | 16 | 3.4 | 88 | 18.5 | 123 | 25.9 | 248 | 52.2 |
| **Subtotal** | (246) | **5.5** | (1,523) | **33.8** | (1,300) | **28.9** | (1,435) | **31.9** |
| 3. Tongue, overlapping lesion (C028)\(^c\) | 152 | 17.7 | 434 | 50.5 | 169 | 19.7 | 105 | 12.2 |
| **4. Total, all HPV-associated anatomic sites** | 1,962 | **4.9** | 13,783 | **34.5** | 15,127 | **38** | 8,785 | **22.5** |

\(^a\)Defined by ICD-O-3 codes for Morphology (M8050–8084) and topography codes for anatomic sites classified as potentially HPV-associated (see text).\(^1-4\)

\(^b\)Tumor grade: “1” (well differentiated), “2” (moderately differentiated), “3” (poorly differentiated), “4” (undifferentiated or anaplastic), and unknown.\(^8\) Data for grades 3 and 4 were combined as high grade (see text).

\(^c\)This site could potentially include some cancers involving the base of tongue (OP); it has been included with HPV-associated sites in some studies\(^1-4\) but not in others.\(^1,11\)

BTT, base of tongue and tonsils; ICD-O-3, International Classification of Diseases for Oncology, third edition; NOS, not otherwise specified; SEER, Surveillance, Epidemiology, and End Results.
The association between tumor grade category and other categorical variables was examined by $\chi^2$ tests (using Excel 2010). Annual age-standardized incidence rates (ASIRs) per 100,000, along with their standard errors (SE) and 95% lower and upper 95% confidence limits (CLs), were obtained using SEER*Stat software, which uses direct standardization to the age distribution of the 2000 US standard population. ASIRs and their SEs were imported into Joinpoint Regression Program Version 4.0.1 (January 2013) (http://surveillance.cancer.gov/joinpoint/revisions.html), to assess changes in direction and/or magnitude of trends. For 2000–2012, the program requirements selected included a minimum of 3 observations (including the joinpoint) from a joinpoint to either end of the data series (13 years), and at least 4 data points between 2 joinpoints. Using all program requirements, the maximum number of joinpoints was 2 (ie, with 3 line segments). To obtain annual percent change (APC) (ie, the slope for each line segment) in the ASIR, the program uses least-squares regression, with the natural logarithm of the ASIR as the dependent variable and calendar year as the independent variable. Lower and upper 95% CLs, which define the 95% confidence interval, for each APC were obtained and tabulated; $P$ values of less than .05 are indicated by an asterisk.

Results

Distribution of Tumor Grade by Anatomic Site for All HPV-Associated OP SCCs in SEER

Tumor grade distributions for all 39,907 SCCs of anatomic sites of OP classified as “potentially HPV-associated” are shown in Table 1. Data for grade 4 (with only 497 or 1.2% of all 39,907) were combined with grade 3 as high-grade. Grade distribution was similar for base of tongue vs lingual tonsil ($P = .849$), with 37%–38% high grade. Grade distribution for base of tongue and lingual tonsil combined (BTT) differed little from that for tonsil and Waldeyer’s ring combined (with 42% high grade), although the association reached statistical significance ($P < .0001$) for the large numbers involved.

In comparison with BTT (n = 34,543), the grade distribution differed ($P < .001$) for the 4,504 cases at all other sites involving the OP, with 26%–30% high grade and 30%–52% unknown grade, and also for the 860 cases of overlapping lesion of tongue (C028). For cases coded to C028, only 20% were high grade and 12% unknown grade (Table 1).

For the selected sites classified as not HPV-associated, the 10,684 oral tongue (C020-023) SCCs included 2,635 (24.3%) grade 1 and 5,290 (48.7%) grade 2, but only 1,590 (14.9%) high grade and 1,318 (12.1%) unknown; the corresponding proportions for the 4,193 tongue NOS (C029) SCC were 22.7%, 43.7%, 14.9%, and 18.7% (data not tabulated). These distributions differed from those for the HPV-associated sites with the sole exception of overlapping lesion of the tongue (C028), as shown in Table 1.

Additional analyses focused on the 34,543 BTT SCCs, or the predominant HPV-associated sites of OP, which also have been most strongly associated with HPV.5,9

Figure 1. Age-Standardized Incidence rate (2000–2012) for Invasive Squamous Cell Carcinomas of Base of Tongue or Tonsils in the White Population of SEER (Surveillance, Epidemiology, and End Results) Registries, by Tumor Grade Category

A. White males

B. White females

Trends in Incidence Rates for BTT SCCs by Race and Ethnicity

Of the 34,543 BTT SCCs diagnosed in 2000–2012, race was coded as white for 29,900; after excluding 4 cases ascertained solely by autopsy or death certificate, 29,776 (99.6%) of all 29,896 were coded as microscopically confirmed. ASIRs showed statistically significant increases for both white males (APC, 3.9%) (n = 24,431) and white females (APC, 1.3%) (n = 5,465) for 2000–2012, although the increase was much smaller for females than males (Table 2, Figure 1).

For white males, the ASIR tended to level off after 2009 but fluctuated (Figure 1A) and there was no joinpoint (Table 2). For white males, the ASIR tended to level off after 2009 but fluctuated (Figure 1A) and there was no joinpoint (Table 2). Trends in ASIRs for 2000–2012 were not statistically significant in blacks/African Americans (n = 3,357), with negative APCs in both males and females, or in Asian–Pacific Islanders (n = 898), with low APCs; only 185 were coded as American Indian/Alaska Natives (n = 185) and 203 as unknown race (data not shown).

For cases coded as Hispanic (n = 2,084, of whom 2,008 or 96.4% were coded as white), the APC was positive for
Further analysis focused on all whites (ie, Hispanic and non-Hispanic combined), as done in previous surveillance reports. Trends were also examined separately for the subgroup of whites coded as non-Hispanic ethnicity; however, for Hispanic whites (1,632 males and 376 females), annual ASIRs fluctuated and annual numbers were too small for analyses of trends by tumor grade for each gender (data not shown).

### Trends in Incidence Rates for BTT SCC in Whites, by Tumor Grade Category

For all 29,896 BTT SCCs in all whites, the numbers of cases and the ASIR were similar for high grade vs grade 2 (moderate) in 2000 but had become larger for high grade in 2012, especially in males (Table 2). For all white males the ASIR increased greatly from 2000–2012, but the increase was smaller after 2005 (Table 2). The ASIR also increased from 2000–2012 for unknown grade in white males, especially in the early-mid 2000s, with smaller increases in recent years (Figure 1). The increase was smaller for grade 2 (vs higher grades) but the APC was statistically significant for white males; ASIRs for grade 1, however, were low and showed no clear temporal increase (Figure 1A, Table 2).

For all white females the ASIR increased from 2000–2012 for high grade, but with a smaller APC vs males; for unknown grade the increase was larger for earlier vs later years, whereas ASIRs did not increase for either grade 2 or 1 (Figure 1B, Table 2).

For the subgroup of whites coded as non-Hispanic, ASIRs were higher than for all whites, but trends (not shown) were similar to all whites. For example, for non-Hispanic white males the ASIR for high grade cancers increased from 1.89 (CL = 1.72, 2.07, n = 480) in 2000 to 3.12 (CL = 2.92, 3.31, n = 972) in 2012, with a joinpoint in 2006; APCs that were similar to those for all whites. For non-Hispanic white females the ASIR for high grade cancers increased from 0.44 (CL = 0.36, 0.52, n = 124) in 2000 to 0.53 (CL = 0.46, 0.62, n = 181) in 2012, with an APC of 1.9% (P = .019) and no joinpoint (not shown); this APC is the same as the APC as for all whites, shown in Table 2.

### Trend in Unknown Tumor Grade in All Whites in Relation to Cancer-Directed Surgery

The increase in ASIR for unknown tumor grade BTT SCCs in whites complicates surveillance and interpretation of trends in ASIRs for high vs lower tumor grade. Tumor

<table>
<thead>
<tr>
<th>Grade</th>
<th>2000</th>
<th>2012</th>
<th>2000–2012</th>
<th>APC (CL)</th>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Males</td>
<td>1,280</td>
<td>4.5 (4.2, 4.7)</td>
<td>2,505</td>
<td>6.7 (6.5, 7.0)</td>
<td>24,431</td>
</tr>
<tr>
<td>1</td>
<td>69</td>
<td>0.2 (0.2, 0.3)</td>
<td>71</td>
<td>0.2 (0.1, 0.3)</td>
<td>967</td>
</tr>
<tr>
<td>2</td>
<td>478</td>
<td>1.7 (1.5, 1.8)</td>
<td>751</td>
<td>2.0 (1.9, 2.2)</td>
<td>8,088</td>
</tr>
<tr>
<td>3 &amp; 4</td>
<td>501</td>
<td>1.7 (1.6, 1.9)</td>
<td>1,037</td>
<td>2.8 (2.6, 3.0)</td>
<td>10,020</td>
</tr>
<tr>
<td>9</td>
<td>232</td>
<td>0.8 (0.7, 0.9)</td>
<td>646</td>
<td>1.7 (1.6, 1.9)</td>
<td>5,356</td>
</tr>
</tbody>
</table>

| All Females | 353 | 1.1 (1,0, 1,2) | 493 | 1.2 (1,1, 1,4) | 5,465 | 1.2 (1,1, 1,2) | 1.3 (0.8, 1,8)* | 2000–2012 |
| 1 | 35 | 0.1 (0.1, 0,2) | 32 | 0.1 (0.1, 0,1) | 380 | 0.1 (0.1, 0,1) | 0.3 (-2.9, +3.8) | 2000–2012 |
| 2 | 134 | 0.4 (0.3, 0,5) | 155 | 0.4 (0.3, 0,5) | 1,947 | 0.4 (0.4, 0,4) | 0.0 (-0.8, +0,9) | 2000–2012 |
| 3 & 4 | 131 | 0.4 (0.3, 0,5) | 194 | 0.5 (0,4, 0,6) | 2,068 | 0.4 (0,4, 0,5) | 1.6 (0,3, 0,6)* | 2000–2012 |
| 9 | 53 | 0.2 (0,1, 0,2) | 112 | 0.3 (0,2, 0,3) | 1,070 | 0.2 (0,2, 0,2) | 9.7 (3,8, 16,0)* | 2000–2004* |

APC, annual percent change in ASIR, from joinpoint regression model (see text); CL, confidence limits, lower and upper, which define the 95% confidence interval; SEER, Surveillance, Epidemiology, and End Results.

*P < .05 for null hypothesis that the APC is zero.
grade was unknown for 1443 (12.0%) of 12,007 with cancer-directed surgery vs 4894 (27.8%) of 17,601 without surgery \((P < .0001)\) (Table 3); this analyses excluded cases coded as surgery recommended but not known if received \((n = 170)\) or as unknown \((n = 118)\).

Use of surgery declined by individual year of diagnosis, from 55%–56% in 2000–2002 (not shown) to 41.2% in 2003 (Table 3); the difference between the proportion in 2000 (56.1%) vs 2003 (4.12%) was statistically significant \((P < .0001)\), but there was little change after 2006 (Table 3) (ie, 34%–37% in 2007–2012). The temporal decline in use of surgery accounted for the increase in unknown grade. For cases with unknown grade that had cancer-directed surgery, the ASIR did not increase over time \((APC = –2.7\%, \text{CL} = –5.0, –0.4\%)\) (data not shown).

### Discussion

A relatively large proportion of all incident cases of BTTCC (ie, the sites most strongly associated with HPV) were coded in SEER as high tumor grade (Table 1). Also, large temporal increases were evident in ASIRs for high-grade BTT SCC in the white population (especially males) covered by the SEER Program (Table 2, Figure 1). In view of the association between high tumor grade and HPV positive status,\(^7\) these findings provide further support for the role of the HPV epidemic in the increasing incidence rates reported for OP SCC in US whites.\(^1-3\)

### Study Limitations

The main limitation of this study is that high tumor grade is only a predictor of (not a surrogate for) HPV-positive status. Also, surveillance and interpretation of trends by tumor grade category in SEER are complicated by the substantial proportion of BTT SCCs coded as unknown grade (Table 1), and ASIRs in whites increased over time for this subgroup. This trend reflected an association of unknown grade with lack of cancer-directed surgery and a temporal decline in the use of cancer-directed surgery (Table 3).

HPV-positive patients with head and neck cancers are more likely to be treated with combined chemotherapy and radiation than with either radiation or surgery alone, as shown for diagnoses in 2010–2013 in a US veteran population.\(^13\) In the National Cancer Database, although not a population-based resource, the use of surgery for OP SCC declined from 41% in 1998 to 30% in 2009 \((P < .001)\) with a slight increase to 35% in 2012, coinciding with the US Food and Drug Administration’s approval for transoral robotic surgery in 2009.\(^14\) The changes in treatment have resulted in less availability of biopsies and/or resections (of the primary site)\(^15,16\) which are required for tumor grading. For a SEER Patterns of Care study of a sample of OP SCC in SEER diagnosed in a single year (ie, 2004), treating physicians were contacted to obtain data on treatment; chemotherapy combined with radiation (but without surgery) was common (ie, 45%), with 37% receiving surgery.\(^16\) This 37% figure for surgery\(^16\) is similar to the 41% figure for diagnoses in 2003 in the present study (Table 3). This suggests that cancer-directed surgery was not under-ascertained in the SEER database\(^8\) used in this study.

The decline in surgery may reflect the advent of HPV testing. OP/SCC frequently presents clinically as an enlarged cervical lymph node,\(^6,16\) but prognosis has generally been reported to be better for HPV-positive (vs HPV-negative) OP SCC.\(^6,17\) Lymph nodes samples obtained by fine-needle aspiration can be used to detect HPV clinically, without a need for biopsy or resection of the potential primary site in some patients.\(^15,18\)

Registry-based studies conducted on HPV in OP tissue for samples of patients, including those with only lymph node tissue available,\(^3\) however, could assess bias in HPV status associated with unknown (vs known) grade. This could improve interpretation of trends in incidence rates by tumor grade in relation to the HPV epidemic.

### Table 3. Incident Invasive Squamous Cell Carcinomas at Anatomic Sites Classified as HPV-Associated\(^{1-4}\) in the White Population of SEER Program Registries, 2000–2012: Cancer-Directed Surgery\(^a\) by Unknown vs Known Tumor Grade and by Year of Diagnosis

<table>
<thead>
<tr>
<th>A. Tumor grade (unknown vs known)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery</td>
<td>Unknown</td>
<td>Known</td>
</tr>
<tr>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Yes</td>
<td>1,443</td>
<td>22.8</td>
</tr>
<tr>
<td>No</td>
<td>4,894</td>
<td>77.3</td>
</tr>
<tr>
<td>Total</td>
<td>6,337</td>
<td>100.0</td>
</tr>
</tbody>
</table>

| B. Year of diagnosis (selected years only) | | |
|---|---|---|---|---|---|
| n  | %  | n  | %  | n  | %  | n  | %  | n  | %  |
| Yes | 907 | 56.1 | 775 | 41.2 | 848 | 38.7 | 957 | 34.6 | 1,094 | 36.8 |
| No  | 710 | 43.9 | 1,104 | 58.8 | 1,343 | 61.3 | 1,811 | 65.4 | 1,880 | 63.2 |
| Total | 1,617 | 100.0 | 1,879 | 100.0 | 2,191 | 100.0 | 2,768 | 100.0 | 2,974 | 100.0 |

\(^a\)Excluded are a total of 288 cases with surgery recommended but not known if given, or with surgery entirely unknown (see text). For results of statistical tests, see text.
Again, high tumor grade OP SCC is not a surrogate for HPV positive status. Another issue for surveillance using tumor grade involves the substantial proportion of BTTSCC in whites that were coded as grade 2 (moderately differentiated), although the ASIR increased over time only for males (Table 2, Figure 1). Moderate grade was combined with low grade in a report on HPV status of OP SCC patients at a large US hospital and also in a report from a clinical trial. Some registry-based studies of HPV in OP tumor tissue for samples of patients have not reported HPV test results by tumor grade. In a sample of 529 OP SCC patients diagnosed in 1994–2005 in 6 registries in the CDC Cancer Registry Sentinel Surveillance System, HPV DNA was detected in 74.4% of 211 poorly differentiated cases, 60.0% of 35 well-differentiated cases, and 65.7% of 198 moderately differentiated cases; however, grade was unknown for 85 (16.1%) of all 529 cases. Therefore, additional studies appear to be needed on HPV status, especially for moderate grade and with smaller proportions of unknown grade.

Other study limitations include potential lack of representativeness of SEER data vs the entire US population; for example, the SEER population resembles the US population on certain socioeconomic status indicators, but the proportion foreign-born is larger. Clearly, studies are needed in non-SEER registries, including those covering Hispanic populations. However, long-term SEER incidence data (ie, since the 1970s) from 9 SEER registries by age and birth cohort have been used for estimating future incidence of OP SCC for the entire United States.

In the present study, ASIRs for BTT SCC in SEER could not be adjusted (statistically) for delayed reporting (of cases to registries) which can affect incidence rates for the most recent few years. Such adjustment is available only for selected cancer sites or site groups such as all oral cavity and pharynx sites combined, and not for base of tongue and tonsil separately. Adjustment has had little effect on trends (APCs) in rates for all oral cavity and pharynx sites combined.

Although SEER maintains quality-control standards for coding of site and tumor grade, the present study did not assess miscoding and misclassification of site and tumor grade. A small proportion of base of tongue cancers, for example, may be miscoded as oral tongue or unspecified tongue. However, the grade distributions for both overlapping lesion of tongue (C028) (Table 1) and tongue NOS (C029) (not tabulated), which could have included some base of tongue cancers, differed from the grade distribution for base of tongue (Table 1); this finding is not consistent with miscoding of site as a major issue.

The specific histopathologic characteristics, possibly including other factors (eg, nuclear features) in addition to amount of differentiation and keratinization, used by pathologists were not documented. Also, nonkeratinizing vs keratinizing typologies may be used for OP SCC instead of conventional grading. Good reliability (ie, reproducibility or agreement) for nonkeratinizing type has been reported among 6 different head and neck pathologists using 40 test cases of OP SCC randomly selected from one hospital in St. Louis, Missouri. Studies are needed that examine actual grading practices, and estimate intra-and inter-observer reliability (consistency) in grading, for samples of pathologists involved with OP SCC in SEER areas.

Ideally, data on actual HPV status or closely related markers should be available for surveillance of incidence of OP SCC. A SEER data item (not in the research database) on the results of any HPV testing of tumor tissue at the primary or metastatic site for selected OP cancers (including BTT) has been collected for diagnoses since 2010. However, in one SEER registry, data were incomplete for this SEER item, which has not been collected by NPCR registries; also, not all OP SCC patients are tested for HPV. Also potentially useful for surveillance is the p16 immunohistochemical tumor marker, which can be tested in most pathology laboratories and is widely used clinically for prognosis. Strong, diffuse p16 staining of most tumor cells predicts HPV positivity of OP SCC, and p16 results may be retrievable from some hospital tumor registries.

In conclusion, findings on trends in ASIRs by tumor grade in SEER registries provide some further support for the role of the HPV epidemic in explaining the increasing incidence rates for OP SCC in US whites, and suggest that tumor grade could be useful in routine surveillance. Studies are needed on completeness of tumor grade data in non-SEER registries. Also needed are studies on the reliability (consistency) in tumor grading within individual pathologists and among different pathologists. Studies on bias in HPV status for the considerable proportion coded as unknown grade in SEER (and possibly non-SEER) registries could use data on HPV in OP cancers from tissue banks and repositories in SEER and non-SEER registries.

References
Finding Your Dream Job is Just a Click Away

Visit NCRA’s Job Bank at www.ncra-usa.org/job!

**VIEW JOBS**
Access the newest jobs available to cancer registrars seeking employment.

**POST AN ANONYMOUS RESUME**
Whether you are actively seeking work or curious about what is available, your online resume is your ticket to great job offers!

**DESIGN A PERSONAL JOB ALERT**
Create job alerts and never let a matching job opportunity pass you by! New jobs that match your search criteria will be e-mailed directly to you.

**CREATE A FREE JOB SEEKER ACCOUNT**
Your personal job seeker account will allow you to find jobs, manage your resume, and set-up job alerts. Get started today!

NCRA’S JOB BANK www.ncra-usa.org/job
Abstract: The cancer registry profession has grown dramatically since its inception in 1926. Certified tumor registrars (CTRs) have become an integral part of the cancer care team by providing quality cancer data for research, statistical purposes, public health, and cancer control. In addition, CTRs have been found to be valuable in other cancer and health-related fields. Based on the need for high-quality, accurate data, the National Cancer Registrars Association (NCRA), the certification body for CTRs, has increased the educational requirement for eligibility for the CTR certification exam. This has resulted in fewer individuals who are able to meet the requirements for CTR certification. In addition, the existing cancer registry workforce is, on average, older than other allied health professions, and therefore will face an increasing number of retirements in the next few years. The high demand for CTRs, the decreased pool of CTR-eligible applicants, and the aging cancer registry workforce has resulted in an existing shortage that will only get worse as the population ages and the incidence of cancer increases. Health information management (HIM) students are well suited to pursuing further training in the cancer registry field and gaining the CTR credential. HIM students or new graduates have the needed skill set and education to pursue a cancer registry career. There are many avenues HIM educational programs can take to encourage students to pursue CTR certification and a cancer registry career. Including cancer registry functions in courses throughout the HIM curriculum, bringing in cancer registry speakers, encouraging networking, and promoting the cancer registry field and profession in general are just a few of the methods that HIM programs can use to raise awareness of and promote a cancer registry career to their students. Illinois State University has used these methods and has found them to be successful in encouraging a percentage of their graduates to pursue cancer registry careers.

Key words: cancer registry, cancer registry shortage, CTR eligibility, education, health information management

Introduction

The Cancer Registry profession was started with 1 registry in 1 hospital in New Haven, Connecticut in 1926.1 Since then, 1 hospital registry has grown to more than “1,500 hospital and 51 central cancer registries… employing over 7,500 specially trained cancer registrars and other staff who collect, manage, and analyze data on persons diagnosed with cancer.”2 The growth of the profession is due to many factors and events that have occurred over the years, including the 1956 requirement for cancer registries in American College of Surgeons’ approved cancer programs, the 1973 establishment of the Surveillance, Epidemiology, and End Results (SEER) program establishing the first national Cancer Registry, and the 1992 establishment of the National Program of Cancer Registries.3 Throughout this same period, the cancer registry job has become a profession with a professional association, the National Cancer Registrars Association (NCRA), and a professional credential, the Certified Tumor Registrar (CTR). Cancer registrars play an invaluable role in the cancer care process. “The wealth of data [collected by registrars] is critical to determining the location of cancer clusters as well as for the development and tracking of the most effective therapies, treatments, and cancer-control interventions. Public health officials also use the data to make decisions on research funding and educational and screening programs.”4 In spite of the valuable role cancer registrars play in cancer care, a definite shortage of CTRs has been noted in recent years; data shows that there will be an even greater shortage in coming years.

Background

The shortage of CTRs is due to many factors, including increased demand and a decreased pool of applicants. In 2005, it was noted, “Currently there are about 3,400 credentialed cancer registrars (CTRs)—a number that grows by only about 300 a year.”5 In the June 2006 NCRA Cancer Registry Workforce Study, it was noted that “it is difficult to know precisely the number of Cancer Registrars currently working in the field….we estimate about 7,280 registrars currently in the workforce.”6 The significant difference in these numbers is due in part to whether or not credentialed registrars or all registrars are considered in these studies. However, as of this year, the American College of Surgeons Commission on Cancer (ACoS CoC) now requires a CTR: “All abstraction of medical records for cancer cases at Commission on Cancer (CoC)-accredited facilities must be performed by cancer registrars who have achieved the Certified Tumor Registrar (CTR) credential.”7 In addition, “hospitals and other cancer care organizations trust the credential and the expertise it represents.”8 “Highly qualified trained cancer registrars...

Jennifer Peterson, MS, RHIA, CTR

---

Address correspondence to Jennifer Peterson, MS, RHIA, CTR, Department of Health Sciences, Illinois State University, Normal, IL 61761. Email: jpete2@ilstu.edu.

---

Footnotes:
8Illinois State University, Normal, Illinois.
are being sought after by regulatory agencies, accrediting organizations, researchers, pharmaceutical firms, software vendors and contract services providers. So, not only are CoC-accredited facilities seeking CTRs, but other organizations are, as well.

CTRs are the preferred professionals for the collection of cancer data due to their training, experience, and understanding of the cancer disease process and treatment. The CTR credential is recognized as the gold standard and is the ACoS CoC requirement for good reason. A study in Wisconsin found that CTRs in CoC- or non-CoC-accredited hospitals had average error rates of 2.5% to 4.6%, while “non CTRs from all types of facilities had...average error rate(s) of 26.2% per submission.” A study of registry staffing in the Northwest found that respondents recommended that registries “hire Certified Tumor Registrars (CTRs) [because] this protects and guarantees the accuracy of the data [and] this practice yields the best results for cancer patients.”

In recent years, both CoC-accredited facilities and other organizations have found great value in the high-quality data that CTRs collect. Cancer registry data has been used “to justify new investments in technology or medical expertise..., to grade ...[a] cancer center’s performance against benchmarks, to determine if patients receive appropriate treatment, ...to educate physicians on what standards of care should be..., [and] to analyze referral patterns or better understand...customer base.” In addition, as hospitals “move into pay for performance, the more data their executives will need to prove quality of care. Abstracted by trained professionals, the cancer registry is a reliable...source of cancer outcomes data.” State laws and other regulations are focusing more closely on the accurate, complete reporting of cancer data. States can penalize institutions that do not report their cancer data in a timely manner. Diane Cleveland, RHIA, CTR, stated, “It can impact Medicare reimbursement or there can be fines. One state told a hospital it couldn’t receive a certificate of need to construct a new building because the facility was behind in its reporting.” “Cancer diagnosis and care” reporting to public health agencies is also now required by the Meaningful Use requirements for hospitals and other providers. These examples make it clear that as more organizations find the need for CTRs, the demand will grow even higher.

In line with the need for highly-qualified, trained individuals who can produce high-quality, useful data, NCRA, the certification body for CTRs, has increased their requirements in recent years for those wishing to sit for the exam. Now, a CTR candidate must have, at minimum, an associate’s degree or equivalent (60 college-level credits) as well as additional training in the field. This requirement is essential due to the level of knowledge and skill needed to contend with the complex field of cancer care. It should be noted however, depending on the eligibility route taken, cancer registry courses are not required, and many never take such classes. These more rigorous requirements have reduced the number of prospective students willing to meet requirements to sit for the exam. In 2011, only 316 candidates took the exam. [Also, only] sixty-two percent of them passed the test; so just 196 credentialed CTRs joined the talent pool after passing the exam in 2011.” While the educational requirements have increased, the lack of cancer registry related courses taken by some CTR candidates could very easily contribute to the low passing rate. This creates a situation in which fewer CTRs are being certified since fewer candidates are sitting for the exam and some may not be adequately prepared academically.

In addition to the high demand for CTRs and a shrinking pool of new CTRs, current CTRs leaving the field and “the retirement of a great many current registrars is a looming issue.” “An average of 135 CTRs leave the profession annually, according to the NCRA.” A recent survey completed by NCRA and the University of California-San Francisco (UCSF) Center for the Health Professions found that “the average age of CTRs is 48, compared to an average age of 35 to 43 years for other allied health professionals. Over the next 5 to 10 years, at least one-third of the cancer registrar workforce is expected to retire.” This will create even more vacant positions to be filled by the already highly-in-demand CTRs.

These factors together have created a situation in which the supply of certified registrars is not meeting the current demand and given projections, seem unlikely to meet the demand in the future. In the coming years, the aging population will place even further demands on the pool of CTRs. The NCRA/UCSF survey found that “in 15 years (2021), the estimated need is projected for at least 800 new registrars (in addition to the pending retirement of older registrars) to meet the needs of a larger and older population with the expected increased incidence of cancer.” With this in mind for the future, it must be realized, however, that vacancies are already an issue for registries. In the NCRA/NPCR Workload and Time Management Study of Central Cancer Registries, it was noted that in 2009 “Central Cancer Registries reported an average vacancy of 1.5 full-time equivalent workers.” While cancer registry contractors used to be a viable alternative for hospitals who couldn’t find qualified CTRs, these companies may not be able to fulfill the hospitals’ needs on a timely basis as “many talented and credentialed consultants are booked solid for the next nine to 12 months, with no apparent signs of a change.” In addition, hiring remote workers from out of state can be problematic as one manager recently found when the hospital attorney found that “the hospital had been paying employee tax to our [sic] own state, [but] it had not paid the taxes due to the state in which each employee was a resident.”

With the additional demand and the growing shortage NCRA and others have been looking for ways to fill the gap. They have completed workforce studies, and have published results of workload and staffing studies for both hospital and central registries. NCRA and the American Health Information Management Association (AHIMA) have teamed up to provide an online training program that can help individuals meet the requirements to sit for the CTR exam. Community colleges and universities have been encouraged to develop certificate or associate degree online programs to provide the needed education and training.
Other initiatives to increase the number of registrars have come from the Workforce Study and include providing education for employers and HR directors as to the value of the certification, increasing recognition and awareness of the field, creating clinical and mentoring opportunities for students, and gaining inclusion in the “US Bureau of Labor Statistics 2006–2007 edition of the Occupational Outlook Handbook.”12

The current cancer registry shortage and the high number of expected CTR retirements opens the door for health information management (HIM) students. With the right education and training, HIM students could be excellently suited to fill cancer registry positions upon graduation or soon after. HIM students have training above and beyond most CTR candidates, including knowledge of other diseases, medical terminology, quality improvement, confidentiality and HIPAA (Health Insurance Portability and Accountability Act of 1996), statistics, and other areas that are not part of registry college classes. This actually makes HIM students superior to most CTR candidates and such knowledge could lead to higher pass rates among HIM CTR candidates. The key is providing HIM students with not only the education, but also the needed training to be eligible to sit for the CTR exam upon graduation or shortly thereafter. This provides an opportunity for HIM educational programs to encourage interest in and pursuit of such training.

Health Information Management Skills and Cancer Registry

Health Information Management students are extremely well suited to move into cancer registry and cancer information management positions based on the natural skills they possess as well as the knowledge they have gained in their HIM programs. HIM students are drawn to the profession due to their inherent organizational skills, interest in health information and data, and desire to make a difference in health care. Through their education, they are taught to be professionals “skilled in data collection, coding and classification of health data, data analysis and use, and management of information and records.”17 These are all skills needed in the cancer registry as well as in HIM: “professionals in both areas are the ‘keepers’ of health information data.”8 In fact, Hebert (2015) states that “health information management (HIM) professionals top the list [of] types of health care professionals …particularly well suited for cancer registry careers.”9 A degree in HIM helps prepare students “to pick up the highly specialized knowledge required for work as registrars.”28 “Their competency…helps them adapt more smoothly” to the cancer registry world.13 The medical background that HIM students have, including courses in medical terminology, anatomy and physiology, and pathophysiology help “new [cancer registry] employee(s) to learn so much faster as they move into the world of anatomy, topography, and morphology codes.”18

HIM students are also well versed in the new electronic health information and information technology (IT) skills now required of registrars. The tremendous growth in recent years of electronic health records as well as electronic health information exchanges and the use of multiple software programs has had a dramatic effect on the IT knowledge and computer skills required by registrars. “Expert usage of computers has also become one of the necessary core competencies of cancer registrars.”19 Joyce Ritter, RHIA, CTR, manager at the Goldston Cancer Registry in Amarillo, Texas, states that, “if [cancer registry workers] are not computer savvy, they will need to be.”13 In addition, Ann Wiedemann, MS, RHIA, FAHIMA, CPEHR, director of professional practice resources at AHIMA, states that “technology will play a major role in the future of cancer registry.”13 Many registrars are now “bridging registry data with electronic health records”,20 using electronic health records to assist in the Rapid Quality Reporting System, and using electronic interchanges to report cancer data to state registries in real time. Accredited HIM educational programs are required by their accreditation agency to teach their students extensive information technology skills related to electronic health records, electronic data, health information exchange, computer systems, networks, and software. The information technology skills that HIM students are gaining in their educational programs are a perfect fit for the computer and IT skills now required of registrars.

It is clear that HIM new graduates and professionals have the skill set needed to become CTRs. These new graduates and young professionals could do much to help fill the current void in the cancer registry world. The question then becomes how can health information management educational programs encourage their students to pursue eligibility for the CTR and cancer registry careers? There are many ways these programs can do just that.

How HIM Educational Programs Can Help Fill the Void

Since the educational requirement has been increased to at least an associate’s degree or equivalent for eligibility for the CTR exam, students in HIM programs are ideally placed to add CTR preparation to their college studies or to pursue such preparation shortly after graduation. The key though, is in introducing and encouraging an interest in the cancer registry field while these students are early in their HIM programs so that they can investigate the field further or possibly complete a professional practice in it. There are a wide variety of methods that HIM programs can use to introduce the cancer registry field. While it would be wonderful if more colleges and universities could provide formal education in cancer registry management, this may not be a feasible option for many. There are, however, a number of ways in which HIM programs can encourage and support student interest in and pursuit of the CTR credential and a cancer registry career.

First, the cancer registry and cancer program management should be introduced in an early HIM class. This can be done through a discussion of various career paths that graduates might be interested in, lab exercises introducing cancer data abstracting, class sessions devoted to the function of the cancer registry and the cancer
program, and introduction to the CTR credential and to the National Cancer Registrars Association. At Illinois State University, we integrate all of these activities into the Data Reporting and Use section of the HIM introductory course. Students are introduced to the cancer registry and cancer program activities in the classroom setting. Also in the classroom students are able to review and evaluate Cancer Program Annual Reports as well as learn more about the CTR credential, paths to obtaining the credential, jobs in the cancer registry field, and the National Cancer Registrars Association. In the lab portion of the course, they are introduced to International Classification of Diseases for Oncology (ICD-O) coding, American Joint Committee on Cancer staging, and abstracting cancer data into the AbstractPlus system. Introducing these concepts and skills early on provides interested students the ability to pursue their interest in the cancer registry field throughout their remaining courses and time at Illinois State University.

In addition to introducing the cancer registry topics early on, HIM programs can also encourage student interest in the field by bringing in cancer registry professionals as speakers in classes or for student associations. The Student Health Information Management Association at Illinois State has at least 1 cancer registry professional speaker each school year. These speakers, both from hospital and central registries, discuss their jobs and the profession with the students which gives students the opportunity to find out more about the profession, ask questions, and network. Interested students may even work with the professionals to procure professional practice or volunteer experiences.

Beyond these experiences, integrating cancer registry examples and case studies into other HIM courses can further stimulate student interest in the field. Utilizing Cancer Registry Management: Principles and Practice and other NCRA publications as required or supplemental reading can further students’ understanding of and interest in the cancer registry field. In addition, utilizing case studies and activities related to the cancer registry or cancer programs enables students to become more aware of and knowledgeable about the profession. At Illinois State, cancer registry case studies are integrated into the Trends in Healthcare course as well as the Organization and Management of Healthcare Services course. A multiple-assignment exercise in the Management course involves the complete hiring process for a cancer registry abstractor. This includes writing a position description, recruiting, interviewing, orientation and training, and performance evaluation for the fictional abstractor. These assignments require the use of additional materials and resources on the cancer registry field so that students gain the requisite knowledge to prepare these materials in a meaningful and accurate manner.

HIM programs can also assist in general promotion of the cancer registry field and profession. NCRA brochures can be made available to students and the public, posters can be displayed in prominent areas, and presentations can be made open to the student body at large or to the public. These are all methods of increasing both HIM student and public knowledge of the cancer registry and cancer programs.

Finally, interested students can complete the AHIMA/NCRA cancer registry management sequence of courses or another online cancer registry management program while a student in the HIM program. These programs provide cancer registry education that can aid in meeting eligibility for the CTR exam for these students. Students who complete these programs during their college career could be placed in cancer programs for their professional practices, thus meeting the experience requirement so that they would be eligible for the Registered Health Information Administrator (RHIA) exam and the CTR exam upon graduation. If that is not a feasible option for the student, certificate programs are available at a number of colleges. These could be completed following graduation, providing the new graduate with the eligibility for the exam.

While there have been discussions regarding a cancer registry track within the HIM program at Illinois State, this is not a financially feasible option at this time. However, through the efforts listed above, Illinois State has had 3% to 5% of graduating students each year go into a cancer registry position either directly upon graduation or shortly thereafter. Encouraging student interest in this field has enabled students to focus on this field through professional practice and classroom activities, thus preparing them for their future careers.

**Conclusion**

The cancer registry field has grown dramatically over the past 90 plus years; however, in recent years, a distinct shortage of trained, qualified registrars has developed. Increasing demand, a decreasing pool of eligible candidates, and a high percentage of pending retirements will result in an even more dramatic shortage of registrars in the next few years. NCRA and others have been working to find ways to fill the gap that is developing. HIM programs are ideally placed to aid in this effort through encouraging students and new graduates to pursue the CTR credential and a cancer registry career.

HIM students are well suited for the cancer registry field. They have the appropriate skill set and the education required to organize and manage a cancer registry or a cancer program. Providing student exposure to the field early on in their HIM programs as well as in an ongoing manner enables interested students to gain even further knowledge and experience. This translates into new graduates seeking jobs in the field and new experienced employees to fill the cancer registrar void. A little effort on the part of HIM programs can go a long way to help ensure that cancer data is collected and analyzed by educated, experienced registrars, resulting in high quality data that is useful for the many entities depending on it.
References


8. Health Information Management and Cancer Registrars: One Road Leads to Another.


How I Do It

How to Build an Abstract

Vicki Hawhee, MEd, CTR

Creating an abstract from the electronic medical record (EMR) can seem overwhelming at first. There is so much information, it is difficult to know where to start and how to work your way through all the documents in a short period of time. This article provides one roadmap to build an abstract that can provide a template for a new abstractor or an alternative way of doing things for an experienced abstractor.

**Key words:** abstracting, beginner, new CTR, roadmap

**Introduction**

You completed your classes and your practical experience, you sat for the Certified Tumor Registrar (CTR) credential, and passed the exam. You have all your reference books and sit down to abstract your first case, all on your own. Now what? Sometimes, the patient’s medical record can be overwhelming with hundreds of documents: where do you begin, and how do you get through it all and accurately record all the pertinent data in the shortest amount of time? There is no single correct way to complete an abstract, and most abstractors have developed their own method that works best for them. This article offers instruction on one way to abstract a case.

**References and Resources**

To get started, you need to have hard copies of your American Joint Committee on Cancer (AJCC) *Cancer Staging Manual* (currently in its seventh edition) and your *International Classification of Diseases for Oncology, 3rd Edition* (ICD-O-3). Most of the other references you need to use can be printed to use as a hard copy, or kept as an electronic copy on the computer. Keeping them as an electronic copy can make finding the information faster (by using the find or search function). The fall 2015 issue of the *Journal of Registry Management* includes a comprehensive list of resources; you can use this list to bookmark all of the references you may need to use or your facility may have provided you with a similar list. These references should be added to your Web browser favorites to expedite finding information when you need it. Abstractors refer to these manuals with every case; this is not an area where you should try to memorize what you need. Even long-time abstractors still consult these manuals daily, so make sure you do the same.

**Where to Start?**

Some facilities employ CTRs that specifically work in casefinding, and those cases end up on each abstractor’s list of cases needing abstraction. Your facility will determine which cases they need you to abstract first. In general, a best practice is to work on those cases included in the Rapid Quality Reporting System (breast and colorectal) first and then by date of first contact, oldest to newest. Open the case in the EMR and open a new Microsoft Word document.

To begin, read the clinic note from the patient’s first visit to your facility. This generally provides a good sense of why the patient came to your facility, along with any pertinent history or other information you may need to look for. This is also a great time to check your EMR in chronological order and pin down the date of first contact with your facility (which may be the date of a scan or biopsy, and not necessarily the first clinic visit).

Then, begin entering information in the Word document. Type the date of this first visit, the name of your facility (which may be the date of a scan or biopsy, and not necessarily the first clinic visit).

Next, look at the outside documents received. Check any outside pathology reports, treatment, letters, etc, and enter all this information into the Word document in chronological order, again using correct text requirements. Many facilities prefer the use of all capital letters in the text. For example, 01/15/2016 – SOUTH HOSPITAL – CT ABD/PELVIS SHOWED A 3.2 CM RUL MASS, NO ADENOPATHY, NO EVIDENCE OF METS. Do not include ADENOPATHY, NO EVIDENCE OF METS. Do not include information that is not pertinent to your case. Any entries that happened before the patient came to your facility are placed before the date of first contact with your facility in your document. It is important to keep your Word document in chronological order.

**Primary Site and Histology**

At this point, try to pin down the primary site and histology, because this makes a difference in the information you need to search for. There is no need to look for tumor markers in the laboratory data if they are not required for this primary site. Additionally, if you have more than one tumor or a history of a prior malignancy, you need to check...
the multiple primary rules at this point to confirm that this is a new primary/single primary.

Once you have determined the working primary site and histology, check the correct chapter in the AJCC to determine what is necessary for clinical and pathologic staging. Are nodes required to be removed to pathologically stage this cancer? Do you need imaging to confirm a clinical N group or is a physical exam enough? Does the entire primary site need to be removed in order to pathologically stage this primary?

The Last Visit

Next, read the last visit at your facility to find out what kind of treatment the patient had, so you know what to look for. Then begin to check through all the documents, such as the operative reports, radiation oncology visits, chemotherapy infusion, etc. By reading the last note, you also get clues about whether you need to check the medication administration record (MAR) to see if the patient was prescribed any oral medications to treat the cancer. It is also important to check the National Comprehensive Cancer Network guidelines to find out what the standard treatment would be for this stage of cancer. If you see that radiation is generally required for this stage/histology cancer, and you do not see notes about it at your facility, then you may need to send letters to some outside references or make some calls to see if the patient indeed had this standard treatment. It is very important to capture all of the first course of treatment for each cancer, not just the treatment provided by your facility.

Entering the Data

Once the Word document is complete and all the notes are in chronological order, then open the registry software and enter the data. Just start at the top of the abstract and work your way down. This should go quickly, since you have all your notes together. When you need to determine the date of diagnosis, the chronological order of the document makes this much easier. Which came first, the scan with the ambiguous term or the biopsy? Enter each code as you work your way through the abstract, referring to the notes, and then copy the entry you need from the Word document and place it in the correct text field on the abstract.

It is important to note as you go through the abstract that everything you have coded in your abstract needs to be supported in your text. For instance, if you are basing the T value on the size of the tumor and extension, then somewhere in the text the size and extension of the tumor need to be documented. If the site-specific values include a laboratory value, make sure that is documented in the correct text area of the abstract. If you are unable to code any value based on the text in your Word document, then you need to go back and look for this information and add it to the text. By determining your codes from the text you have in the Word document, you are ensuring your text will support your codes.

When you have completed entering all the data in your abstract, copy the entire Word document and add it to your abstract in an area that allows for a large amount of text. The registry software utilized by my facility has a “Comment Text” in the “Follow-Up Notes” area that allows for 32,767 characters. This field is not transmitted when submitting the abstract to the state/National Cancer Data Base, but by placing the entire Word document there, all the notes for the case are available in one place and can also easily be utilized by the tumor board or cancer conference.

Summary

Abstracting can be an overwhelming task to complete as a new CTR. There is so much information available in the EMR, so how can you possibly get through it all and accurately record all the correct information? Creating a document in chronological order that includes all things pertinent to this case is an excellent way to do this. This provides many advantages:

1. You can see at a glance the information that needs to be included in the abstract.
2. You are easily able to determine the sequence of events that happened in this patient’s cancer history in order to accurately record all the information.
3. You can double check to make sure your text includes all information necessary to support the codes assigned in this case.

I also recommend a visit to National Cancer Registrars Association (NCRA)’s Center for Cancer Registry Education. This site offers a wealth of information, including informational abstracts for many primary sites written by the NCRA Education Committee. These can be found here: http://www.cancerregistryeducation.org/rr.

Vicki Hawhee, MEd, CTR is the QC/Education Specialist in the Cancer Registry at Moffitt Cancer Center and Research Institute in Tampa, Florida.
Raising the Bar

The Privilege of Being Invited In

Michele Webb, CTR

As I rounded the corner to my office I saw her. A frail-looking elderly woman dressed in a pink jogging suit, sitting on a white bench with a peanut butter and banana sandwich and can of diet cola beside her. She was hanging on to her walker with both hands and coughing (or choking; it was hard to tell which). I quickly went to her and asked if she was okay. After a few more deep coughs she pushed herself up from her walker and looked at me with red, watery eyes and tears running down her cheeks.

“No, dearie,” she said, coughing and clearing her throat again. “Thank you for asking, but I am okay.” She pointed to her sandwich and said, “It’s my favorite, but it is very dry. I knew I should have brought some water with me instead of that dad-burned soda.”

Chuckling at her sense of humor despite violent coughing, I stepped inside my office and grabbed a small bottle of water. She gratefully took it and began taking small sips. Despite her protests, I stayed with her until she was breathing easier.

What happened next was completely unexpected. As I sat there, she began to ask me questions, starting with who I was, what I did, and why. She extended a hand to me and introduced herself, and I to her. She seemed fascinated with the work of a cancer registrar and told me her husband of 56 years had died from lung cancer. She spoke of his many years of cigarette and cigar smoking and how, despite her pleading with him to stop, he was so addicted that he just could not. She told me how they had moved from their residential home to an assisted-living center when his disease progressed and she could no longer care for him. With tears in her eyes, she talked about his last days and how, after over five decades of marriage, he still made her heart go “pitter pat” when he looked at her. Her story was so simple and so sweet.

“So, Lillian,” I asked. “Why are you out here and eating lunch by yourself? Do you have anyone with you?”

“Oh, no,” she said, shaking her head. “I may look old and feeble with this thing,” she said patting her walker. “But, I am actually quite mobile. The bus from the home had to drop me off early today and I’ll get a ride home later this afternoon. I have to see my cancer doctor first and then they are going to take some pictures of my chest and head,” she said pointing to her chest. “I am going to have surgery next week.”

She surprised me a bit with the news of her surgery but I was amazed at how strong and positive she seemed about it all. She even made jokes about whether they would find anything “between her ears” when she had the CT of her brain. I told her I would be thinking of her and hoping for the best possible outcome.

She picked up her sandwich and, before taking another bite, said, “Well, you’ve been so kind and I don’t want to take you away from your work. I’ll be okay and back on my feet in no time,” she beamed.

I did think of Lillian a number of times that following week and wondered how her surgery had gone. She was quite spunky and there was something about her that made a strong impression.

Several weeks later as I again turned the corner to my office, there sat Lillian! She was in a wheelchair this time but dressed in a snappy, red knit pantsuit, hair beautifully combed and bright red lipstick to boot. As she saw me, her eyes lit up and she clapped her hands together.

“Well, there she is,” she said looking at her female companion. “I told you about this nice gal who works here in the Cancer Center. She’s the one who brought me water when I was choking.”

She grabbed my hand and gave it a firm squeeze as she introduced me to her daughter. We talked for a few minutes and Lillian explained that she had undergone surgery (thankfully with no complications) and been diagnosed with early-stage lung cancer.

“No lymph nodes or anything like that,” she proclaimed. “And no spots in my head or anywhere else. I’m going to outlive them all!”

I was so happy for Lillian and told her that I had been concerned for her and was pleased to have seen her again.

“You are so kind,” she gushed. And this is where she really got me! “You just go tell folks my story. Tell them to stop smoking those nasty cigarettes. My husband’s smoking almost got me,” she said. “You tell them to stop before it’s too late!”

I waved good-bye to Lillian and her daughter and felt so blessed to have met her and gotten to know her in a small way.

And now, you’ve guessed it: I am telling her story. It’s not earth-shattering or emotional. Just a short-and-sweet message straight from this dear woman’s heart. It’s almost like a page out of American history.

As cancer registrars, we have the privilege of connecting with patients, health care providers, and others every day. When, during the course of that connection, one or more of
us feels that they have been seen, heard, or valued, when we can give and receive without judgement, and when we get some sort of happiness or strength from that encounter, it is a privilege.

We have the opportunity every day to let, and be let, into someone’s life and to tell their story. As we comb through the detailed medical record information, we are really putting together the pieces of the puzzle that make up their life. It is rich with facts about their personal and work life, medical history, family history, risk factors, and how cancer has crept into their life to rob them of their health and wellness. Sometimes we are afraid to tell the whole story, afraid that it is too much detail, or not relevant, or that it is sad or emotionally compelling. But withholding that information, or not telling the story, is neither healthy nor good.

Through storytelling, we also share a little of ourselves and why we do this work. Sharing our dreams, vulnerabilities, and life experiences can be frightening, or empowering. As this author sees it, if the cancer patient can derive strength and hope from their experience, then we as cancer registrars can certainly do the same, if not more!

Our goal in being a storyteller is to affect the world in a profound way, because this life should be lived in greater health and happiness than what those who have gone before us experienced. Our mission as cancer registrars should be to model that purpose and to help others to do the same.

What an incredible privilege we have to be held in sacred trust with the intimate details and knowledge of someone’s life. And, if we are ever so lucky, we will see into someone’s story and realize that their information can be used to lift the dark cloud and devastating hold of cancer for generations to come. With their information we contribute to a lasting hope for a better life, new cancer treatments, improved outcomes, survivorship and patient care experiences.

As one of your peers, I encourage you to look for that invitation into someone’s life. Enjoy the privilege of getting to know them, beyond the medical facts, to love on them and conduct your work in such a way that you tell their story accurately and eloquently. Appreciate and value the privilege of being invited in and honor their life with your best work!

Journal of Registry Management Continuing Education Quiz—SPRING 2016

LINKING THE NATIONAL HEALTH INTERVIEW SURVEY WITH THE FLORIDA CANCER DATA SYSTEM: A PILOT STUDY

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:
- Describe the advantages of database linkage
- Identify uses of the enriched data set resulting from linked data
- Explain the challenges associated with database linkage

1. Root causes of health disparities along the cancer continuum:

<table>
<thead>
<tr>
<th>progression and consequences of cancer</th>
<th>the number of cancer data items collected</th>
<th>cancer development</th>
<th>stage at presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Y Y Y N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Y Y Y N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Y N Y Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) N Y Y Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. According to the authors, missing data was an insignificant challenge to the accuracy of this linkage since gold certification of the FCDS by North American Association of Central Cancer Registries (NAACCR) ensures that no more than:
   - a) 2% are missing data on age
   - b) 3% are missing information on sex
   - c) 4% are missing data on county
   - d) 5% are missing information on race

2. Goals of linking the Florida Cancer Data System (FCDS) with the National Center for Health Statistics (NCHS) include:

<table>
<thead>
<tr>
<th>determining the feasibility of linking databases</th>
<th>establishing the administrative burden</th>
<th>assessing the utility of the linked data</th>
<th>limiting the number of data items collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Y Y Y N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Y Y Y N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Y N Y Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) N Y Y Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. According to Table 3, Number of Cancer Cases (Florida Residents Only) in the 1986–2009 National Health Interview Survey (NHIS) and 1981–2010 Florida Cancer Data System Linkage, Sociodemographic and Cancer Diagnosis Factors by State of Survey Participation, most linked cancer cases occurred in:

<table>
<thead>
<tr>
<th>patients aged 65 years and older</th>
<th>patients with less than a high school education</th>
<th>non-Hispanic white patients</th>
<th>patients who rated their health as fair/poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. The National Health Interview Survey (NHIS):
   - a) takes place every 10 years
   - b) surveys institutionalized individuals
   - c) is representative of the international population
   - d) collects information on health insurance

8. Linking data from the NHIS and FCDS provides a resource for evaluating which variable not typically available in cancer registries?
   - a) Health insurance coverage
   - b) Marital status
   - c) Self-rated health
   - d) Occupation and industry

4. NHIS participants included in this linkage were those who:

<table>
<thead>
<tr>
<th>shared their Social Security number directly</th>
<th>shared their Social Security number via proxy</th>
<th>had personally identifiable information available</th>
<th>provided consent to record linkages</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Y Y Y N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Y Y Y N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Y N Y Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) N Y Y Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. According to Table 4, Number of Linked Cancer Cases (Florida Residents Only) in the 1986–2009 National Health Interview Survey (NHIS) and 1981–2010 Florida Cancer Data System Linkage, and Estimated Number of Cancer Cases Expected if a Linkage was Conducted with Cancer Registries Nationwide, which of the following statements is true regarding the estimated number of cancer cases from a potential nationwide linkage?
   - a) Incidence is higher than prevalence for all sites
   - b) Prevalence is higher than incidence for all sites
   - c) The proportion of Florida cancer cases in the United States is 11%
   - d) The proportion of Florida cancer cases in the United States is 21%

10. According to the authors, cancer registry linkage to other data sources: is unavailable due to cost and effort required
    provides data sets for hypothesis-driven research
    enriches the value of national health surveillance systems
    should immediately be expanded to all states

<table>
<thead>
<tr>
<th>is unavailable due to cost and effort required</th>
<th>provides data sets for hypothesis-driven research</th>
<th>enriches the value of national health surveillance systems</th>
<th>should immediately be expanded to all states</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Y Y N N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) N N N N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Y N N N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) N Y N N</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.CancerRegistryEducation.org/jrm-quizzes. Download the article, complete the quiz and claim CE credit all online.
Instructions: Mark your answers clearly by filling in the correct answer, like this ■ not like this X. Passing score of 70% entitles one (1) CE clock hour per quiz. Please use black ballpoint pen.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

For Internal Use Only
Date Received: ____________
Amount Received: ____________
Notification Mailed: ____________
TRAUMA REGISTRIES
CALL FOR PAPERS

Journal of Registry Management

Journal of Registry Management, the official journal of the National Cancer Registrars Association, announces a call for original manuscripts for a Fall 2016 Special Focus on TRAUMA REGISTRIES. Invited papers should cover a broad range of topics related to trauma registry management, including the collection, quality review, reporting, and use of trauma registry data. We encourage authors to report on the special challenges associated with the collection and management of data, and the development and operation of trauma registries, as well as the benefits to patients, care institutions, and research that accrue through the operation of hospital-based trauma registries. We invite practitioners, researchers, registrars, and interested others to submit manuscripts on these topics and on results of original research studies using registry data.

Manuscripts for this issue will be accepted through June 1, 2016, and should be submitted to Guest Editor Michelle Pophrey, MLT, RN, CSTR, President, Pophrey Consulting, 2683 Springhill Road, Staunton, VA 24401. Telephone: (540) 448.2770. Please email articles to Michelle.Pophrey@PophreyConsulting.com and cc: JRMEditor@NCRA-USF.org with the subject line: JRM Trauma Registry Article.

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. All papers will be subject to peer review procedures.
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 Venetta L. Williams, PhD, MPH, CTR, Editor-in-Chief at JREditor@ncra-usa.org or (813)745-1783.

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. Email 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 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, email address, and telephone number of the corresponding author. For electronic submission, files should be in 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 (excluding 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.


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 emailed. 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 through 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:


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.