Visit NCRA’s Center for Cancer Registry Education!
www.CancerRegistryEducation.org

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

Use your NCRA Website member log-in and password!

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

The CCRE Website is partially funded by the Centers for Disease Control and Prevention, Cooperative Agreement U50CFO03842.
Contents

Original Articles

Enhancing Central Cancer Registry Treatment Data Using Physician Medical Claims: A Florida Pilot Project
Monique N. Hernandez, PhD; Jill A. MacKinnon, PhD, CTR; Lynne Penberthy, MD, MPH; Judy Bonner, RN, MS, CTR; Youjie X. Huang, MD, MPH, DrPH

The Value of Billing Data from Oncology Practice to Supplement Treatment Information for Cancer Surveillance
Lynne Penberthy, MD, MPH; Valentina Petkov, MD, MPH; Donna McClish, PhD; Steven Peace, BS, CTR; Sandra Overton, BS, CTR; Soundarya Radhakrishnan, MS; Chris Gillam, BS

Reda J. Wilson, MPH, CTR, RHIT; M.E. O’Neil, MPH; E. Ntekop, MD, MPH; Kevin Zhang, PhD; Y. Ren, PhD

Relative Survival Analysis Using the Centers for Disease Control and Prevention’s National Program of Cancer Registries Surveillance System Data, 2000–2007
Reda J Wilson, MPH, CTR, RHIT; A. Blythe Ryerson, PhD, MPH; Kevin Zhang, PhD; Xing Dong

Recent Improvement in Completeness of Incidence Data on Acute Myeloid Leukemia in US Cancer Registries
Anthony P. Polednak, PhD

K. Masood, T. Zafar, J. Zafar

Features and Other Journal Departments:

Raising the Bar: How to Strut Your Cancer Registry Stuff
Michele Webb, CTR

Edits Impact Statement: North American Association of Central Cancer Registries
Edits Impact Workgroup

Summer 2014 Continuing Education Quiz
Deborah C. Roberson, MSM, CTR; Denise Harrison, BS, CTR

Call for Papers

Information for Authors

© 2014 National Cancer Registrars Association
Editors
Vicki Nelson, MPH, RHIT, CTR, Editor-in-Chief
Vonetta Williams, MPH, CTR, Associate Editor
Reda J. Wilson, MPH, RHIT, CTR, Editor Emeritus
Ginger Yarger, Copy Editor
Janice Ford, Production Editor

Contributing Editors
Faith G. Davis, PhD—Epidemiology
April Fritz, BA, RHIT, CTR—Cancer Registry
Denise Harrison, BS, CTR—CE Credits
Deborah C. Roberson, MSM, CTR—CE Credits
Michele A. Webb, CTR

NCRA 2014-2015 Board of Directors
President: Terri M. Richardson, RHIA, CTR
President Elect/Secretary: Leah Kiesow, MBA, CTR
Immediate Past President: Shirley Jordan Soay, PhD, OCN, CTR
Treasurer Senior: Janet Reynolds, CTR
Treasurer Junior: Kathleen Hess, RHIT, CTR
Educational Director: Paulette Zinkann, CTR
Professional Development Director: Sharmen Dye, CTR
Public Relations Director: Kyle L. Ziegler, CTR
Recruitment & Retention Director: Linda Corrigan, MHE, RHIT, CTR
ATP Director—East: Pamela Moats, RHIT, CTR
ATP Director—Midwest: Mindy Young, CTR
ATP Director—West: Kendra Hayes, RHIA, CTR

2014-2015 JRM Editorial Advisory Board
Tim Aldrich, PhD
Tammy Berryhill, MA, RHIA
Betty Gentry, BS, CTR
Robert German, PhD
Donna Gress, BS, CTR
Alison Hennig, BS, CTR
Monique Hernandez, PhD
Amy Kahn, MS, CTR
Leah Kiesow, MBA, CTR
Herman Menck, MBA, CDP
Kathleen Thoburn, CTR
Kevin Ward, PhD, MPH, CTR
Pamela Warren, PhD, CTR
Reda Wilson, MPH, RHIT, CTR (Editor Emeritus)
Vonetta Williams, MPH, CTR (Associate Editor)

Production and Printing
The Goetz Printing Company

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

For your convenience, the Journal of Registry Management is indexed in the 4th issue of each year and on the Web (under “Resources” at http://www.ncra-usa.org/jrm). The 4th issue indexes all articles for that particular year. The Web index is a cumulative index of all JRM articles ever published.
Enhancing Central Cancer Registry Treatment Data Using Physician Medical Claims: A Florida Pilot Project

Monique N. Hernandez, PhD; Jill A. MacKinnon, PhD, CTR; Lynne Penberthy, MD, MPH; Judy Bonner, RN, MS, CTR; Youjie X. Huang, MD, MPH, DrPH

Abstract: Background: To capture the complete first course of therapy and cancer incidence, given the shift in cancer care from the hospital to the private physician practice, central cancer registries (CCRs) in the United States are actively pursuing cancer reporting from ambulatory providers. The 837 medical health claim is a national standard which CCRs can use to capture and translate data into standardized cancer reporting for surveillance. Methods: The Florida Cancer Data System conducted a pilot project with a large medical oncology practice to transmit electronic claims from 2011 to 2013. Using the logic and platform developed under a previous National Cancer Institute (NCI) contract, claims were consolidated and translated into standardized cancer registry codes. Consolidated physician claims were compared against gold standard data from the practice electronic health record (EHR) and evaluated for enhancement to registry data. Results: A total of 623 patient tumor cases were collected from the practice EHR and matched to the physician claims data, and to the original cancer registry record. The claims captured 256 cases (41%) with chemotherapy, compared to 28% in the registry data set, and 45% in the gold standard EHR data set. Combining physician claims with registry data produced 280 cases (45%) with chemotherapy. The physician claims plus the registry cancer chemotherapy treatment data produced 92% agreement, 92% sensitivity, and 91% positive predictive value. Claims added 103 cases, or 16.5%, to the total chemotherapy received. Conclusions: Physician medical claims data capture chemotherapy information not otherwise reported by hospitals, and is a standardized and efficient mechanism for cancer reporting.

Key words: cancer surveillance, data linkage, medical oncology, physician medical claims, treatment enhancement

Introduction

For decades, central cancer registries (CCRs) have been legislatively charged with the full capture of cancer incidence and first course of cancer-directed treatment for public health surveillance purposes. Increasingly, these data are being utilized for outcomes research, which relies on complete and detailed treatment data. While CCRs typically receive cancer reports from hospitals, radiation centers, and surgery centers, a shift in the diagnosis and treatment of some cancers to the private physician’s office is compelling departments of health to pursue active reporting from private physicians more than any time previously.

Given the growing requirements placed on physician practices through national incentive programs that encourage meaningful use of certified electronic health technology, the challenges in expecting physicians to report to CCRs, albeit mandated by law, grows exponentially. This is where leveraging existing mechanisms that capture cancer diagnosis and treatment procedures already utilized by the private physician office, such as electronic medical claims data, can improve cancer surveillance without burdening physicians.

Medical claims, typically from hospital discharge records and Medicare, have been used by CCRs for case ascertainment and treatment data enhancement. However, these sources rely predominantly on data from hospitals that primarily serve Medicare patients aged 65 years and older. Nevertheless, studies continue to demonstrate the robustness of claims data, with a high accuracy for identifying unreported cancer cases as well as supplementing profiles for surgical, radiation, and chemotherapy treatment data. As the era of patient-centered outcomes research looms large, the linkage of cancer reports with administrative health data becomes increasingly important, and provides data that are more generalizable than clinical trials representing less than 4% of the cancer population, with patients who are healthier and younger than cancer patients in general. Additionally, these data capture information on cancer diagnoses and treatments from all major medical facilities.

Most recently, claims have been used successfully for case ascertainment from private physicians’ offices, particularly for hematologic cancers, which are more likely to be diagnosed outside a hospital setting. The utility of claims is also broadly demonstrated through its use in capturing comorbid conditions, evaluating disease-free cancer survival, and assessing the use of services. For the CCR, it has been clearly demonstrated that the use of claims contributes to a more complete population cancer data set and also enables longitudinal tracking of conditions providing records at each patient encounter.
The benefits of linking claims data with public health data are significant. The value relates to the highly standardized and translatable coding systems and the universal electronic transmission of medical claims in a standardized format. Integral to sustainable use of medical claims as a consistent source of public health data is the ability to automate and translate data from the standard 837 claim for professionals, or the standardized electronic transaction format, into the standard coding values and layouts established for national cancer reporting.

Background

In 2010, the Florida Department of Health’s Bureau of Epidemiology was awarded funding from the Centers for Disease Control and Prevention (CDC) for enhancing specialized cancer registries for comparative effectiveness research (CER) to collect detailed cancer treatment data. A portion of the funding targeted the expansion of physician cancer reporting activities through novel methodologies to minimize reporting burden on physicians. Under the direction of the Florida Department of Health, the Florida Cancer Data System (FCDS) designed and implemented a physician cancer reporting pilot project that used electronic physician medical claims as the standard for reporting. Partnerships were established with the Florida Cancer Specialists (FCS) to serve as the provider of medical claims data, and the FCDS implemented a National Cancer Institute (NCI)-funded software that included an automated algorithm for processing of claims, known as MDoffice. Processed data were reviewed by certified tumor registrars (CTRs) and compared to the gold-standard treatment data collected from the FCS electronic health record (EHR) system for the CER project. The CER data was used as the gold standard given that all patient chemotherapy treatments were captured from the practice EHR. The gold standard data set was used to validate the accuracy of the treatment mapping table in MDoffice and to measure the treatment enhancements to the original registry records submitted by hospital abstractors. The strength of this study is that it leverages data collected for comparative effectiveness research combined with data from a physician practice EHR to provide accurate and robust treatment information.

The Florida Cancer Data System and Pilot Partners

The FCDS (http://fcds.med.miami.edu) is Florida’s statewide, population-based cancer registry and has been collecting incident cancer data since 1981. FCDS is wholly supported by the State of Florida Department of Health, the National Program of Cancer Registries of the CDC, and the Sylvester Comprehensive Cancer Center at the University of Miami School of Medicine. Under existing Florida statutes, all licensed hospitals, radiation therapy centers, laboratories, and ambulatory surgical facilities are required to report annual cases of cancer to the FCDS. A new initiative to collect cases from physicians was supported through the CER funding from the CDC, whereby mandated reporting from physicians who diagnose and/or treat cancer in the state of Florida was initiated.

In 2012, the FCDS partnered with the Florida Cancer Specialists (FCS), one of Florida’s largest private medical oncology practices, which employs more than 120 physicians and has more than 45 locations in the state. The FCDS worked with FCS and their third party claims vendor, Unlimited Systems, to set up a trigger mechanism by which copies of the 1500 Health Insurance Claim Form,23 using the 837 electronic transaction file format, were transmitted to the FCDS via a secure file transfer protocol (SFTP) on a nightly basis. The 1500 medical claim for professionals form is standardized to meet all the requirements set forth by the Health Insurance Portability and Accountability Act (HIPAA).24 Triggers were activated by the presence of a reportable cancer code using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) coding system.25 Transmission of records to the FCDS began in June 2012 and consisted of medical claims from FCS with service dates beginning in January 2011.

The FCDS also partnered with Lynne Penberthy, who directed the original development of the MDoffice software under an NCI contract. The MDOffice algorithm feeds data from medical claims, consolidates the data for treatment within a cancer case, and translates ICD-9-CM, Healthcare Common Procedure Coding System (HCPCS),26 and Current Procedural Terminology (CPT)27 codes to populate 56 fields in the standardized layout and coding schemas developed by the North American Association of Central Cancer Registries (NAACCR) in the Version 12.2.28 The resultant NAACCR record contains consolidated treatment received by the patient for each ICD-9-CM tumor.

Materials and Methods

Using the MDoffice algorithm, we undertook a pilot study to test and validate the transmission and processing methods of electronic physician medical claims into the standardized central cancer registry layout. Through the use of a gold standard data set, collected as part of the CER project, we evaluated the agreement, sensitivity, and specificity of chemotherapy treatment from medical claims and the enhancement to the cancer registry chemotherapy data. We expected that medical claims would provide additional chemotherapy information not reported by traditional hospital sources.

Study Design and Analysis: Validation of Chemotherapy Treatment using the Gold Standard

The medical claim records from FCS with the encounter dates from January 1, 2011 to September 1, 2013 were imported into an Oracle database environment and linked to the central registry’s patient database using patient first name, last name, and date of birth. Social Security number was also used when available. As part of the CER project, the central registry staff collected patient tumor records with detailed treatment information from the physician practice EHR for patients with a breast, colon, rectum, or chronic myeloid leukemia invasive cancer diagnosed in 2011. These cases served as the gold standard for comparison of data from the medical claims, and included the originally reported information from hospital registries to produce a comprehensive patient treatment profile. The additional
Table 1. Chemotherapy Treatment Category Recodes Using NAACCR Data Item #1390 RX-Summ-Chemo*

<table>
<thead>
<tr>
<th>Category recodes</th>
<th>NAACCR code and description for #1390 RX Summ Chemo</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Chemotherapy Received</td>
<td>00 None, chemotherapy was not part of the first course of therapy; not customary therapy for this cancer.</td>
</tr>
<tr>
<td>Yes Chemotherapy Received</td>
<td>01 Chemotherapy, Not Otherwise Specified (NOS)</td>
</tr>
<tr>
<td></td>
<td>02 Chemotherapy, single agent</td>
</tr>
<tr>
<td></td>
<td>03 Chemotherapy, multiple agents (combination regimen)</td>
</tr>
<tr>
<td>No Chemotherapy Received</td>
<td>82 Chemotherapy was not recommended/administered because it was contraindicated due to patient risk factors (comorbid conditions, advanced age, etc.).</td>
</tr>
<tr>
<td></td>
<td>85 Chemotherapy was not administered because the patient died prior to planned or recommended therapy.</td>
</tr>
<tr>
<td></td>
<td>86 Chemotherapy was not administered; it was recommended by the patient’s physician, but was not administered as part of first-course therapy. No reason was noted in the patient record.</td>
</tr>
<tr>
<td></td>
<td>87 Chemotherapy was not administered; the patient’s physician recommended it, but this treatment was refused by the patient, the patient’s family member, or patient’s guardian. The refusal was noted in the patient record.</td>
</tr>
</tbody>
</table>


Results

A total of 15,273 medical claims were transmitted from FCS that, after consolidation and matching, linked to a total of 623 CER patient tumors. A majority of the service records contained either anti-neoplastic agents (chemotherapy) or ancillary therapies. A very small percentage of claims captured hormone agents. There were no surgical procedures listed on the claims. The total percent of patients who received chemotherapy treatment in the FCS medical claims data set was 41% compared to 28% in the original registry data set, and 45% in the gold standard CER data set (Table 2 and Figure 1). When combining treatment information from claims with the registry data, a total of 45% of cases with chemotherapy were captured. This approximates the percent of cases where chemotherapy was received in the CER data set (Figure 2).

There were no unknown chemotherapy values in the gold standard CER data set. A comparison of detailed chemotherapy categories is shown in Table 3. The distribution of single agent chemotherapy treatment was 5% in the claims data set, 1% in the registry data set, and 5% for the CER data set; multiagent distributions were 35% in claims, 19% in the registry, and 41% in CER. The registry data set contained 14% of records where chemotherapy treatment was unknown, and 9% of records as chemotherapy not otherwise specified (NOS).
Results for the percent agreement, sensitivity, PPV and $k$ statistics are presented in Table 4. Comparing the medical claims data set against the gold standard CER data set, there was agreement 91% of the time, with 86% sensitivity, 93% PPV, and a $k$ of 0.81. The registry data set agreed with the CER data set 76% of the time, with 60% sensitivity, 93% PPV, and a $k$ of 0.56. After integrating the medical claims with the registry data, the level of agreement was 92%, with 92% sensitivity, 91% PPV, and a $k$ of 0.85. When comparing the medical claims against the registry data there was only a 73% agreement with 86% sensitivity and 0.52 $k$.

There were a total of 25 false negative cases where no chemotherapy treatment was ever recorded on a claim with the same primary site diagnosis code. An additional 7 records contained chemotherapy information for the patient but did not have a corresponding cancer diagnosis code recorded in 1 of the diagnosis fields; 6 cases were mapped to a different cancer subsite ICD-9-CM code; and 1 case had a generic chemotherapy procedure code that is not included in the treatment mapping table. There were a total of 11 false positive cases where claims included chemotherapy procedures for subsequent treatment of disease progression; 6 cases were mapped to a different primary site, where patients had multiple primary diagnoses; and there were 2 cases where patients refused treatment. Table 5 shows the scenarios for which cases were categorized incorrectly.

**Limitations**

There were limitations to the medical claims pilot project and validation study in the form of data completeness. First, the FCS transmitted claims files incrementally, resulting in an incomplete patient data set. To ensure more accurate comparison to the CER data set, the FCS resubmitted a complete set of 2011 claims records to the FCDS. However, the 2012 and 2013 claims records may have contained gaps in services. Additionally, the FCS was undergoing a rapid expansion of services throughout the project period, acquiring existing practices and incorporating patient data into the FCS system. The integration
of new patients and patient records into the FCS system resulted in incomplete capture of reported patient services on the physician medical claims. Therefore, the sensitivity measured may have been underestimated.

Additionally, linkage and capture of cancer-directed therapy is contingent on having a valid ICD-9-CM reportable cancer code. There were claims where procedures for chemotherapy infusions were listed in either the diagnosis or procedure fields, but no corresponding reportable cancer was recorded that linked to a CCR cancer. In these cases, the cancer-directed treatments would not be used in assigning treatment. Another issue we encountered was the coding of drugs on claims that were not specific to an antineoplastic agent, such as J3490 “Unclassified Drug” and 96523 “Irrig drug delivery device.” For these cases, the procedure date matched the chemotherapy treatment date recorded in the gold standard record, but the generic code was not enough information to translate into a definitive chemotherapy treatment. Therefore, administrative coding errors and/or incomplete reporting on claims do affect the capture rate of cancer-directed treatments.

One of the major limitations of claims data in this study is that there was not the ability to identify which treatment agents were used toward the patient’s first course of treatment as opposed to second course, other than by visual comparison to the gold standard record. This could potentially be addressed in the future using interval timing rules from the date of diagnosis as a proxy to define first course of therapy. Additionally, there is no way to positively distinguish between subsequent treatment and progression of disease status. These are clinically important elements to capture when evaluating standards of care, treatment effectiveness, and disease status. The MDoffice algorithm, however, can include timing rules to help distinguish between first course and subsequent course of treatment. But in the current analysis, there were no such timing rules implemented. Therefore, the comparison of detailed treatment agents between the claims output and the CCR record is limited by these issues.

It is also important to note that when there are multiple primaries for a patient, the claim record may include a treatment code that corresponds to the first but not the secondary primary, resulting in missed treatment information. However, this is an issue external to claims and is related more to the coding of the primary tumor in the registry data. Additionally, the high discordance of values when comparing to the registry data set was most likely due to the high number of unknown cases (n = 84) in the registry data set. A final limitation of the claims data is that oral antineoplastic agents are not typically billed, with the exception of a few agents covered by Medicare. Orally administered anti-neoplastic treatments were similar to the hormonal therapy in that they were unlikely to be captured in claims data, with a few exceptions representing oral equivalents to infusion agents that are covered under Medicare Part D. However, most oral medications are dispensed by pharmacies and would not be captured on the physician practice claim. This is a growing issue in capturing complete treatment information as oral agents are becoming more widely prescribed to cancer patients.

**Discussion**

The claims pilot project demonstrates proof of concept that electronic medical claims can be securely transferred using a viable trigger mechanism, provide a high level of sensitivity with regard to capturing chemotherapy treatments, and enhance existing sources of cancer-directed therapies that more closely approximate full treatment profiles. Previous studies demonstrated the ability of Medicare and inpatient claims to supplement and enhance registry-collected data. Using physician medical claims allows collection of information over a greater length of time for each patient and from additional reporting sources not currently captured in the hospital setting or in Medicare claims files. Incorporating medical claims reporting from physicians as an additional source of cancer surveillance is effective and has the potential for broader implementation.

As the diagnosis and treatment of cancer are more commonly taking place in the private physician setting, it becomes increasingly important to capture information from these sources to prevent underreporting of cases. As demonstrated by Penberthy and colleagues, there is already evidence of underreporting for hematologic and urologic malignancies. Furthermore, we have shown that reliance on hospital abstracts for chemotherapy treatment information falls short of a complete first-course profile. Supplementing treatment information from physician medical claims not only provides a higher level of case completeness, but can also be used to capture granular treatment information for comparative effectiveness research, and patterns of care studies.

While the resources involved in the design and implementation of this medical claims pilot project were funded through the CDC’s CER study, the long-term operation of
physician reporting through medical claims is considered to be a sustainable activity. This is not to minimize the initial resources involved for implementation, physician outreach, and integration of medical claims into a central cancer registry database. However, in large part, the methods and tools have already been established and continue to be updated and validated through its use. Through collaborative efforts and national partnerships, the dissemination of these practices can minimize the cost to other central cancer registries significantly, while providing an additional mechanism to capture both treatment and newly diagnosed cases not otherwise captured from traditional reporting sources.

**References**

Abstract: Background: Cancer treatment information is often underreported in cancer registries due to the shift in cancer care to ambulatory settings. Incomplete treatment information for central registries limits the usefulness of these data for understanding disparities in outcomes. The objective of this study was to evaluate the added value and validity of medical billing data to supplement treatment information for incident cases within a central cancer registry. Methods: Billing data using standardized structure and nomenclature as submitted by all practices was evaluated using an automated software (MDOffice, MDO) process that captures and processes these data and submits the information in a standardized format. A validation of the billing reported treatment was performed using data from 3 community oncology practices. Results: The accuracy of treatment data captured was 100% for both chemotherapy and radiation therapy among the 313 cases validated. Chemotherapy (36% and 5% respectively for solid tumors and hematologic cancers) and radiation therapy (46% and 20% respectively for solid tumors and hematologic cancers) information was added to 738 known incident cases using billing data. Conclusion: Automated reporting based on billing data from community specialty providers is likely to markedly enhance the completeness of treatment data among known cancer cases as these community providers render significant amounts of treatment for cancer patients.

Key words: automation, cancer registry, claims data

Background
Gaps in reporting of initial course of therapy for cancer surveillance limits the usefulness of the cancer surveillance data maintained by central cancer registries, including both the National Cancer Institute’s (NCI’s) Surveillance Epidemiology and End Results Program (SEER) and the Centers for Disease Control and Prevention’s (CDC’s) National Program of Cancer Registries (NPCR). The data are incomplete for several reasons. The primary source for case and treatment reporting has traditionally been manual abstraction via hospital cancer registries. Hospital-based cancer registrars lack access to information on treatments provided to patients outside their health system, such as in community-based oncology practices. This lack of access, coupled with a simultaneous migration of systemic treatments for cancer to the ambulatory setting, has further increased the gap in treatment information maintained in registries. Further, when treatment data are captured, the information recorded is limited to whether or not the patient received 1 or multiple chemotherapies with no detail on the specific agents, regimens, number of cycles, or dosage received. This information gap represents a critical limitation in the ability to use these data for evaluating the generalizability of clinical trial proven therapies, comparative effectiveness research, and understanding differences in outcomes among population subgroups.

One source for providing more complete data on systemic therapy may be medical claims (billing) data. Medicare claims linked with SEER data have been used extensively to study the use of chemotherapy for specific cancer sites and have been found to be both sensitive and specific. The Medicare claims have also been demonstrated to be valid for identifying initial (induction therapy) and subsequent therapy based on Medicare data. However, there remain many unanswered questions related to the supplementation of treatment reporting from billing data for patients outside the Medicare-insured population. First, Medicare covers primarily patients aged 65 years and older and thus does not provide information on patients under age 65 years. Patterns of treatment for Medicare elderly may be substantially different than for cancer patients under age 65 years. Further, the

Recent advances in treatment regimens for some cancers have been demonstrated to increase survival in these patients. However, incomplete information on the more than 95% of patients not enrolled in clinical trials makes it impossible to evaluate the dissemination of these new therapies and their effectiveness in the general population. With the increasing costs of cancer care, it is even more important to have a population-based system providing detailed information on treatment, as there may be substantial variation in costs and effectiveness represented by variations in treatment protocols.
validity of treatment reporting from claims data has not been demonstrated across a wide variety of cancer sites. To address some of these questions, we evaluated the ability of medical claims data from community oncology practices to supplement systemic treatment reporting among registry-reported, incident cancer cases.

This manuscript describes the validity and potential benefit to central cancer registries of automated capture of claims data directly from community oncology practices. Claims submitted from the practice provide data to all payers. This represents information on the entire set of patients rather than those represented by a single insurer that may be biased. We performed an assessment of completeness of reporting of systemic therapy and illustrated the additional benefit of information detail that might be provided by claims data from community oncology practices.

**Methods**

This pilot study was conducted in 3 community medical oncology practices using a novel automated software application that imports and extracts key data elements from standardized electronic billing data from community physician offices (MDO). Details on the software and the added value in casefinding and treatment for urologic cancers has been described elsewhere. The software uses a sophisticated algorithm that maps standardized International Classification of Diseases 9th Edition (ICD-9) diagnosis codes, Healthcare Common Procedure Coding System (HCPCS) codes, and Current Procedural Terminology (CPT) codes to standardized data items (patient demographics, diagnosis, treatment, and provider information) in the North American Association of Central Cancer Registries (NAACCR) record layout. The software leverages the standardized transmission format for billing data (837 professional) which includes the standardized nomenclature related to diagnosis and treatment (ICD-9 CM, HCPCS, and CPT codes) described above.

**Data Collection**

Twenty months (November 1, 2008 through June 30, 2010) of billing data were captured and processed from 3 oncology practices in North Carolina representing 22 oncologists and 11 nurse practitioners.

**Match with Central Cancer Registries**

Personnel at the North Carolina Central Cancer Registry matched the software-generated NAACCR records for patients identified as having cancer from billing with existing cancer cases in the central cancer registry database. Details of the matching process have been reported previously. Matching for hematologic malignancies was performed based on 6 categories as reported previously. Matching for solid tumors was based on the 2-digit International Classification of Diseases for Oncology (ICD-O) topography (primary site) code. Because the match rates (proportion of cases known to the registry) are substantially different for hematologic malignancies vs solid tumors, the contribution of the claims data to treatment for known cases was calculated separately for hematologic malignancies and solid tumors.

Incident cases were defined as registry-matched cases whose initial date of diagnosis was between November 1, 2008 and June 30, 2009.

**Treatment**

All cancer-directed treatment (chemotherapy or radiation therapy) was captured for a 20-month period from billing data from the 3 oncology practices from November 2008 through June 30, 2010. The extended period for capture of treatment assured that we had a minimum of 12 months of follow-up data for all matched incident cancer cases identified in the central registry. This provided data for the entire initial treatment period. The mapping for the HCPCS and CPT codes to NAACCR standard treatment codes used in this study is available on request. In addition to maintaining the code, text, and administration date for all cancer treatments, treatment captured from the billing data was automatically coded by the software to the SEER/CRC RX-Summ (treatment summary) fields and populated in the NAACCR record according to the rules for those fields. Also, the total number of unique chemotherapeutic agents for each case was calculated from the billing data. All dosages of the same drug, even if represented by a different HCPCS code, were considered to be a single agent.

Chemotherapy treatment as reported from the registry was defined using the RX_Summ_Chemo field (NAACCR Item # 1390); see Table 4 for codes and definitions. Radiation treatment in the registry was defined using the RX_Summ_Radiation (NAACCR Item # 1360) standard codes and coding instructions.

Only 2 practices provided radiation treatment as a single modality. We examined whether or not radiation therapy was reported in the registry and compared the added number of patients treated with radiation represented by billing. The analysis was performed for patients over and under age 65 years and overall for incident cases.

**Validation**

The purpose of the validation in this study was to confirm the administration of chemotherapy agents and radiation therapy reported in billing in the medical record of the billing provider. Cases for validation were randomly selected from among all nonmatched cases as part of a larger study to validate casefinding. For the 313 cases in the validation sample, all chemotherapy and radiation therapy treatments and dates from the oncology practice medical record were captured and recorded. The abstracted information was then compared with the treatments reported in the claims data for the study interval (November 1, 2008 through June 30, 2010). Information on cancer treatments was abstracted without knowledge of whether or not the agent was reported in the billing.

Treatments were not included in the validation when delivered at locations other than the billing practice (n patients = 47) or as part of a research protocol (n patients = 3). Only treatments received within the study interval (November 1, 2008 through June 30, 2010) were included for comparison with the billing data.
Discrepancies in which billing data did not include treatment reported from the initial practice medical record abstraction were reabstracted by an independent reviewer to validate whether or not the treatment was received.

Analysis

The ability of the billing data to supplement chemotherapy reporting among incident registry-matched cases was evaluated using several approaches. First, we determined how often a cancer case had any chemotherapy reported in the registry according to the NAACCR RX_Summ_Chemo field codes (01, 02 or 03) (See Table 4 for definitions). We also determined how often a cancer case had any radiation therapy reported in the registry according to the NAACCR RX_Summ_Radiation field codes (codes 1 through 5 indicating that radiation was received). We calculated the number of patients with chemotherapy and radiation therapy reported only in the billing data to measure the contribution provided by the billing data to measure existing cancer registry data. This was calculated as the number of patients with chemotherapy or radiation therapy identified only in billing divided by the number of patients with chemotherapy or radiation therapy already in the registry. Comparisons were performed for patients under and over age 65 years using a chi-square test.

In addition to determining the added value for missed chemotherapy as a dichotomous variable, we compared the registry-reported code with the information available in the billing data based on 10 categories within the RX_Summ_Chemo codes. The purpose was to determine how often the billing data provided more specific and/or more accurate information than that captured in the registry (See Table 4 for the codes and descriptions). For example, the registry reported having 1 agent, but the billing data identified more than 1 agent.

Finally, we evaluated the number of unique agents that were captured through the billing data for incident matched cases overall and by age groups (<65 years and ≥65 years).
The majority of patients were female reflecting the high proportion of breast cases in this set of practices.

Validation

Among the 313 cases included in the validation sample, 23.0% received chemotherapy and 5.1% received radiation therapy during the study period. This represented 33 unique chemotherapeutic agents. Of the 142 chemotherapy administrations reported from billing, 100% were validated in the oncology practice medical records as correct. Two patients receiving 3 agents were unable to be accessed in the third practice due to change in practice ownership and were excluded from the analysis.

Three patients had chemotherapy recorded in the original validation abstraction but did not have chemotherapy coded in the billing data. A reabstraction by an independent certified tumor registrar revealed that in all 3 cases the patient did not receive chemotherapy at the practice although the initial intent (as described in the medical record) was to provide onsite chemotherapy (patient died, chemotherapy was refused, and chemotherapy received as inpatient). Of the 15 patients who were reported to have had radiation therapy, 100% were validated as having received radiation on site at the practice.

Added Chemotherapy and Radiation Therapy

The added benefit of chemotherapy captured from billing data for incident hematologic malignancies was relatively small at 5.4% overall. The added benefit was slightly greater for patients aged <65 years (6.7%) vs those aged 65 years and older (3.8%). For incident solid tumors, billing data added 36.2% of chemotherapy overall among patients for whom no chemotherapy had been previously reported to the central cancer registry. The added value was significantly higher among patients aged <65 years (39.4%) vs patients aged 65 years and older (33.3%) (Table 2).

We also examined the percent of patients with reported chemotherapy among nonmatched patients with solid tumors (1,522) and hematologic malignancies (832). The percent of patients who received chemotherapy was 16.2% and 13.0%, respectively.

For radiation therapy (Table 3), the proportion of hematologic cases where billing data added information was low with a 20.0% increase overall, ranging from 16.7% for patients under age 65 years to 25% for patients aged 65 years and older. The added proportion of radiation therapy to incident matched solid tumors was higher with a 46.2% increase overall, 44.6% for patients under 65 years and 47.7% for patients over 65 years.

As for chemotherapy, we examined the percentage of unmatched patients who were reported to have received radiation therapy which was 3.4% and 2.0% respectively for solid tumors and hematologic malignancies.

Table 4 represents the impact of the billing data to provide more specific and/or additional information for incident cases with chemotherapy reported as “no” or “unknown” (RX_Summ_Chemo codes 00, 87, 88, or 99), or whose chemotherapy may have been underreported (RX_Summ_Chemo codes 01 or 02). Of the 415 solid tumor patients with no chemotherapy or unknown chemotherapy reported in the registry, 85 (20.5%) had chemotherapy identified from billing data. For the 69 patients with known receipt of chemotherapy, 29 (42.0%) had additional information from billing data adding specificity to the number

### Table 3. Added Value of Billing Data from 2 Oncology Practices in Supplementing Radiation Therapy Capture by Cancer Registry for Matched Incident Cases

<table>
<thead>
<tr>
<th></th>
<th>&lt;65</th>
<th>≥65</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Solid tumors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of incident cases</td>
<td>282</td>
<td>378</td>
<td>660</td>
</tr>
<tr>
<td>Total number of incidents with radiation in registry</td>
<td>65</td>
<td>65</td>
<td>130</td>
</tr>
<tr>
<td>Total number of incidents with radiation in billing only</td>
<td>29</td>
<td>31</td>
<td>60</td>
</tr>
<tr>
<td>Percent with radiation in registry</td>
<td>23</td>
<td>17.2</td>
<td>19.7</td>
</tr>
<tr>
<td>Percent with radiation in billing only</td>
<td>10.3</td>
<td>8.2</td>
<td>9.1</td>
</tr>
<tr>
<td>Total percent with radiation (registry and/or billing)</td>
<td>33.3</td>
<td>25.4</td>
<td>28.8</td>
</tr>
<tr>
<td>Added percent from billing</td>
<td>44.6</td>
<td>47.7</td>
<td>46.2</td>
</tr>
</tbody>
</table>

<p>| | | | |
|                      |     |     |       |
| <strong>Hematologic malignancies</strong> |     |     |       |
| Total number of incident cases | 37  | 41  | 78   |
| Total number of incidents with radiation in registry | 6   | 4   | 10   |
| Total number of incidents with radiation in billing only | 1   | 1   | 2    |
| Percent with radiation in registry | 16.2| 9.8 | 12.8 |
| Percent with radiation in billing only | 2.7 | 2.4 | 2.6  |
| Total percent with radiation (registry and/or billing) | 18.9| 12.2| 15.4 |
| Added percent from billing | 16.7| 25  | 20   |</p>
<table>
<thead>
<tr>
<th>Registry code*</th>
<th>Description</th>
<th>Solid tumors and FORDs code distribution from billing</th>
<th>Hematologic malignancies and FORDs code distribution</th>
<th>% with chemotherapy information added</th>
<th>% with chemotherapy information added</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>No chemotherapy</td>
<td>363</td>
<td>6 (1.7)</td>
<td>67 (18.5)</td>
<td>73</td>
</tr>
<tr>
<td>01</td>
<td>Chemo administered but type and number of agents not documented.</td>
<td>31</td>
<td>5 (16.1)</td>
<td>19 (61.3)</td>
<td>24</td>
</tr>
<tr>
<td>02</td>
<td>Single agent chemo administered.</td>
<td>38</td>
<td>n/1</td>
<td>5 (13.2)</td>
<td>5</td>
</tr>
<tr>
<td>82</td>
<td>Chemo not recommended/ administered due to contraindications.</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>Chemo not administered because the patient died before the planned chemo administration</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>86</td>
<td>Chemo was recommended but not administered. No reason documented in patient record</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>87</td>
<td>Chemo was not administered. It was recommended but refused.</td>
<td>16</td>
<td>1 (6.3)</td>
<td>0 (0)</td>
<td>1</td>
</tr>
<tr>
<td>88</td>
<td>Chemo recommended but unknown if administered.</td>
<td>15</td>
<td>0 (0)</td>
<td>7 (46.7)</td>
<td>7</td>
</tr>
<tr>
<td>99</td>
<td>Unknown if chemo was recommended or administered.</td>
<td>21</td>
<td>0 (0)</td>
<td>4 (19.0)</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>494</td>
<td>12 (2.3)</td>
<td>102 (20.7)</td>
<td>114</td>
</tr>
</tbody>
</table>

*03 code (more than 1 agent) excluded because no additional benefit from billing could be added for patients with a FORDs code 03 in registry.
of agents received based on billing data. It is worth noting that for over two-thirds (77.4%) of solid tumor patients who had code 01 in the central registry (chemotherapy, not otherwise specified), the billing data was able to provide more complete information on the number of chemotherapeutic agents the patient received. For the 22 hematologic malignancies where the registry reported no chemotherapy or unknown chemotherapy status, 3 (13.6%) had chemotherapy identified from billing. For the 13 patients with hematologic malignancy who had some chemotherapy reported in the registry (codes 01, 02), the billing data provided information for 1 patient (8%).

Table 5 provides the distribution of the number of chemotherapeutic agents that were provided during the initial treatment period. The distribution of number of agents (categorized as 1, 2, and ≥3 unique agents) varies by age group with younger patients (age <65 years) more likely to receive 3 or more agents for both hematologic malignancies (17.6% in <65 years vs 13.3% in ≥65 years (P-value NS) and for solid tumors (45.6% in <65 years vs 25% in ≥65 years, P = .001). Thirteen percent of patients received 3 or more agents during the initial treatment window as identified from claims alone.

### Discussion

The goal of this study was to evaluate the potential to supplement treatment information among known cancer cases using billing data from community oncology practices. The importance of capturing treatment information from these practices has become increasingly relevant as the patterns of care for cancer shift further to the ambulatory care setting and as community practices increase the menu of services available.7,26,29 In this study, using billing data from 3 community oncology practices, we found that the billing data added substantial treatment information to incident cancer cases previously reported to the central cancer registry. The data from billing were highly accurate with all billing-reported antineoplastic agents validated. This study supports previous analyses in which claims have been demonstrated to have a high sensitivity, specificity and positive predictive value for the capture of cancer information.25-28 Further, in 3 instances where chemotherapy was reported through the initial medical record abstraction but for which there was not a bill, a reabstraction showed that for each of those instances the patient did not receive the chemotherapy as initially planned.

The benefit was shown for both chemotherapy and radiation therapy and was greatest for solid tumors. Over one-third of patients for whom the registry did not have any chemotherapy reported as part of the initial course of treatment were identified as having received treatment using the billing data. Billing data increased the proportion of patients having received chemotherapy from 35.6% to 48.5% of patients for solid tumors. For hematologic malignancies, the added benefit was smaller—5.4% overall. The lower proportion of added benefit among hematologic malignancies may reflect several factors. A previous study demonstrated that hematologic malignancies are substantially underreported to central cancer registries. The registry-matched incident cases used in this study may represent a biased subset of the entire population of patients with hematologic malignancies, namely cases reported with a hospital admission. For these hospitalized patients, treatment is more likely to be identified and reported by hospital-based registrars.

In addition to the potential for added information for known (matched) cases, we also found patients with chemotherapy administration among patients not reported to the registry for both hematologic malignancies (13.0%) and solid tumors (16%). Given the high level of accuracy in reporting of chemotherapy, these cases may represent cases unreported to the Registry. Billing data (including treatment data) may serve as a useful tool for targeted casefinding (new case identification). However, not all cases identified through this manner will be reportable.25-26

Billing data provide an opportunity to capture detailed information on the specific agents and dosages received. As shown in the analysis of number of agents, it is clear that there are significant differences in how patients are treated, with age being only 1 factor potentially influencing treatment. The number of agents that were used during the initial 12 months post diagnosis for these patients ranged from 1 to 9 unique agents. While not currently a component of cancer case reporting, this information might easily be maintained for research purposes as an adjunct to a central registry. The availability of more detailed information on these treatments would be an important addition for outcomes studies in particular.

The use of claims data from community provider practices offers a readily available, effective and efficient opportunity to implement routine and rapid reporting of

---

<table>
<thead>
<tr>
<th>Number of agents</th>
<th>Hematologic malignancies*</th>
<th>Solid tumors**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;65</td>
<td>≥65</td>
</tr>
<tr>
<td>n = 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 (5.9)</td>
<td>3 (20.0)</td>
</tr>
<tr>
<td>2</td>
<td>1 (5.9)</td>
<td>1 (6.7)</td>
</tr>
<tr>
<td>≥3</td>
<td>15 (88.2)</td>
<td>11 (73.3)</td>
</tr>
</tbody>
</table>

*P < .473. **P < .001.
treatment delivered in a nonhospital setting. While meaningful use criteria include cancer case reporting, there is a lag in the capacity of oncology electronic medical record systems to do so. Thus, a system building on existing standardized nomenclature, coding system(s), and record layout that is not reliant on customized reporting by electronic medical record systems vendors may provide a more consistent and rapidly implementable tool for filling an important gap in cancer treatment reporting. Such an automated system would likely represent only modest impact on the workload for both the provider and the registry largely related to the initial implementation. Further, the ability to capture discrete and accurate data from billing is an efficient method that does not require extraction of specific information from free text documents.

We also compared the benefit of claims-reported treatment among older and younger patients using age 65 years as the division since it represents the age at which Medicare coverage typically begins. Because Medicare data have been used extensively in research studying patterns of care, it is important to determine if there were differences among the 2 age groups in the proportion of patients who were treated as well as the intensity of treatment. For chemotherapy, among solid tumors, there were significant differences in treatment by age. Younger patients (aged <65 years) were more likely to be treated with any chemotherapy and to receive more agents. The analysis by age supports the need to capture information on patients of all ages from the community setting rather than relying only on data from the Medicare population for understanding patterns of care and outcomes.

While chemotherapy has been known to be under-reported, the radiation therapy identified as missing in this sample of oncology practices was greater than anticipated. Although radiation therapy has been considered to be more completely reported than chemotherapy, studies have suggested that there is a gap in reporting of this therapeutic modality as well. Historically, it has been a problem for registries to identify all independently operated radiation therapy facilities. As the number of community specialty practices (such as oncology and urology) that provide the full spectrum of treatment services increases, the gap in unreported radiation therapy may also increase. In fact, there has been growth in the number of radiation oncologists by 26% from 1995 to 2008. Thus, identification of community specialty practices with on-site radiation services to enhance reporting may be important for registries to enhance completeness of radiation treatment reporting.

Limitations

The first limitation of this study was that it is based on only 3 practices in 1 state, thus the generalizability may be limited. Other previously published studies representing other states have demonstrated the potential benefit of community provider billing data to supplement both case-finding and treatment reporting.

Second, it is important to recognize that the billing data from the community providers do not represent the entirety of treatment for these patients. Not all chemotherapy may be identified by a single data source such as the physician practice billing. In this study, 68.5% and 51.8% of patients with solid tumor and hematologic cancer, respectively, with registry-reported chemotherapy also had billing for chemotherapy during the study interval. Patients whose chemotherapy was not captured in the billing data may represent patients who received chemotherapy either as an inpatient, hospital outpatient, or at other practice locations/infusion centers (n = 47 for this study based on the validation). This suggests that data capture from multiple sources is required to provide the most complete information on multimodal cancer-directed treatments. It also illustrates the continued necessity of reporting from hospital facilities to provide information on inpatient and hospital outpatient treatments.

Conclusion

Community oncology providers render significant amounts of treatment of varying modalities. This study supports other prior studies that demonstrated the usefulness of automated reporting using medical claims/billing data from these providers for case-finding. It also quantified the ability of claims to enhance the completeness of cancer-directed treatment for known cancer cases. The billing data offer additional advantages as the information captured is nearly universally available in a standardized format and using standardized coding. They provide highly detailed information including the name and dose of individual antineoplastic agents, and the data are highly accurate as shown in this and previous studies. Thus, the automated use of these data combined with ongoing reporting by hospital cancer registries offers an opportunity to efficiently supplement treatment information without increasing manual data collection efforts on already stressed oncology practices and cancer surveillance programs.

Acknowledgments

The authors would like to thank the US Oncology—Cancer Centers of North Carolina, Raleigh, Durham and Asheville practices for providing support and data for this study.

References

Coding Completeness and Quality of Relative Survival-Related Variables in the National Program of Cancer Registries Cancer Surveillance System, 1995–2008

Reda J. Wilson, MPH, CTR, RHIT; M.E. O’Neil, MPH; E. Ntekop, MD, MPH; Kevin Zhang, PhD; Y. Ren, PhD

Abstract: Background: Calculating accurate estimates of cancer survival is important for various analyses of cancer patient care and prognosis. Current US survival rates are estimated based on data from the National Cancer Institute’s (NCI’s) Surveillance, Epidemiology, and End Results (SEER) program, covering approximately 28% of the US population. The National Program of Cancer Registries (NPCR) covers about 96% of the US population. Using a population-based database with greater US population coverage to calculate survival rates at the national, state, and regional levels can further enhance the effective monitoring of cancer patient care and prognosis in the United States. The first step is to establish the coding completeness and coding quality of the NPCR data needed for calculating survival rates and conducting related validation analyses. Methods: Using data from the NPCR-Cancer Surveillance System (CSS) from 1995 through 2008, we assessed coding completeness and quality on 26 data elements that are needed to calculate cancer relative survival estimates and conduct related analyses. Data elements evaluated consisted of demographic, follow-up, prognostic, and cancer identification variables. Analyses were performed showing trends of these variables by diagnostic year, state of residence at diagnosis, and cancer site. Results: Mean overall percent coding completeness by each NPCR central cancer registry averaged across all data elements and diagnosis years ranged from 92.3% to 100%. Results showing the mean percent coding completeness for the relative survival-related variables in NPCR data are presented. All data elements but 1 have a mean coding completeness greater than 90% and was the mean completeness by data item group type. Statistically significant differences in coding completeness were found in the ICD revision number, cause of death, vital status, and date of last contact variables when comparing diagnosis years. The majority of data items had a coding quality greater than 90%, with exceptions found in cause of death, follow-up source, and the SEER Summary Stage 1977, and SEER Summary Stage 2000. Conclusion: Percent coding completeness and quality are very high for variables in the NPCR-CSS that are covariates to calculating relative survival. NPCR provides the opportunity to calculate relative survival that may be more generalizable to the US population.

Key words: cancer, data quality, National Program of Cancer Registries, relative survival rates

Background

Estimation of cancer survival is an important part of assessing the overall strength of cancer care and the success of prevention programs. Relative survival is a measure that can be used to describe the survival of a cohort of cancer patients by removing the effect of competing death events of a comparable general population. The measure is the ratio of observed survival among cancer patients divided by the expected survival of the general population that is comparable to the cancer patients with respect to covariates including age, sex, and year of diagnosis. Population-based cancer relative survival rates are important for medical and public health efforts, including measuring the survivorship of cancer patients after diagnosis and monitoring the impact of intervention and early detection programs.

Current US cancer survival rates are estimated based on data from the National Cancer Institute’s (NCI’s) Surveillance, Epidemiology, and End Results (SEER) program, which covers approximately 28% of the US population. The National Program of Cancer Registries (NPCR), established by Congress in 1992 and administered by the Centers for Disease Control and Prevention (CDC), is conducted in 45 states, the District of Columbia, Puerto Rico, and the Pacific Island jurisdictions, and covers approximately 96% of the US population. Data submitted annually to the NPCR-Cancer Surveillance System (NPCR-CSS) may also be used to calculate survival rates and provide greater coverage at national, regional, and state levels so clinicians, public health practitioners, and researchers can effectively monitor cancer patient care and prognosis in the United States.

NPCR-CSS collects data on the occurrence of cancer including the type, extent, and anatomic location of the cancer and the type of initial treatment by providing funding and technical assistance to the central cancer registries (CCRs) within the program. Population-based CCRs are...
data systems that collect, manage, and analyze data about cancer cases. In each state, medical facilities (including hospitals, physicians’ offices, therapeutic radiation facilities, free-standing surgical centers, and pathology laboratories) are required to report demographic and clinically-related data to their CCR. Each year, CDC supports efforts to link registry data with the National Death Index (NDI), Indian Health Service data, and state vital records; receives data from NPCR registries; and assesses the completeness and accuracy of the data. The annual data submissions from CCRs add a new year of data, update data from prior diagnosis years, and include the variables needed to calculate survival rates (eg, age, sex, year of diagnosis, date of last contact) as well as variables that are important surrogates of the quality of the follow-up information obtained (eg, type of reporting source, follow-up source, ICD revision number, cause of death) or can be used to stratify analyses (eg, stage, county or state, race, ethnicity).

Before using NPCR-CSS data to assess cancer survival, it is necessary to understand the coding completeness and coding quality of the data elements used in survival analyses across all the participating cancer registries. NPCR rigorously evaluates the completeness of case ascertainment and data quality for each annual data submission. Other studies have evaluated completeness, accuracy, and data quality of some, but not all, of the NPCR-CSS data required for conducting and validating survival analyses. For example, German and colleagues looked at the quality of breast and prostate cancer variables; Hall and colleagues compared SEER and NPCR incidence rates of cutaneous melanoma; McDavid and colleagues assessed breast, prostate, and colon variables; and Singh and colleagues investigated the quality of the census tract 2000 variable.

Using 1998–2001 data from 34 CCRs, Thoburn and colleagues compared NPCR-CSS incidence data to medical record data for 13 data elements among 4 primary cancer sites (lungs/bronchus, colorectal, prostate, and female breast) to assess the case completeness and data accuracy. The elements investigated included date of birth, race, sex, state of residence at time of diagnosis, diagnosis date, primary site, histology, behavior, grade, and SEER summary stage. The authors found data accuracy in 95% of the cases and case completeness in 96% of the cases; individual site-specific data element accuracy ranged from 81.2% to 100%, with a median accuracy of 98.1%.

The purpose of the current study is to build upon these previous studies by evaluating the coding completeness and quality of 26 data elements for all primary cancer sites, that are covariates to survival calculations and those used to assess the calculations across the 46 funded CCRs.

### Materials and Methods

Using NPCR-CSS data from 46 CCRs inclusive of 1995–2008 diagnosis years, 26 data elements used in survival calculations and validation were examined for coding completeness and quality. Coding completeness was assessed by calculating the proportion of nonmissing values by data element and by central registry; the numerator was the number of nonmissing values and the denominator was the total number of values (Table 1). Coding quality was calculated through the proportion of known values; the numerator was the number of known values (excluding unknown or blank values) and the denominator was the total number of values (Table 1).

The data elements consisted of survival analysis variables, demographic variables, cancer identification variables, follow-up and death variables, and cancer stage and prognostic variables. They specifically included:

- **Survival analysis:** date of birth, date of diagnosis, date of last contact/death, and sex.
- **Demographics:** age at diagnosis, US state of residence, county of residence at diagnosis, race, ethnicity (Hispanic), Indian Health Service (IHS) linkage (used to better classify cases of American Indian/Alaska Native heritage), North American Association of Central Cancer Registries’ (NAACCR) Hispanic Identification Algorithm (NHIA) derived Hispanic origin and NAACCR Asian/Pacific Islander Identification Algorithm (or NAPIIA, which is used to better classify cases of Asian Pacific Islander origin).
- **Cancer identification:** behavior (benign, in situ, or malignant), diagnostic confirmation, histology, SEER primary site group, sequence number central (number of primary cancers), and type of reporting source (primary source from which original cancer incidence report received).
- **Follow-up, recurrence, and death:** cause of death code, follow-up source central (2006–2008 diagnosis years for which the variable was captured by CCRs), follow-up source (1995–2005 diagnosis years for which the variable was captured by CCRs) (source from which follow-up information received), International Classification of Diseases (ICD) revision number (cause of death coding system version; in the analyses, this variable was divided into separate groups as versions 7–9 and version 10 or combined into 1 variable, versions 7–10), and vital status.

### Table 1. Coding Completeness and Coding Quality: Example Using the County of Residence Variable

<table>
<thead>
<tr>
<th>County of residence at diagnosis—codes</th>
<th>Coding completeness</th>
<th>Coding quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerator</td>
<td>000, 001–840, 999</td>
<td>001–840</td>
</tr>
<tr>
<td>Denominator</td>
<td>000, 001–840, 999, invalid/blank</td>
<td>000, 001–840, 999, invalid/blank</td>
</tr>
</tbody>
</table>

Figure 1. Mean Percent Coding Completeness Averaged over All Data Elements Combined (N = 26) and over All Diagnosis Years Combined (1995–2008) by National Program of Cancer Registries (NPCR) Central Cancer Registry (N = 46)

- **Cancer stage and prognostic:** SEER summary stage 1977 (1995–2000 diagnosis years for which the variable was in use), SEER summary stage 2000 (2001–2003 diagnosis years for which the variable was in use), and collaborative stage (CS) derived SEER summary stage 2000 (2004–2008 diagnosis years for which the variable was in use).

The mean percent coding completeness for each of the 46 central cancer registries was calculated and averaged over the 26 data elements combined and all diagnosis years (1995–2008) combined; the mean percent coding completeness was also calculated for each of the 26 data elements averaged over all the diagnosis years combined and all central cancer registries combined. Mean percent coding quality was assessed by each of the 26 data elements, taking the average of all diagnosis years and all central cancer registries combined. General linear modeling was performed to assess statistical differences for each data element by diagnosis year (year was coded as a categorical variable; the latest year available, 2008, was used as the referent year) and NPCR central cancer registry (assessed individually; the referent variable was a state that has maintained high coding quality and stability over time) \( (P = .05) \). Coding completeness of each data element was modeled as the outcome variable with diagnosis year and NPCR central cancer registry as the independent variables, respectively, in a least squares linear model. All analyses were conducted using SAS (version 9.2, SAS Institute).

**Results**

**Coding Completeness**

The mean overall coding completeness by each NPCR central cancer registry averaged across all 26 data elements and all diagnosis years ranged from 92.3% to 100% (Figure 1). Twenty-one of the central cancer registries (46%) had a mean overall coding completeness greater than or equal to 99% (Figure 1). All but 1 of the 26 data elements’ mean coding completeness by combined diagnosis year was greater than 90%; the completeness for the elements ranged from 91.9% to 100% (Table 2). Follow-up source (1995–2005) had a mean completeness of 42.2% (Table 2). Similarly, the data elements’ mean coding completeness by all central cancer registries combined was high; the majority ranged from 96.8% to 100% (Table 2). Only 2 of the elements were lower than that range: mean coding completeness by central cancer registry for Indian Health Service linkage was 88.9% and follow-up source (1995–2005) was 64.9% (Table 2). When we examined the data elements grouped by variable type, the mean coding completeness was above 90% for
each group; 99.3% over the 4 survival analysis variables combined, 98.6% over the 8 demographic variables, 100% over the 6 cancer identification data elements combined, 92.1% for the 5 follow-up/death data elements, and 100% for the 3 cancer stage/prognostic (not shown in tables).

In assessing the statistical difference in mean coding completeness of each data element by diagnosis year through the general linear models procedure, we found that date of birth, SEER summary stage 2000, CS derived SEER summary stage 2000, SEER summary stage 1977, cause of death, and follow-up source (2006–2008) showed no statistically significant differences by diagnosis year (Table 3). For ICD revision number, there was statistically significant difference comparing 2008 to 1996, but there were no statistically significant differences for the other diagnosis years. The mean coding completeness rates for the vital status data element are statistically significantly different comparing 2008 to the 13 years, 1995–2007 diagnosis years. Date of last contact’s coding completeness rates for 1995 and 2000–2003 were significantly different compared to 2008.

<table>
<thead>
<tr>
<th>Table 2. Mean Percent Coding Completeness for Relative Survival Data Elements Averaged Over All Diagnosis Years (1995–2008) and Over All National Program of Cancer Registries (NPCR) Central Cancer Registries (N = 46), NPCR Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Coding Completeness</strong></td>
</tr>
<tr>
<td>(%) [range]</td>
</tr>
<tr>
<td><strong>Survival analysis variables</strong></td>
</tr>
<tr>
<td>Date of birth</td>
</tr>
<tr>
<td>Date of diagnosis</td>
</tr>
<tr>
<td>Date of last contact or death</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td><strong>Demographic variables</strong></td>
</tr>
<tr>
<td>Age at diagnosis</td>
</tr>
<tr>
<td>County of residence at diagnosis</td>
</tr>
<tr>
<td>Ethnicity (Hispanic)</td>
</tr>
<tr>
<td>Indian Health Service linkage</td>
</tr>
<tr>
<td>NHIA (Hispanic origin)</td>
</tr>
<tr>
<td>NAPIIA (Asian Pacific Islander origin)</td>
</tr>
<tr>
<td>Race</td>
</tr>
<tr>
<td>State of residence at diagnosis</td>
</tr>
<tr>
<td><strong>Cancer identification variables</strong></td>
</tr>
<tr>
<td>Behavior</td>
</tr>
<tr>
<td>Diagnostic confirmation</td>
</tr>
<tr>
<td>Histology</td>
</tr>
<tr>
<td>SEER primary site group</td>
</tr>
<tr>
<td>Number of primary cancers</td>
</tr>
<tr>
<td>Type of reporting source</td>
</tr>
<tr>
<td><strong>Follow-up/recurrence/death variables</strong></td>
</tr>
<tr>
<td>Cause of death (ICD v.7–10)</td>
</tr>
<tr>
<td>Follow-up source (1995–2005)</td>
</tr>
<tr>
<td>Follow-up source (2006–2008)</td>
</tr>
<tr>
<td>ICD revision number</td>
</tr>
<tr>
<td>Vital status</td>
</tr>
<tr>
<td><strong>Cancer stage/prognostic variables</strong></td>
</tr>
<tr>
<td>SEER Summary Stage 1977 (1995–2000)</td>
</tr>
<tr>
<td>SEER Summary Stage 2000 (2001–2003)</td>
</tr>
<tr>
<td>SEER Summary Stage 2000 (CS derived) (2004–2009)</td>
</tr>
</tbody>
</table>
Table 3. General Linear Models Procedure (GLM) to Assess Percent Mean Coding Completeness Differences for Relative Survival Data Elements by Diagnosis Year and by NPCR Central Cancer Registry using NPCR-CSS Data (1995–2008)

<table>
<thead>
<tr>
<th>Data Elements</th>
<th>Diagnosis Year</th>
<th>NPCR Central Cancer Registry (CCR)*</th>
<th>Statistical Difference**</th>
<th>P-value*</th>
<th>Statistical Difference**</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survival analysis variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date of birth</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date of diagnosis</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Demographic variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County of residence at diagnosis</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity (Hispanic)</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHIA (Hispanic origin)</td>
<td>2008 SSD from 1995</td>
<td>.01 (1995)</td>
<td>2 NPCR CCRs SSD from all other NPCR CCRs</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAPIIA (Asian Pacific Islander origin)</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State of residence at diagnosis</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cancer identification variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis confirmation</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEER primary site group</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of primary cancers</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of reporting source</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Follow-up/recurrence/death variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cause of death (ICD v.7–10)</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up source (1995–2005)</td>
<td>2005 NSD from other diagnosis years</td>
<td>2005 NSD from other diagnosis years</td>
<td>.01 (1995)</td>
<td>.02 (2000)</td>
<td>20 NPCR CCRs SSD from all other NPCR CCRs</td>
<td>&lt;.02</td>
</tr>
<tr>
<td>Follow-up source (2006–2008)</td>
<td>NSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 NSD from other diagnosis years</td>
<td>.04 (1996)</td>
<td>2008 SSD from 1996</td>
<td>.04 (1996–2000)</td>
<td>2 NPCR CCRs SSD from all other NPCR CCRs</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Vital status</td>
<td>2008 SSD from 1995–2007</td>
<td>.02 (1995-2007)</td>
<td>2 NPCR CCRs SSD from all other NPCR CCRs</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cancer stage/prognostic variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*P-value calculated at a = .05 level of significance.
**SSD, statistically significant difference; NSD, no statistically significant difference.
†Statistically significant National Program of Cancer Registries (NPCR) central cancer registries all had a mean completeness less than that of the referent NPCR central cancer registry.
We also found that the percent coding completeness for the follow-up source variable for 20 NPCR central cancer registries were less than, and statistically significantly different from, the remaining NPCR central cancer registries (Table 3). For variables ICD revision number, date of birth, cause of death, and SEER summary stage 2000, only 1 central cancer registry has a less than statistically significant difference in mean coding completeness percent for each of these data elements. CS derived SEER summary stage 2000, IHS linkage, SEER summary stage 1977, vital status, follow-up source central, and date of last contact/death data elements all had 2 central cancer registries with a less than mean coding completeness percent statistically significantly different from all other NPCR central cancer registries.

**Coding Quality**

All of the survival analysis variables (date of birth, date of diagnosis, date last contact/death, and sex) achieved 100% coding quality (Figure 2). The majority of other variables also had a mean percent coding quality greater than 90%. The exceptions to this high percent were cause of death (78% ICD versions 7, 8, or 9 and 81% ICD version 10), follow-up source (1995–2005) (33%), SEER Summary Stage 1977 (85%), and SEER Summary Stage 2000 (87%).

**Discussion**

The results show a high level of coding completeness and quality for the 4 survival analysis variables (date of diagnosis, date of birth, date of last contact or death, and sex) across central cancer registry sites and diagnosis years. The majority of variables used to assess quality of the follow-up information and to stratify analyses also had high averaged means of coding completeness and quality by central registry sites and diagnosis years. These findings may be indicative of the training and support CDC-NPCR provides in monitoring and improving coding completeness.

The mean coding completeness percent for the relative survival data elements examined is relatively high compared with the previous studies, with an increase in the 2006 diagnosis year, and is similar among NPCR CCRs. This increase may result from the increase in NPCR CCRs conducting linkage processes with the NDI, including additional data editing and record updating, as well as the availability of training sessions and other resources. Demographics (some of which are evaluated annually to determine compliance with NPCR program standards), cancer identification, cancer stage/prognostic, and the majority of the follow-up/death data elements have high mean coding completeness percentages.
Even though the results are very promising, additional work may be needed for some of the data elements (e.g., date of last contact/death, follow-up source, IHS linkage). Some NPCR central cancer registries have concerns regarding releasing the full date of last contact/death due to confidentiality, while other central cancer registries may not be updating this data element following death certificate clearance procedures and/or NDI linkages. Additional discussions to assure confidentiality or resources to facilitate automatic record updates may be needed to improve the completeness of date of last contact/death. Our analysis also showed how competing priorities can affect coding completeness, as exhibited with the date of last contact/death variable. Starting in 2001, NPCR established a linkage agreement with NDI, which facilitated improved linkages and date of last contact/death information. However, in 2004, when Collaborative Stage activities became a priority for CCRs, the linkages were not completed as frequently and the completeness of the variable was affected.

The data element follow-up source (1995–2005 diagnosis years) has not been required by CDC for the NPCR registries, so the low level of completeness is not surprising. However, this data element is important; it makes it possible to identify records with information resulting from NDI linkages and, when necessary, release of that information can be recorded and reported back to NDI. The data item can also serve as a surrogate for the quality of the follow-up information.

As shown in the results section of this article, the cause-of-death variable has a low percent of coding quality, 78% and 81% for the different ICD versions. The cause of death is dependent on information recorded on death certificates, available through vital statistics linkages, and data quality issues have been identified in other evaluations. For this reason, researchers generally rely on relative survival rates for cancer rather than cause-specific survival rates.

Not all NPCR CCRs link with the IHS administrative database on an annual basis, but all do link every 5 years. If a record is not sent for IHS administrative database linkage, this data element is not coded. This most likely explains the lower percent completeness for the IHS linkage variable. Additional analyses may be needed, limiting the analyses to only those CCRs that conduct the IHS linkage annually or to the years where all CCRs conduct the linkage. The coding quality evaluation, however, showed that for records that were sent for linkage, the percent with a coded known value is very high.

Our results show that the NPCR-CSS data can be a complete source of information for researchers interested in using population-based cancer data to study cancer relative survival in the United States. Another strength of the NPCR data is the potential to calculate relative survival by race and ethnicity, which may assist researchers and comprehensive cancer control coalitions in making decisions about the type of cancer care and cancer programs they provide to the various ethnicities in the United States, thereby having the potential to reduce disparities in cancer incidence and survival.

More work is needed to improve coding completeness for cancer case follow-up, the information with the lowest mean coding completeness percentages in this database. A limitation of this study is that we did not assess data accuracy for these relative survival variables. Evaluating the data accuracy requires an audit of the source documents and the assigned codes. Other projects are conducting this evaluation and include some, but not all, of the data elements assessed in this project.

Survival analysis estimates are critical for many prevention, control, and treatment activities, including evaluation of the impact of screening and comprehensive cancer control programs and assessing the progress in cancer treatments. Because NPCR provides data for approximately 96% of the US population, it has the potential to provide near-national estimates as well as regional and state-based measures that have not before been available to researchers, clinicians, and public health decision makers. Our analyses demonstrate the high coding completeness and quality of the NPCR-CSS variables that are needed to calculate relative survival estimates and variables used to validate and stratify the estimates.

References

Relative Survival Analysis Using the Centers for Disease Control and Prevention’s National Program of Cancer Registries Surveillance System Data, 2000–2007

Reda J. Wilson, MPH, CTR, RHIT; A. Blythe Ryerson, PhD, MPH; Kevin Zhang, PhD; Xing Dong

Abstract: Background: Cancer survival rates are important in evaluating cancer care, identifying disease patterns, and estimating the probability of death due to cancer. To date, survival rates have been calculated using other data sets with limited population coverage that may not be able to fully identify differences by treatment, geographic region, and racial or ethnic group. Data from the Centers for Disease Control and Prevention’s (CDC’s) National Program of Cancer Registries (NPCR) have not been used previously to calculate relative survival rates within the United States. Methods: Data from CDC’s November 2011 submission for 21 state population–based central cancer registries, representing 50% of the US population, were included in this analysis. This paper presents relative survival rates for diagnosis years 2000–2007 with follow-up through 2008. Results: The relative survival rate for all cancers and races combined was 65.0% (65.3% for male and 64.8% for female patients). Black patients had a lower relative survival rate than white patients, except for lung and bronchus. For all cancers, the <45 age groups had the highest relative survival rates, except for black males. Discussion: For all cancer primary sites combined for 2000–2007, the CDC NPCR 5-year relative survival rate is comparable to that reported by the National Cancer Institute and the Canadian Cancer Registry. This analysis presents, for the first time, relative survival rates for half of the total US population and demonstrates that reliable survival rates can be calculated using CDC’s NPCR data now and in the future.

Key words: cancer surveillance, cancer surveillance system, National Program of Cancer Registries, population-based central cancer registries, relative survival rates

Introduction

Other publications1–7 have shown the importance of cancer survival rates in evaluating cancer care, identifying disease patterns, and estimating the probability of death due to cancer. This information is of interest to clinicians, patients, public health practitioners, and researchers, as well as to the public. To date, survival rates have been calculated using other data sets with limited population coverage that may not be able to fully identify differences by treatment, geographic region, and racial and ethnic groups.1

Estimates of current US survival rates are based on data from the National Cancer Institute’s (NCI’s) Surveillance, Epidemiology and End Results (SEER) program or from special studies.18–20 The NCI data set covers approximately 14% or 28% of the US population, depending on the table referenced.1 Until the preparation of this report, data from the Centers for Disease Control and Prevention’s (CDC’s) National Program of Cancer Registries (NPCR) have not been used to calculate relative survival rates. This report demonstrates that researchers can also use NPCR data to calculate reliable relative survival rates that help provide a more comprehensive picture of cancer survival in the United States.

CDC’s NPCR was established in 1992 to support the implementation and enhancement of population-based central cancer registries in US states and territories.21 CDC currently supports 45 states, the District of Columbia, Puerto Rico, and the Pacific Island jurisdictions covering 96% of the US population. CDC and NCI jointly support 5 states, and 5 states are funded by NCI alone.

Beginning in 2001, NPCR-funded grantees began reporting incidence data to CDC.12 Currently, CDC and NCI cancer surveillance data together are the foundation of national data on cancer trends, including risks in special populations, the design and evaluation of public health interventions, many clinical and etiologic epidemiologic studies, and quality of care evaluation studies.13

In order to understand cancer prognosis and improve control of all types of cancer in the United States, the survival rates for specific cancer sites and subpopulations may be monitored over time. Survival data are critical for evaluating the progress and impact of early detection and screening programs, comprehensive cancer control plans, and interventions from other sources.14 As one available resource, survival data assists in identifying high-risk
population groups, which may lead to more effective and comprehensive program planning.

**Methods**

The relative survival rate is the ratio of the observed survival in a patient cohort divided by the expected survival of a comparable cancer-free group from the general population in the absence of the specific cancer. A data set of invasive cancer cases for diagnosis years 2000–2007 from people ages 20–99 years was created using the NPCR November 2011 data submission file. We excluded data from 23 registries that either did not meet NPCR data quality standards for inclusion in the United States Cancer Statistics or did not conduct active case follow-up or linkage with the National Center for Health Statistics' National Death Index (NDI) through the 2008 death file. In 2006, the NPCR entered into an intra-agency agreement with NDI, allowing registries to link with that database on a regular basis. An additional 4 registries requested their data not be used in survival analyses at this time. The quality and completeness of individual data items used in this analysis were previously conducted and reported elsewhere. Due to difficulties with data collection following hurricanes Katrina and Rita in 2005, we excluded data for July 1 through December 31, 2005 for Alabama, Louisiana, Mississippi, and Texas, according to established practices. Twenty-one NPCR registries were included in the analysis representing 50% of the US population.

Where incomplete, the full date of diagnosis and date of last contact were imputed. Additionally, cases not linking to the state death files and NDI were presumed to be alive and the date of last contact was set to December 31, 2008. Cases with multiple primary cancers were included in the data set, though only the first primary was included in calculating survival rates for all cancer sites combined. Where a patient had multiple primaries of the same site, only the first primary was included in the calculations for that primary site. We excluded cancer cases with unknown age or sex (1,873; 0.03%) and those identified solely on the basis of a death certificate or autopsy (75,253; 1%).

Using state-specific life tables by race (black and white) and sex, provided by the National Center for Health Statistics, we calculated 5-year survival rates for cases diagnosed in 2000–2007 with follow-up through 2008 following a process described by Paul Dickman that includes the Ederer II relative survival methodology. Rates were calculated for all cancer primary sites combined and 23 specific primary sites, grouped by the SEER Site Recode, by survival interval, sex, race, age group, and SEER Summary stage (local, regional, distant). The 23 specific primary site groups were selected due their frequency, screening amenability, and modifiable risk factors. All primary sites combined and the top 10 primary site groups are shown in Table 1. Life tables are not available for races other than whites and blacks. The age groups are defined as <45, 45–54, 55–64, 65–74, and ≥75 years. All analyses were performed using SAS software version 9.3 (SAS Institute, Inc). Five-year relative survival rates are presented in Tables 1, 2 and 3 for all primary cancer sites combined and select specific primary sites by site, sex, age, and stage groups.

**Results**

More than 5 million cancer cases from diagnosis years 2000–2007 were included in this data set, and were evenly distributed between male and female patients. Cases for white patients represented 85% of the total cases with 10% for black patients. The population from the 21 states was more than 300 million, with a similar distribution for the
white and black populations (81% and 12%, respectively; data not shown).

We present 5-year relative survival rates for all sites combined and specific cancer primary sites by sex and race in Table 1. The 5-year relative survival rate for all cancer primary sites combined, all races combined, was 65.0% (65.3% for male and 64.8% for female patients). For all cancers and each specific cancer site, black patients had a lower 5-year relative survival rate than white patients, except lung and bronchus, where the rate for black males was higher than white males and males of all races (17.0%, 14.8%, and 14.7%, respectively). Of the specific cancer sites analyzed, thyroid cancer had the highest relative 5-year survival rate for male and female patients combined (96.4%) and for female patients alone (97.6%). Prostate cancer had the highest relative 5-year survival rate for male patients (98.5%). The lowest rates were seen for lung and bronchus, ranging from 12.5% for black women to 20.4% for white women.

Table 2 shows 5-year relative survival rates by age at diagnosis and for all primary sites combined and the top 4 primary sites. For all cancer primary sites combined, the <45 age group had the highest 5-year relative survival rates, except black males, where the 65–74 age group had the highest rate. The 45–54 age groups had the highest relative survival rates for colon and rectum, except black males, where the 65–74 age group had the highest rate. For lung and bronchus, the <45 age groups also had the highest relative survival rates. Age group 65–74 has the highest relative survival rates for prostate and female breast for both races combined (white and black).

Five-year relative survival rates by cancer stage at diagnosis are presented in Table 3. The highest rates are seen in the localized stage group and lowest rates in the distant stage group. This pattern is seen for all cancers combined and each specific cancer primary site analyzed.

<table>
<thead>
<tr>
<th>Age at Diagnosis, y</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>All sites combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45</td>
<td>76.5</td>
<td>84.2</td>
</tr>
<tr>
<td>45–54</td>
<td>66.7</td>
<td>78.0</td>
</tr>
<tr>
<td>55–64</td>
<td>68.4</td>
<td>69.8</td>
</tr>
<tr>
<td>65–74</td>
<td>67.8</td>
<td>61.2</td>
</tr>
<tr>
<td>≥75</td>
<td>55.7</td>
<td>49.4</td>
</tr>
<tr>
<td>Colon and rectum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45</td>
<td>67.1</td>
<td>71.0</td>
</tr>
<tr>
<td>45–54</td>
<td>69.1</td>
<td>71.2</td>
</tr>
<tr>
<td>55–64</td>
<td>67.7</td>
<td>69.8</td>
</tr>
<tr>
<td>65–74</td>
<td>66.1</td>
<td>67.3</td>
</tr>
<tr>
<td>≥75</td>
<td>57.7</td>
<td>58.3</td>
</tr>
<tr>
<td>Lung and bronchus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45</td>
<td>23.8</td>
<td>29.2</td>
</tr>
<tr>
<td>45–54</td>
<td>17.0</td>
<td>26.0</td>
</tr>
<tr>
<td>55–64</td>
<td>16.5</td>
<td>24.1</td>
</tr>
<tr>
<td>65–74</td>
<td>15.6</td>
<td>21.5</td>
</tr>
<tr>
<td>≥75</td>
<td>11.0</td>
<td>14.4</td>
</tr>
</tbody>
</table>

Table 2. CDC NPCR-CSS 5-Year Relative Survival (Percent) by Age at Diagnosis, SEER Cancer Primary Site Recode,¹ Race, and Sex, 2000–2007

CDC, Centers for Disease Control and Prevention; NPCR-CSS, National Program of Cancer Registries – Cancer Surveillance System.
Black patients had lower 5-year relative survival rates than white patients for all cancers combined, colon and rectum, lung and bronchus (except distant stage), and female breast. The relative survival rates for prostate for black patients were equal to or slightly higher than those seen for white patients.

**Discussion**

This report provides 5-year relative survival rates for the largest proportion of the total US population ever presented. While existing survival estimates cover 14% to 28% of the US population, the NPCR data represents nearly double this population coverage and, when combined with NCI’s data in the future, the 2 programs together have the capability of covering 100% of the nation.

For all cancers combined for 2000–2007, the NPCR 5-year relative survival rate is 65.0%, comparable to that available from the NCI for the same time period, 67.1%, and the Canadian Cancer Registry, 62.0% for 2004–2006. As expected, lower age groups have the highest survival as does localized stage at diagnosis. These trends are also comparable to those reported by NCI.

While SEER data have been a trusted and valued source of cancer survival information for decades, only a few states within the United States have previously been able to monitor their local cancer survival rates, an essential tool in evaluating local cancer control initiatives. The addition of survival estimates from NPCR programs for national estimates will also provide more generalizable data, particularly for certain currently underrepresented minority populations. The NPCR’s expansive geographic coverage will also allow survival rates to be calculated for many rural populations and by geographic region. Although, through exclusions, the current report is not derived from the entire NPCR program, the data reported here provide for the most complete description of the cancer burden in the United States to date. Furthermore, the NPCR data set supports the analysis of rare cancers, such as retinoblastoma or primary fallopian tube cancers. As the number of NPCR registries that link to the NDI increases, the quality and completeness of data on relative survival in the US population will continue to grow as will our understanding of the impact of geographic, racial/ethnic, and medical care variations.

CDC’s NPCR registries currently meet established standards for high quality data and most, but not all, conduct either active case follow-up activities or linkages with the NDI in order to obtain and update vital status information. This means that NPCR data are now in a position where valid and reliable relative survival rates can be produced and published. However, additional work continues so that all NPCR registries will conduct data linkages, update their database following those linkages prior to the NPCR data submission, and allow inclusion of their data in these analyses.

---

**Table 3. CDC NPCR-CSS 5-Year Relative Survival (Percent) by SEER Cancer Primary Site Recode, SEER Summary Stage at Diagnosis, Race, and Sex, 2000–2007**

<table>
<thead>
<tr>
<th>SEER Summary Stage at diagnosis</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>All sites combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized</td>
<td>91.7</td>
<td>91.0</td>
</tr>
<tr>
<td>Regional</td>
<td>57.0</td>
<td>66.5</td>
</tr>
<tr>
<td>Distant</td>
<td>24.8</td>
<td>25.7</td>
</tr>
<tr>
<td>Colon and rectum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized</td>
<td>88.1</td>
<td>88.7</td>
</tr>
<tr>
<td>Regional</td>
<td>67.8</td>
<td>68.5</td>
</tr>
<tr>
<td>Distant</td>
<td>11.2</td>
<td>13.0</td>
</tr>
<tr>
<td>Lung and bronchus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized</td>
<td>45.1</td>
<td>55.5</td>
</tr>
<tr>
<td>Regional</td>
<td>19.6</td>
<td>25.5</td>
</tr>
<tr>
<td>Distant</td>
<td>3.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Regional</td>
<td>98.6</td>
<td></td>
</tr>
<tr>
<td>Distant</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>Female breast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized</td>
<td>98.1</td>
<td></td>
</tr>
<tr>
<td>Regional</td>
<td>84.1</td>
<td></td>
</tr>
<tr>
<td>Distant</td>
<td>25.6</td>
<td></td>
</tr>
</tbody>
</table>

CDC, Centers for Disease Control and Prevention; NPCR-CSS, National Program of Cancer Registries – Cancer Surveillance System.
Conclusion

The NPCR data are a robust data set available for calculating reliable 5-year relative survival rates that are not likely to be affected by small case counts. CDC will be including relative survival rates for a subset of NPCR programs in subsequent editions of the United States Cancer Statistics Web-Based Report. The combination of survival estimates from both NCI and CDC will allow for more complete monitoring of any upcoming Healthy People objectives to increase the percentage of persons with cancer living 5 years or longer after diagnosis.

References

Recent Improvement in Completeness of Incidence Data on Acute Myeloid Leukemia in US Cancer Registries

Anthony P. Polednak, PhD

Abstract: A limitation of data prior to 2010 on incidence of leukemia in US population-based cancer registries is that acute myeloid leukemia (AML) diagnosed as progression (transformation) from a previously diagnosed myelodysplastic syndrome (MDS), myeloproliferative neoplasm (MPN), or chronic myeloid leukemia (CML) was not reportable. Data were used from a research database for the 18 cancer registries of the Surveillance, Epidemiology and End Results (SEER) Program, and from all registries in the US Cancer Statistics (USCS) database. Analyses compared the age-standardized incidence rate (ASIR) per 100,000 for AML before (ie, 2000–2009) vs after (ie, 2010) the new reportability rules for AML. The ASIR for all ages combined fluctuated until increasing from 3.60 (95% CI, 3.47–3.73; N = 3,068) in 2009 to 3.89 (95% CI, 3.76–4.03; N = 3,355) in 2010 in SEER, and from 3.64 (95% CI, 3.58–3.71; N = 11,488) in 2009 to 3.89 (95% CI, 3.82–3.96; N = 12,351) in 2010 in USCS. The increase from 2009 to 2010 was limited to ages 60+ years (from 13.87 to 15.59 in SEER and from 14.13 to 15.34 in USCS). The SEER research database allowed analysis by the number of cancers per person, which showed that the increase in AML cases and rates for age 60+ years from 2009 to 2010 was due to an increase in cases with a previous cancer(s) largely representing newly-reportable post-MDS, post-MPN and post-CML AML cases. Continued surveillance is needed to address the eventual impact of delayed reporting of diagnoses in 2010 on estimates and projections of AML incidence in the US population.

Key words: acute myeloid leukemia, cancer mortality, cancer registries, cancer surveillance, chronic myeloid leukemia, leukemia, myelodysplastic syndrome, myeloproliferative neoplasms

Introduction

A limitation of data on leukemia in all US population-based cancer registries, as recognized in the literature,1,2 is that acute myeloid leukemia (AML) occurring in individuals as transformation (progression) from another myeloid neoplasm was not reportable as a separate primary cancer until diagnoses in 2010. This is important because myelodysplastic syndrome (MDS) and myeloproliferative neoplasm (MPN) are myeloid neoplasms reportable to US cancer registries starting with diagnoses in 2001, and are associated with a risk of progression to leukemia (predominantly AML) estimated at about 5% to >50% (“preleukemic states”) for MDS1 but lower for MPN.4

Post-MDS and post-MPN myeloid AML diagnoses starting in 2010, however, were defined as multiple primaries in the National Cancer Institute’s 2010 Hematopoietic and Lymphoid Neoplasm Case Reportability and Coding Manual and the accompanying Hematopoietic Database for use in coding.5 These resources3 were cited as the basis for new rules on reportability and coding, replacing the February 2001 rules, by the North American Association of Central Cancer Registries (NAACCR).6 Both the National Cancer Institute’s Surveillance, Epidemiology and End Results (SEER) Program and the Centers for Disease Control and Prevention’s National Program of Cancer Registries (NPCR) use NAACCR uniform standards for collection and coding.

As also suggested elsewhere,2 these changes in reportability of leukemia should result in improved completeness of registration of incident AML cases in all US central cancer registries starting with year of diagnosis 2010.2 Detailed analyses relevant to this potential improvement in registration of AML, however, appear to be lacking. The 2013 Annual Report to the Nation on the Status of Cancer included data on the recent trend (2001–2010) in total age-standardized incidence rate for leukemia, using the combined data (covering about 85% of the US population) from registries funded by the SEER Program and/or NPCR.7 Data were not provided, however, for AML as a subgroup of leukemia.7 For AML, a SEER Program report included data on annual age-standardized incidence rates for 9 SEER areas combined for each year from 1975–2010, including rates for males and females separately, but detailed analyses and interpretation of temporal trends (eg, by age category) were not included.7

The incidence data for 2010 (but not for 1975–2009) in that report6 would have included the newly reportable AML cases diagnosed in 2010 after a previously diagnosed myeloid neoplasm (including MDS and MPN) in the same person, but the impact of this change in reporting was not addressed.8

The present report examined recent trends (2000–2010) in AML incidence rates for all SEER areas included in a research database6 that allowed identification of incident cases diagnosed with AML as the second or subsequent reportable cancer. The representativeness of AML incidence rates and trends in SEER is a concern,8 however, and was assessed by comparisons with AML incidence data from a public-use Centers for Disease Control and Prevention

____________________


Address correspondence to Anthony P. Polednak, PhD. Email: appoled7@yahoo.com.
The SEER and USCS databases are based on the same tumor reportability rules, and the same definition of AML based on codes in the International Classification of Diseases for Oncology Version 3 (ICD-O-3), but USCS does not allow some of the detailed analyses performed using the SEER research database. Both the SEER and USCS databases, however, do allow comparison of AML incidence rates for years prior to the adoption of new reportability rule for AML (ie, for diagnoses in 2010) vs earlier years (ie, 2000–2009) for which data were collected under the old rules.

Methods

Databases and Analytic Methods Used

De-identified research databases available from the SEER Program include files with incident cancer diagnoses in 2000–2010 in the 18 SEER registries (SEER-18), along with population estimates and software (SEER*Stat version 8.1.2) for analyses. The 18 registries include 9 with long-term incidence data (the states of Connecticut, Hawaii, Iowa, New Mexico, and Utah, and the metropolitan areas of Detroit, Michigan; Atlanta, Georgia; San Francisco–Oakland, California; and Seattle–Puget Sound, Washington). The 9 others, with incidence data for 2000–2010, are Los Angeles, California; San Jose–Monterey, California; Alaska Native Registry; rural Georgia; Kentucky; Greater California (ie, California excluding the other SEER areas in the state); New Jersey; Louisiana; and Greater Georgia). SEER-18 covers about 28% of the US population.

Using ICD-O-3, the SEER Site and Morphology Recode group for AML is defined by morphology (M) codes 9840, 9861, 9865–9867, 9869, 9871–9874, 9895–9897, 9898, 9910–9911, and 9920. This group excludes sarcoma, myelofibrosis, and certain other rare cancers that have been included with AML in some reports.

Historically, prior to the changes in reportability starting with diagnoses in 2010, SEER rules for multiple hematopoietic primaries involved coding a new primary “if the physician clearly states ... a new primary,” but “if there is no clear information from the physician” then a table of definitions of single vs subsequent primaries was to be used. Malignancies of the lymphoid series (eg, chronic lymphocytic leukemia) and the myeloid series were considered to be different and counted as multiple primary cancers for the same patient, but a myeloid cancer diagnosed after a previous myeloid malignancy “or MDS or myeloproliferative disorder” would not be counted as a subsequent primary. Myeloid leukemia diagnosed after polycythemia vera (PCV), a major subgroup of MPN, was reportable as a separate primary cancer, but not AML diagnosed after any other MPN or after MDS.

The NAACCR hematopoietic and lymphoid neoplasm reportability and coding rules in 2010, referencing 2 resources (the Hematopoietic and Lymphoid Neoplasms Case Reportability Coding Manual and the Hematopoietic Database), replaced the February 2001 rules for multiple primary cancers. The Manual included detailed criteria and highlighted changes from previous rules, with consideration of biopsy documentation and length of time between any 2 diagnoses. An acute (blast) myeloid cancer (eg, AML) diagnosed more than 21 days after a chronic phase myeloid cancer (eg, chronic myeloid leukemia or CML) was regarded as second primary. The accompanying Hematopoietic Database indicates specific transformations as multiple primaries, such as MDS to AML, chronic myeloproliferative disease (CMD) (ie, MPN) to acute leukemia, and essential thrombocytopenia (ET) (a subgroup of MPN) to AML. The Hematopoietic Manual and Database are periodically updated.

The reportability of post-MDS, post-MPN, and post-CML AML cases to SEER should have resulted in an increase in AML (starting in 2010) with a SEER tumor sequence code “02” (ie, second of 2 or more primary tumors) or higher. Therefore, trends in AML incidence rates for SEER-18 were examined separately for AML coded as sequence 00 (no other primary) or 01 (first of 1 or more primaries) combined, vs AML coded as sequence number 02 or greater.

MDS and MPN, included in the SEER Site Recode ICD-O-3/World Health Organization 2008 Group labeled “miscellaneous,” were defined by using ICD-O-3 morphology codes M9980–9989 (MDS) and 9950–9969 (MPN); CML (M9863, 9875–9876, 9945–9946) is a separate SEER Site Recode group. For incident cancers diagnosed in 2000–2010 in the SEER-18 database, the proportion diagnosed at ages 60+ years was: 30,815 (87.7%) of all 35,150 MDS, including 1,607 (4.6% of 35,815) diagnosed at age 60–64 years; 66.7% (14,053/21,102) of MPN cases, including 2,105 (10.0% of 21,102) diagnosed at age 60–64 years; and 58.7% (8,346/14,208) CML cases, including 1,081 (8.0%) for age 60–64 years. Leukemia diagnosed after PCV was already reportable as a new primary (since 2001), and MDS is more common than MPN; also, the risk of leukemia is higher after MDS than MPN. Therefore, the impact of the 2010 reportability rules on AML incidence rates should be most apparent for post MDS AML and for ages 60+ years.

Age-standardized incidence rates (ASIRs) and their 95% CIs for AML were analyzed for all ages (using 19 age groups from <1, and 1–4 through 80–84 years, plus 85+ years), and also within selected age groups including 60+ years, using SEER*Stat.

For the SEER-18 registries, statistical analysis of temporal trends (2000–2010) for AML included the annual percent change (APC), estimated as the slope of the line(s) obtained by fitting regression models (using weighted least squares) to the natural logarithm of each annual rate. APCs were obtained by exporting data files with annual rates and their standard errors obtained from SEER*Stat into a SEER Joinpoint Regression Program (version 4.0.1, January 2013) to identify changes in the magnitude and/or direction of trends over time. The requirements of the program include at least 4 data points (calendar years) between consecutive joinpoints. APCs for each of 2 segments (ie, time periods) are obtained when the model discloses 1 joinpoint (ie, a change in direction or magnitude of the trend). The average annual percent change (AAPC) is a
Summary of the trend for a prespecified number of recent years (5 or 10), even if the joinpoint model indicates changes in trends during those years; it is useful in comparisons (eg, by individual cancer registry) and is estimated as the weighted average of the joinpoint APCs. Statistically significant (P < .05) APCs or AAPCs, with zero as null hypothesis, and P-values were obtained from the joinpoint program (2-tailed tests).

Two ASIRs with CIs that do not overlap may be considered as different (a conservative indicator, not a formal statistical test). The percentage difference or change (PC) between 2 rates may be useful in assessing importance. PC is calculated as [(final rate – initial rate) divided by the initial rate] x 100, and is also useful in examining the magnitude of delayed reporting, which requires data on longer-term incidence trends. The present report, however, compared incidence data from previous SEER submission files (eg, November 2011) to the November 2012 SEER submission, for recent years of diagnosis, to estimate the overall impact of delayed reporting on data for years prior to the 2010.

Assessing the Representativeness of SEER Data: Comparison with USCS Database

Detailed analyses of incidence trends for AML by tumor sequence number, and identification of individual persons with multiple cancers, were limited to the SEER-18 population. Representativeness of SEER-18 incidence data (2000–2010) on AML was assessed, however, by comparing SEER rates with AML incidence rates for the all US registries included in the USCS public information database (1999–2010). The USCS database uses the SEER Site and Morphology Recode groups (eg, for AML, as defined above), but does not allow examination of tumor sequence code or listing of individual records, and does not include software for computing APCs in rates.

NPCR registries had to meet specific data quality criteria for all cancer sites combined, including: estimated completeness of ascertainment of unduplicated cases, using methods developed by NAACCR; missing information on specific sociodemographic variables (eg, age at diagnosis); computerized edits designed by the SEER Program but expanded into NAACCR edits software; and ascertainment by death certificate only (DCO) (ie, 5% or less). Quality-control issues for SEER registries are addressed by SEER contract requirements (eg, DCO rate 1.5% or less as a standard) and by the SEER Quality Control Profile system.

Among all the 51 areas (50 states plus the District of Columbia), USCS data were missing for 1999 for 9, including 2 with missing data for 1999 but not for 2000–2010. The present report used the USCS database for 2000–2010; for this period, data were missing for 1 year for 3 areas, 2 years for 1 area, 3 years for 2 areas, 4 years for 1 area, 7 years for 1 area, and all 11 years for 1 area.

Results

Incidence Rates for AML Using SEER and USCS Databases

Using the SEER-18 database, the ASIR for AML for all ages was 3.81 in 2000 and 3.92 in 2001, declining to 3.43 in 2004 but increasing (Figure 1) from 3.60 (3,068 cases) in 2009 to 3.89 (3,355 cases) in 2010; the increase from 2009 to 2010 was statistically significant (Table 1), and reflected a PC of +9.4% (287/3068) in the number of cases and +8.1% (0.29/3.60) in the rate.

Using the USCS database for diagnoses in 2000–2010, the ASIR for all ages combined (Figure 1) declined from 2000 (3.92; 95% CI, 3.85-4.00; N = 9,721) to 2009 (3.64; 95% CI, 3.58-3.71; N = 11,488) but increased to 3.89 (95% CI, 3.82-3.96; N = 12,351) in 2010; the CIs for the ASIRs in 2009 and 2010 did not overlap, and the PC in number of cases was +7.5% (863/11,488).

By age group, the total number of AML cases diagnosed at all ages in 2000–2010 in the SEER-18 database reached almost 1,000 by age 35–39 years (age-specific rate, 1.43; 95% CI, 1.34-1.52), and cases age 35+ years comprised 28,160 (87.7%) of all 32,125 cases. For SEER, the ASIR for age 35–59 years showed no overall temporal increase (Figure 1) from 2000 (2.53) to 2010 (2.50) (APC, −0.0%; 95% CI, −1.3 to +1.3%) or from 2009 (2.47) to 2010 (2.50). Within this broad age group, there was also no increase in the ASIR for age 50-59 years from 2000 to 2010 or from 2009 to 2010 (data not shown).

In the USCS database, the number of incidence AML cases (2000–2010) for all ages reached >3,000 (3,243) by age 35–59 (age-specific rate, 1.47; 95% CI, 1.42-1.52), and cases age 35+ years comprised 104,519 (88.9%) of all 117,622 cases.
The USCS age-specific rate for age 35–59 years showed little overall increase (Figure 1) from 2000 (2.54; 95% CI, 2.43–2.65) to 2010 (2.67; CI, 2.57–2.77), and USCS rates in each calendar year were similar to those for the subgroup of SEER-18 registries (Figure 1). Within the age group 50–59 years in USCS, the ASIR also did not increase from 2000 to 2010, with only a small increase from 2009 (3.78; 95% CI, 3.59–3.97) to 2010 (4.04; 95% CI, 3.85–4.24).

Additional analyses focused on AML diagnoses at ages 60+ years. Using the SEER database the ASIR for this age group declined from 2000–2009, but increased sharply (Figure 1) from 13.87 (95% CI, 13.24–14.51; N = 1,911) in 2009 to 15.59 (95% CI, 14.94–16.27; N = 2,192) in 2010 (Table 1) or a PC of +12.4% (1.72/13.87) in the rate and +14.7% (281/1911) in cases.

Using the USCS database, the ASIR for AML at age 60+ years (Figure 1, Table 2) increased from 14.13 (95% CI, 13.81–14.46) in 2009 to 15.34 (95% CI, 15.00–15.68) in 2010; the PC was +8.6% (1.21/14.13) for the ASIR, and +9.6% (714/7413) for the number of cases (Table 2).

Within the broad age group of 60+ years, the ASIR increased from 2009 to 2010 for each age group of 60–69, 70–79 and 80+ years for both SEER and USCS (Table 2, Figure 2). The PC was lowest for age 80+ years (Table 2). CIs on rates for SEER and USCS overlapped (Table 2). The PC was larger for SEER than for USCS in each age group, but for both age groups 60–69 and 70–79 years in USCS, the CIs for the rates in 2009 and 2010 did not overlap (Table 2).
but the CIs overlapped and numbers were small (data not shown). Additional analyses focused on the age group 60+ years, using only the SEER database (Table 3).

The increase in the ASIR for age 60+ years in SEER-18 was due to AML cases coded as tumor sequence 02+ (Table 3, Figure 3); the APC for 2008–2010 in joinpoint regression was +27.1% (95% CI, +10.3 to 46.4%, \( P = .006 \)) and the ASIR for 2010 (6.48) had a 95% CI that did not overlap with the CIs for the ASIRs in any previous year (Table 3).

An increase from 2009 to 2010 in the ASIR for age 60+ years coded as sequence 02+ was found for each SEER registry, although the magnitude of the change (PC) varied (Table 4); as in SEER reports,\(^a\) data are tabulated only for the 13 registries with rates based on >15 cases. Rates for 2009 were based on small numbers for many registries, and data for the last 5 years (2006–2010) give a better idea of the recent trend; the AAPC for 2006–2010 was positive for all 13 registries and the CIs did not include zero for 11 (Table 4).

---

### Table 2. Age-Standardized Incidence Rate per 100,000 for 2009 and 2010 for Acute Myeloid Leukemia Diagnosed at Age 60–69, 70–79 and 80+ Years Using Databases from 18 SEER Registries Combined vs All Registries in the United States Cancer Statistics Database, and Percentage Change (PC) in Rate from 2009 to 2010

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Database</th>
<th>2009 No.</th>
<th>Rate (95% CI)</th>
<th>2010 No.</th>
<th>Rate (95% CI)</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>60–69</td>
<td>SEER</td>
<td>597</td>
<td>8.19 (7.55–8.88)</td>
<td>706</td>
<td>9.34 (8.66–10.06)</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>USCS</td>
<td>2,303</td>
<td>8.43 (8.09–8.78)</td>
<td>2,597</td>
<td>9.24 (8.89–9.61)</td>
<td>9.6</td>
</tr>
<tr>
<td>70–79</td>
<td>SEER</td>
<td>659</td>
<td>15.82 (14.64–17.08)</td>
<td>778</td>
<td>18.45 (17.18–19.80)</td>
<td>16.6</td>
</tr>
<tr>
<td></td>
<td>USCS</td>
<td>2,594</td>
<td>16.12 (15.50–16.75)</td>
<td>2,914</td>
<td>18.08 (17.43–18.75)</td>
<td>12.2</td>
</tr>
<tr>
<td>80+</td>
<td>SEER</td>
<td>665</td>
<td>22.84 (21.12–24.66)</td>
<td>708</td>
<td>24.23 (22.48–26.09)</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>USCS</td>
<td>2,516</td>
<td>23.12 (22.22–24.04)</td>
<td>2,616</td>
<td>23.86 (22.96–24.80)</td>
<td>3.2</td>
</tr>
<tr>
<td>60+</td>
<td>SEER</td>
<td>1,911</td>
<td>13.87 (13.24–14.51)</td>
<td>2,192</td>
<td>15.59 (14.94–16.27)</td>
<td>12.4</td>
</tr>
<tr>
<td></td>
<td>USCS</td>
<td>7,413</td>
<td>14.13 (13.81–14.46)</td>
<td>8,127</td>
<td>15.34 (15.00–15.68)</td>
<td>8.6</td>
</tr>
</tbody>
</table>

CI, confidence interval; SEER, Surveillance, Epidemiology and End Results Program.

\(^a\)95% confidence intervals (in parentheses) do not overlap for the rates in 2009 and 2010.

PC: \[\frac{\text{difference between the rate in 2009 and 2010}}{\text{the rate in 2010}}\] times 100.

### Table 3. Age-Standardized Incidence Rate per 100,000 for Acute Myeloid Leukemia Diagnosed at Age 60+ Years in 2000–2010 for 18 SEER Cancer Registries Combined, by Tumor Sequence Number

<table>
<thead>
<tr>
<th>Year</th>
<th>Sequence 00, 01</th>
<th>Sequence 02+</th>
<th>Sequence 02+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Rate (95% CI)</td>
<td>No.</td>
</tr>
<tr>
<td>2002</td>
<td>1,350</td>
<td>11.23 (10.64–11.84)</td>
<td>379</td>
</tr>
<tr>
<td>2003</td>
<td>1,319</td>
<td>10.76 (10.19–11.36)</td>
<td>447</td>
</tr>
<tr>
<td>2005</td>
<td>1,220</td>
<td>9.98 (9.43–10.56)</td>
<td>427</td>
</tr>
<tr>
<td>2006</td>
<td>1,277</td>
<td>9.98 (9.44–10.55)</td>
<td>499</td>
</tr>
<tr>
<td>2007</td>
<td>1,323</td>
<td>10.18 (9.64–10.75)</td>
<td>514</td>
</tr>
<tr>
<td>2008</td>
<td>1,425</td>
<td>10.64 (10.09–11.21)</td>
<td>511</td>
</tr>
<tr>
<td>2009</td>
<td>1,338</td>
<td>9.65 (9.13–10.18)</td>
<td>573</td>
</tr>
</tbody>
</table>

Joinpoint regression analysis

<table>
<thead>
<tr>
<th>Segment</th>
<th>APC (95% CI)</th>
<th>Segment</th>
<th>APC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000–2010</td>
<td>–2.3 (–3.2, –1.4)</td>
<td>2000–2008</td>
<td>+1.9 (–0.1, +3.9)</td>
</tr>
<tr>
<td>2008–2010</td>
<td>+27.1 (10.3, +46.4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval; SEER, Surveillance, Epidemiology and End Results Program.

\(^a\)95% confidence intervals (in parentheses) do not overlap for the rates in 2009 and 2010.

\(^b\)The 95% CI on the APC does not include zero.

APC, annual percent change in rate, from joinpoint regression model. Two segments (ie, time periods) are shown when the model disclosed 1 joinpoint (ie, a change in direction or magnitude of the trend) (see text).
Using the unique de-identified patient identification number in the database, the number of cases increased from 553 in 2009 to 845 in 2010 for these 13 registries combined (Table 4). This increase of 292 cases included 186 with MDS, 28 with MPN (not including PCV), 1 with both MDS and MPN, and 23 with CML, and 1 with both MDS and CML, or a total of 239 newly reportable transformations (81.8% of 292). After including the data for the other 5 registries (not shown in Table 4), there were 254 AML cases that involved newly reportable post-MDS, post-MPN, and/or post-CML transformations in SEER-18, which accounted for 78.2% of the increase from 2009 to 2010 in the number of AML cases (325, using the data in Table 1) age 60+ years with sequence 02+ for SEER-18.

If these 254 cases (transformations) were excluded, the ASIR for AML at age 60+ years for 2010 would have been reduced to 13.80, which is not higher than the ASIR for age 60+ years in 2009 (13.87).

These analyses did not take into account a small number of AML cases that were transformations from myeloid sarcoma (ICD-O-3 M9930), newly reportable in 2010, or MDS/MPN unclassifiable (M9975, not reportable until 2010). There were also some post-MDS, post-MPN and post-CML AML cases coded as diagnosed in 2009 (prior to the 2010 rules); these cases may reflect a 2001 SEER coding rule (“if the physician clearly states... a new primary”), but this could not be assessed from the database available.

The Potential Impact of Delayed Reporting of AML Cases Diagnosed at Age 60+ Years in SEER

For all AML diagnoses at age 60+ years, a comparison of the SEER 2012 submission file vs the 2011 submission file (with incidence data for 2000–2009 for SEER-18) showed increases of only 0–13 cases for each year of diagnosis.

### Table 4. Number and Age-Standardized Incidence Rate per 100,000 for 2009 and 2010 for Acute Myeloid Leukemia Cases Age 60+ Years Coded with Tumor Sequence 02 or Greater, and Percentage Change (PC) in the Rate, by SEER Registry

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater California</td>
<td>119</td>
<td>204</td>
<td>62.6</td>
<td>+4.2 (+0.1, 8.5)</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>67</td>
<td>92</td>
<td>36.0</td>
<td>+15.6 (+0.2, 33.3)</td>
</tr>
<tr>
<td>New Jersey</td>
<td>66</td>
<td>80</td>
<td>25.0</td>
<td>+3.0 (~0.5, +6.6)</td>
</tr>
<tr>
<td>Seattle–Puget Sound</td>
<td>41</td>
<td>67</td>
<td>62.8</td>
<td>+17.8 (~7.0, +49.1)</td>
</tr>
<tr>
<td>Detroit</td>
<td>40</td>
<td>71</td>
<td>77.1</td>
<td>+16.1 (+5.0, 28.3)</td>
</tr>
<tr>
<td>Greater Georgia</td>
<td>35</td>
<td>53</td>
<td>47.5</td>
<td>+6.3 (+13.1, 11.5)</td>
</tr>
<tr>
<td>Louisiana</td>
<td>32</td>
<td>43</td>
<td>31.4</td>
<td>+10.1 (+3.6, 17.0)</td>
</tr>
<tr>
<td>San Francisco–Oakland, California</td>
<td>31</td>
<td>41</td>
<td>25.1</td>
<td>+8.8 (+3.3, 14.5)</td>
</tr>
<tr>
<td>Kentucky</td>
<td>31</td>
<td>56</td>
<td>69.2</td>
<td>+13.8 (+8.9, 18.8)</td>
</tr>
<tr>
<td>Connecticut</td>
<td>29</td>
<td>41</td>
<td>41.0</td>
<td>+19.4 (+8.0, 32.0)</td>
</tr>
<tr>
<td>Iowa</td>
<td>28</td>
<td>43</td>
<td>44.7</td>
<td>+6.2 (+1.7, 10.9)</td>
</tr>
<tr>
<td>Atlanta, Georgia</td>
<td>17</td>
<td>27</td>
<td>65.1</td>
<td>+33.9 (+8.1, 65.9)</td>
</tr>
<tr>
<td>San Jose–Monterey, California</td>
<td>17</td>
<td>27</td>
<td>59.5</td>
<td>+13.7 (+8.0, 19.8)</td>
</tr>
<tr>
<td>Total (13 registries)</td>
<td>553</td>
<td>845</td>
<td>49.4</td>
<td>+13.5 (+6.0, 21.6)</td>
</tr>
</tbody>
</table>

CI, confidence interval; SEER, Surveillance, Epidemiology and End Results Program

*Only SEER registries with at >15 cases for both 2009 and 2010 are shown, ranked from largest to smallest number of cases in 2009. For definitions of SEER areas, see text.7,8

†[(Difference between the rate in 2009 and 2010) divided by the rate in 2010] times 100. All PCs were positive.

AAPP, average annual percent change in rate, 2006–2010 (see text).

95% CIs (in parentheses) for the rates in 2009 and 2010 do not overlap for this registry or group of registries.
from 2000 through 2007, indicating little or no additional delayed reporting between the 2 files. For year of diagnosis 2007, the 2009 submission file was used for the initial routine reporting of incidence data for diagnoses in 2007; the increase in numbers between the 2009 and the 2012 submission files was from 1,637 to 1,837 or a PC of +12.2% (200/1637).

For AML cases diagnosed at age 60+ years coded with sequence 02+, the increase from the 2011 to the 2012 submission files was from 502 to 514 for diagnoses in 2007, with smaller or no increases for each diagnosis year from 2000–2006. For year of diagnosis 2007, the total increase in cases from the 2009 to the 2012 submission files was from 453 to 514 (PC, +13.5%). These analyses provide historical estimates of the magnitude of delayed reporting of AML cases diagnosed at age 60+ years, prior to year of diagnosis 2010.

**Discussion**

Evidence from SEER-18 and USCS databases indicates that the adoption of the 2010 rules for reportability of hematologic cancers has addressed at least part of the problem of underascertainment of AML by US cancer registries. The similarity in ASIRs and in the temporal trends using the SEER database vs all registries in the UCSC database supports the utility of the SEER data for detailed analyses using tumor sequence number. In SEER, the increase from 2009 to 2010 in the ASIR at age 60+ years was due to an increase in AML cases coded as sequence 02+ that largely reflected the newly reportable post-MDS, post-MPN, and post-CML AML cases diagnosed in 2010. A limitation is that those AML cases that had been diagnosed with MDS or MPN before 2001 (when these cancers were made reportable) could not be identified; this is relevant mainly to MPN, in view of earlier average age at diagnosis and longer survival for MPN (vs MDS) patients.

The completeness of ascertainment by US registries of AML cases diagnosed in 2010, however, requires further study. Delayed reporting of AML cases diagnosed at age 60+ years using previous SEER submission files was estimated at about 12%, but the figure for diagnoses in 2010 could prove to be higher (with the advent of new reportability rules) and must be determined from future submission files to SEER (and also to USCS).

Delayed reporting cannot address cancer diagnoses never reported to central cancer registries (ie, underreporting). A study of hematologic cancers linked the SEER-9 database (2000–2005) with Medicare databases that cover most of the US elderly population also included analysis of a 5% sample of all Medicare beneficiaries age 65+ years, whether or not registered with a hematologic cancer in SEER. Using a validated algorithm requiring 2 Medicare claims for AML plus a Medicare record of a blood count and a bone marrow biopsy to define putative AML cases, an estimated 20.9% of all AML cases in Medicare were not registered with any hematologic neoplasm in SEER; pathology reports could not be obtained for verification of the AML diagnosis.

These SEER-Medicare linkage results support the need for additional efforts to improve ascertainment of incident AML cases diagnosed in the population, beyond the improvements resulting from the 2010 hematopoietic reporting rules.

After accounting for delayed reporting of AML cases diagnosed in 2010, individual central cancer registries with relatively small numbers of post-MDS, post-MPN, and post-CML AML cases reported for 2010 may want to consider special quality-control efforts, such as querying physicians about any newly reportable AML cases that may not have been registered. Some registries have explored enhanced hematologic cancer surveillance using electronic billing data from oncology practices, or direct reporting from hematology practices in order to improve physician reporting to the central cancer registry.

Nevertheless, the present findings have implications for estimating the burden of AML in the US population. For all ages combined, 12,351 incident AML cases for 2010 (as cited above) were in the USCS database (which excluded 2 states). The American Cancer Society has projected that 18,860 incident AML cases (36.0% of all 52,860 leukemias) may be diagnosed in the US population in 2014. These projections took into account expected delays in reporting, using incidence data for 1996–2010 submitted to NAACCR. The present findings indicate that AML rates for years prior to 2010 would be underestimated, however, and delayed reporting for 2010 may prove to be greater than in the past. In addition, therapy-related MDS (ICD-O-3 M9987), involving small numbers of incident cases (eg, 66-79 per year from 2007-2010) in the SEER database, was recoded to therapy-related myeloid leukemia (ICD-O-3 M9920, or AML) in the 2014 Hematopoietic and Lymphoid Coding Manual (consistent with World Health Organization classifications). This change will need to be accounted for in future analyses of time trends in AML incidence.

**Conclusions**

The substantial increases in AML incidence rates from 2009 to 2010 (especially at age 60+ years) should be taken into account in comparing AML incidence data over time for the United States and with countries that have excluded post-MDS and post-MPN leukemia as reportable incident cases. Post-MDS and post-MPN AML have been associated with poorer prognosis (vs “de novo” AML) and need to be adequately identified in registries, because temporal changes in the proportion of AML cases that are post-MDS or post-MPN could affect trends in overall survival rates for all AML patients.

Future efforts should include assessing the eventual impact of delayed reporting on AML incidence rates for 2010, to estimate and project the burden of incident AML cases in the US population (especially at ages 60+ years). Completeness of ascertainment of AML and related myeloid neoplasms such as MDS in cancer registries is also relevant to conducting population-based studies of risk factors for myeloid neoplasms.
References


K. Masood; T. Zafar; J. Zafar

Abstract: Background: Due to the unavailability of a common cancer incidence database in Pakistan, the cancer incidence data from nuclear medicine and oncology institutes have been gathered and presented. Methods: The cancer incidence data for the last 27 years (1984–2011) is included to present a data set of male and female patients. The data analysis concerning occurrence, trends of common cancers in male and female patients, stage-wise distribution, and mortality/follow-up cases is also incorporated for the last 7 years (2004–2011). Results and Conclusions: The cancer incidence rates for head and neck (13.41%), brain tumors (10.90%), and non-Hodgkin lymphoma (NHL, 9.70%) were found to be the highest in male patients, whereas breast cancer (45%), ovary tumors (6.6%), and head and neck (6.21%) cancer incidence rates were observed to be the most common in female patients. The age range distribution of diagnosed and treated patients in conjunction with the percentage contribution of cancer-treated patients from different cities of Punjab at the Institute of Nuclear Medicine and Oncology, Lahore is also included. Leukemia has been found the most common cancer for the age group of 1–12 years. It has been identified that the maximum number of diagnosed cases were found in the age range of 51–60 years for male and 41–50 years for female cancer patients.

Key words: breast cancer, cancer diagnosis, cancer occurrence, cancer risk, cancer survival

Introduction

Cancer is a major cause of elevating death tolls across the globe, and even the most developed countries are continuously experiencing a surge in cancer incidence trends. Continuous efforts have been made to alleviate the risk of various cancerous diseases in the Asia-Pacific region. Recently, particular interest has been seen in the detailed study of cancer incidence trends by race/ethnicity, gender, and age group. In Pakistan, an improvement has been experienced in the oncology infrastructure over a period of time, but low cancer survival rates are still prevalent due to lack of awareness about cancer among the population, poor health facilities and infrastructure, social stigmas, difficult economic conditions, and cancer detection at advanced stages. At present, a national level cancer registry database is nonexistent in Pakistan, and the only data available is from hospitals or institutional databases. The significance of cancer incidence is critical in ensuring effective implementation of cancer control programs. To obtain a deeper insight into this cancer epidemic, implementation of breast cancer screening with stress on public health education is a major responsibility of any government. However, at the same time, the projected number of new cancer cases and cancer deaths should be interpreted cautiously, because these estimates are model-based and may vary considerably from year to year for reasons other than changes in cancer occurrence. Cancer risk continues to increase in the developing countries of South Asia, including Pakistan, India, Sri Lanka, Bangladesh, Nepal, and Bhutan, where cancer is presently responsible for about 25% of all deaths. Therefore, the coordination of activities within South Asia must be a high priority for cancer control in this region. There are a number of active registries in the region and population-based data are now available for a considerable number of countries. However, the presentation and availability of data from Pakistan is virtually nonexistent at a global scale. This hinders substantially the epidemiological research into the causes of cancer and the enactment of effective control programs, as the complete cancer incidence data helps in providing valuable information for upgrading the initiation of cancer early diagnostic, prevention, and treatment programs.

According to the population census of 1998, Pakistan has a population of 130,580 million, and approximately 320,000 new cancer cases are diagnosed each year. Among these, roughly 60,324 patients are being treated at Pakistan Atomic Energy Commission (PAEC) cancer institutes. The Institute of Nuclear Medicine and Oncology (INMOL) is one of the PAEC cancer institutes that handles thousands of new cancer cases per year. The institutional data on cancer incidence at INMOL have been registered and properly maintained since 1984, and are frequently used to report on cancer cause statistics, cancer management, and identification of risk factors. The aim of this study was to disseminate cancer incidence data, common cancer trends, and cancer occurrence by age group for male and female cancer patients in a developing country. This could be helpful in understanding the problems and assessing the challenges.

*Institute of Nuclear Medicine and Oncology (INMOL), Lahore, Pakistan. †Department of Economics, Government College University, Lahore, Pakistan. ‡Faculty of Engineering, Department of Electrical Engineering, Government College University Lahore, Pakistan.

Address correspondence to Dr. Junaid Zafar, Government College University, Katchery Road, Lahore, Pakistan.

Email: chairperson.engineering@gcu.edu.pk

Journal of Registry Management 2014 Volume 41 Number 2
that need to be addressed. The patient registration record maintained at INMOL, Lahore has all the key information covering a wide range of demographic and medical information. The demographic information includes age, gender, race/ethnicity, birthplace, and residence, and the medical information includes physical findings, screening information, and diagnostic and follow-up procedures, including annual information about treatment, recurrence, and patient status.

The institutional cancer incidence data includes 4,120 male and female patients. These also include patients treated at INMOL, Lahore from other cities of Punjab (Gujranwala, Sialkot, Kasur, Faisalabad, Gujrat, Okara, Narowal, and Rahim Yar Khan). In 1990, only 1,463 cancer patients were registered at INMOL from the local population. This data set had 665 male and 798 female patients. The number of patients continued to increase every year and reached up to 2,220 patients in 2000 with 975 male and 1,245 female patients. In 2011, a total of 4,087 new cancer patients were registered (1,613 male and 2,474 female patients).

The 10 most common cancers for male and female patients have been identified for the last 7 years (2004–2011) and the trends followed by the most common cancers both in male and female patients have been presented against a time variable. The most common male cancers are head and neck cancers, affecting 13.41% of the diagnosed patients. In female patients, breast cancer is the most common, affecting 45% of the diagnosed female patients, with incidence increasing every year.

The trends suggest that cancers in male and female patients increased up to 64% and up to 98.4%, respectively. The results exhibiting the typical cancer occurrence profile for different age groups have also been elaborated. The trend profile illustrates that cancer incidence among Lahore patients aged 31-40 years is also increasing rapidly. Leukemia has been identified as the most common cancer in patients aged 1–12 years, constituting 21% of total childhood malignancies. In patients aged 15-44 years, melanoma and breast cancer are the most common cancers while breast, melanoma, prostate, and lung cancers are predominant in people over 45 years of age. The data analysis reveals that in the year 2011, the overall cancer occurrence was 37% among the patients aged 20-40 years. In the next section, the institutional setup and data collection and analysis mechanism at INMOL is briefly described.

**Materials and Methods:**

Currently there are 13 nuclear medical centers in PAEC, extending excellent diagnostic and therapeutic facilities all over Pakistan. INMOL has been declared a Center of Excellence in Southeast Asia by the International Atomic Energy Agency (IAEA), where oncologic diagnosis and treatment facilities of various cancerous and noncancerous diseases are being provided using unsealed and sealed radioactive sources, respectively, in nuclear medicine and clinical oncology. INMOL has 80 beds for inpatients and 20 for a chemotherapy day ward. On average, approximately 300 specific chemotherapy drugs are dispensed per week in these wards. In addition, it has leukemia bays that can accommodate 11 acute leukemia patients concurrently. For diagnostic purposes, there are 3 gamma cameras, computed tomography scanner, mammography units, an ultrasound unit, and an X-ray machine at INMOL. A new pediatric hematological malignancy unit with 20 beds and a dedicated clinical laboratory and blood transfusion facilities has also become functional. The commissioning of a 16.5MeV Cyclotron, PET-CT and a complete surgery department have further facilitated the local and surrounding population with the latest diagnostic facilities. Institutes are also conducting research in collaboration with national and international scientific bodies such as IAEA and the World Health Organization. The treatment of cancer is carried out using radiotherapy and chemotherapy modalities. By observing high standards of professional and clinical practice at INMOL, the College of Physicians and Surgeons Pakistan has recognized INMOL as the training center for Fellow of College of Physicians and Surgeons Pakistan (FCPS) students in the fields of nuclear medicine and oncology. INMOL is providing services in nuclear medicine, oncology (medical oncology and radiotherapy), radioimmunoassay, radiology (radiography, mammography, and ultrasonography), medical physics, clinical pathology laboratory, radiopharmacy, and social welfare.

The institute is ISO 9001:2008 certified and it especially emphasizes maintaining and updating all cancer patients’ records for review and data analysis. Records of malignant cancers granted registration numbers have been maintained and identified annually. The patient files and computerized records are simultaneously maintained for future reference. The individual patient records contain all necessary information (name, address, age, gender, contact, origin, disease history, treatment details, etc). The disease classification and details of organ-wise male and female

---

**Figure 1. A Schematic Flow of the Procedural Effort Used to Maintain Data Records of Cancer-Diagnosed Patients**

1. **Demographics & Screening Information**
2. **Dignostic Findings & Cancer Information**
3. **Cancer Therapy & Follow-Up Procedures**
4. **Post Treatment Cancer Surveillance & Calculate Survival Rates**
5. **Assessment Concerning the Efficacy of Treatment Modalities**
cancer patient data are maintained through locally developed computer software (CANSOFT I and CANSOFT II) that contain all necessary information. The privacy for the individual cancer patient’s record is maintained in the use of information on cancer data for statistical analysis and research. The institutional cancer statistics are compiled from the yearly data set summary records of cancer patients, hospital inpatient records, and other statistical collections such as diagnostic statistics. The clinical oncology protocol books contain the organ-specific cancer details and records stored at INMOL are protected by PAEC rules. The cancer incidence record has been maintained by employing the procedure highlighted in Figure 1.

Results and Discussion

Many cancers are initially recognized either by their symptoms or through screening. Although a number of cancers share risk factors, most cancers have a unique set of risk factors. The risk of death due to particular cancers may be reduced through engaging intensive monitoring of individuals at high risk, reducing external risk factors, detecting and treating cancers early in their development, and treating them in accordance with the best available evidence.

The averages of 10 common cancers in male patients for the 2004–2011 are described in Figure 2. The most common cancers are head and neck with 13.41%, followed by brain (10.90%), NHL (9.70%), metastasis of unknown origin (MUO, 8.50%), lung (7.10%), and leukemia (6.30%). The other common cancers include prostate and urinary bladder cancers with a contribution of 6.0% and 5.60%, respectively. The other common cancers are distributed among larynx, acute lymphocytic leukemia (ALL), bone, and miscellaneous cancer diseases; together these account for 32% of the remaining percentage. These percentages are distributed among larynx cancer (4.30%), ALL (4.0%), and bone cancer (3.0%). Other miscellaneous types of cancers account for the remaining 20%.

Figure 2 also illustrates the last 7-year averages of 10 common cancers in female patients in the population of Lahore and the surrounding area treated at INMOL. The most common cancer in female patients is breast cancer, with an incidence of 45% and increasing every year. The other common registered cancers after breast are ovary (6.6%), head and neck (6.21%), cervix (5.63%), and MUO (5.7%). The rest of the cancers percentages are distributed among uterine cancer (4.2%), brain (3.8%), thyroid (3.2%), NHL (2.9%), tongue (1.8%), and other miscellaneous types of cancer (17%). In contrast to western data of female patients with lung cancer, it is not present in the 10 most common diseases among our female population; it is, however, the fourth most common cancer disease among males.

The cancer trends for male and female patients over the last 27 years and the proportion of cancer incidence data attributed to male and female cancer patients is illustrated in Figure 3. An overall cancer incidence rate of 7.5% has been noted during the last 10 years. The total number of cancer patients with staging at the time of prognosis is illustrated in Table 1. Admission/readmission data for male and female cancer patients are presented in Figures 4 and 5.

The last 7-year trend of the 3 most common male cancers is illustrated in Figure 6. The most common among male patients, head and neck cancer, shows a slowly increasing trend, whereas NHL has become the second most common disease with an elevating pattern. Similarly, brain cancer has shown an elevating trend among the male
Table 1. Cancer Staging of Patients at Time of Prognosis, Institute of Nuclear Medicine and Oncology, 1984–2011

<table>
<thead>
<tr>
<th>Year</th>
<th>No. cancer patients</th>
<th>Stage 0 Carcinoma in situ</th>
<th>Stage II Localized</th>
<th>Stage III Locally advanced</th>
<th>Stage IV Late locally advanced</th>
<th>Stage V Metastasized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>95</td>
<td>6</td>
<td>11</td>
<td>23</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>1985</td>
<td>407</td>
<td>21</td>
<td>52</td>
<td>113</td>
<td>144</td>
<td>77</td>
</tr>
<tr>
<td>1986</td>
<td>637</td>
<td>37</td>
<td>103</td>
<td>242</td>
<td>169</td>
<td>86</td>
</tr>
<tr>
<td>1987</td>
<td>762</td>
<td>41</td>
<td>117</td>
<td>193</td>
<td>218</td>
<td>193</td>
</tr>
<tr>
<td>1988</td>
<td>984</td>
<td>46</td>
<td>234</td>
<td>357</td>
<td>315</td>
<td>272</td>
</tr>
<tr>
<td>1989</td>
<td>1,224</td>
<td>51</td>
<td>174</td>
<td>406</td>
<td>291</td>
<td>302</td>
</tr>
<tr>
<td>1990</td>
<td>1,453</td>
<td>67</td>
<td>196</td>
<td>568</td>
<td>419</td>
<td>203</td>
</tr>
<tr>
<td>1991</td>
<td>1,595</td>
<td>79</td>
<td>237</td>
<td>724</td>
<td>338</td>
<td>217</td>
</tr>
<tr>
<td>1992</td>
<td>1,821</td>
<td>72</td>
<td>256</td>
<td>689</td>
<td>389</td>
<td>415</td>
</tr>
<tr>
<td>1993</td>
<td>1,872</td>
<td>64</td>
<td>278</td>
<td>574</td>
<td>444</td>
<td>512</td>
</tr>
<tr>
<td>1994</td>
<td>2,019</td>
<td>83</td>
<td>352</td>
<td>801</td>
<td>389</td>
<td>392</td>
</tr>
<tr>
<td>1995</td>
<td>2,040</td>
<td>65</td>
<td>297</td>
<td>762</td>
<td>492</td>
<td>424</td>
</tr>
<tr>
<td>1996</td>
<td>2,060</td>
<td>77</td>
<td>209</td>
<td>678</td>
<td>467</td>
<td>629</td>
</tr>
<tr>
<td>1997</td>
<td>1,921</td>
<td>81</td>
<td>231</td>
<td>534</td>
<td>579</td>
<td>496</td>
</tr>
<tr>
<td>1998</td>
<td>2,029</td>
<td>72</td>
<td>343</td>
<td>609</td>
<td>393</td>
<td>612</td>
</tr>
<tr>
<td>1999</td>
<td>2,054</td>
<td>88</td>
<td>405</td>
<td>422</td>
<td>638</td>
<td>501</td>
</tr>
<tr>
<td>2000</td>
<td>2,220</td>
<td>92</td>
<td>337</td>
<td>716</td>
<td>586</td>
<td>489</td>
</tr>
<tr>
<td>2001</td>
<td>2,260</td>
<td>69</td>
<td>373</td>
<td>423</td>
<td>538</td>
<td>857</td>
</tr>
<tr>
<td>2002</td>
<td>2,380</td>
<td>102</td>
<td>356</td>
<td>577</td>
<td>408</td>
<td>937</td>
</tr>
<tr>
<td>2003</td>
<td>2,619</td>
<td>94</td>
<td>503</td>
<td>468</td>
<td>547</td>
<td>1,007</td>
</tr>
<tr>
<td>2004</td>
<td>2,752</td>
<td>117</td>
<td>532</td>
<td>471</td>
<td>434</td>
<td>1,198</td>
</tr>
<tr>
<td>2005</td>
<td>2,972</td>
<td>109</td>
<td>398</td>
<td>759</td>
<td>592</td>
<td>1,114</td>
</tr>
<tr>
<td>2006</td>
<td>3,137</td>
<td>119</td>
<td>585</td>
<td>643</td>
<td>756</td>
<td>1,034</td>
</tr>
<tr>
<td>2007</td>
<td>3,516</td>
<td>146</td>
<td>672</td>
<td>821</td>
<td>549</td>
<td>1,328</td>
</tr>
<tr>
<td>2008</td>
<td>3,647</td>
<td>133</td>
<td>497</td>
<td>732</td>
<td>1,098</td>
<td>1,187</td>
</tr>
<tr>
<td>2009</td>
<td>3,800</td>
<td>114</td>
<td>563</td>
<td>697</td>
<td>1,287</td>
<td>1,139</td>
</tr>
<tr>
<td>2010</td>
<td>3,940</td>
<td>147</td>
<td>658</td>
<td>549</td>
<td>1,183</td>
<td>1,403</td>
</tr>
<tr>
<td>2011</td>
<td>4,087</td>
<td>129</td>
<td>498</td>
<td>777</td>
<td>995</td>
<td>1,688</td>
</tr>
</tbody>
</table>

population during the last 7 years. The profile of the 3 most common female cancers is also shown in Figure 6, indicating that breast cancer is continuing to increase among the female population. The second and third most common cancers found were head and neck cancer and ovarian cancer. For the sake of clarity, the cancer incidence and mortality/follow-up cases are reported for the years 2007–2011 in Figures 7 and 8. The age-group cancer occurrence data is shown in Figure 9. Leukemia is the most common cancer in patients aged 1–12 years, contributing to 21% of total childhood malignancies. The incidence of leukemia has been reported in patients who are exposed to radiation, exposed to benzene, pesticides, or herbicides. Breast cancer is dominant in female patients, contributing to about 45% of total cancer diagnoses in female patients. The exact causes of breast cancer are not known, but research has shown that women with certain risk factors are more likely to develop breast cancer. Figure 9 illustrates that cancer incidence among the young population of Lahore between the ages of 41–50 years is also increasing rapidly. Breast cancer is mostly seen among women between the ages of 41–60 years and can be controlled by an early diagnosis (all women over the age of 40 years should have mammography every year). In 2011, the average age for the maximum number of diagnosed cases in males was 51–60 years, followed by 61–70 years. The average age for the maximum number of diagnosed cases in females was 41–50 years, followed by 31–40 years.
Figure 4. The Admission and Readmission of Male Cancer Patients at the Institute of Nuclear Medicine and Oncology, 2011

Figure 5. The Admission and Readmission of Female Cancer Patients at the Institute of Nuclear Medicine and Oncology, 2011

Figure 6. Trends Followed by the 3 Most Common Cancers in Male and Female Patients at the Institute of Nuclear Medicine and Oncology, 2004–2011

Figure 7. Cancer Incidence and Mortality Rates at the Institute of Nuclear Medicine and Oncology, 2007–2011

Figure 8. Cancer Incidence and Details of Follow-up Cases at the Institute of Nuclear Medicine and Oncology, 2007–2011

Figure 9. The Percentage Distribution of Various Cancers by Age Group at the Institute of Nuclear Medicine and Oncology, 2004–2011

Figure 10. The Percentage of Patients Treated at the Institute of Nuclear Medicine and Oncology, Lahore Compared to Other Cities of Punjab, 2004–2011
A typical cancer occurrence profile of the local population is presented in Figure 10. Almost 25.2% of patients from other remote cities have been treated at INMOL. As illustrated in Figure 10, INMOL is serving a massive population and the percentage of treated patients here is far greater than other cities of Punjab. The figure reveals that over 70% of cases are being treated at INMOL, and this can be attributed to lack of diagnosis and treatment facilities in other cities of the province. This is also relevant in terms of increase in death toll due to nonexistent cancer detection practices at early stages, and lack of awareness in rural areas and other cities of Pakistan.

Conclusions
The cancer registry incidence data has been presented to share the information and challenges to be addressed in third world countries like Pakistan. This may help in devising effective cancer control strategies for trends shown by common prevalent cancers by focusing energies in the right direction to detect cancer at early stages. For cancer alleviation, there is also a dire need to organize public awareness workshops and training sessions to assist the oncologists in cancer treatment practices in Pakistan, by encompassing elements of surgical, medical, and radiation oncology. In assessing these trends, it is important to recognize that small changes seen in the most common cancers mean a substantial shift in the numbers of new cases, whereas the same shift in less common cancers can have a relatively small impact. The data reveal that risk of cancer incidence increases with age.

References
Special Feature

Raising the Bar: How to Strut Your Cancer Registry Stuff

Michele Webb, CTR

“I know you are a cancer registrar, but what is it, exactly, that you do?” The question was sincere, but I found myself at a loss for words to describe my work and how it could add value to the research and publication activities for the hospital’s new thoracic surgeon and cancer committee chair. After a few awkward moments, I finally managed to find my voice and ended the conversation on a positive note by promising my help in delivering high-quality data and analysis that would support his research and publication needs. It was obvious that day that I needed to quickly find a compelling way to “strut my stuff” as a cancer registrar.

Perception disconnect happens everywhere. You may not realize how it can hold you back until you miss a perfect opportunity because someone has a misperception about you or your work. Perhaps you have experienced the frustration and disappointment that comes from learning that a colleague, physician, or administrator did not seek you out because he or she misunderstood your passion or vision, didn’t know you had the skills or background to perform a special task, or had been given misinformation about you or your abilities.

From time to time, we need to recalibrate what we do and where we want to be by adding new opportunities that extend the boundaries of our current knowledge and the services we provide in our registries. Here are 3 steps you should take to align your public persona with new opportunities to demonstrate your extraordinary potential.

Step #1: Add a sentence to your profile that describes what you are passionate about.

Add this line to your biography, email signature, publicly available cancer registry websites, your Facebook, LinkedIn or Twitter profiles, and your hospital’s website. For example, not long ago I created a new tagline for several cancer registry websites that said, “Reinventing excellence in cancer data management.” The tagline is powerful and thought provoking because it implies not only that we are already doing good work, but that we are passionate about raising the bar to a higher level of service and performance.

Step #2: Share your passion in separate conversations in your hospital or cancer registry community at least 3 times each week.

Don’t be obnoxious or boastful, but find fun and creative ways to bring your tagline into conversation so it adds value and lets people know who you are and what you really care about. There’s a term in marketing I like: “Rinse and repeat.” Use this in Step #2 by starting the process over each week with 3 new conversations so others learn about what is important to you and motivates you to do your best work. This will anchor your purpose and vision not only in your own mind and actions, but also in the minds of those you work with and serve. You may be surprised at how new opportunities to strut your stuff come your way when people know more about your passion and vision as a registrar.

Step #3: Gather at least 10 interesting facts about your work and passion.

They may include surprising data points, funny stories, personal stories or breakthroughs, or challenges you are working to solve in your registry. You can write these down on a 3x5 index card and keep in your pocket or on your smartphone. Use the facts as talking points to pique people’s curiosity and make them want more. Using your facts in this way gives other people just enough insight to see you in a different light and to seek you out for a quick cup of coffee or to ask for your help on important projects.

How you perceive yourself and your value as a cancer registrar is important to your well-being. But how others perceive you will determine the opportunities given to you at work and in life. I personally know this to be true as I would not be writing this column each quarter had I not stepped out of my comfort zone and begun sharing my passion for cancer registry over a decade ago.

As you create your own taglines and 10 interesting facts to illustrate your vision and passion for cancer registry, I hope you will share them with your colleagues and me. Chris Widener said it so well, “The world needs people like you to dream of something great and then pursue it with all your heart.”

Michele is a cancer registry speaker, educator, coach, and independent contractor living in Rancho Cucamonga, California. She is the founder of www.CancerRegistrar.com, www.RegistryMindset.com, and http://MicheleWebb.com, offering cancer registry leadership, mentoring, and continuing education opportunities. Your comments are welcomed by email to michele@michelewebb.com.
Concerns around new or revised edits affecting large numbers of cases were identified during the 2013 annual calls for data for the central cancer registries. As a result, the North American Association of Central Cancer Registries (NAACCR) Edits Workgroup, which includes representation from the Commission on Cancer, Centers for Disease Control and Prevention, National Cancer Institute, and several central cancer registries, formed the Edits Impact Workgroup in 2013. The directive for this smaller group is to review, discuss, and disseminate information about the impact of new or revised edits.

Issues are usually identified at the central cancer registry level first and brought to the group’s attention. A critical item addressed is whether the new or revised edits are having a significant impact on the cancer registry. This is especially true if the changes cause previously cleared cases to now fail edits. The workgroup established guidelines to resolve these issues, disseminate information as soon as possible, identify default values wherever possible, identify appropriate diagnosis years for implementation of changes, and identify the potential impact of the changes on the registries.

To disseminate information based on these guidelines, the group created Edit Impact Statements. Those statements describe the purpose, description, impact, and resolution as well as the impact on stage derivation (where appropriate). The recommended resolution includes a manual process as well as a global fix process. The manual process is the preferred method, while the global fix process is intended for central cancer registries only. In some instances, a global fix may not be enough to correct every affected case and further instructions are provided when appropriate.

Many issues are being discussed by the Edits Impact Workgroup such as the January 2014 changes in the hematopoietic diseases. The group is currently developing a set of Edit Impact Statements focusing on these diseases and the associated updated or revised edits planned for the NAACCR v15 Edits metfile. It is anticipated that the edit impact statements, including global changes where applicable, will ensure the changes are as smooth as possible.

The first set of Edits Impact Statements, primarily relating to Collaborative Stage (CS) edits, was distributed in September 2013. As a reminder, a summary of those statements is provided here.

**Edit Name:** CS SSF 1, SSF 2, SSF 15, SSF 16, Breast (CS)

**Purpose of Edit:** This edit verifies that for cases using the Breast schema, SSF 1, SSF 2, SSF 15, and SSF 16 are coded consistently.

**Description of Edit:** ER, PR, and HER2 results are coded in the first, second, and third digits (respectively). Based on CS SSF 1, 2, and 15, the edit expects SSF 16 to be a particular value.

**Impact of Edit:** This edit was added to the NAACCR v13A metafile and updated in the v14 metafile to include only those cases with a diagnosis date of 1/1/2010 and forward.

**Resolution of Edit:** Depending on the number of edit failures and available resources, the cases should be manually reviewed and corrected, or a global fix could be instituted (Table 1). Only central cancer registries should apply the global fix.

**Impact on Stage Derivation:** Changing these data items will not affect stage. The CS algorithm will not need to be rerun due to these changes.

**Edit Name:** Lymph-vascular Invasion, Penis and Testis (CS)

**Purpose of Edit:** Lymph-vascular invasion (LVI) is necessary for American Joint Committee on Cancer (AJCC) staging of Penis and Testis cases. This edit was developed to ensure correct coding so that the correct stage could be derived.

**Description of Edit:** This edit was introduced in the NAACCR v13A metafile. For the CS schemas Penis and Testis, the case will fail the edit if LVI is coded to 8 (not applicable).

**Impact of Edit:** Cancers of the affected primary sites are relatively rare. In the National Cancer Data Base, for 2011 diagnoses, 276 of 5,082 (~5%) testis cases have code “8”, and 161 of 1,227 (~13%) of penis cases.

<table>
<thead>
<tr>
<th>SSF 1</th>
<th>SSF 2</th>
<th>SSF 15</th>
<th>SSF 16</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>20</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>10</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>20</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>988</td>
<td>988</td>
<td>988</td>
<td>988</td>
<td>Code 988 is not appropriate for these variables for diagnosis years 2010 and later.</td>
</tr>
</tbody>
</table>

Note: If no combinations exist in the table, populate SSF 16 with code “999".
Resolution of Edit: Depending on the number of edit failures and available resources, the cases should be manually reviewed and corrected, or a global fix could be instituted. Change all cases with a code “8” to code “9”. Only central cancer registries should apply the global fix.

Impact on Stage Derivation: Changing these data items could potentially affect stage. The CS algorithm will need to be rerun due to these changes.

Edit Names: CS Mets at DX-Bone, CS Mets at DX (CS); CS Mets at DX-Brain, CS Mets at DX (CS); CS Mets at DX-Liver, CS Mets at DX (CS); and CS Mets at DX-Lung, CS Mets at DX (CS)

Purpose of Edits: Logic was added to the edits for coding CS Mets-Bone, Brain, Liver, and Lung when the Primary Site is not coded to “C809”. If CS Mets at DX is coded to “98” (not applicable) and the Primary Site is not coded to “C809” (unknown primary site), then CS Mets at DX-Bone/Brain/Liver/Lung must be coded to “8” (not applicable).

Impact of Edits: These edits were added to the NAACCR v13A metafile. Approximately 10%-15% of cases evaluated were miscoded to code “9” (unknown).

Resolution of Edits: Depending on the number of edit failures and available resources, the cases should be manually reviewed and corrected, or a global fix could be instituted. For those cases where the primary site is not coded to “C809” and CS Mets at DX is coded to “98”, then CS Mets at DX for Bone, Brain, Liver, and Lung must be coded to “8”. Only central cancer registries should apply the global fix.

Impact on Stage Derivation: Changing these data items will not affect stage. The CS algorithm will not need to be rerun due to these changes.

Edit Name: CS TS/Ext Eval, Surgery, Bladder Schema (CS)

Purpose of Edit: The edit checks to make sure that bladder cancer cases indicating surgical procedures less extensive than partial cystectomy (eg, transurethral resections (TURBTs)) do not have CS Tumor Size Ext/Eval codes that would meet the criteria for pathologic staging. According to the AJCC staging manual, pathologic staging requires at least partial cystectomy.

Description of Edit: Where the Surgery of Primary Site is coded “10” through “27” (invasive procedures less extensive than partial cystectomy) and the Extension data item maps to “T4b” (codes “700” through “805”), the case will fail the edit if the Eval data item is coded to “5”, “6”, or “8” (consistent with pathologic staging). According to the AJCC staging manual, TURBT procedures do not meet the criteria for pathologic staging. These cases should have an Eval data item coded as “1” (clinical). Also, if the Eval data item is coded to “3” and the Surgery of Primary Site is coded “10” through “27”, regardless of the extension code, the case will fail the edit.

Impact of Edit: As many as 20% of bladder tumor cases may have been miscoded as meeting criteria for pathologic staging based on TURBTs.

Resolution of Edit: Depending on the number of edit failures and available resources, the cases should be manually reviewed and corrected, or a global fix could be instituted. If, after review, the surgery is coded correctly, the CS TS/Ext Eval data item should be recoded to “1”. Only central cancer registries should apply the global fix.

Impact on Stage Derivation: Changing these data items could potentially affect stage. The CS algorithm will need to be rerun due to these changes.

Edit Name: Primary Site, Hemato Morphology, Date of Dx (SEER)

Purpose of Edit: This edit ensures that the Primary Site rules documented in the Hematopoietic Database and Manual are followed. This edit is effective for all cases with a diagnosis date of 1/1/2010 and forward.

Description of Edit: The edit enforces the primary site for the histologies shown in Table 2. An override is possible if documentation supports another primary site.

Impact of Edit: From Center for Disease Control and Prevention’s (CDC’s) National Program of Cancer Registries – Cancer Surveillance System 2013 data submission, for diagnosis year 2010, there were 235 of 37,350 (~3%) Myelodysplastic Syndromes cases failing this edit. There were 240 of 8,538 (~3%) Chronic Myeloproliferative Diseases cases and 147 of 13,866 (~3%) Myelodysplastic Syndromes cases failing this edit.

Resolution of Edit: Depending on the number of edit failures and available resources, the cases should be manually reviewed and corrected, or a global fix could be instituted. After verifying the histology code, change the primary site code to “C42.1”. A change in histology may result in a change to the schema which, in turn, changes other CS fields. Only central cancer registries should apply the global fix.

Impact on Stage Derivation: Changing these data items will affect stage if the histology code is changed to a histology that is not in the HemeRetic schema. For those cases, the CS algorithm will need to be rerun.

Edit Name: CS SSF 10, SSF 11, Breast (CS)

Purpose of Edit: This edit verifies that for cases using the Breast schema, SSF 10 (HER2: FISH Lab Value) and SSF 9989, 9980, 9982-9983, 9985-9987, 9950, 9960-9964, 9975

9733, 9800, 9801, 9805-9809, 9820, 9826, 9831-9836, 9840, 9860, 9861, 9863, 9865-9867, 9869-9876, 9891, 9895-9898, 9910, 9911, 9920, 9931, 9940, 9945, 9946, 9948

9950, 9960-9964, 9975

9980, 9982-9983, 9985-9987, 9989, 9991-9992

*An override is possible if documentation supports another primary site. See the 2012 Hematopoietic Database for more information: http://www.seer.cancer.gov/seertools/hemelymph/
11 (HER2: FISH Test Interpretation) are coded consistently.

**Description of Edit:** For cases with a SSF 11 coded as “020” (negative), this edit enforces that SSF 10 must be coded “500”, “991”, or “997”. See Table 3 for further information regarding consistency between SSF 10 and SSF 11 codes.

**Impact of Edit:** In 1 large state central cancer registry, for diagnosis years 2004 and later and CS version 02.04, there were approximately 2,500 abstract-level records (~1.8%) and 1,400 consolidated records (~1.4%) where the CS SSF 11 and CS SSF 10 codes were inconsistent.

**Resolution of Edit:** Depending on the number of edit failures and available resources, the cases should be manually reviewed and corrected, or a global fix could be instituted. After verifying the appropriate value was recorded for CS SSF 11, change CS SSF 10. Only central cancer registries should apply the global fix.

**Impact on Stage Derivation:** Changing these data items will not affect stage. The CS algorithm will not need to be rerun due to these changes.

<table>
<thead>
<tr>
<th>SSF 10</th>
<th>SSF 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>998 (Test not done)</td>
<td>Must = 998 (Test not done)</td>
</tr>
<tr>
<td>991 (Ratio &lt; 1.00)</td>
<td>Must <strong>NOT</strong> = 010 (positive/elevated; amplified)</td>
</tr>
<tr>
<td>&gt;500 &lt;981</td>
<td>Must <strong>NOT</strong> = 020 (negative/normal; within normal limits; not amplified)</td>
</tr>
<tr>
<td>500, 991, or 997</td>
<td>20</td>
</tr>
</tbody>
</table>
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
CODING COMPLETENESS AND QUALITY OF RELATIVE SURVIVAL-RELATED VARIABLES IN THE NATIONAL PROGRAM OF CANCER REGISTRIES CANCER SURVEILLANCE SYSTEM, 1995–2008

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

After reading this article and taking the quiz, the participants will be able to:
- Identify why calculating accurate estimates of cancer survival is important for various analyses of cancer patient care and prognosis
- Explain how current US survival rates are estimated
- Discuss how using data from the National Program of Cancer Registries (NPCR) could improve accuracy of cancer survival data

1. This study conducted an analysis of specified variables, showing trends by:
   a) stage of disease at diagnosis
   b) stage of disease at death
   c) state of residence at diagnosis
   d) state of residence at death

2. Mean overall percent coding completeness by each NPCR central cancer registry averaged across all data elements and diagnosis years ranged from approximately:
   a) 68% to 76%
   b) 76% to 84%
   c) 84% to 92%
   d) 92% to 100%

3. Statistically significant differences in coding completeness were found in:
   a) American Joint Committee on Cancer (AJCC) edition number
   b) cause of death
   c) Collaborative Stage
   d) date of first contact

4. Which of the following may provide the opportunity to calculate relative survival that may be more generalizable to the US population?
   a) National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) program
   b) Commission on Cancer's (CoC's) National Cancer Data Base (NCDB)
   c) NPCR
   d) Centers for Disease Control and Prevention (CDC)

5. The estimation of cancer survival is an important part of assessing the:
   a) overall strength of cancer care
   b) cost-benefit ratio of treatment regimens
   c) need for survivorship programs
   d) accuracy of stage at diagnosis

6. Population-based cancer relative survival rates are important for medical and public health efforts, such as:
   a) monitoring the impact of insurance status
   b) monitoring the impact of early detection programs
   c) measuring the cost of treatment
   d) measuring the benefits of CoC accreditation

7. NPCR covers approximately what percentage of the US population?
   a) 28%
   b) 36%
   c) 78%
   d) 96%

8. According to Figure 2, Mean Percent Coding Quality of Known Value for Relative Survival Data Elements Averaged over All Diagnosis Years (1995–2008) and over All NPCR Central Cancer Registries (N = 46) Data, the variable with the lowest coding quality percent is:
   a) cause of death (ICD v.10)
   b) cause of death (ICD v.7-9)
   c) follow-up source (1995–2005)
   d) follow-up source (2006–2008)

9. According to the study authors, competing priorities such as _________ can affect completeness of coding.
   a) linkages with the National Death Index (NDI)
   b) linkages with the Indian Health Service (IHS)
   c) implementation of Collaborative Stage
   d) implementation of SEER Summary Stage 2000

10. One of the strengths of the NPCR data is the ability to calculate relative survival by:
    a) age and tobacco use
    b) height and weight
    c) religion and marital status
    d) race and ethnicity

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.

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

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

This original quiz answer sheet will not be graded, no CE credit will be awarded, and the processing fee will be forfeited unless postmarked by: September 30, 2016

Quiz Identification Number: 4102.01

JRM Quiz Article: CODING COMPLETENESS AND QUALITY OF RELATIVE SURVIVAL-RELATED VARIABLES IN THE NATIONAL PROGRAM OF CANCER REGISTRIES CANCER SURVEILLANCE SYSTEM, 1995–2008

☐ Processing Fee: Member $25 Nonmember $35

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

Please check one:

☐ Enclosed is check #_____________ (payable to NCRA)

☐ Charge to the following card:

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

Card Number ___________________ Exp. Date ______

Signature ________________________________

Print Cardholder’s Name ________________________________

Telephone ________________________________

Mail to: NCRA Executive Office
JRM CE Quiz
1330 Braddock Place
#520
Alexandria, VA 22314
National Cancer Registrars Association
CALL FOR PAPERS

Vicki G. Nelson, MPH, RHIT, CTR | EDITOR-IN-CHIEF, JRM

The *Journal of Registry Management*, official journal of the National Cancer Registrars Association (NCRA), announces a call for original manuscripts on registry methodology or research findings related to the 5 subjects listed below and related topics.

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

Contributed manuscripts are peer-reviewed prior to publication.

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


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](http://www.ncra-usa.org/jrm).
INFORMATION FOR AUTHORS

Journal of Registry Management (JRM), the official journal of the National Cancer Registrars Association, invites submission of original manuscripts on topics related to management of disease registries and the collection, management, and use of cancer, trauma, AIDS, and other disease registry data. Reprinting of previously published material will be considered for publication only when it is of special and immediate interest to the readership. JRM encourages authorship by Certified Tumor Registrars (CTRs); special value is placed on manuscripts with CTR collaboration and publication of articles or texts related to the registry profession. CTR continuing education (CE) credits are awarded; a published chapter or full textbook article equals 5 CE hours. Other published articles or documents equal CE hours. All correspondence and manuscripts should be addressed to the Editor-in-Chief, Vicki Nelson, MPH, RHIT, CTR at: izo4@cdc.gov, or at: CDC/NCCDPHP/DCCPC/CB, 4770 Buford Drive, MS K-53, Atlanta, GA 30341-3717, 770-488-6490 (office), 770-488-4799 (fax).

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

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

Submission Requirements

Manuscripts. The terms manuscripts, articles, and papers are used synonymously herein. E-mail only submission of manuscripts is encouraged. If not feasible, submit the original manuscript and 4 copies to the Editor. Manuscripts should be double-spaced on white 8-1/2” x 11” paper, with margins of at least 1 inch. Use only letter-quality printers; poor quality copies will not be considered. Number the manuscripts consecutively with the (first) title page as page one, followed by the abstract, text, references, and visuals. The accompanying cover letter should include the name, mailing address, e-mail address, and telephone number of the corresponding author. For electronic submission, files should be 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, Discussion), or for expository 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 e-mailed. Each visual must have a brief, self-explanatory title. Submit each visual on a separately numbered page at the end of the manuscript, following the references.

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

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


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.
Revised and Improved!


The new edition has been revised to reflect the updates to the 2014 exam. Five case studies have been added with answers and rationales provided. Companion CD contains additional study resources. NCRA member price: $85; non-member price: $105.

Go to www.ncra-usa.org/CTRPrep to order.

Special Compilation of Collaborative Stage Data Collection System Coding Instructions
(For 2014 CTR Exam Candidates Only)

NCRA has compiled the sections from the Collaborative Stage Data Collection System Coding Instructions (Version 02.04) that can be used in the open-book portion of the 2014 CTR Exam. This easy-to-use special edition includes Part 1: Sections 1&2 and the five specific sites: bladder, breast, colon, lung, and prostate. NCRA member price: $65; non-member price: $85.

Go to www.ncra-usa.org/CTRPrep to order.

Fall 2014 CTR Exam Prep Online Workshop

NCRA has created a CTR Exam Prep Online Workshop consisting of nine, 90-minute live webinars on critical exam topics. Workshop member price: $365; non-member price: $485. Individual webinars can be purchased separately. Member price: $45; non-member price: $60. Go to the “CTR Prep” tab of the CCRE at www.CancerRegistryEducation.org to learn more about the content, presenters, dates, and to register.

FROM THE PEOPLE WHO KNOW CTRs BEST!

NCRA: Your Source for CTR Exam Prep

ONLINE CTR EXAM PRACTICE TEST

CancerRegistryEducation.org/ctr-prep

NCRA offers a 125-question online CTR Exam Practice Test through its Center for Cancer Registry Education (CCRE). Designed to provide candidates with a tool to assess understanding and determine where further study is needed. Pricing options: One-time access: NCRA member - $65; non-member - $105. Unlimited access for 60 days: NCRA member - $180; non-member - $220. Go to the “CTR Prep” tab of the CCRE at www.CancerRegistryEducation.org for more detailed information on the online test.

2014 CTR Exam

The CTR Credential is nationally recognized in the recruitment and retention of cancer registry personnel. LAST TESTING DATE IN 2014!

October 18 - November 8, 2014

Application deadline: September 19

Learn more about the CTR Exam and download the 2014 CTR Exam Candidate’s Handbook at www.CTRExam.org.

Membership Matters...

NCRA benefits are designed to support your career and ensure your success as a cancer registry professional.

- Get and Stay Certified
  Use NCRA Educational Products to Earn and Maintain the CTR
- Keep Current
  Gain Access to Professional Resources
- Support the Profession
  There is Strength in Numbers! NCRA represents members’ collective voice with standard setters, national organizations, and federal agencies.
- Make Connections and Network
  Meet Colleagues at Live Events and Through NCRA's Online Communities
- Advance Your Career
  Develop Leadership and Management Skills
- Save Money
  Get Exclusive Discounts on Products and Services

LEARN MORE!

Website: www.ncra-usa.org/membership
Phone: 703-299-6640 Ext. 319
E-mail: member@ncra-usa.org.
From the People Who Know CTRs Best!

NCRA: Your Source for CTR Exam Prep

Revised and Improved!
NCRA Study Guide for the CTR Exam:
2014 Edition
The new edition has been revised to reflect the updates to the 2014 exam. Five case studies have been added with answers and rationales provided. Companion CD contains additional study resources. NCRA member price: $85; non-member price: $105. Go to www.ncra-usa.org/CTRPrep to order.

Special Compilation of Collaborative Stage Data Collection System Coding Instructions (For 2014 CTR Exam Candidates Only)
NCRA has compiled the sections from the Collaborative Stage Data Collection System Coding Instructions (Version 02.04) that can be used in the open-book portion of the 2014 CTR Exam. This easy-to-use special edition includes Part 1: Sections 1&2 and the five specific sites: bladder, breast, colon, lung, and prostate. NCRA member price: $65; non-member price: $85. Go to www.ncra-usa.org/CTRPrep to order.

Fall 2014 CTR Exam Prep Online Workshop
NCRA has created a CTR Exam Prep Online Workshop consisting of nine, 90-minute live webinars on critical exam topics. Workshop member price: $365; non-member price: $485. Individual webinars can be purchased separately: Member price: $45; non-member price: $60. Go to the “CTR Prep” tab of the CCRE at www.CancerRegistryEducation.org to learn more about the content, presenters, dates, and to register.

Online CTR Exam Practice Test
NCRA offers a 125-question online CTR Exam Practice Test through its Center for Cancer Registry Education (CCRE). Designed to provide candidates with a tool to assess understanding and determine where further study is needed. Pricing options: One-time access: NCRA member - $65; non-member - $105. Unlimited access for 60 days: NCRA member - $180; non-member - $220. Go to the “CTR Prep” tab of the CCRE at www.CancerRegistryEducation.org for more detailed information on the online test.

Online CTR Exam Practice Test
CancerRegistryEducation.org/ctr-prep

2014 CTR Exam
The CTR Credential is nationally recognized in the recruitment and retention of cancer registry personnel.

LAST TESTING DATE IN 2014!
October 18 - November 8, 2014
Application deadline: September 19
Learn more about the CTR Exam and download the 2014 CTR Exam Candidate’s Handbook at www.CTRExam.org.
Journal of Registry Management

NCRA Executive Office
1330 Braddock Place
#520
Alexandria, VA 22314

Printed on SFI fiber sourcing paper with non-petroleum,
vegetable based inks and manufactured with renewable electricity