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 U50/0003842.
Original Articles

Enhancing Cancer Registry Data for Comparative Effectiveness Research (CER) Project: Overview and Methodology ....................................................103
Vivien W. Chen, PhD; Christie R. Eheman, PhD; Christopher J. Johnson, MPH; Monique N. Hernandez, PhD; David Rousseau, BS; Timothy S. Styles, MD, MPH; Dee W. West, PhD; Meichin Hsieh, MSPH, CTR; Anne M. Hakeneworth, PhD; Maria O. Celaya, MPH, CTR; Randi K. Rycroft, MSPH, CTR; Jennifer M. Wike, MBA, MPH; Melissa Pearson, CTR; Judy Brockhouse, MPH, CTR; Linda G. Mulvihill, RHIT, CTR; Kevin B. Zhang, PhD

Surveillance of the Frequency and Results of Testing of Incident Oropharyngeal Cancers for Human Papillomavirus: The Potential Role of Population-Based Cancer Registries ..........................................................113
Anthony P. Polodnak, PhD; Cathryn Phillips, CTR

Misclassification of Sex in Central Cancer Registries ..................................................................................................................120
Recinda L. Sherman, MPH, PhD, CTR; Francis P. Boscoe, PhD; David K. O’Brien, PhD, GISP; Justin T. George, MPH; Kevin A. Henry, PhD; Laura E. Soloway, PhD; David J. Lee, PhD

Medicaid and Cancer Disparities: The Need for Comprehensive Research Data Sets .................................................................125
Sujha Subramaniana

Comprehensive Assessment of Population-Based Cancer Registries: An Experience in Colombia ..............................................................128
Constanza Pardo, MsC; Luis Eduardo Bravo, MsC; Claudia Uribe, MD; Guillermo López, MD; María Clara Yépez, MsC; Edgar Navarro, MsC; Esther de Vries, PhD; Marion Piñeros, MsC

Trends in Reporting Delay in United States Central Cancer Registries ............................................................................................135
Francis P. Boscoe, PhD

Role of Outpatient Physicians in Reporting of Hematopoietic Malignancies for Statewide Surveillance ............................................................139
Sue-Min Lai, PhD, MS, MBA; John Keighley, PhD; Sarma Garmella, MBBS, MPH

An Assessment of the Reliability of Race, Hispanic Ethnicity, Birthplace, and Tobacco History Data in the Massachusetts Cancer Registry, 2005–2009 ..................................................146
R. Knowlton; S. Gershman; A. Solis; B. Das

Colorectal Cancer Care at Falmouth Hospital: An Assessment of Progress Through Registry Statistics .............................................................151
Amanda Doodlesack, BS

Breast Cancer Incidence Among Nebraska Women: Early- and Late-Stage Trends, 1995–2009 ........................................................................154
Kristin Yeoman, MD, MPH; A. Blythe Ryerson, PhD, MPH; Natalie Koziol, MA; Bryan Buss, DVM, MPH; Thomas Safranek, MD

Features and Other Journal Departments

Raising the Bar: When Hope Vanishes .................................................................................................................................158
Michele Webb, CTR

NAACCR EDITS Impact: Release of NAACCR Edits Metafile v14A..............................................................................................159
How I Do It: Monthly Quality Review – The Tools I Use ..............................................................................................................160
Kathleen Hess, RHIT, CTR

Fall 2014 Continuing Education Quiz ..................................................................................................................................164
Deborah C. Roberson, MSM, CTR; Denise Harrison, BS, CTR

Call for Papers .............................................................................................................................................................................166
Information for Authors .........................................................................................................................................................167
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
Kevin Ward, PhD, MPH, CTR
Pamela Warren, PhD, CTR
Reda Wilson, MPH, RHIT, CTR (Editor Emeritus)
Vonetta Williams, MPH, RHIT, 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 Cancer Registry Data for Comparative Effectiveness Research (CER) Project: Overview and Methodology

Vivien W. Chen, PhD*; Christie R. Ehemann, PhD†; Christopher J. Johnson, MPH‡; Monique N. Hernandez, PhD§; David Rousseau, BS∥; Timothy S. Styles, MD, MPH¶; Dee W. West, PhD; Meichin Hsieh, MSPH, CTR; Anne M. Hakenewerth, PhD; Maria O. Celaya, MPH, CTR*; Randi K. Rycroft, MSPH, CTR; Jennifer M. Wike, MBA, MPH; Melissa Pearson, CTR; Judy Brockhouse, MPH, CTR; Linda G. Mulvihill, RHIT, CTR; Kevin B. Zhang, PhD

Abstract: Following the Institute of Medicine’s 2009 report on the national priorities for comparative effectiveness research (CER), funding for support of CER became available in 2009 through the American Recovery and Reinvestment Act. The Centers for Disease Control and Prevention (CDC) received funding to enhance the infrastructure of population-based cancer registries and to expand registry data collection to support CER. The CDC established 10 specialized registries within the National Program of Cancer Registries (NPCR) to enhance data collection for all cancers and to address targeted CER questions, including the clinical use and prognostic value of specific biomarkers. The project also included a special focus on detailed first course of treatment for cancers of the breast, colon, and rectum, as well as chronic myeloid leukemia (CML) diagnosed in 2011. This paper describes the methodology and the work conducted by the CDC and the NPCR specialized registries in collecting data for the 4 special focused cancers, including the selection of additional data variables, development of data collection tools and software modifications, institutional review board approvals, training, collection of detailed first course of treatment, and quality assurance. It also presents the characteristics of the study population and discusses the strengths and limitations of using population-based cancer registries to support CER as well as the potential future role of population-based cancer registries in assessing the quality of patient care and cancer control.

Key words: cancer treatment, CER support, methodology, population-based registry

Introduction

In June 2009, the Institute of Medicine (IOM) published a report entitled Initial National Priorities for Comparative Effectiveness Research (CER) and listed 100 CER priorities, including cancer-related objectives. The IOM defined CER as “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both individual and population levels.” Funding for CER was provided to the Department of Health and Human Services (HHS) through the American Recovery and Reinvestment Act (ARRA) of 2009, and projects were coordinated through the Agency for Healthcare Research and Quality (AHRQ).

The Centers for Disease Control and Prevention (CDC) received funding to enhance the infrastructure of population-based cancer registries and to support collection of data for CER. CDC developed a project entitled Enhancing Cancer Registry Data for Comparative Effectiveness Research (CER) and in 2010 awarded funds to ICF International and state cancer registries for expanded data collection and 6 special projects. Ten states receiving National Program of Cancer Registries (NPCR) funding that met the eligibility criteria...
were selected for the CER project based on competitive proposals and were established as specialized cancer registries. These registries had the potential and capability to enhance their infrastructure for "additional data collection, training, methodological development, and expansion of electronic reporting with the goal of supporting comparative effectiveness research and to develop sustainable methods to enhance registry data for public health and research." As part of this CER project, participating NPCR specialized registries expanded data collection to include additional data variables, such as height, weight and smoking status, for all cancers, and they performed linkages with secondary data sets including census data, National Death Index files, hospital discharge data, and the state’s breast and cervical cancer early detection programs to enhance registry data. The project also included a special focus on detailed treatment information for cancers of the breast, colon, and rectum as well as chronic myeloid leukemia (CML). An outcome of the project included a dataset that could be used for CER and other research. This paper provides a description of the methodology and the work conducted by the CDC and the NPCR specialized registries in collecting data for the 4 special focused cancers. It also presents the characteristics of the study population and a discussion of the strengths and limitations of using population-based cancer registries to support CER.

Objectives

The primary objectives of the CER project were to enhance the registry infrastructure and to obtain data that would support CER for cancers of the breast, colon, and rectum as well as CML. The collected data included tumor characteristics (predictive and prognostic biomarkers), stage at diagnosis, first course of treatment (both neoadjuvant and adjuvant), and patient sociodemographic factors and other factors, such as comorbidities and insurance coverage that may influence the choice of treatment options for these patients.

To ensure that current CER needs could be addressed by the project, the CDC in collaboration with AHRQ identified a series of questions, along with the data variables needed to address those and other questions in comparative effectiveness research. The targeted CER issues were:

- Are colon and rectum (colorectal) cancer patients tested for KRAS and are the results used appropriately to determine treatment? What impact does KRAS testing have on 2–3 year survival among colorectal cancer patients?
- Are rectal cancer patients receiving radiotherapy and what is the timing of radiotherapy? Are disparities apparent in the appropriate neoadjuvant use of radiotherapy among these patients?
- Are CML patients being tested for the BCR-ABL gene and receiving appropriate treatment according to those results?
- Are women with breast cancer being tested appropriately for human epidermal growth factor receptor 2 (HER2), progesterone receptor (PR), and estrogen receptor (ER) status and treated accordingly?

Methods

Case Eligibility

Cases included in the CER project were male and female patients diagnosed in 2011 with either in situ or malignant tumors of the breast or colon/rectum or with CML (see Table 1 for ICD-O-3 site codes and histologies). The 10 participating states included the entire states of Alaska, Colorado, Idaho, Louisiana, New Hampshire, North Carolina, Rhode Island, and Texas, as well as 13 counties of the Sacramento region of California and 5 metropolitan counties of Miami, Florida.

Data Variables

Population-based cancer registries routinely collect North American Association of Central Cancer Registries (NAACCR) standard data variables such as patient demographics, tumor characteristics, cancer stage, treatment, address of residence, and comorbidities. However, these data are insufficient to support contemporary comparative treatment research. Working together with AHRQ, the CDC identified additional data items (Table 2) that were crucial to address CER questions. The final data variables included in the CER project were: expanded patient information such as height and weight, comorbid conditions, and smoking status; area-based (census tract) socioeconomic status; stage of disease (all Collaborative Stage version 2 [Csv2] data items that are necessary to derive the American Joint Commission on Cancer [AJCC] TNM and Stage for both 6th and 7th editions); tumor biomarkers of prognostic and predictive significance listed under Csv2 site-specific factors (SSFs); and detailed first course of cancer-directed treatment. For breast cancer, the SSFs included: ER status, PR status, number of positive ipsilateral level I–II axillary lymph nodes, presence of isolated tumor cells in regional lymph nodes, size of invasive component of the tumor, Nottingham or Bloom-Richardson tumor grade/score, HER2, response to neoadjuvant therapy, and multigene signature testing. For colorectal cancer, the SSFs collected were: carcinoembryonic antigen (CEA), clinical assessment of regional lymph nodes, presence of tumor deposits, tumor regression grade, circumferential resection margin (CRM), microsatellite instability (MSI), perineural invasion, KRAS testing, and loss of heterozygosity (LOH). As for CML, Janus Kinase 2 (JAK2) gene mutation was collected in addition to BCR-ABL testing.

In order to evaluate comparative effectiveness of treatments for cancers of the breast, colon, rectum, and CML, complete and accurate first course of cancer-directed treatment was essential. First course of cancer treatment was defined as the therapy regimen that was given or planned at the time of initial diagnosis, prior to disease recurrence or progression. While central cancer registries routinely collect detailed information on surgery (which often is performed in hospital at inpatient or outpatient settings) and radiation (hospital-affiliated and freestanding), other adjuvant treatments have often been missing or incomplete. Therefore, the CER project focused on collecting complete adjuvant treatment occurring within 12 months of diagnosis, particularly...
detailed chemotherapy data. To identify each drug consistently across the registries, abstractors used the Cancer Chemotherapy National Service Center (NSC) number. The NSC is a numeric identifier for substances submitted to the National Cancer Institute (NCI) for testing and evaluation during its investigational phase and a registration number for the Developmental Therapeutics Program (DTP) repository. Abstractors obtained NSC numbers from the Web-based version of the SEER*Rx Interactive Antineoplastic Drugs Database\(^8\) that provides detailed drug information including generic and brand name, abbreviation, NSC number, drug category and subcategory and is readily available on the NCI website (http://seer.cancer.gov/seertools/seerrx).

The CER chemotherapy variables included each chemotherapy agent’s name and NSC number, number of chemotherapy cycles planned, total dose planned, number of doses received, and total dose received, start and end dates of chemotherapy, and whether chemotherapy was completed as planned, as well as growth factor agents (granulocyte, erythrocyte, and thrombocyte) that might have been given. Patient’s height and weight were also collected, especially for CER patients receiving chemotherapy. In addition, hormonal therapy and biological response modifier use were documented. If any subsequent treatment was given within 12 months of diagnosis as a result of recurrence, all subsequent treatment modalities were also recorded when available.

A CDC work group developed a data dictionary for the non-NAACCR standard variables, with an emphasis on capturing all cancer therapies received within 12 months of diagnosis. An oncologist provided consultation on the selection and definitions of additional variables of CER significance. Once a variable was identified as necessary and available in medical records, it was added to the data dictionary with a description, rationale, codes, and coding instructions. The work group used established, recognized guidelines for cancer data collection whenever possible. As an example, the SEER*Rx database was used to assign the NSC number for systemic therapy because it provides a consistent method to identify each drug. A few drugs that were either not listed in SEER*Rx or did not have an assigned NSC number were assigned an identifying number to be used by all registries.

### Data Collection Tool and Software

The participating states use different registry software for routine data collection and have various database management systems. There was no existing electronic abstraction tool for the CER project because the non-NAACCR standard data variables had not previously been collected. The CER data dictionary was provided to the registries’ software vendors for modification of their software so that these data items could be abstracted and be incorporated into the registry database. Different approaches were used by the specialized registries based on the capability of their software vendors, experience of their information technology (IT) staff, reporting requirement notification process, and CDC technical support. As a result, each registry had its own CER-specific data abstracting tools, edit programs and consolidation logic. However, all registries used a standard edits metafile and the same record layout for data submission.

To ensure data quality and consistency CDC staff developed a set of single field and inter-field edits that all CER participating registries ran prior to data submission.

### Institutional Review Board Approvals

The collection of additional data items for the CER project was authorized by existing cancer reporting laws and regulations in each state. No patients were contacted for the project and the collection of enhanced registry data was considered by most states to be public health surveillance and, as such, exempt from institutional review board (IRB) approval requirements. One state amended its cancer registry reporting regulations to add the nonstandard data items to its surveillance reporting requirements. Because the CER project title included the word “research,” some cancer treatment facilities required further explanation or, in some cases, sought full IRB review.

It should be noted that while the project activities described herein did not include research, but focused on collection of the data needed to support CER, all subsequent data analyses, conducted after the completion of project, must be covered under applicable IRB requirements of the investigators.

### Training

Training was conducted by the CDC and its contractors before data collection began. The objectives of the CER

---

**Table 1. Eligibility Criteria for Comparative Effectiveness Research Cases**

<table>
<thead>
<tr>
<th>Site</th>
<th>ICD-O-3 Site Code</th>
<th>Histology</th>
<th>Behavior</th>
<th>Sex</th>
<th>Diagnosis Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>C50.0–C50.9</td>
<td>All except 9050–9055, 9140, and 9590–9992</td>
<td>In situ, malignant</td>
<td>Male and female</td>
<td>2011</td>
</tr>
<tr>
<td>Colon</td>
<td>C18.0–18.9*</td>
<td>All except 9050–9055, 9140, and 9590–9992</td>
<td>In situ, malignant</td>
<td>Male and female</td>
<td>2011</td>
</tr>
<tr>
<td>Rectum</td>
<td>C19.9, C20.9</td>
<td>All except 9050–9055, 9140, and 9590–9992</td>
<td>In situ, malignant</td>
<td>Male and female</td>
<td>2011</td>
</tr>
<tr>
<td>Chronic Myeloid Leukemia</td>
<td>C42.1</td>
<td>Include 9863, 9875, 9876, 9945, and 9946</td>
<td>malignant</td>
<td>Male and female</td>
<td>2011</td>
</tr>
</tbody>
</table>

*Including Appendix (site code = C18.1).
project were explained, and the data variables and data dictionary, especially those items not routinely collected by the registries, were reviewed and discussed. Trainers provided instructions on data collection and coding rules, including CSv2 stage; SSFs, especially biomarkers such as KRAS and HER2; complete first course of treatment which included surgery, chemotherapy, hormonal therapy, biological response modifiers (including bone marrow transplant, stem cell harvest, surgical or radiation endocrine therapy); growth factors; and BCR-ABL tests for CML.

In addition to CDC central training, all participating specialized registries offered subsequent in-house training and ongoing education, depending on the needs and experience of their abstractors. Various training modes included workshops, webinars, webcasts, teleconferences, and presentations at the state cancer registries’ associations meetings. All states conducted one-on-one, in-person and/or webcast training when needed, and some also offered onsite demonstration. The intensity and extent of training varied by registry and depended on their available resources. For example, one registry housed at the health sciences center received training, from a clinical oncology nurse, on the National Comprehensive Cancer Network (NCCN) treatment guidelines for each cancer by tumor size, nodal involvement, and distant metastasis, as well as by biomarker/prediction of treatment response (eg, HER2, ER/PR and KRAS) and menopausal status (breast only). This training detailed first course of treatment, sequence of treatment modalities, common chemotherapy regimens, chemotherapy drug names, “standard” cycle and dose, and calculation of total dose (planned and received). In addition, the training included explanation and discussion on the chemotherapy “flow chart,” common side effects, and toxicities. Another registry created an abstracting workbook for data collection which provided information not contained in the data dictionary, such as treatment guidelines based on stage of disease and tips for interpreting laboratory reports for biomarkers. Other registries used highly experienced certified tumor registrars (CTRs), or combined CTR and registered nurse, to conduct trainings with hospital-based and non–hospital-based CER data collection staff. One registry had its abstractors practice CER data collection on cases diagnosed prior to 2011, interacted with hospital and nonhospital staff, and established remote access to some facilities. All participating registries conducted refresher trainings, after data collection was in progress, to ensure accurate and consistent interpretation of the information and coding of the data. Every registry designated a contact staff person to address CER data collection issues, answer questions, and provide clarifications.

The CDC data dictionary work group reviewed all training materials developed by the states to ensure accuracy and consistency. Many states shared and customized training materials which were made available through the CDC CER Information-Sharing Portal.

Data Collection

As noted previously, all NPCR specialized registries have legislative rules on cancer reporting which provide authorization to access medical records and collect cancer-related data variables from hospitals and nonhospital settings. Because of the vast number of cancer cases (more than a quarter of the annual caseload) and the large number of additional data variables collected, the majority of specialized registries hired and trained additional abstractors for this project; many reassigned existing staff and some used a combination of new and existing staff.

a. Hospital: When CER data collection started in 2012, a majority of the 2011 cases were already reported to the central registries. Building on the routine NAACCR abstracts they had on the CER cases, the hospital registrars (or central registry abstractors who visited the hospitals where the cancer was initially diagnosed) collected as many of the additional CER data items as possible. Most adjuvant treatment was given after the patient was discharged from the hospital; it was therefore necessary for abstractors to identify the facilities and/or physicians who followed the patients and provided adjuvant treatments. Some registries requested that hospital abstractors delay their submission of CER cancer cases until after the first course of treatment was complete and all information that could be obtained was abstracted. However, most CER cases required data collection from all source documents, including both hospital and nonhospital settings. On very rare occasions, where hospitals used a unified chart system that provided inpatient and outpatient medical records in one single database, abstractors were able to record all therapy modalities from a single setting.

b. Nonhospital sources: Adjuvant radiation was generally obtained from hospital outpatient or freestanding radiation centers, while information on chemotherapy was obtained mostly from hospital-based or independent hematology/oncology practice groups. Most specialized registries hired additional data specialists to collect CER data at physician offices and treatment facilities. This included identifying noncancer registry reporting sources to be targeted for data collection, developing methods of data collection and coordinating with the reporting source to carry out case identification and data collection.

Registries used 3 primary methods of data collection: onsite visits for abstracting, which was the most common method, obtaining hard copies or securing remote access to the facility’s electronic medical record, and, on rare occasions, transmission of data by the facility. Abstraction at nonhospital sources usually required an initial onsite visit to learn how to navigate each practice’s electronic or paper medical record systems, and follow-up visits to capture missing information. Follow-up was also done by phone, fax, or email whenever possible. Multiple visits to more than one physician office were frequently necessary to complete the first course of chemotherapy and other adjuvant treatment. Ascertaining the dose for each cycle of every chemotherapy agent was very challenging, and information was not always available, making the calculation of total chemotherapy dose received extremely difficult, often necessitating multiple visits.

Once data were obtained from all sources, CER staff reviewed and edited the cases for accuracy, consistency,
Socio-Economic Status Indicators

Tobacco Use: cigarette, smoking tobacco products other than cigarettes (eg, pipes, cigars, kreteks), smokeless tobacco, not otherwise specified

Socio-Economic Status Indicators

A wide range of indicators linked to cancer incidence data from multiple US Census Bureau data files, including area-level indicators for:

- Urban/rural
- Poverty level
- Health insurance coverage
- Income ranges
- Employment status
- Occupation and Industry data
- Class of worker (blue collar, white collar)

Diagnostic work up for CML

- BCR-ABL Cytogenetic Result and Date
- BCR-ABL FISH Result and Date
- BCR-ABL Qualitative Result and Date
- BCR-ABL Quantitative and Date

First course treatment information:

- Chemotherapy: agent and NSC number (for up to 6 chemotherapy agents)
- Chemotherapy: number of doses planned and received (for up to 6 chemotherapy agents)
- Chemotherapy: dose amount and units planned and received (for up to 6 chemotherapy agents)
- Chemotherapy: administration start and end dates (for up to 6 chemotherapy agents)
- Chemotherapy: completion status
- Chemotherapy: Granulocyte CSF Status
- Chemotherapy: Erythrocyte Growth Factor Status and Thrombocyte Growth Factor Status
- Hormone: agent and NSC number (for up to two hormone agents)
- Biologic Response Modifier (BRM): agent and NSC number (for up to 2 hormone agents)

Subsequent/second course treatment information:

- Reason Subsequent Treatment
- Subsequent treatment date
- Subsequent Surgery
- Subsequent Radiation
- Subsequent Chemotherapy and Chemo NSC (for up to 6 chemo agents)
- Subsequent Hormone and hormone NSC (for up to 2 hormone agents)
- Subsequent BRM and BRM NSC (for up to 2 BRM agents)
- Subsequent Transplant/Endocrine
- Subsequent Other

<table>
<thead>
<tr>
<th>Table 2. Non-NAACCR Standard Data Variables* Defined and Collected for the Comparative Effectiveness Research Project</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient characteristics</strong></td>
</tr>
<tr>
<td>• Height</td>
</tr>
<tr>
<td>• Weight</td>
</tr>
<tr>
<td>• Comorbidities: up to 10 standard NAACCR comorbidities</td>
</tr>
<tr>
<td>• Tobacco Use: cigarette, smoking tobacco products other than</td>
</tr>
<tr>
<td>cigarettes (eg, pipes, cigars, kreteks), smokeless</td>
</tr>
<tr>
<td>tobacco, not otherwise specified</td>
</tr>
<tr>
<td>• Socio-Economic Status Indicators</td>
</tr>
<tr>
<td>A wide range of indicators linked to cancer incidence data</td>
</tr>
<tr>
<td>from multiple US Census Bureau data files, including area</td>
</tr>
<tr>
<td>level indicators for:</td>
</tr>
<tr>
<td>• Urban/rural</td>
</tr>
<tr>
<td>• Poverty level</td>
</tr>
<tr>
<td>• Health insurance coverage</td>
</tr>
<tr>
<td>• Income ranges</td>
</tr>
<tr>
<td>• Employment status</td>
</tr>
<tr>
<td>• Occupation and Industry data</td>
</tr>
<tr>
<td>• Class of worker (blue collar, white collar)</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Diagnostic work up for CML</strong></td>
</tr>
<tr>
<td>• BCR-ABL Cytogenetic Result and Date</td>
</tr>
<tr>
<td>• BCR-ABL FISH Result and Date</td>
</tr>
<tr>
<td>• BCR-ABL Qualitative Result and Date</td>
</tr>
<tr>
<td>• BCR-ABL Quantitative and Date</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>First course treatment information:</strong></td>
</tr>
<tr>
<td>• Chemotherapy: agent and NSC number (for up to 6 chemotherapy</td>
</tr>
<tr>
<td>agents)</td>
</tr>
<tr>
<td>• Chemotherapy: number of doses planned and received</td>
</tr>
<tr>
<td>(for up to 6 chemotherapy agents)</td>
</tr>
<tr>
<td>• Chemotherapy: dose amount and units planned and received</td>
</tr>
<tr>
<td>(for up to 6 chemotherapy agents)</td>
</tr>
<tr>
<td>• Chemotherapy: administration start and end dates</td>
</tr>
<tr>
<td>(for up to 6 chemotherapy agents)</td>
</tr>
<tr>
<td>• Chemotherapy: completion status</td>
</tr>
<tr>
<td>• Chemotherapy: Granulocyte CSF Status</td>
</tr>
<tr>
<td>• Chemotherapy: Erythrocyte Growth Factor Status and Thrombo-</td>
</tr>
<tr>
<td>cyte Growth Factor Status</td>
</tr>
<tr>
<td>• Hormone: agent and NSC number (for up to two hormone agents)</td>
</tr>
<tr>
<td>• Biologic Response Modifier (BRM): agent and NSC number</td>
</tr>
<tr>
<td>(for up to 2 hormone agents)</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Subsequent/second course treatment information:</strong></td>
</tr>
<tr>
<td>• Reason Subsequent Treatment</td>
</tr>
<tr>
<td>• Subsequent treatment date</td>
</tr>
<tr>
<td>• Subsequent Surgery</td>
</tr>
<tr>
<td>• Subsequent Radiation</td>
</tr>
<tr>
<td>• Subsequent Chemotherapy and Chemo NSC (for up to 6 chemo</td>
</tr>
<tr>
<td>agents)</td>
</tr>
<tr>
<td>• Subsequent Hormone and hormone NSC (for up to 2 hormone</td>
</tr>
<tr>
<td>agents)</td>
</tr>
<tr>
<td>• Subsequent BRM and BRM NSC (for up to 2 BRM agents)</td>
</tr>
<tr>
<td>• Subsequent Transplant/Endocrine</td>
</tr>
<tr>
<td>• Subsequent Other</td>
</tr>
</tbody>
</table>

*Only variables defined and collected exclusively for Comparative Effectiveness Research (CER) activities are included above. There are many other variables included in the CER data collection activities that are routinely collected by National Program of Cancer Registries (NPCR) and defined by North American Association of Central Cancer Registries’ (NAACCR’s) Standards for Cancer Registries, Volume II: Data Standards and Data Dictionary, Fifteenth Edition, Record Layout Version 12.1.
Table 3. Demographic and Tumor Characteristics of Comparative Effectiveness Research Patients by Cancer Site, 2011

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>48,456</td>
<td>64.6%</td>
<td>18,413</td>
<td>24.5%</td>
<td>7,129</td>
<td>9.5%</td>
<td>1,044</td>
<td>1.4%</td>
<td>75,042</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

**Sex**

<table>
<thead>
<tr>
<th>Sex</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>372</td>
<td>0.8%</td>
<td>9,361</td>
<td>50.8%</td>
<td>4,186</td>
<td>58.7%</td>
<td>609</td>
<td>58.3%</td>
<td>14,528</td>
<td>19.4%</td>
</tr>
<tr>
<td>Female</td>
<td>48,079</td>
<td>99.2%</td>
<td>9,044</td>
<td>49.1%</td>
<td>2,942</td>
<td>41.3%</td>
<td>435</td>
<td>41.7%</td>
<td>60,500</td>
<td>80.6%</td>
</tr>
</tbody>
</table>

**Other (hermaphrodite)**

<table>
<thead>
<tr>
<th>Other (hermaphrodite)</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>3</td>
<td>0.0%</td>
<td>1</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>4</td>
<td>0.0%</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>0.0%</td>
<td>1</td>
<td>0.0%</td>
<td>1</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>4</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not State/Unknown</td>
<td>0</td>
<td>0.0%</td>
<td>6</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>6</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Age at Diagnosis (years)**

<table>
<thead>
<tr>
<th>Age at Diagnosis (years)</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;50</td>
<td>10,068</td>
<td>20.8%</td>
<td>1,709</td>
<td>9.3%</td>
<td>1,049</td>
<td>14.7%</td>
<td>273</td>
<td>26.1%</td>
<td>13,099</td>
<td>17.5%</td>
</tr>
<tr>
<td>50–59</td>
<td>11,883</td>
<td>24.5%</td>
<td>3,226</td>
<td>17.5%</td>
<td>1,849</td>
<td>25.9%</td>
<td>160</td>
<td>15.3%</td>
<td>17,118</td>
<td>22.8%</td>
</tr>
<tr>
<td>60–69</td>
<td>13,024</td>
<td>26.9%</td>
<td>4,597</td>
<td>25.0%</td>
<td>1,844</td>
<td>25.9%</td>
<td>197</td>
<td>18.9%</td>
<td>19,662</td>
<td>26.2%</td>
</tr>
<tr>
<td>≥70</td>
<td>13,479</td>
<td>27.8%</td>
<td>8,881</td>
<td>48.2%</td>
<td>2,387</td>
<td>33.5%</td>
<td>414</td>
<td>39.7%</td>
<td>25,161</td>
<td>33.5%</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

**Race/Ethnicity**

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic White</td>
<td>33,502</td>
<td>68.8%</td>
<td>12,263</td>
<td>66.6%</td>
<td>4,722</td>
<td>66.1%</td>
<td>697</td>
<td>67.1%</td>
<td>51,184</td>
<td>68.2%</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>6,456</td>
<td>13.3%</td>
<td>2,777</td>
<td>14.9%</td>
<td>947</td>
<td>13.0%</td>
<td>132</td>
<td>13.2%</td>
<td>10,312</td>
<td>13.7%</td>
</tr>
<tr>
<td>Non-Hispanic Other</td>
<td>1,183</td>
<td>2.3%</td>
<td>373</td>
<td>1.9%</td>
<td>187</td>
<td>2.5%</td>
<td>18</td>
<td>1.5%</td>
<td>1,761</td>
<td>2.3%</td>
</tr>
<tr>
<td>Hispanic (all races)</td>
<td>6,679</td>
<td>13.3%</td>
<td>2,772</td>
<td>14.7%</td>
<td>1,166</td>
<td>16.2%</td>
<td>180</td>
<td>15.7%</td>
<td>10,797</td>
<td>14.4%</td>
</tr>
<tr>
<td>Oth/Unk/Missing</td>
<td>636</td>
<td>2.3%</td>
<td>228</td>
<td>2.0%</td>
<td>107</td>
<td>2.2%</td>
<td>17</td>
<td>2.4%</td>
<td>988</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

**Health Insurance**

<table>
<thead>
<tr>
<th>Health Insurance</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Insurance</td>
<td>1,820</td>
<td>3.8%</td>
<td>1,011</td>
<td>5.5%</td>
<td>470</td>
<td>6.6%</td>
<td>80</td>
<td>7.7%</td>
<td>3,381</td>
<td>4.5%</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>26,335</td>
<td>54.3%</td>
<td>6,982</td>
<td>37.9%</td>
<td>3,173</td>
<td>44.5%</td>
<td>411</td>
<td>39.4%</td>
<td>36,901</td>
<td>49.2%</td>
</tr>
<tr>
<td>Public: Medicaid</td>
<td>4,620</td>
<td>9.5%</td>
<td>1,865</td>
<td>10.1%</td>
<td>770</td>
<td>10.8%</td>
<td>112</td>
<td>10.7%</td>
<td>7,367</td>
<td>9.8%</td>
</tr>
<tr>
<td>Public: Medicare</td>
<td>12,228</td>
<td>25.2%</td>
<td>6,793</td>
<td>36.9%</td>
<td>2,031</td>
<td>28.5%</td>
<td>322</td>
<td>30.8%</td>
<td>21,374</td>
<td>28.5%</td>
</tr>
<tr>
<td>Public: Other</td>
<td>901</td>
<td>1.9%</td>
<td>640</td>
<td>3.5%</td>
<td>290</td>
<td>4.1%</td>
<td>18</td>
<td>1.7%</td>
<td>1,849</td>
<td>2.5%</td>
</tr>
<tr>
<td>Unknown</td>
<td>2,446</td>
<td>5.0%</td>
<td>1,078</td>
<td>5.9%</td>
<td>379</td>
<td>5.3%</td>
<td>100</td>
<td>9.6%</td>
<td>4,003</td>
<td>5.3%</td>
</tr>
<tr>
<td>Missing/Blank</td>
<td>106</td>
<td>0.2%</td>
<td>44</td>
<td>0.2%</td>
<td>16</td>
<td>0.2%</td>
<td>1</td>
<td>0.1%</td>
<td>167</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

**Comorbidity Data**

<table>
<thead>
<tr>
<th>Comorbidity Data</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>None documented</td>
<td>23,796</td>
<td>49.1%</td>
<td>6,540</td>
<td>35.5%</td>
<td>2,754</td>
<td>38.6%</td>
<td>459</td>
<td>44.0%</td>
<td>33,549</td>
<td>44.7%</td>
</tr>
<tr>
<td>At least one</td>
<td>24,233</td>
<td>50.0%</td>
<td>11,737</td>
<td>63.7%</td>
<td>4,335</td>
<td>60.8%</td>
<td>579</td>
<td>55.5%</td>
<td>40,884</td>
<td>54.5%</td>
</tr>
<tr>
<td>Blank</td>
<td>427</td>
<td>0.9%</td>
<td>136</td>
<td>0.7%</td>
<td>40</td>
<td>0.6%</td>
<td>6</td>
<td>0.6%</td>
<td>609</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

**Census Tract Residence**

<table>
<thead>
<tr>
<th>Census Tract Residence</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>28,060</td>
<td>57.9%</td>
<td>10,358</td>
<td>56.3%</td>
<td>3,935</td>
<td>55.2%</td>
<td>586</td>
<td>56.1%</td>
<td>42,939</td>
<td>57.2%</td>
</tr>
<tr>
<td>Rural</td>
<td>3,871</td>
<td>8.0%</td>
<td>1,653</td>
<td>9.0%</td>
<td>665</td>
<td>9.3%</td>
<td>77</td>
<td>7.4%</td>
<td>6,266</td>
<td>8.3%</td>
</tr>
<tr>
<td>Mixed</td>
<td>15,746</td>
<td>32.5%</td>
<td>5,988</td>
<td>32.5%</td>
<td>2,385</td>
<td>33.5%</td>
<td>359</td>
<td>34.4%</td>
<td>24,478</td>
<td>32.6%</td>
</tr>
<tr>
<td>Missing</td>
<td>779</td>
<td>1.6%</td>
<td>414</td>
<td>2.2%</td>
<td>144</td>
<td>2.0%</td>
<td>22</td>
<td>2.1%</td>
<td>1,359</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

1 Appendix cases are included from Colon.
2 CML = Chronic Myeloid Leukemia.
3 Non-Hispanic Other includes American Indian/Alaskan Native /Asian or Pacific Islander.
4 Public: Other includes Tricare, Military, VA, Indian Health Service (IHS).
5 A census tract residence was considered “Urban” if all households in that census tract were considered to be in an urban setting as defined by the Census Bureau; A census tract residence was considered “Rural” if all households in that census tract were considered to be in a rural setting as defined by the Census Bureau; A census tract residence was considered “Mixed” if some of the households in the census tract were considered to be in an urban setting and some in a rural setting as defined by the Census Bureau.
Table 3, cont.

<table>
<thead>
<tr>
<th></th>
<th>Breast</th>
<th>Colon¹</th>
<th>Rectum</th>
<th>CML²</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td><strong>Census Tract Poverty⁶</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in poverty</td>
<td>40,920</td>
<td>84.4%</td>
<td>14,785</td>
<td>80.3%</td>
<td>5,721</td>
</tr>
<tr>
<td>Poverty</td>
<td>6,744</td>
<td>13.9%</td>
<td>3,203</td>
<td>17.4%</td>
<td>1,262</td>
</tr>
<tr>
<td>Blank/Missing</td>
<td>792</td>
<td>1.6%</td>
<td>426</td>
<td>2.3%</td>
<td>146</td>
</tr>
<tr>
<td><strong>Cigarette Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Used</td>
<td>20,692</td>
<td>42.7%</td>
<td>7,047</td>
<td>38.3%</td>
<td>2,489</td>
</tr>
<tr>
<td>Current user</td>
<td>4,078</td>
<td>8.4%</td>
<td>1,696</td>
<td>9.2%</td>
<td>927</td>
</tr>
<tr>
<td>Former user</td>
<td>7,092</td>
<td>14.6%</td>
<td>3,141</td>
<td>17.1%</td>
<td>1,311</td>
</tr>
<tr>
<td>Unknown</td>
<td>16,594</td>
<td>34.2%</td>
<td>6,529</td>
<td>35.5%</td>
<td>2,402</td>
</tr>
<tr>
<td><strong>SEER Summ Stg 2000⁷</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Situ</td>
<td>8,796</td>
<td>18.2%</td>
<td>824</td>
<td>4.5%</td>
<td>288</td>
</tr>
<tr>
<td>Localized</td>
<td>24,090</td>
<td>49.7%</td>
<td>6,289</td>
<td>34.2%</td>
<td>2,649</td>
</tr>
<tr>
<td>Regional</td>
<td>11,312</td>
<td>23.3%</td>
<td>5,991</td>
<td>32.5%</td>
<td>2,225</td>
</tr>
<tr>
<td>Distant</td>
<td>2,368</td>
<td>4.9%</td>
<td>3,731</td>
<td>20.3%</td>
<td>1,256</td>
</tr>
<tr>
<td>Unknown</td>
<td>1,890</td>
<td>3.9%</td>
<td>1,578</td>
<td>8.6%</td>
<td>711</td>
</tr>
<tr>
<td><strong>Derive AJCC Stg 7th Ed⁸</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 0</td>
<td>8,804</td>
<td>18.2%</td>
<td>1,132</td>
<td>6.1%</td>
<td>443</td>
</tr>
<tr>
<td>Stage I</td>
<td>18,037</td>
<td>37.2%</td>
<td>3,514</td>
<td>19.1%</td>
<td>1,717</td>
</tr>
<tr>
<td>Stage II</td>
<td>11,714</td>
<td>24.2%</td>
<td>4,218</td>
<td>22.9%</td>
<td>1,240</td>
</tr>
<tr>
<td>Stage III</td>
<td>4,215</td>
<td>8.7%</td>
<td>4,122</td>
<td>22.4%</td>
<td>1,587</td>
</tr>
<tr>
<td>Stage IV</td>
<td>2,291</td>
<td>4.7%</td>
<td>3,568</td>
<td>19.4%</td>
<td>1,170</td>
</tr>
<tr>
<td>NA/Unknown</td>
<td>3,395</td>
<td>7.0%</td>
<td>1,859</td>
<td>10.1%</td>
<td>972</td>
</tr>
<tr>
<td><strong>State of Residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AK</td>
<td>530</td>
<td>1.1%</td>
<td>214</td>
<td>1.2%</td>
<td>68</td>
</tr>
<tr>
<td>CA-Sacramento⁹</td>
<td>3,328</td>
<td>6.9%</td>
<td>1,104</td>
<td>6.0%</td>
<td>440</td>
</tr>
<tr>
<td>CO</td>
<td>4,198</td>
<td>8.7%</td>
<td>1,354</td>
<td>7.4%</td>
<td>520</td>
</tr>
<tr>
<td>FL-Metro Miami¹⁰</td>
<td>6,734</td>
<td>13.9%</td>
<td>2,976</td>
<td>16.2%</td>
<td>1,019</td>
</tr>
<tr>
<td>ID</td>
<td>1,231</td>
<td>2.5%</td>
<td>468</td>
<td>2.5%</td>
<td>187</td>
</tr>
<tr>
<td>LA</td>
<td>3,915</td>
<td>8.1%</td>
<td>1,721</td>
<td>9.3%</td>
<td>705</td>
</tr>
<tr>
<td>NC</td>
<td>9,177</td>
<td>18.9%</td>
<td>2,970</td>
<td>16.1%</td>
<td>1,165</td>
</tr>
<tr>
<td>NH</td>
<td>1,454</td>
<td>3.0%</td>
<td>464</td>
<td>2.5%</td>
<td>193</td>
</tr>
<tr>
<td>RI</td>
<td>1,057</td>
<td>2.2%</td>
<td>340</td>
<td>1.8%</td>
<td>151</td>
</tr>
<tr>
<td>TX</td>
<td>16,832</td>
<td>34.7%</td>
<td>6,802</td>
<td>36.9%</td>
<td>2,681</td>
</tr>
</tbody>
</table>

⁶Classification of Census Tract Poverty based on Krieger et al (ref 10). Not in poverty defined as: <20% of census tract families had income below poverty line in last 12 months; Poverty: ≥20% of census tract families had income below poverty line in last 12 months.
⁸AJCC = American Joint Commission on Cancer.
⁹California-Sacramento includes Alpine, Amador, Calaveras, El Dorado, Nevada, Placer, Sacramento, San Joaquin, Sierra, Solano, Sutter, Yolo, and Yuba counties.
¹⁰Florida-Miami includes Broward, Hillsborough, Miami-Dade, Orange, and Palm Beach counties.
Approximately half (48.2%) of the colon cancer cases were diagnosed in patients 70 years and older whereas for breast cancer, only about a quarter (27.8%) were aged 70 or older. The majority of patients were non-Hispanic whites (68.2%), non-Hispanic blacks, and Hispanics (of all races) each represented about 14%, and less than 4% were of other or unknown races. The racial/ethnic pattern was very similar for all 4 cancer groups.

Overall, approximately half of the patients had private insurance (including Medicare with private supplement or Medicare Advantage Plans), 29% had Medicare only, 10% had Medicaid, 2.5% had other public insurance such as Tricare, Veterans’ Affairs, or Indian Health Services, and less than 5% had no insurance. Breast cancer patients (54%) were most likely to have private insurance whereas colon cancer patients were the least likely (38%). The large proportion of Medicare coverage among colon cancer patients (37%) reflects the older age at diagnosis compared with other cancers. Over half (54.5%) of all patients had at least 1 comorbidity, with the comorbid conditions being more prevalent in colorectal than breast cancer patients. The majority of CER patients lived in census tracts designated as urban or urban/rural mixed, and less than 10% lived in rural area. About 83% resided in census tracts with less than 20% of families having income below federal poverty guidelines. Information on cigarette smoking was missing in 34.5% of the patients. Among those with known information, 62% never smoked cigarettes, 24% were former users, and 14% were current smokers.

There was considerable variation of tumor stage at diagnosis by cancer type. About two-thirds (68%) of breast cancer patients were diagnosed with early stage disease (in situ and localized) based on the SEER Summary Stage 2000, contrasting with only 40% among colorectal cancer patients. Similar tumor stage pattern was observed for AJCC TNM Stage Group.

The distribution of cancer cases also varied greatly by state (ranging from 819 to 26,752), reflecting the population size of the participating states and geographic areas.

Discussion

When the National Cancer Policy Board of the IOM, National Academy of Sciences issued a report on Ensuring Quality of Cancer Care in 1999, it concluded that “…for many Americans with cancers, there is a wide gulf between what could be construed as the ideal and the reality of their experience with cancer.” Its subsequent report of Enhancing Data Systems to Improve the Quality of Cancer Care in 2000 recommended the need for comprehensive cancer data systems that could be used to gauge the status of cancer care and measure quality. It identified 3 existing national cancer surveillance programs that could be used for quality improvement of cancer care. They were: the NPCR of CDC; the Surveillance, Epidemiology and End Results (SEER) program of the NCI; and the National Cancer Data Base (NCDB), sponsored by the American College of Surgeons’ Commission on Cancer (ACoS-CoC) and the American Cancer Society. While each of these data systems has its own limitations, they also each hold great potential and could be enhanced to assess and improve quality of cancer care in the nation.

In response to the IOM recommendations, CDC-NPCR initiated a pattern of care study, comparing the observed patterns of care with the accepted guidelines for localized breast cancer, localized prostate cancer and Stage III colon cancer, in conjunction with the international CONCORD Study. Subsequently the CDC, in collaboration with researchers from 7 central cancer registries, conducted the Breast and Prostate Cancer Data Quality and Patterns of Care (POC-BP) Study in 2007–2009. The study examined first course of cancer treatment received and how the patterns of care varied by patient, provider and other health system level factors. It also evaluated whether the care was concordant with nationally recognized treatment guidelines, given the presence and severity of comorbidities.

The NCI has also conducted patterns of care/quality of care (POC) studies since 1987 under a congressional mandate [Public Law 100-607, Sec. 413 (a)(2)(C) adopted November 4, 1988]. The collection of NCI POC data is coordinated jointly by the Division of Cancer Control and Population Sciences and the Division of Cancer Treatment and Diagnosis. Using the infrastructure of SEER registries, population-based samples of cases are selected every year for various cancer types to evaluate the dissemination of cancer therapy or guideline care into community practice and to identify possible determinants of dissemination and variations in therapy. Findings from POC studies provide educational or training opportunities for professional societies and public health groups to improve the quality of cancer care and reduce disparities in treatment and survival among different population groups. A wide range of cancer care has been studied by NCI over the years, including the trends in using adjuvant multi-agent chemotherapy and tamoxifen for breast cancer; age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer; and clinical trial participation and time to treatment among adolescents and young adults with cancer. Though it is not population-based, researchers have also used the NCDB to assess treatment patterns and their determinants among patients treated in facilities participating in the CoC-accredited cancer programs. Examples include assessing the importance of socioeconomic status and treatment institution in neo-adjuvant therapy for Stage IIIA non–small cell lung cancer and the impact of facility volume on therapy and survival for advanced cervical cancer.

In 2009, CDC received funding to establish 10 specialized cancer registries to collect additional data to support CER. This CER project is the largest and most comprehensive data collection effort ever conducted by population-based cancer registries in the United States to obtain complete and detailed first course of cancer-directed treatment for breast, colon, and rectal cancers and CML. This project covers more than a quarter (27.3%) of the US population, including a very high representation among minority populations. In fact, the combined catchment areas of the CER project include approximately 25% of African Americans, 37% of Asian/Pacific Islanders, 32% of American Indians/Alaska.
natives, and 44% of Hispanics living in the United States.\textsuperscript{31} This provides a unique opportunity to examine the patterns of care for these cancers in a racially and ethnically diverse population, to compare the benefits and risks of alternative treatments and care deliveries as well as to assess their outcomes. The project includes all newly-diagnosed cancer patients residing in the participating states in care settings that are representative of contemporary practice in the 10 geographic areas across the United States. Previous projects of this nature were often conducted in single major medical centers or in clinical trial groups where the patients were not representative of all US patients, and the treatments were not representative of contemporary practice of both major urban cancer centers and small rural hospitals.\textsuperscript{32-34}

Using population-based central cancer registries to collect data for CER has numerous strengths. Existing registry infrastructure and authorizing laws and regulations make it feasible to expedite the process, implement the procedures and collect additional cancer data in a timely manner. This registry infrastructure provided a baseline of well-established standard definitions and codes for cancer reporting, tumor staging and treatment/drugs. Because central cancer registries have been in place for several decades, registry staff are familiar with cancer reportability and the standard rules and codes. They have extensive experience with medical records abstraction and are knowledgeable of various cancer treatment modalities, requiring fewer staff and less training. Over time, central cancer registries have built excellent rapport with cancer care facilities and providers, allowing them to gain access to the facilities and/or connect remotely to the electronic medical records, resulting in an efficient and cost-effective means of obtaining the additional data for CER. In addition, population-based registries include all newly-diagnosed cancer patients; therefore findings using their data or a representative sample can be generalized to the US general population. There are also, however, limitations. Hospitals are required to submit cancer cases to central registries within 6 months after diagnosis, when first course of cancer treatment is often still ongoing; thus only partial treatment information is recorded at the time of initial report. While some states ask hospitals to submit update records which contain the additional treatment information, most state registries do not include these update records due to limited staff and competing registry priorities. Information on adjuvant therapies, especially hormonal and chemotherapy that are provided in nonhospital settings, is sometimes missing or incomplete. Collecting additional data that are not standard registry variables but are relevant to address CER topics requires special trainings. Furthermore, efforts and resources spent on project-specific activities such as software modification, training, data edits and added data submissions compete with regular registry operations and functions, potentially causing delays in routine registry data collection.

Despite the limitations, the CDC and the 10 NPCR specialized cancer registries have demonstrated the feasibility of utilizing population-based cancer registries to support comparative effectiveness investigations and other research in a cost-effective and timely manner. Moving forward, the CDC and the specialized cancer registries will continue to expand the uses of registry data beyond measuring cancer burden and evaluating stage shift by examining diagnostic procedures and cancer treatments and assessing how different procedures and treatment modalities impact patient outcomes of recurrence and survival. Central cancer registries provide a unique population base for cancer research. In addition, the consolidation of data from multiple locations of care is unique to the central cancer registry. No other disease surveillance system in the United States is as standardized and comprehensive as cancer surveillance, allowing researchers to track the outcomes of changes in the medical care system and assess the effectiveness of public health interventions. The CDC’s specialized cancer registries and the entire cancer registration program are well positioned to conduct comparative effectiveness research, evaluate patient outcome, and inform future medical practice.

References


Original Article

Surveillance of the Frequency and Results of Testing of Incident Oropharyngeal Cancers for Human Papillomavirus: The Potential Role of Population-Based Cancer Registries

Anthony P. Polednak, PhD; Cathryn Phillips, CTR

Abstract: Temporal increases in incidence rates for certain cancers of the oropharynx (OP), especially the base of tongue and tonsil (BTT), have been interpreted in relation to the epidemic of human papillomavirus (HPV) infection, but data on the actual presence of HPV in these tumors are limited. Data on the frequency and results of testing for HPV in OP cancers in defined populations also can be useful to clinicians. This study used the American Joint Committee on Cancer Collaborative Staging System’s Site-Specific Factor 10 (SSF 10) for HPV status of OP tumors, collected by some registries for diagnoses since 2010. The study included 483 incident invasive BTT cancers diagnosed in 2010–2012 and reported to the Connecticut Tumor Registry of the Surveillance, Epidemiology and End Results (SEER) Program. Of the 483 cancers, 45.8% were reportedly tested for HPV in tumor tissue; the proportion coded as unknown declined from 54.6% for 2010 to 34.3% for 2012. The 153 cases reported as HPV-positive comprised 69.2% of the 221 cases with a known HPV test result, which is consistent with the proportions reported in the literature. Trends (2000–2010) in BTT cancer incidence rates in Connecticut were representative of trends in all 18 SEER registries combined. Similar studies are needed from other US central cancer registries that are collecting or want to start collecting HPV status of OP tumors, along with data on the specific types of HPV testing, for surveillance of the frequency and results of HPV testing of OP cancers.

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

Introduction

Incidence rates for all oral cavity-pharynx cancers combined in the US population have been declining, but rates for certain oropharyngeal (OP) cancers (especially base of tongue and tonsil) associated with human papillomavirus (HPV) infection have been increasing, mainly for middle-aged white males. These trends in incidence rates have been reported for the combined cancer registries of the Surveillance, Epidemiology and End Results (SEER) Program registries and the National Program of Cancer Registries (NPCR) for a group of OP squamous cell carcinomas (SCC) regarded as potentially HPV-associated.

Incidence data on HPV-related cancers of the OP, along with other HPV-related sites, are of special interest because incidence rates are projected to increase but there is also the potential for primary prevention with HPV vaccines that protect against some of the HPV types regarded as “high risk” for cancer. The interpretation of OP cancer incidence rates in relation to HPV infection, however, has been hampered by the lack of data on HPV status of tumors.

Using cancer tissue repositories for 3 SEER registries, data were obtained on detection of HPV in formalin-fixed paraffin-embedded specimens of 271 non-randomly selected incident invasive squamous cell carcinomas of the OP, extrapolating the data to the populations covered by these registries, an increase in HPV-positive cases appeared to account for the temporal increases in incidence rates for OP cancers. While such studies using tissue repositories are valuable, they do not allow estimation of the actual frequency of testing for HPV in tumors of all OP cancer patients in the population, and involve only small samples for specific sites within the OP. Data on the actual extent of HPV testing was suggested for all OP cancer patients in the population can be useful to clinicians and researchers. Also useful to researchers would be data on the proportion of cancers that are HPV-positive for specific sites in the OP, for interpreting trends in incidence rates.

A potentially useful source for data on HPV status of OP tumors is the American Joint Committee on Cancer (AJCC) Collaborative Staging (CS) system which includes various site-specific factors (SSFs) on extent of disease and certain prognostic tumor markers, starting with diagnoses in 2010. The CS SSF 10 for OP sites is the HPV status of the tumor. While this item has not been a required SSF for OP cancers in the NPCR, it is required by the SEER Program, and also by US facilities with American College of Surgeons’ Commission on Cancer (CoC) accredited cancer programs; it is required for base of tongue, pharyngeal tonsil and other parts of OP, as an “additional clinically significant” SSF. Data on CS SSF 10 for OP cancers in the SEER-wide research databases.

The present report used data from the Connecticut SEER registry, to explore the potential for using CS SSF 10 on OP cancers to estimate the proportion of OP cancer patients
that have been tested for HPV status of their tumors, and also the proportion of OP cancers that were reported as HPV-positive. First, comparisons between Connecticut and the entire SEER-wide registry data, to assess the similarity of trends in incidence rates for OP cancers.

**Methods**


Age-standardized incidence rates (ASIRs) for invasive OP cancers, and recent trends (2000–2010) in ASIRs, for Connecticut were compared with data for the entire SEER-18 registries. ASIRs are routinely directly adjusted to the age distribution of the 2000 US standard population, using 19 age groups in SEER*Stat. For this report, ages <15 years were excluded.

Annual percent change (APC) in the ASIR was 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, and 95% confidence limits (CLs, lower and upper) defining the confidence interval (CI) were obtained, by using SEER Joinpoint Regression Program (version 4.0.1, January 2013), which examines changes in the direction and/or magnitude of trends over time. Statistically significant ($P < .05$) APCs (with zero as the null hypothesis) were identified by using $P$ values (2-tailed tests) obtained from the joinpoint program.

The latest SEER research database available included data from the November 2013 submission file, with diagnoses in 2000–2010 for all 18 SEER registries. Using the International Classification of Diseases for Oncology third edition (ICD-O-3), the OP cancers selected were base of tongue (BT) (site code C01.9), and tonsils including lingual tonsil (C02.4), pharyngeal tonsil (C09), and Waldeyer ring (C14.2). Base of tongue and tonsil (BTT) are the most common HPV-related sites in the OP, with clear temporal increases in incidence rates and the strongest evidence for a role for HPV in causation. Also examined separately were trends in incidence rates for a combination of sites including lateral wall of OP (C10.2), overlapping lesion of OP not otherwise specified (NOS) (C10.9), pharynx NOS (C14.0), and overlapping lesion of lip, oral cavity, and pharynx (C14.8), which have been regarded as also potentially HPV-associated.

For OP cancer sites, ICD-O-3 Morphology codes <9590 were selected (as in SEER reports); these codes exclude certain Morphologic types in OP (mainly lymphomas). Some analyses examined data for the specific subgroup of squamous cell carcinomas (SCC), defined as ICD-O-3 Morphology codes 8050–8084 and 8120–8131. Data on SSF 10 for OP Cancers (2010–2012) in the Connecticut SEER Registry

Data on SSF 10 for the selected OP cancers were available from the Connecticut SEER registry. The registry covers the entire population of the state and receives (by state regulations) reports from all licensed acute-care hospitals and all licensed clinical laboratories in the state; reciprocal reporting agreements with registries in all adjacent states and several other states (including Florida) cover Connecticut residents diagnosed and/or treated in these states.

Analyses of SSF 10 data for the Connecticut registry included data on invasive BTT cancers diagnosed not only in 2010 but also in 2011 and 2012, in order to examine the most recent incidence data available since the adoption of SSF 10 for OP cancers. At the time the analyses were done, incidence data for 2011 had been submitted (in November 2013) to SEER, for incorporation into analyses of SEER statistics on cancer incidence for 2011. Incidence data for 2012, however, were incomplete and should be regarded as preliminary. Also, all pathology reports for diagnoses in 2012 received by the Connecticut registry had not yet been uploaded into the database management system. Available pathology reports, however, were reviewed for cases with SSF 10 coded as unknown, and recoding was done when additional information was available.

For this report, codes for SSF 10 on the result of any HPV testing done on pathologic specimens from the primary tumor or a metastatic site were combined into 3 groups for some analyses. One group was HPV negative (SSF 10 code 000). The second group (codes 010–070) includes HPV positive subgroups with specific codes for HPV types referred to as “high-risk” (ie, for cancer), including HPV16 and 18 (the highest risk types), with or without mention of “low-risk” types. The third group included cases with HPV testing not known to have been done, which included HPV test ordered but results not in chart (code 997), test not ordered and not done (code 998), and unknown or no information on any HPV testing (code 999). Distributions of these SSF 10 code groups were analyzed by age at diagnosis, gender, and year of diagnosis.

The presence of a pathology report in the Connecticut registry was examined as a potential factor affecting testing for HPV. Of all the 32 hospitals in Connecticut required to report cases to the Connecticut registry, 4 do not routinely provide pathology reports, and this also applies to some out-of-state facilities or sources. For most of the cases without a pathology report in the Connecticut registry, detailed information abstracted from the pathology report (eg, relevant biomarkers) is available in the synoptic summary. At the time of the study, only 1 Connecticut hospital submitted pathology reports electronically.

Statistical analyses included a chi-square test (using Excel) for differences by gender and year of diagnosis in the distribution of SSF 10 findings.

**Results**

**Recent Trend in OP Cancer Incidence Rates in the Connecticut Registry vs SEER-18 Registries**

The ASIR for invasive BTT cancers diagnosed in 2000–2010 at age 15+ years increased for males, with a smaller recent increase for females, in the SEER-18 registries combined and also in the Connecticut SEER registry analyzed separately (Table 1, Figure 1). Rates in the early years were lower for Connecticut than for all SEER-18 registries combined, but trends for recent years (2003–2010)
Table 1. Age-Standardized Incidence Rate (Age 15+ Years) per 100,000 and Annual Percent Change (APC) for Invasive BTT Cancers in All 18 SEER Registries Combined vs the Connecticut SEER Registry Alone

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEER-18</td>
<td>1,560</td>
<td>5.72 (5.44, 6.02)</td>
<td>2,674</td>
<td>7.56 (7.27, 7.86)</td>
<td>+3.4 (+2.7, +4.1)</td>
</tr>
<tr>
<td>(SCC)b</td>
<td>(1,490)</td>
<td>5.46 (5.18, 5.75)</td>
<td>(2,602)</td>
<td>7.36 (7.07, 7.65)</td>
<td>+3.5 (+2.8, +4.2)</td>
</tr>
<tr>
<td>CT Only</td>
<td>49</td>
<td>3.75 (2.77, 4.97)</td>
<td>124</td>
<td>7.88 (6.53, 9.44)</td>
<td>+3.4 (+0.9, +5.9)</td>
</tr>
<tr>
<td>(SCC)b</td>
<td>(46)</td>
<td>3.54 (2.59, 4.72)</td>
<td>(121)</td>
<td>7.62 (6.30, 9.15)</td>
<td>+3.4 (+1.1, +5.8)</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEER-18</td>
<td>435</td>
<td>1.39 (1.26, 1.53)</td>
<td>595</td>
<td>1.54 (1.42, 1.67)</td>
<td>+0.9 (+0.5, +1.3)</td>
</tr>
<tr>
<td>(SCC)b</td>
<td>(416)</td>
<td>1.33 (1.20, 1.46)</td>
<td>(561)</td>
<td>1.45 (1.33, 1.58)</td>
<td>+0.9 (+0.4, +1.4)</td>
</tr>
<tr>
<td>CT Only</td>
<td>14</td>
<td>0.92 (0.50, 1.55)</td>
<td>27</td>
<td>1.48 (0.97, 2.17)</td>
<td>+1.0 (-3.3, +5.6)</td>
</tr>
<tr>
<td>(SCC)b</td>
<td>(13)</td>
<td>0.87 (0.46, 1.49)</td>
<td>(27)</td>
<td>1.48 (0.97, 2.17)</td>
<td>+0.7 (-3.3, +4.8)</td>
</tr>
</tbody>
</table>

bBTT is base of tongue (ICD-O-3 site code C019) and tonsils (C09, C024, C142), with ICD-O-3 Morphology codes <9590 (excluding lymphatic and hematopoietic) (see text).

CT, confidence limits (lower and upper [95%] which define the confidence interval [for the APC, a CL can be a negative number]).

SEER, Surveillance, Epidemiology, and End Results Program.

Figure 1. Trend in Age-Standardized Incidence Rate (Age 15+ Years) per 100,000 for Base of Tongue and Tonsils, for the SEER-18 Registries vs Connecticut (CT) SEER Registry

were similar; 95% CIs on APCs were wider for Connecticut (reflecting the smaller numbers of cancers) than for SEER-18 (Table 1, Figure 1). Trends were similar for the subgroup coded as SCC, which comprised 93%+ of all cases with the selected site and Morphology codes, and APCs were similar for the Connecticut and SEER-18 registries (Table 1).

For the group of OP sites other than BTT regarded as potentially HPV-associated (ie, as defined above) in Connecticut, numbers of cases by gender in some calendar years were small (<10) and precluded statistical analysis of trends in rates. Therefore, analyses of data on SSF 10 in Connecticut were limited to BTT cancers (as defined above), for this preliminary report.

Distribution of SSF10 (HPV Status) for BTT Cancers in the Connecticut Registry

Almost all the 483 invasive BTT cancers (477 or 97.3%) diagnosed in Connecticut residents in 2010–2012 were coded as SCC (as also evident for diagnoses in 2010, Table 1). Mean age at diagnosis of the 483 cases was 60.5 (median 59.0) years, with 66 cases (13.7%) <50 years, and the largest subgroups were 50-59 years (180 cases or 37.3%) and 60–69 years (146 cases or 30.2%) (data not tabulated). Of all 483 cases, 221 (45.8%) were reportedly tested (ie, 153 as positive and 68 as negative) for HPV in tumor tissue, 54 cases (11.2%) were not tested, and 208 (43.1%) were coded as unknown with regard to testing (Table 2).

The distribution of codes for SSF 10 differed little by gender, but differences were evident by year of diagnosis. The proportion of all 483 cases coded as HPV test not done declined from 15.8% in 2010 to 6.5% in 2012, and the proportion coded as unknown with regard to any HPV testing (SSF 10 code 999) declined from 54.6% in 2010 to 34.3% in 2012 (Table 2). The 153 HPV positive cases comprised 31.7% of all 483 cases, but the proportion positive increased from 20.3% in 2010 to 40.8% in 2012 (Table 2).

There were 262 cases with SSF 10 coded as 997–999 (HPV test reportedly not done or unknown if done) and 221 coded as 000-070 (HPV reportedly tested and HPV test result known) (Table 2). The distribution of age at diagnosis (<50, 50–59, 60–69, 70–79 and 80+ years) differed slightly (median age 61 vs 58 years) but not statistically significantly between the 2 groups (P = .111 for chi-square test) (data not tabulated).

HPV-Positive Cases as a Proportion of All Cases with a Known HPV Test Result in Connecticut Registry

The 153 HPV-positive cases comprised 69.2% of all 221 cases coded with a known HPV test result (either positive or negative) (Table 3). This proportion was slightly higher in males (71.2%) than females (61.4%), and declined from...
77.8% for diagnoses in 2010 to 64.5% in 2011 and 69.0% in 2012, but these associations were not statistically significant (Table 3).

Completeness of data on high vs low risk types of HPV was also examined. The 153 HPV-positive cases included: 88 cases (57.5%) coded as SSF 10 code 30 (only HPV16 type detected); 10 cases with code 50 (both HPV16 and 18); 7 cases as code 20 (high-risk type other than 16 or 18); and 16 cases as high-risk type not otherwise specified (NOS) for a total of 121 cases (79.0%) coded as high-risk types; the remainder were 7 cases coded as 010 (low-risk types only) and 25 (16.3%) as 070 (HPV positive, unknown or unspecified risk type) (data not tabulated).

Table 2. Distribution of Data on AJCC Collaborative Stage Site-Specific Factor 10 (SSF 10) on Human Papillomavirus (HPV) Testing of Tumor, for Incident Invasive BTT* Cancers Diagnosed in 2010–2012 in the Connecticut Tumor Registry

<table>
<thead>
<tr>
<th>SSF 10 (codes)</th>
<th>Total</th>
<th>Gender</th>
<th>Year of Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>Male</td>
</tr>
<tr>
<td>Negative (000)</td>
<td>68</td>
<td>14.1</td>
<td>51</td>
</tr>
<tr>
<td>Positive (010-070)</td>
<td>153</td>
<td>31.7</td>
<td>126</td>
</tr>
<tr>
<td>Not done (997,998)</td>
<td>54</td>
<td>11.2</td>
<td>42</td>
</tr>
<tr>
<td>Unknown (999)</td>
<td>208</td>
<td>43.1</td>
<td>174</td>
</tr>
<tr>
<td>Total</td>
<td>483</td>
<td>100.0</td>
<td>393</td>
</tr>
</tbody>
</table>

P value: gender, P = .377; year, P < .0001.

*BBT is base of tongue (ICD-O-3 site code C019) and tonsils (C09, C024, C142), with ICD-O-3 Morphology codes <9590 (excluding lymphatic and hematopoietic) (see text).

Table 3. Human Papillomavirus (HPV) Test Result for BTT* Cases in the Connecticut Tumor Registry Coded as Either HPV Positive or HPV Negative Using Site-Specific Factor 10 (SSF 10)*

<table>
<thead>
<tr>
<th>HPV test result</th>
<th>Total</th>
<th>Gender</th>
<th>Year of Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>Male</td>
</tr>
<tr>
<td>Positive</td>
<td>153</td>
<td>69.2</td>
<td>126</td>
</tr>
<tr>
<td>Negative</td>
<td>68</td>
<td>30.8</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>221</td>
<td>100.0</td>
<td>177</td>
</tr>
</tbody>
</table>

P value: gender, P = .207; year, P = .308.

*BBT is base of tongue (ICD-O-3 site code C019) and tonsils (C09, C024, C142), with ICD-O-3 Morphology codes <9590 (i.e., excluding lymphatic and hematopoietic).

Table 3. Human Papillomavirus (HPV) Test Result for BTT* Cases in the Connecticut Tumor Registry Coded as Either HPV Positive or HPV Negative Using Site-Specific Factor 10 (SSF 10)*

<table>
<thead>
<tr>
<th>HPV test result</th>
<th>Total</th>
<th>Gender</th>
<th>Year of Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>Male</td>
</tr>
<tr>
<td>Positive</td>
<td>153</td>
<td>69.2</td>
<td>126</td>
</tr>
<tr>
<td>Negative</td>
<td>68</td>
<td>30.8</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>221</td>
<td>100.0</td>
<td>177</td>
</tr>
</tbody>
</table>

P value: gender, P = .207; year, P = .308.

*BBT is base of tongue (ICD-O-3 site code C019) and tonsils (C09, C024, C142), with ICD-O-3 Morphology codes <9590 (i.e., excluding lymphatic and hematopoietic).

See Table 2 and text for details on SSF 10 coding.

Chi-square test for differences in distribution of SSF 10 codes by gender or by year.

Availability of Pathology Reports in Relation to SSF 10 Codes in Connecticut Registry

At the time this study was done, a full pathology report was available in the Connecticut registry, for 264 (54.7%) of all 483 cases; multiple reports were available for the OP tumor of certain patients, resulting in about 600 pathology reports available for review. Missing reports were due in part to the Connecticut hospitals and out-of-state reporting sources that do not routinely providing full pathology reports and the Connecticut registry; together, these sources accounted for about 16% of the BTT cancer cases.

For the 262 cases with SSF 10 codes indicating that HPV testing was not known to have been done (codes 997–999, but predominantly code 999 for unknown), a pathology report was available for 148 (56.5%). This included 42 (77.3%) of 54 cases coded as 997 or 998, and 106 (51.0%) of 208 cases coded as 999. Availability of a pathology report(s) varied by year of diagnosis: 63.6% (68 of 107 cases) for 2010, 69.8% (60 of 86 cases) for 2011, and only 29.0% (20 of 69 cases) for 2012; as noted above, Connecticut registry data for 2012 were incomplete at the time of this study.

The availability of pathology reports for SSF codes 997–999 combined (56.5%) differed little from that for cases with SSF 10 coded as having HPV status known (codes 000–070) (116 of 221, or 52.5%).
Quality Control Study of a Sample of Connecticut Registry Cases with SSF 10 Coded as Unknown

A sample of 27 BTT cancer cases (10 diagnosed in 2010, 11 in 2011 and 6 in 2012) with SSF 10 coded as 999 were reviewed by 1 of the authors (CP, a certified tumor registrar).

These cases represent 25% of all 106 cases with SSF 10 coded as unknown that had a pathology report. Four (14.8%) of the 27 were found to have information on HPV testing (ie, positive or negative) either in a pathology report or (for 1 of the 4 cases) a radiotherapy consult report; 3 of the 4 cases, however, were diagnosed in 2010.

Discussion

Prospects for Estimating the Proportion of Incident OP Cancers Tested for HPV

Invasive BTT cancer incidence rates for Connecticut were representative of all SEER-18 registries for recent years, and temporal increases in rates for both males and females were evident for Connecticut as well as in all SEER areas combined (Table 1, Figure 1). The well-known gender difference in OP cancer rates and in the magnitude of recent increases (Table 1, Figure 1) are poorly understood, although differences in behaviors and in prevalence of oral HPV16 infection may be involved.

Analyses of SSF 10 data on HPV testing of tumor tissue among 483 incident invasive BTT cancers in the Connecticut SEER registry showed a decline from 2010 to 2012 in the proportion reportedly not tested for HPV and in the proportion unknown with regard to testing (Table 2), and no clear trend in the proportion positive among those with a known HPV test result across these three years (Table 3). These findings suggest improvement in reporting of SSF 10 data to the Connecticut registry and/or a rise in the rate of HPV testing of tumors in Connecticut.

The extent of clinical testing for HPV in all OP patients may depend largely on the clinical utility of such testing (eg, for prognosis). With regard to clinical utility, a tissue-repository study of a sample of OP patients in three SEER registries found higher survival rates for HPV-positive (vs HPV-negative) OP cancer patients, a findings that is consistent with reports from clinical series and from population-based registries in other countries.

Some head and neck oncology groups reportedly have concluded that HPV-related testing should be part of the standard work-up for patients with SCC of OP. In the National Comprehensive Cancer Network (NCCN) clinical practice guidelines for OP cancers (BTT, soft palate, and posterior pharyngeal wall) published in 2011, clinical “tumor HPV testing” was “suggested.” The 2 specific tests mentioned were HPV DNA testing by in situ hybridization (ISH) and p16 protein immunohistochemistry (IHC) staining of tumor tissue. Overexpression of p16, evidenced by intense staining of cells, results from the effect of an oncoprotein produced by HPV, and is a proposed surrogate marker for HPV status.

This p16 marker cannot distinguish HPV type (eg, HPV16) but has been associated with better prognosis in some studies. Recommendations by oncology groups may have affected recent clinical practice regarding testing for HPV in tissues of OP cancer patients.

Furthermore, testing of tumors for HPV or using surrogate markers has been proposed as an adjunct to TNM staging, for use in prognosis of patients with OP SCC. HPV-positive OP cancer has been associated with reduced risk of second primary cancer diagnosis (including synchronous cancers), in comparison to OP cancers associated with a history of smoking and alcohol use. HPV-positive OP cancer patients tend to have a better prognosis, despite more extensive regional lymph-node involvement at diagnosis. A higher risk for HPV-positive (vs HPV-negative) OP cancer patients has been reported for distant metastases after a long time interval of follow-up after radiotherapy or chemoradiotherapy; a small number may have prolonged survival after treatment. Such reports may contribute to increased clinical interest in testing for HPV status of newly diagnosed OP cancers.

With regard to understanding the completeness of information on SSF 10 for HPV testing of OP tumor tissue, one factor examined was the presence of a pathology report for BTT cancer cases in the Connecticut registry. The proportion of cases with a pathology report available in the central registry did not differ between cases with SSF 10 coded as known (codes 000–070) vs unknown or uncertain (codes 997–999) with regard to HPV testing of tumor tissue. This suggests that hospital tumor registrars tend to use information available on HPV testing from pathology reports in their coding of SSF and provide such information with the summary of pathology findings in case reports routinely submitted to the Connecticut central registry.

Review of a sample of 27 cases coded with SSF 10 as 999 (unknown), however, found that all information on HPV status may not be captured by the hospital and by routine quality-control efforts at the Connecticut central registry. Assuming a random sample, the estimated error rate would be 14.8% (95% CI, 1.4%–28.2%); thus, an error rate <1.4% is statistically unlikely. Noteworthy, however, is that finding that 3 of the 4 cases with coding errors were diagnosed in 2010, suggesting improvement over time.

Similar assessments of quality of SSF data on OP cancers are needed from other US central cancer registries that collect this item. Hospital and central cancer registries will differ in such factors as the availability of pathology reports (including supplemental reports) that may contain information on HPV testing, and also in the extent of electronic pathology reports that would permit searching of text (relevant to HPV status) in quality control. In addition, completeness of central registry data on HPV testing being ordered by clinicians could be evaluated by patterns-of-care type studies; such studies could query physicians regarding their knowledge of any HPV-testing of tissues, for samples of OP patients randomly selected from the registry database.

Such information on the actual extent of HPV testing of newly-diagnosed OP cancer patients, as estimated from population-based cancer registries, should be of interest to clinicians using (or considering) HPV testing for prognostic purposes In addition, identification of OP cancers actually
tested for HPV, with known results, in central cancer registries could provide a centralized resource for researchers seeking to identify persons with HPV-positive and/or HPV-negative tumors for potential enrollment in clinical and epidemiologic studies of risk factors or prognostic factors.

**Findings on the Proportion of BTT Cancer Cases in Connecticut with an HPV-positive Test Result**

The potential utility of SSF 10 data on OP cancers for surveillance and interpretation of future OP cancer incidence rates by a central cancer registry would depend on the ability to estimate the proportion of all cancers that are HPV-positive. On the basis of CS SSF 10 data for 483 incident BTT cancers in 2010-2012 in Connecticut, the minimum proportion HPV-positive (assuming accuracy of the reporting and coding) would be 31.7% (Table 1). For the subgroup of 221 cases coded as having been tested for HPV with a known result, the proportion reported as HPV positive was 69.2% (Table 3). Thus, the actual proportion of BTT tumors that are HPV positive may be within 32-69% (Tables 2 and 3).

Limitations of the present study must be recognized. The 69.2% figure may be an overestimate, because cancer patients may have been selected for HPV testing based on characteristics (eg, smoking history) and/or clinical characteristics (eg, stage at diagnosis) that predict the likelihood of being HPV-positive. The true proportion of OP cancers that are HPV positive also would depend on the proportion HPV positive among all tumors with SSF 10 coded as uncertain or unknown (codes 997-999) with regard to testing for HPV. A study limitation is that not all full pathology reports (vs summaries) received for diagnoses in 2012 had been processed at the time of the study. Also, the availability of a pathology report in the central cancer registry for some cases with SSF 10 coded as 997-999 does not rule out the possibility of an HPV-positive tumor, because supplemental pathology reports may have been missing, especially from facilities without electronic reporting of pathology reports. This issue needs to be addressed in future studies.

With regard to data on subtypes of HPV based on SSF 10 codes, 79.0% of HPV-positive BTT cases in the Connecticut registry involved high-risk HPV types, mostly HPV16 with or without HPV18, whereas the literature indicates >90% HPV16 among HPV-positive OP cancers. A study limitation, however, is that 16.3% of HPV-positive BTT cases in Connecticut were coded as NOS (type unspecified), which may have included cases with HPV16.

Despite the limitations, the estimate of 69% of BTT tumors that are HPV positive (Table 3) does not differ greatly from estimates reported in the literature, which have varied, due to such factors as differences in OP cancers selected and in the techniques used to assess HPV status in tumor tissue. A figure of 72% was reported for polymerase chain reaction (PCR) genotyping of OP cancers (2000-2004) in 3 SEER registries. An estimate of 63% (95% CI, 50%-75%) for HPV-associated OP cancers attributable to HPV, from a literature review, was used in a CDC report. In a review of North American studies with HPV DNA measured by sensitive PCR-based assays, HPV prevalence was 47.0% (95% CI, 41.1%-53.0%) for cancers of the vallecula, walls of oropharynx, and tonsils combined.

Limitations of the coding rules for CS SSF 10 on HPV status of OP cancers also must be recognized. There is no information on the specific test used to detect HPV in tumor tissue. Coding rules exclude using HPV detected in patients by blood tests (serology) alone. Although IHC staining from pathology reports has been mentioned in AJCC coding instructions for SSF 10, a recent AJCC ruling is that cases tested with only the p16 IHC marker should be coded as 999 (unknown). Finally, the presence of HPV DNA in pathologic tissue is not synonymous with an actual role for HPV infection in the pathogenesis of cancer, and specific tests differ in analytical performance as evaluated against various proposed “gold standard” tests for oncogenic HPV infection such as the presence of intact HPV16 mRNA molecules amplified by PCR in fresh tumor samples.

In view of these limitations, studies of SSF 10 data on OP cancers from other SEER registries should also collect information on the specific HPV test(s) done for random samples of individual patients in the geographic area.

**Conclusions**

Data from the Connecticut SEER registry show temporal increases in incidence rates for BTT cancers that are also evident for all 18 SEER registries combined. Data from the Connecticut registry suggest the potential utility of CS SSF 10 on HPV status of BTT cancers, for monitoring current and future rates of actual testing for the presence of HPV in tumor tissues of incident OP cases in defined populations. Studies are needed from other US cancer registries that have collected data on SSF 10 for OP cancers, such as those in the >1500 CoC-accredited facilities contributing to the American College of Surgeons–American Cancer Society National Cancer Data Base, which is not population-based but covers an estimated 70% of all incident cancers in the United States. Central cancer registries not already collecting SSF 10 for OP cancers may want to consider initiating collection.

Special efforts would be desirable to confirm cases reported as not known to have been tested for HPV in tumor tissue, and also to identify the specific pathologic tests used for those reportedly HPV positive or negative. Estimating the proportion of all registered OP cancers that are HPV-positive is challenging, but could enhance the interpretation of surveillance data on trends in incidence of “HPV-related” OP cancers in defined populations.

**References**


7. American Joint Committee on Cancer, Collaborative Stage Work Group of the American Joint Committee on Cancer. Collaborative Stage Data Collection Coding Instructions, Version 02.00.00. Incorporates updates through January 1, 2010. Chicago, IL.


29. Collaborative Stage Work Group of the American Joint Committee on Cancer. Collaborative Stage Data Collection System Coding Instructions, version 02.00.00. Chicago, IL: American Joint Committee on Cancer; 2010: I-2-7.


Misclassification of Sex in Central Cancer Registries

Recinda L. Sherman, MPH, PhD, CTR; Francis P. Boscoe, PhD; David K. O’Brien, PhD, GISP; Justin T. George, MPH; Kevin A. Henry, PhD; Laura E. Soloway, PhD; David J. Lee, PhD

Abstract: Background: Intrarecord edits on site-sex combinations are a standard tool to identify errors in the coding of sex in cancer registry data. However, the percentage of sex-specific cancers, like cervix, is low (20% of total invasive cases). Visual review and follow-back to improve the quality of the sex coding is labor intensive and typically only performed as a special project on subsets of data. The New York State Cancer Registry (NYSCR) created an edit for identifying potential sex misclassification in cancer registry data and has made its components available for use through the North American Association of Central Cancer Registries (NAACCR). The edit uses the most popular male and female first names based on decade of birth to identify potentially miscoded cases. This paper provides a summary of 3 independently conducted assessments of the sex edit at the central cancer registry level and includes a focus on misclassification of sex for breast cancer. Methods: The sex edit was applied in 3 state cancer registries: Alabama, Alaska, and Florida. Alabama applied the edit to their entire database for 1996–2004 (N = 190,614) and compared the results to external databases available to most cancer registries. Alaska applied the edit to their entire database (N = 46,645) and were able to compare the results to 2 unique, state-based databases (Alaska Permanent Fund Dividend database and State Troopers database). Florida applied the sex edit to a sample of sites (n = 953,074) with particular attention to breast cancer. Results for breast cases were compared to results from an a priori quality control project on Florida male breast cancer cases. Using the Florida data, issues specific to male breast cancer were evaluated. Results: In Alabama, 45% of 977 cases flagged as potentially miscoded sex were determined to be miscodes. In Alaska, 19% of 88 cases flagged as potentially miscoded sex were determined to be miscodes but the percent of miscoded cases identified by the edit more than doubled in the most recent years of data. For the Florida male breast cancer comparison, the sex edit correctly identified 729 of 903 cases known to be miscoded (81%) and was unable to assign a potential sex on the remaining 174 cases—but did not incorrectly flag any cases as miscodes. Implications: The sex edit is a useful tool for identifying cases that require further review to confirm the reported sex code is correct. However, it only asesses 69%–84% of cases based on name and, of those flagged, only 19%–45% are true misclassifications. But for breast cancer, a site with a skewed male to female ratio, the verified misclassification rate was 100% of the male breast cancer cases flagged as potential females. The proper application of the sex edit can improve the quality of the sex variable and can greatly reduce the impact of miscoded sex on gender-skewed sites like male breast cancer.

Key words: automated edit, bias, breast cancer, cancer registries, data quality

Introduction

Missing data in cancer registries is due to either the absence of the data in the clinical assessment, such as an unstaged case due to contraindicated comorbidities, or the failure of the surveillance system to capture the information.\(^1\) In the case of a demographic variable like sex, it is rarely missing but may be miscoded due to a clerical error either during patient intake or when the certified tumor registrar (CTR) abstracts the case. Although there is little documentation of the impact of missing or miscoded data on research results,\(^1\) it is likely that miscoded data, even due to a random clerical error versus the more problematic systemic error that can occur with software problems, can have profound impacts on research results.

For instance, a CTR will generally abstract a comparable number of male and female cases. A miscoded sex for a sex-specific tumor, like cervical or prostate, can be easily identified using an automated edit, but a clerical error on the sex code for other cancer sites may never be identified once the case is sent to the cancer registry. If the miscode rate is low, say 1 in 500, and the miscode is a random clerical error and not a systemic error, there will be essentially no impact on rates, except in the cases of cancers with a skewed male to female ratio.

Breast cancer, for instance, is 0.5%–1% male with about 2,000 male breast cancer cases diagnosed annually in the United States.\(^2\) With approximately 200,000 new female breast cancers a year and 2,000 male, and a theoretical random miscode rate of 1 in 500, 400 women would be errantly coded as men but only 4 men would be miscoded as women. Male breast cancer cases would actually be 2,396 and the female cases would be 199,602. That is a 20% increase in the crude male breast cancer rate but only a 0.2% decrease in the crude female breast cancer rate.

Such miscodes can produce invalid rates as well as bias research results. But visual review and follow-back to improve the quality of the sex coding is labor intensive and so is typically only performed as a special project on subsets.
of data. Most registries use intrarecord edits on site-sex combinations as a standard tool to identify errors in the sex code in cancer registries, which only apply to sex-specific sites. But the New York State Cancer Registry (NYSCR) created an edit for identifying potential sex miscodes for all sites and has made its components available for use through the North American Association of Central Cancer Registries (NAACCR). The edit uses the most popular male and female first names based on decade of birth to identify potentially miscoded cases. For specific names that are gender-specific to the opposite gender in the United States compared to other countries (eg, Jean, Carmen, Andrea, Angel), the edit is not used if the person was born outside the United States.

Florida was one of the first states to apply the NYSCR sex edit to their registry. The Florida registry is one of the largest cancer registries in the world, with more than 115,000 incident cancer cases collected annually. The Florida registry relies heavily on automated processes to ensure data integrity including automated site-sex edits. Manual quality control projects to improve data accuracy are considered impractical for standard registry operations in Florida.

In 2002, researchers in Florida were concerned that the rates of breast cancer among Floridian men were higher than for men nationally. The data indicated that breast cancer incidence rates were increasing at a faster, statistically significant, rate in Florida males compared to the SEER-9* males. Studying such a high-risk population could be important in advancing etiologic knowledge about the disease. However, before drawing research conclusions, potential spuriousness of results that can occur due to underlying data errors, such as miscoded sex, should be evaluated.

This paper describes the application of the NYSCR automated sex edit to improve the coding of sex in registry data using 3 example states: Alaska, Alabama, and Florida. Extensive detail is given to the issue of sex miscodes of breast cancer cases in Florida.

Methods

New York State Sex Edit

The sex edit was developed by the NYSCR. It evaluates names that are highly correlated with gender and flags suspicious name-sex combinations. Many names have been gender-specific for centuries (eg, Elizabeth and Charles), but occasionally the gender associations of names change over time (eg, Rosario was a typical male name in 1900 but became typically female in 1940 forward). The edit uses the Social Security Administration database of the 1,000 most popular male and female names for each decade from 1890-2008. Names with at least a 49:1 ratio of one sex to the other were branded as sex-specific. This list of sex-specific names is matched against the names in the cancer registry and potentially miscoded name-sex combinations can be identified for review.

Multi-State Assessment

The sex edit was evaluated for use in 3 uncoordinated efforts in the Alabama, Alaska, and Florida central cancer registries as part of state-specific quality control (QC) registry operations.

The Alabama assessment tested the edit against the database for cases diagnosed from 1996-2004. If a case was flagged as a potentially miscoded sex, descriptive text from all the original source records was first reviewed. If the text review was inconclusive, the prefix, suffix and spouse name fields were also reviewed for confirmation of sex. If no determination of sex could be made, the patient’s vital status was checked. If the patient was alive, sex was confirmed when possible based on an external, prior linkage with state Medicaid data. If the patient was deceased, sex was confirmed using state and national death files. If the sex code still remained unresolved, the code was determined based on staff judgment using primary site and name.

The Alaska assessment tested the edit against the entire database from 1996-2009. If a case was determined to be a potentially miscoded sex, descriptive text from all the original source records was first reviewed. If the text review was inconclusive, the patient’s sex was manually looked up in the Alaska Permanent Fund Dividend (PFD) database. The PFD database contains demographic information, updated annually, on approximately 95% of Alaska residents who submit an application to share the interest on royalties paid by the petroleum industry to the Alaska state government for transporting oil thorough the Trans-Alaska Pipeline. If the PFD review was inconclusive, then the record was reviewed using the Alaska Department of Public Safety’s State Troopers database. The Troopers database includes demographic and driver’s license-related information for individuals with a driver’s license or other state issued identification, who have been fingerprinted, or who have had contact with state or local law enforcement.

The Florida assessment tested the edit against breast, thyroid, liver, and colorectal cancers diagnosed from 1981-2008. The Florida assessment also compared the results of the sex edit with the results of a quality control (QC) project conducted in 2003 on the accuracy of reported sex of male breast cancer cases. The QC project was prompted by concerns regarding an increase in male breast cancer among Floridian men, and it assessed male breast cancer cases diagnosed in Florida from 1981-2000. For the QC project, cases determined by visual review to be female names were followed back with a letter to the hospital to confirm the sex of the patient.

Results

Alabama

In Alabama, 190,164 cases were evaluated; about 0.5% (977) were flagged as potentially miscoded sex and 44%

*SEER (Surveillance, Epidemiology, and End Results) registries are population-based cancer registries funded by the National Cancer Institute. SEER data was used as the comparison to more closely match the available years of diagnosis in Florida. NPCR (National Program of Cancer Registries) funded by the Centers for Disease Control and Prevention covers the majority of the nation but start with diagnosis year 1995. The SEER-9 registries are Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco–Oakland, Seattle–Puget Sound, and Utah.
(429) were confirmed as miscodes and corrected in the registry data. Sixty-six percent of potentially miscoded cases were reported as female. Of the 429 changed, registry personnel could find corroborating evidence for the change for 283 (66%) and the remainder were changed based on visual review by the registry personnel. Of the 548 cases where sex remained unchanged, registry personnel could find corroborating evidence in support of the reported sex for 388 (71%). The remainder could not be confirmed, so the codes were left unchanged and the cases will eventually be re-reviewed with additional years of death data. There were 332 of 738 (45%) non-Hispanic whites confirmed as having miscoded sex, but only 4 of 14 (29%) Hispanics and 81 of 204 (40%) non-Hispanic blacks were confirmed as having miscoded sex. These groups are more likely to have unique names or names that have a less common gender affiliation, such as Angel being a common name for Hispanic males but a common name for non-Hispanic white females. There were an additional 4 of 4 (100%) cases of unknown sex that were updated with a known sex after being flagged with a potentially miscoded sex. The number of cases for which a potential sex could not be determined by the edit was not recorded.

Alaska

Of the 46,645 consolidated cancer cases in the Alaska databases, 16% (7,303) could not be assigned a potential sex by the edit because their first names were not gender-specific or were not common enough to be ranked. There were 88 cases that were flagged as a potentially miscoded sex and underwent manual review. During manual review, it was determined that several names were either misspelled or were nicknames and were corrected to their formal first name. The corrected names were appropriate for the sex, such as Louis vs Lois and Marty vs Martha. Using either the accompanying text in the source abstracts, the PFD database or Troopers database, sex was confirmed for all 88 cases. Of the 88 cases flagged with potentially miscoded sex, 19% (17) were confirmed to be miscoded and their coded sex was corrected in the registry data. An assessment by year indicated that the percent of cases truly misclassified by sex is higher for more recent years with 31% of the potentially miscoded sex cases identified as true miscodes for cases diagnosed in the last 2 years, 2008–2009. Increasing demands on the CTR may be resulting in increased clerical error, but it is likely that a small registry like Alaska is able to identify most miscoded sex cases through visual review and use of the data over time.

Florida

A data quality project was undertaken in Florida to evaluate the sex coding of breast among males. The first name of male breast cancer patients diagnosed from 1981–2000 were visually reviewed. A total of 904 of approximately 3,800 male cases of breast cancer were identified as potentially female based on first name. All but 3 were confirmed female by the hospitals, and the sex code was corrected in the registry data.

Figure 1 illustrates the number of breast cases that were misclassified as men (“fake men”) by year of diagnosis from 1981–2000. It is clear that sex misclassification for breast is more problematic with later diagnosis years. This is likely due to changes in International Classification of Diseases for Oncology (ICD-O) coding. Prior to the 1990s, the ICD-O classification system was similar to the International Classification of Diseases (ICD)-9 classification with separate codes for female (174) and male (175) breast cancers. Starting with ICD-O-2 in the early 1990s, breast cancer became a single code (C50) regardless of sex. This level of misclassification can significantly inflate breast cancer rates in males because it is a rare cancer while only negligibly altering rates in females.

The sex edit was tested against the original 904 cases manually followed back to hospitals in the 2003 QC project. The edit correctly identified 729 (81%) of the “fake men” as female plus 1 of the 3 “real men” breast cancer cases as male. The remaining cases were not assessable because the name was not gender-specific. Although the edit could not determine a potential sex code for 175 of the cases, the edit did not misclassify the sex of any of the male breast cancer cases.

Most (648,769, or 68%) of the 953,074 cases in the site-specific evaluation of the sex edit agreed with the edit’s potential sex. About 31% were indeterminate: 298,888 had non–gender-specific first names, 68 had a missing year of

<table>
<thead>
<tr>
<th>Site</th>
<th>% Potentially Miscoded</th>
<th>% Not Determined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>21.00%</td>
<td>31%</td>
</tr>
<tr>
<td>Thyroid</td>
<td>1.30%</td>
<td>29%</td>
</tr>
<tr>
<td>Liver</td>
<td>0.30%</td>
<td>29%</td>
</tr>
<tr>
<td>Colorectal</td>
<td>0.50%</td>
<td>33%</td>
</tr>
</tbody>
</table>

Table 1. Percent Identified as Potential Miscoded Sex and Percent Not Assessed by Edit, in Florida
birth, 595 had a reported year of birth born prior to 1890, and 90 cases were coded in the registry as hermaphrodite or transgender. There were 4,519 (0.5%) cases that were identified as potential sex miscodes. Additionally, 145 cases were coded as unknown sex in the registry but the edit identified a potential sex.

Results varied by site (Table 1). Over a fifth, 21%, of breast cancer patients reported as male were identified as potentially miscoded sex (0.2% for breast cancer cases reported as female). Breast cancer is about 100 times more common in women than men, so the count of potential miscodes for each sex were close; 1,076 cases reported as male were identified as miscodes and 984 reported as female were identified as miscodes. For thyroid, a site 3 times more common in women than in men, 1.3% of the thyroid cases reported as male were identified as potential miscodes (vs 0.4% for thyroid cases reported as females). For liver, a site 3–8 times more common in men than women, 1.1% of the liver cases among females and 0.3% among males were identified as a potential miscoded sex. For these sex-skewed sites, the sex ratio of cases identified as potential miscoded sex is exactly inversely proportional to the sex ratio of the cancers themselves. For colorectal cancer, a site with similar rates for both sexes, the percent of cases identified with a potential miscoded sex was similar for cases reported as men (0.5%) and as women (0.6%).

The utility of the edit was higher for non-Hispanic whites than other race/ethnicity categories (Table 2). As with Alabama, a greater proportion of Hispanics and non-whites were not identified with a potential sex by the edit—meaning the date of birth was missing, the decade of birth was prior to 1890 (72 cases), or the name was not gender-specific or not popular enough to be ranked in the top 1,000 most common names by decade (55,393 cases).

### Table 2. Percent of Cases With Potential Sex Not Determined by Edit by Race, Ethnicity in Florida

<table>
<thead>
<tr>
<th>Race</th>
<th>% Not Determined</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>31%</td>
</tr>
<tr>
<td>Black</td>
<td>34%</td>
</tr>
<tr>
<td>American Indian</td>
<td>37%</td>
</tr>
<tr>
<td>All Others</td>
<td>55%</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>46%</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>30%</td>
</tr>
</tbody>
</table>

**Discussion**

It is clear the sex edit can be used to improve the quality of sex coding in cancer registry data. The extent of misclassification of sex is low as evidenced by the results from the 3 registries. But even a few cases of miscoded sex a year can potentially impact rates of rare cancers or in small-area analyses.

Alaska is one of the least populated states with one of the smallest registries, making it reasonable to conduct manual follow-back. The difference of 19% true misclassification in the early years compared to 31% in the more recent years may reflect the on-going QC efforts of registry staff in Alaska. So even in states like Alaska, the sex edit can be effectively applied proactively to address misclassification of sex. In all states, the sex edit can reduce the extent of manual follow-back, which is a significant barrier to QC efforts in many registries.

Using the sex edit can reduce the impact of miscoded sex on male breast cancer rates, as we can demonstrate using the Florida data. Specifically, the results of the comparison of sex edit to the QC project indicates the edit is reliable enough to reclassify all reported male breast cancer cases indicated as potential miscodes to female, a total of 1,076 cases. Ideally, we would like to conduct manual follow-back of the Florida data to determine which potential miscodes are truly miscoded. But Florida is a large registry and relies heavily on automated algorithms and edit checks with very little follow-back, unlike Alaska. If we apply the 45% correct miscoded proportion from Alabama for the potentially miscoded females, we would recode 440 of the female breast cancer cases as male. So we would move 1 of 78 previously unknown sex cases to male based on the potential sex identified by the edit. When we compare the results in Figure 2, we see that the rates of male breast cancer actually increase for diagnosis years 1981–2000 and decrease for more recent diagnosis years in Florida. This is because the historical QC project only resolved miscoded breast cases reported as male and did not review cases that were reported as female for any potential miscodes. Focusing on improbable sex for male breast cancers only removes female cases miscoded as male, which falsely suppresses the rate because no misclassified female breast cancer cases are added back into the male category. For more recent years, no manual resolution of male breast cancer patients was conducted so the impact of recoding sex based on the edit was an overall decrease in male breast cancer rates in more current years.

One limitation of the edit is an inability to determine a potential sex for many cases due to the lack of assessment for unusual names, for cases lacking date of birth, and for
decades for which the data is not incorporated into the edit. There is a current NAACCR effort to incorporate more recent years of data (2009 year of birth forward) to update the edit. But the edit will continue to be less effective for minorities with names that are less likely to be popular. In addition, first generation males may not be accurately determined or not determined at all. For instance, Andrea, Angel, Carmen, and Jean are common names for females in much of the United States but are male names among Hispanics, Haitians, and Italians. These names are excluded from the edit for patients who are foreign-born. But they had to be removed completely from the edit when applied in Florida based on a preliminary review of the results. Similar adjustments might need to be made that could be informed based on other state’s demographic profile.

Also, a registry may be tempted to automatically change the sex of all cases identified as potentially miscoded rather than committing the fiscal resources and personnel needed for follow-back or confirmation from secondary sources. However, the edit was intended to be used to flag cases for further follow-back only. The results from Alabama and Alaska indicate that, overall, the sex edit correctly flags a true miscoded sex as a potentially miscoded sex less than 50% of the time. If the sex edit is implemented at the central registry level, sex must be confirmed through an external source—not automatically updated. However, for cases where sex is unknown, registries that are unable to perform manual review, either due to large size, like Florida, or lack of access to useful outside data, might consider assigning the potential sex identified by the edit. This should be documented so that if additional information is reported to the registry it is assessed appropriately.

Breast cancer, however, is a special circumstance. When applying the sex edit, the percentage of cases flagged as a potential miscoded sex that are truly miscoded is significantly higher than sites with more similar male to female rates. In fact, the edit did not have any cases falsely identified as a miscode. Although we have no “gold standard” to use in a formal calculation, we can consider the edit to have 100% sensitivity but a modest specificity for male breast cancer. It may be efficient for larger registries, for which manual review is impossible on all cases, to automatically recode male breast cancer cases flagged as female without confirmation from secondary sources. However, if the registry only resolves breast cases that are potentially miscoded as male but not the reverse (reported female cases identified as potentially male), the registry will be falsely suppressing male breast cancer rates.

Conclusions
Overall, the extent of miscoding of sex appears minimal in cancer registries, less than 1%. But miscoding disproportionately affects sex-skewed sites, like liver, thyroid, and breast. The problem is highlighted in male breast cancer and artificially inflates male breast cancer rates to the point that can cause unwarranted alarm, as occurred in Florida, or might misdirect public health resources. However, subset-specific quality control projects on male breast cancer alone artificially suppresses rates because such projects only remove miscoded males and do not add in miscoded females. The use of the NYSCR sex edit can improve quality of sex codes by significantly reducing the number of cases requiring manual follow-back.

Acknowledgements
We would like to acknowledge the contributions of Jackie Button, Megsys Herna, and Stuart Herna for their work on the original QC project while at the Florida Cancer Data System.

References
Abstract: Individuals with low socioeconomic status and racial/ethnic minorities experience significant cancer disparities. They are more likely to be diagnosed with late-stage disease and experience lower survival rates, even when stage at diagnosis is taken into account. Low-income individuals, who are disproportionately racial minorities, are historically more likely to be enrolled in Medicaid and this proportion is increasing with the implementation of the provisions of the Patient Protection and Affordable Care Act. Only a limited number of studies have attempted to assess cancer screening and treatments provided to Medicaid enrollees. Creating standardized, prelinked cancer registry and Medicaid administrative data, similar to data available for the Medicare program, is essential to spur on research to understand the role of Medicaid in eliminating cancer disparities.

Key words: cancer, health care disparities, low income, Medicaid, minority health

Cancer is the leading cause of death for people under the age of 85 in the United States and there are substantial disparities as the burden from cancer is not shared equally across population groups. Cancer health disparities exist due to differences in incidence, prevalence, morbidity, and mortality among subgroups of the population beyond what would be expected under equitable conditions.

Individuals with low socioeconomic status and racial/ethnic minorities are more likely to be diagnosed with late-stage disease and have lower overall survival rates. Racial and ethnic differences in cancer mortality rates have been decreasing over time but significant disparities still remain. Among racial and ethnic groups, African Americans experience the starkest disparities. Although white women have the highest incidence rate for breast cancer, African American women are more likely to die from the disease. African American men are about 1.5 times as likely to be diagnosed with prostate cancer compared to white men but more than twice as likely to die from it. Even when outcomes are compared for those diagnosed at the same stage of disease, African Americans are less likely than whites to survive 5 years after diagnosis with most forms of cancer. Insurance coverage does reduce some of the differences seen across racial/ethnic groups but it does not eliminate disparities in cancer care. Cancer disparities have been less thoroughly researched among children than adults but minority children, specifically Hispanic, African-American, and American Indian, do appear to experience worse outcomes than white children.

Socioeconomics and race/ethnicity are interrelated and interlinked with health insurance status. Low-income individuals, who are disproportionately racial/ethnic minorities and the disabled, are more likely to be enrolled in Medicaid. In 2012, 31% of non-elderly adults with family incomes less than $20,000 were covered by Medicaid or public programs, compared with 3% of individuals with family incomes over $40,000. In the same year, 28% of African Americans and 27% of Hispanics were enrolled in Medicaid (Figure 1) compared to 11% of whites. The Medicaid program provides insurance coverage for nearly 30 million racial and ethnic minorities.

Medicaid therefore plays a critical role in providing insurance coverage for many low-income minorities who are diagnosed with cancer. The importance of Medicaid coverage for cancer treatment was further enhanced with the implementation of the Breast and Cervical Cancer Prevention and Treatment Act of 2000, through which Medicaid programs offer treatment services for women diagnosed with cancer in the National Breast and Cervical Cancer Early Detection Program. The ongoing implementation of the provisions of the Patient Protection and Affordable Care Act will allow for more comprehensive research on the role of Medicaid in addressing cancer disparities.

Figure 1. Sources of Insurance Coverage by Race and Ethnicity, 2012

These estimates are derived from the 2012 US Census Bureau, Income Poverty and Health Insurance Coverage in the United States. The bars indicate the proportion of individuals in each racial and ethnic group that have specific types of insurance coverage. The white category only includes non-Hispanic whites. The other group includes those covered by military health plans. The total for each racial and ethnic group may not equal 100% because some individuals have more than one type of coverage.

Abstract: Individuals with low socioeconomic status and racial/ethnic minorities experience significant cancer disparities. They are more likely to be diagnosed with late-stage disease and experience lower survival rates, even when stage at diagnosis is taken into account. Low-income individuals, who are disproportionately racial minorities, are historically more likely to be enrolled in Medicaid and this proportion is increasing with the implementation of the provisions of the Patient Protection and Affordable Care Act. Only a limited number of studies have attempted to assess cancer screening and treatments provided to Medicaid enrollees. Creating standardized, prelinked cancer registry and Medicaid administrative data, similar to data available for the Medicare program, is essential to spur on research to understand the role of Medicaid in eliminating cancer disparities.

Key words: cancer, health care disparities, low income, Medicaid, minority health

Cancer is the leading cause of death for people under the age of 85 in the United States and there are substantial disparities as the burden from cancer is not shared equally across population groups. Cancer health disparities exist due to differences in incidence, prevalence, morbidity, and mortality among subgroups of the population beyond what would be expected under equitable conditions.

Individuals with low socioeconomic status and racial/ethnic minorities are more likely to be diagnosed with late-stage disease and have lower overall survival rates. Racial and ethnic differences in cancer mortality rates have been decreasing over time but significant disparities still remain. Among racial and ethnic groups, African Americans experience the starkest disparities. Although white women have the highest incidence rate for breast cancer, African American women are more likely to die from the disease. African American men are about 1.5 times as likely to be diagnosed with prostate cancer compared to white men but more than twice as likely to die from it. Even when outcomes are compared for those diagnosed at the same stage of disease, African Americans are less likely than whites to survive 5 years after diagnosis with most forms of cancer. Insurance coverage does reduce some of the differences seen across racial/ethnic groups but it does not eliminate disparities in cancer care. Cancer disparities have been less thoroughly researched among children than adults but minority children, specifically Hispanic, African-American, and American Indian, do appear to experience worse outcomes than white children.

Socioeconomics and race/ethnicity are interrelated and interlinked with health insurance status. Low-income individuals, who are disproportionately racial/ethnic minorities and the disabled, are more likely to be enrolled in Medicaid. In 2012, 31% of non-elderly adults with family incomes less than $20,000 were covered by Medicaid or public programs, compared with 3% of individuals with family incomes over $40,000. In the same year, 28% of African Americans and 27% of Hispanics were enrolled in Medicaid (Figure 1) compared to 11% of whites. The Medicaid program provides insurance coverage for nearly 30 million racial and ethnic minorities.

Medicaid therefore plays a critical role in providing insurance coverage for many low-income minorities who are diagnosed with cancer. The importance of Medicaid coverage for cancer treatment was further enhanced with the implementation of the Breast and Cervical Cancer Prevention and Treatment Act of 2000, through which Medicaid programs offer treatment services for women diagnosed with cancer in the National Breast and Cervical Cancer Early Detection Program. The ongoing implementation of the provisions of the Patient Protection and Affordable Care Act will allow for more comprehensive research on the role of Medicaid in addressing cancer disparities.
Affordable Care Act (ACA) will now substantially increase Medicaid enrollment as about half of the states have elected to expand Medicaid coverage. Substantial proportions of the African American and Hispanic populations are uninsured (Figure 1) and the majority of these individuals would qualify for Medicaid if they live in a state that is expanding coverage.

Assessing cancer care provided to Medicaid enrollees is essential to critically evaluate the role of Medicaid coverage in eliminating cancer disparities. Although there is substantial literature on cancer, poverty and race/ethnicity, only a limited number of studies have attempted to assess cancer screening and treatments provided to Medicaid enrollees.12–14 Some researchers have studied single state cancer registry data linked with Medicaid administrative data but these analyses have limited generalizability given that state Medicaid programs differ in racial/ethnic composition and policies related to eligibility, recertification processes, copayment requirements, coverage of medical services, and reimbursement rates to providers. Other studies have used cancer registry data that do contain some information on insurance status at the time of diagnosis and treatment. The payer details available in cancer registry data are not consistently of high quality and are particularly problematic for Medicaid enrollment status as this can change over short periods of time.12 These data sets do not capture the timing of Medicaid enrollment and the continuity of enrollment which have both been shown to be of importance in determining cancer outcomes. Individuals can frequently move back and forth between being uninsured and enrolled in the Medicaid program, and with the implementation of ACA provisions, more transitions between Medicaid and plans offered through the health insurance Marketplace are expected.13 Therefore it is important to understand differences between those who enroll in Medicaid before and after their cancer diagnosis. In addition, the length of time enrolled in Medicaid after cancer diagnosis is essential to identify potential access to care issues related to treatment and survivorship. All-payer claims data initiatives that are in various stages of development are a promising new source of data that can be linked to cancer registry data but this data is only available for limited number of years and for a few states at this time.

Currently, no standardized linked dataset such as the Surveillance, Epidemiology and End Results database linked with Medicare administrative data (SEER-Medicare) exists for the Medicaid program. The SEER-Medicare data contain clinical, demographic and cause of death information from SEER cancer registries and the Medicare claims for covered health care services from the time of a person’s Medicare eligibility until death. Over the past 15 years, more than 1,000 scientific articles that have been produced using the linked SEER-Medicare data focusing on cancer in the elderly.14 In comparison, a much smaller pool of evidence exists that assess the impact of Medicaid coverage on cancer outcomes.

One approach to increase research on Medicaid cancer patients is to create a similar de-identified linked database that contains multiple (if not all) state Medicaid administrative datasets linked with cancer registry data. Cancer registries operate in all states with support from the Centers for Disease Control and Prevention’s National Program of Cancer Registries and the National Cancer Institute funded SEER. Medicaid administrative datasets for all states are available in a standardized format from the Centers for Medicare and Medicaid Services (CMS) and historic data beginning with the year 1999 can be obtained. The feasibility of linking these data sources has been established but challenges remain as approvals and data use agreements need to be established with multiple agencies including CMS and state cancer registries. The common shared goal of eliminating health disparities can serve as the catalyst to overcome any potential administrative challenges and the proposed database will address one of the key recommendations of the Institute of Medicine’s report on Enhancing Data Systems to Improve the Quality of Cancer Care15 by expanding support for cancer research using existing databases.

Creating a Medicaid and cancer registry prelinked database would encourage important disparities research and provide multiple years of data and adequate sample sizes to perform meaningful research focused on nonelderly adults and children. As with other data collected for administrative and billing purposes, there are constraints with using Medicaid claims but greater awareness of potential limitations and data anomalies can help researchers tailor and present analysis to minimize potential bias. It is important to acknowledge and understand that state Medicaid programs differ across multiple dimensions and control for these differences whenever possible in the analysis performed. Several states have been increasing enrolling in managed care programs and in the past these enrollees have been excluded from most analysis but recent studies have shown encounter data submitted for managed care enrollees to be of adequate quality to support research in many states.16 The overall quality of the Medicaid data should improve in the future as CMS has launched an initiative that requires states to submit “transformed” Medicaid Statistical Information System data to improve accuracy, completeness and timeliness of the data. In addition, information for several critical variables, such as initial course of treatment, is available from both the cancer registry and the Medicaid administrative data which creates complementary data sources to reduce potential missing or unknown information. Linkages to additional data sources can identify comparison cohorts and also provide socioeconomic, behavior, access to care and genetic factors to perform multifactorial assessment of health disparities in order to explore complex interrelationships.

Linked Medicaid and cancer registry data from multiple state programs will allow researchers and policy makers to understand the impact of state and Medicaid policies related to the implementation of ACA on cancer outcomes along the continuum of care from screening, to diagnosis, treatment and survivorship. States with large uninsured minority populations are among those not planning to expand Medicaid coverage and therefore the ongoing implementation of health reform creates the situation where natural
experiments exist to understand the impact of Medicaid coverage in reducing racial/ethnic disparities. Medicaid programs policies can be systematically assessed to inform future planning with evidence-based strategies to eliminate cancer disparities.

References
Comprehensive Assessment of Population-Based Cancer Registries: An Experience in Colombia

Constanza Pardo, MsC; Luis Eduardo Bravo, MsC; Claudia Uribe, MD; Guillermo López, MD; María Clara Yépez, MsC; Edgar Navarro, MsC; Esther de Vries PhD; Marion Piñeros, MsC

Abstract: Background: In the global context, the establishment of population-based cancer registries, particularly in less developed regions, has become of strategic importance. The factors influencing the operation and sustainability of registries can be determinants for their success, despite the existence of uniform quality indicators in the cancer incidence information. Our objective was to determine the current state of the structure, organization and operation of population-based cancer registries in Colombia, obtain information on their degree of development and identify specific problems that affect their operation and sustainability. Methods: We developed a descriptive study in 5 population-based cancer registries (Barranquilla, Bucaramanga, Cali, Manizales, and Pasto). The analysis included 7 broad categories: general characteristics, operational procedures, scientific production, completeness, validity, comparability and continuing education. To establish the validity of the information we used the available incidence databases. Results: All registries were based in a university (3 public, 2 private). The 5 registries covered 11.8% of the Colombian population. Four registries published their results on cancer incidence. Financing came from different sources and costs varied significantly. Cancer incidence rates ranged from 94.1 to 189.2 per 100,000. The coverage of information sources ranged from 60 to 90%. Validity indicators were within acceptable limits while comparability parameters showed variations between registries. All registries participated in regular workshops and congresses. Conclusions: Operation of cancer registries in a model with universities and with several financial sources seems to provide sustainability; follow-up, training and assistance are critical to motivation and quality, costs vary significantly and determinants of costs of registry activities need to be further assessed.

Key words: Colombia, data collection, neoplasms, quality control, registry

Introduction

Population-based cancer registries are the only entities that provide unbiased estimates of cancer incidence, and therefore are a vital tool to evaluate the cancer burden, organize cancer control programs, evaluate effects of interventions and generate hypotheses.1,2

On a global scale, the relevance of population-based cancer registries is increasingly recognized; recently, cancer incidence by type of cancer was included as one of the 25 indicators in the comprehensive global monitoring framework for the prevention and control of non-communicable diseases.3 However, establishing and maintaining cancer registries in order to determine cancer incidence is complex and costly, resulting in relatively low levels of cancer registry coverage in developing countries. While in the United States 95% of the population is covered by a cancer registry included in Volume X of Cancer Incidence in 5 Continents (CI5X), this proportion is only 8% in Latin America, 6% in Asia and 2% in Africa.4

For many cancer registries in lower income settings it is difficult to comply with established quality standards5-6, resulting in low acceptance rates for CI5X: while 44% of all population-based registries from Africa which submitted their data for publication met the quality standards it was 90% for North America and Europe.7 In constraint resource settings, aspects like the logistics of identifying all cancer cases in a population, differentiating residents from non-residents, and ensuring continuity of staff member contracts are difficult and can threaten the continuity of the registry and its data quality.

In order to provide technical assistance to a cancer registry, there is a need to understand aspects which are heavily context-dependent such as the organizational structure, the networking capacity, the operational costs, the capacity to disseminate of information, research and number and training of staff of registries, among others. Evaluation procedures of these aspects, unlike those for data quality, are more difficult to standardize. A better understanding of the degree to which these factors determine the implementation, functioning and sustainability of cancer registries, as well as their data quality, is necessary, as they may affect the sustainability of registries, even before they manage to produce valid information on cancer incidence. Some of these aspects were incorporated in a recent Latin American proposal of external evaluation of cancer registries.8

1Instituto Nacional de Cancerología, Bogotá, Colombia. 2Population-Based Cancer Registry, Cali, Colombia. 3Population-Based Cancer Registry, Area Metropolitana Bucaramanga, Colombia. 4Population-Based Cancer Registry, Manizales, Colombia. 5Population-Based Cancer Registry, Pasto, Colombia. 6Population-Based Cancer Registry, Barranquilla, Colombia. 7International Agency for Research on Cancer, Section Cancer Information, Lyon, France.

Address correspondence to Constanza Pardo, MsC Instituto Nacional de Cancerología, Bogota, Colombia, Calle 1 # 9-85. Telephone: +571 3341111, Ext. 4111. Email: cpardo@cancer.gov.co.

This study was funded by the Colombian Government through resources administered by the National Cancer Institute, Bogotá, Colombia. The Colombian Ministry of Health provided additional funds through the Contract Agreement No. 199-2011.
The systematic and standardized procedures of cancer registries consist of 4 major procedures, which are also useful as conceptual framework to evaluate the general situation of the registries: data collection, data processing, data analysis and dissemination of information.8,9

In this manuscript we evaluate and describe the situation of the 5 population-based cancer registries in Colombia, which receive technical and financial support from Colombia’s National Cancer Institute (NCI). All these registries function as a special department within universities. The NCI evaluates the registries’ performance and publishes national estimates of incidence and mortality.10,11 A previous, unpublished, evaluation of these 5 registries formed the base to develop and finance specific technical support for this current evaluation.

The objectives of this work were to determine the current situation of the structure, organization and functioning of the population-based cancer registries, obtain information on the degree of improvement of the registries’ performance compared to the previous evaluation, and identify concrete impediments for efficient functioning of the registries.

We think that our methodology and results provide useful parameters for the evaluation of population-based cancer registries in limited resource settings and thereby improve technical support to such registries.11

**Materials and Methods**

A descriptive study was realized including the population-based cancer registries of Barranquilla, Bucaramanga (Metropolitan Area), Cali, Manizales and Pasto.

Site visits of NCI staff to the registries were organized. Information was collected based on interviews with registry employees, inspection of specific documents and stored in structured questionnaire forms. The following 7 parameters were examined: general characteristics; operational indicators; completeness, validity and comparability of the data; scientific production and continued education. Some criteria and quality indicators of cancer registries as specified by the International Agency for Research on Cancer (IARC) were used.5,6 The data was stored in an excel sheet, data were analyzed in SPSS version 19.0. We included the following aspects of each of the parameters mentioned above:

**General Characteristics**

Date of initiation of the registry, time of continued functioning, year of affiliation to the International Association of Cancer Registries (IACR), coverage of the registries’ source population, population size in 2011,12 and incidence data for the most recent 5 years.

**Operational Aspects**

Type of organization hosting the registry (public versus private universities), financial resources for 2011 (excluding those for research), average number of new cases detected, average cost per registered case, human resources, type of contract of the human resources (permanent or temporary), annual number of data sources per registrar, and the average annual number of new cases identified per source.

**Completeness Criteria**

Number of information sources determined at the moment of the census, number of sources per 100,000 inhabitants, and coverage of sources.

**Validity Criteria**

Validity of the information was determined inspecting the databases of the most recent available years. Eight validity indicators were included with limits of acceptable values within the registry databases: proportion of microscopically verified cases (%MV: 75%–98%), proportion of cases notified by death certificate only (DCO ≤ 10%), proportion of cases diagnosed based on clinical information only (CIO ≤ 10%) and proportion of cases with missing information on identification, age and date of birth, sex and date of diagnosis (Missing basic data, MBD ≤ 10%).

We included the absolute number and percentage of cases without specific information on subsite (fourth digit of the ICD-O C code 9, C_ _,9) for stomach (C16), colon (C18), breast (C50), cervix uteri (C53), and brain cancers (C71).

**Comparability Criteria**

Case definition per registry, definition for date of diagnosis, classifications used for general codification of tumors and rules for multiple primary tumors applied.

**Dissemination of Information**

Information generated by the registries in the 5 years prior to the evaluation (2011), including monographs, national publications with information generated by the registries, and research projects. Articles resulting from research projects (defined as an activity to investigate a specific research question with a beginning and end date) were excluded.

**Continued Education**

Training of staff through congresses and workshops. Workshops (16 hours) covered a combination of theory and practice, in which registry staff applied the theory to their own datasets.

**Results**

The 5 registries represent different geographical zones of Colombia: Barranquilla at the Caribbean cost in the North, and in the inland there are Bucaramanga in the east, Cali in the west, Manizales in the central zone, and Pasto in the south (Figure 1). Together they covered 11.8% of the 2011 Colombian population.12 All were active in 2011 with a continuous activity of 4 to 49 years since their date of initiation. Only 1 registry included more than 1 municipality/city. They registered an average of 8,499 new cases of cancer annually, with incidence rates varying between 94.1 and 189.2 cases per 100,000 inhabitants (Table 1).

**Operational Aspects**

All registries were hosted at universities, 3 public and 2 private, and each registry had some permanent staff members (Table 2).

All registries received financial support by their host university and the NCI. Two registries received additional
financial support from other sources, one of them being international. The total costs of performing the primary routine tasks of the registries (excluding research) were $636,429 (based on market exchange rate September 6, 2011), 68.7% of which were borne by the Cali registry. The average cost per registered case was $65 (varying between $28 and $115). Cali and Barranquilla had most affiliated staff members. In 4 of the registries, 80% of staff was hired on a temporary (contract) basis (Table 2).

The average number of registrars per registry was 9 (range, 2–30). Cali had the highest number of registrars, including 25 medical students assigned to visit the medical specialists and extract cases. The average number of sources per registrar was 16, ranging between 7 and 25. Bucaramanga and Pasto covered the highest number of sources per registrar. The average number of new cases identified per source was between 13 (Pasto) and 36 (Cali) (Table 2).

Completeness

The number of information sources according to census information varied between 38 and 444, being highest in Cali and Bucaramanga. The number of sources per 100,000 inhabitants was between 6 and 20, being highest in Cali and Pasto (Table 2). Coverage of these sources was 60%–90%, with highest coverage in Manizales and Pasto.

The Cali registry provided an indicator of efficiency per source (number of cancer cases registered per source divided by the total number of cases revised per source per time period), which allows establishing which sources provide most cancer cases in order to prioritize case-finding (results not shown).

Validity

The proportions of MV and DCO cases were within acceptable limits in all registries, except in one where the DCO indicator could not be established as the death registry was not covered. The proportion of CIO cases varied between 0.6% and 16.1%, with the highest percentage reported by Barranquilla (Table 3).

The most frequently identified problems in demographic variables were lack of personal identification number, date of birth, and age. One registry exceeded the limits established for 2 of these indicators.

The inspection of the topography information for the 5 specified sites showed lowest proportions of unspecified subsite cases for colon and brain. The Cali registry showed the lowest proportions of unspecified information on subsite. This indicator could not be established for the registry which does not use ICD-O-3 for coding and therefore does not have information on subsite.

Table 1. General Characteristics of the 5 Included Population-Based Cancer Registries in Colombia, 2011

<table>
<thead>
<tr>
<th>General characteristics</th>
<th>Barranquilla</th>
<th>Bucaramanga</th>
<th>Cali</th>
<th>Manizales</th>
<th>Pasto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of continuous functioning (years)</td>
<td>4</td>
<td>11</td>
<td>49</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Coverage area</td>
<td>Urban part of municipality</td>
<td>Metropolitan Area</td>
<td>Urban part of municipality</td>
<td>Municipality</td>
<td>Municipality</td>
</tr>
<tr>
<td>Population size 2011*</td>
<td>1,189,503</td>
<td>1,084,699</td>
<td>2,233,016</td>
<td>390,084</td>
<td>417,484</td>
</tr>
<tr>
<td>Incidence rate all cancersb</td>
<td>94.1c</td>
<td>162.7b</td>
<td>189.2b</td>
<td>174.8b</td>
<td>169.6b</td>
</tr>
</tbody>
</table>

Source: National Cancer Institute, Bogotá-Colombia, results based on databases of the population-based cancer registries.

*Information of National Statistics Office (DANE) of Colombia.


cFor 2008.
### Table 2. Operational and Completeness Indicators, Population-Based Cancer Registries in Colombia, 2011

<table>
<thead>
<tr>
<th>Operational indicators</th>
<th>Barranquilla</th>
<th>Bucaramanga</th>
<th>Cali</th>
<th>Manizales</th>
<th>Pasto</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of organisation hosting the registry</strong></td>
<td>Private University</td>
<td>Private University</td>
<td>Public University</td>
<td>Public University</td>
<td>Public University</td>
</tr>
<tr>
<td><strong>Resources 2011 ($)&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td>36,150</td>
<td>48,930</td>
<td>437,443</td>
<td>32,335</td>
<td>81,567</td>
</tr>
<tr>
<td><strong>Average costs per registered case (USD)</strong></td>
<td>32</td>
<td>28</td>
<td>103</td>
<td>47</td>
<td>115</td>
</tr>
<tr>
<td><strong>Number of financing sources</strong></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Fixed staff positions (n and %)</strong></td>
<td>2 (20.0)</td>
<td>2 (28.5)</td>
<td>6 (42.8)</td>
<td>1 (16.7)</td>
<td>1 (14.3)</td>
</tr>
<tr>
<td><strong>Staff on temporary contract base</strong></td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td><strong>Registrars</strong></td>
<td>6</td>
<td>3</td>
<td>30</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td><strong>Number of sources per registrar&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td>7</td>
<td>25</td>
<td>15</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td><strong>Average number of new cases per source</strong></td>
<td>16</td>
<td>15</td>
<td>36</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of sources according to census</strong></td>
<td>70</td>
<td>119</td>
<td>444</td>
<td>38</td>
<td>56</td>
</tr>
<tr>
<td><strong>Coverage of sources (%)</strong></td>
<td>60</td>
<td>63</td>
<td>...</td>
<td>90</td>
<td>80.4</td>
</tr>
<tr>
<td><strong>Number of sources per 100,000 inhabitants</strong></td>
<td>6</td>
<td>11</td>
<td>20</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

Source: National Cancer Institute, Bogotá-Colombia, results based on databases of the population-based cancer registries.  
<sup>a</sup>Does not include resources for research. Market exchange rate, September 6, 2011: 1,782,80 COP  
<sup>b</sup>Annual.

### Table 3. Indicators of Validity of the Population-Based Cancer Registries in Colombia, 2011

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Barranquilla&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Bucaramanga&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Cali</th>
<th>Manizales</th>
<th>Pasto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microscopically verified cases (%MV)</td>
<td>76.9</td>
<td>84.4</td>
<td>79</td>
<td>85.5</td>
<td>85.8</td>
</tr>
<tr>
<td>Death certificate only cases (%DCO)</td>
<td>NA</td>
<td>9</td>
<td>4.4</td>
<td>6.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Clinical information only cases (%CIO)</td>
<td>16.1</td>
<td>0.9</td>
<td>9.4</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td><strong>Cases with missing basic data (%MBD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases without identifying document number</td>
<td>35</td>
<td>1.6</td>
<td>7.3</td>
<td>0.5</td>
<td>1.3</td>
</tr>
<tr>
<td>Cases without age at diagnosis</td>
<td>8</td>
<td>0.5</td>
<td>2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Cases without date of birth</td>
<td>30</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Cases without information on sex</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cases without incidence date</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Codification of 4th topography digit ICD-O (%)&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C16.9 Stomach</td>
<td>NA</td>
<td>73</td>
<td>65.7</td>
<td>87.8</td>
<td>92.5</td>
</tr>
<tr>
<td>C18.9 Colon</td>
<td>NA</td>
<td>52</td>
<td>34</td>
<td>50</td>
<td>85.8</td>
</tr>
<tr>
<td>C50.9 Breast</td>
<td>NA</td>
<td>92.1</td>
<td>82.5</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>C53.9 Cervix uteri</td>
<td>NA</td>
<td>86</td>
<td>85.9</td>
<td>95.8</td>
<td>96.8</td>
</tr>
<tr>
<td>C71.9 Brain</td>
<td>NA</td>
<td>26</td>
<td>13.9</td>
<td>8.3</td>
<td>30.5</td>
</tr>
</tbody>
</table>

Source: National Cancer Institute, Bogotá-Colombia, results based on databases of the population-based cancer registries  
<sup>%CIO</sup>, percentage of cases with only clinical information; <sup>%DCO</sup>, percentage of cases with death certificate as only source of diagnosis; ICD-O, International Classification of Diseases—Oncology; <sup>%MDC</sup>, percentage of cases with missing information on basic data; <sup>%MV</sup>, percentage of microscopically verified cases; NA, not applicable.  
<sup>a</sup>Percentage of the total number of cases of the specific site.  
<sup>b</sup>Since Barranquilla does not have the death certificate registry as one of the information sources, <sup>%DCO</sup> could not be estimated, and since the coding system in this registry is ICD-10 there is not information on the fourth digit.
Comparability

With the exception of Pasto, all registries followed IARC recommendations for date of diagnosis, with a prioritization of different dates in descending order (Table 4). All registries included malignant cases for all sites with the exception of basal cell carcinoma of the skin; and 2 included information on in situ cases for all sites. The ICD-O3 coding system was used by 4 registries; the fifth used ICD-10 for coding of the tumors. IARC rules for the inclusion of second primaries were used by 4 of the 5 registries.

Visibility, Dissemination, and Scientific Production

Four registries published their incidence data (per lustrum), 2 published their mortality data, and 1 also published survival information. Four of the registries published their data in national journals, with on average 2 articles per registry. During the latest 5 years, in total 36 research projects were realized. Four registries were member of the IACR, with different years of affiliation (Table 5).

Continued Education

During the evaluation period, NCI organized 3 workshops for registry staff and 2 congresses with presentations by each registry. The workshops were aimed at coordinators and directors of the registries. All consisted of a combination of theoretical classes and practical sessions, in which each registry processed its own data, and suggestions were made according to the findings, discussing common findings or points to improve among participants. The topics covered included general aspects of descriptive epidemiology, age standardization, time trend analysis, survival analysis, operative indicators, data quality checks using different tools, use of CanReg, and preparation and analysis of information for submission to IARC for inclusion in CI5X (following editorial process and tables).

Discussion

The 5 registries combined covered 11% of the Colombian population, which is a low coverage compared to countries in Northwestern Europe and the United States. However, in Colombia, like in Brazil, the existing registries cover...
populations which are geographically and culturally quite distinct, and the obtained information has been the base for subnational and national incidence estimates, necessary for decision-making related to cancer control.\textsuperscript{18,20} This model of establishing few registries, covering representative geographical areas of the country, with high quality standards and ensuring continuity has proved efficient under circumstances of limited resources. Ultimately, countries will need to decide whether achieving a national coverage is needed, and if so, whether it is feasible and sustainable. If national coverage is desired, learning from the experiences of a few and small scale registries is useful. The Colombian operating model with population-based cancer registries functioning as part of a university has shown a higher stability and sustainability than other organizational settings. Previous experiments where cancer registries were established under the sole responsibility of a district health directorate resulted in a loss of support and continuity with changes in the political landscape. Being part of a unit and academic program allows for teaching personnel of the university to dedicate part of their time to registry tasks (in Columbia directors and coordinators of registries get a reduction in their required hours of teaching or have their registry tasks included in their annual activity plans). This provides more continuity. Being in a university also allows to involve students in data collection, resulting in lower salary costs and transferring knowledge and the value of cancer registries to future generations of health specialists, which is in turn valuable for future continuity.

The youngest registries had more difficulties accessing sources, resulting in lower coverage and leading to imprecise incidence numbers. With the exception of the Cali registry, a very high proportion of staff is hired on a temporary (contract) base, and human resources are scarce. The majority of staff on a contract base performed vital tasks for the primary functioning of the registry, and was financed by the NCI. This represents a threat as administrative delays in transfer of funds can jeopardize the continuity of the registries' functioning and lead to high turnover of personnel which, in addition to the lack of qualified staff and training opportunities, impedes optimal functioning. The Colombian registries have 5–6 times fewer personnel hired for routine activities than other registries in different parts of the world\textsuperscript{21}, and a budget which is 9-fold lower than the one used for personnel of the National Program of Cancer Registries (NPCR) of the United States.\textsuperscript{22} It would be useful to determine and evaluate the annual costs of registry activities, with standard methods which facilitate comparisons with previous financial evaluations of other cancer registries. This would allow identifying factors influencing cost-effectiveness of a registry, improve budget allocation, and contemplate potential expansions of the coverage by establishing new registries in other parts of the country.\textsuperscript{23}

Technical assistance, follow-up on quality and training are particularly relevant and therefore critical factors for sustainability.\textsuperscript{24} All registries participated in at least one workshop annually, as recommended by NCI after the previous evaluation. The workshops aimed to standardize knowledge, improve skills and the quality of the information of the registries; these efforts are reflected in progressive improvements of data quality. A key factor for successes of these workshops were the application of the theory to the registries own databases during the workshop. We believe that the success of the registries in passing the evaluations for CI5X is partly due to the improved use of existing tools and methods and a better understanding of the editorial process and the general context of data quality and content. Despite general standards, each registry has its own specificities, influencing possibilities for analysis, corrections and improvement.

The fact that the standard quality indicators were within acceptable values (\%MV, \%DCO, \%CIO, \%MBD), may be a positive result of the training provided to the registries in the previous years. In fact, besides Cali, which has been present in all editions of Cancer Incidence in Five Continents, data of 3 more registries were accepted for publication in CI5X.\textsuperscript{4} Some of the differences between registries may be due case-finding methods: if case-finding is centralized in databases of pathology laboratories or not, efforts to obtain additional information on DCO cases, level of access to the sources, completeness of the medical records and the “age” since establishment and therefore experience of the registries itself.\textsuperscript{5,6} For some sites we observed a high proportion of cases without specified subsite information which indicates either a lack of training of registrars or scarce information in the source and/or medical record, representing problems beyond cancer registry efforts which have their origins in the healthcare system.\textsuperscript{25}

The population coverage of the evaluated registries, although variable, was within recommended limits,\textsuperscript{7} comparable with other reported coverage levels for South and Central American cancer registries\textsuperscript{4} and resulted in good levels of completeness and data quality. Although there is a continuous increase in the amount of information available through electronic sources, this does not make site visits to the sources and active case finding redundant. The necessity of active case finding is a determining factor of staff resources in order to cover the registries source population properly.

We need to establish key operative indicators for registries, with continuous evaluation in order to facilitate planning, coordinating and supervising operative activities. Good examples of such indicators are (1) the coverage of the sources, (2) the proportion of cancer cases identified out of the number of potential cases evaluated in order to prioritize visits to sources contributing many cases, and (3) the number of sources per case, which is an indirect indicator of the integrity of the data.\textsuperscript{8} The registries have not defined such indicators and implementing them as performance standards will require special efforts. However, these measures will allow a permanent contact with the sources, ensuring complete coverage of incident cases and diminishing difficulties in access to information sources.\textsuperscript{25}

Currently, the more precisely defined quality parameters, better technical support over the past 5 years and financing by local governments, are reflected in improved functioning, with more continuity; we expect that this will positively influence publications and dissemination of
information by the registries, as already indicated by the most recent publication of Cancer Incidence in Five Continents. A limitation of this study is that the role of NCI with regard to technical assistance for the registries was not evaluated within the framework of the cancer information system, as the evaluation was very much oriented toward the registries. However, future evaluations should be designed to include the role of NCI as well. In addition evaluations of specific skills for administration of chronic disease programs in aspects like program and financial management, human resources management, strategic leadership and resource mobilization would be desirable.

Acknowledgments
The authors would like to acknowledge the collaborating teams from Barranquilla, Bucaramanga, Cali, Manizales, and Pasto population-based cancer registries for their active participation in this study.

References
Abstract: Timeliness is one of the key indicators of cancer surveillance data quality, as delayed reporting of cases results in an underestimation of the cancer rate in a population. The purpose of this paper is to assess temporal trends in reporting delay by cancer site from 1999–2010. Using data from the Surveillance, Epidemiology, and End Results (SEER) 9 cancer registries and the New York State Cancer Registry, I calculated short-, medium-, and long-term delay for the most common cancer sites for each year and identified the linear trend. Nearly all sites showed a decrease in delay over the period, and many showed a statistically significant decrease. The decrease in delay was more pronounced in the New York State data. These findings reflect long-term improvement in the timeliness of cancer reporting, but there remains room for improvement. Leukemia and myeloma are especially problematic, as these sites are heavily dependent on reporting by private physicians.

Key words: cancer registration, reporting delay, timeliness

Introduction

Since first described by Clegg et al in 2002, delay-adjusted cancer rates have been widely used in cancer statistics reported by the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) program, and are under consideration for use by registries belonging to the Centers for Disease Control and Prevention’s National Program of Cancer Registries (NPCR), as well as by Canadian registries. Delay-adjusted rates anticipate the small percentage of cases that have yet to be reported at the time cases are originally submitted. An example of how delay adjustment has impacted national cancer reporting may be seen in a recent Annual Report to the Nation on the Status of Cancer, which included both delay-unadjusted and delay-adjusted rates for the most common cancers for 13 SEER areas covering 14% of the US population. For all sites combined among males, without delay adjustment there was a significant 0.6% per year decrease in incidence from 2005–2009, but with delay adjustment there was a nonsignificant 0.1% per year increase. Nearly every sex–site combination showed a less negative (or more positive) trend for the 2005–2009 period when delay adjustment was applied: lung cancer among females changed from a significant 1.2% per year decrease to a still-significant 0.3% per year decrease; leukemia among males changed from a significant 0.3% per year decrease to a nonsignificant 0.2% per year increase. From these examples, it is clear how the prevailing direction of cancer incidence in the United States—arguably the single piece of information from this report that is of greatest interest to the public—is sensitive to the timeliness of case reporting.

The purpose of this paper is to assess whether reporting delay has been diminishing over time, as might reasonably be expected based on technological improvements in case reporting. Evidence of decreasing delay would represent a success story for cancer surveillance, as timeliness is one of the fundamental elements of data quality. A comparison of the delay-adjustment factors originally reported for 1998 cases by Clegg et al with those for 2010 suggests that this is indeed the case. For 1998 cases, it was projected that 3% of colorectal cancers, 4% of female breast and lung cancers, 12% of prostate cancers among white men, 14% of prostate cancers among black men, and 14% of melanoma of the skin among whites were not reported in a timely manner. For 2010 cases, the values were 1% for female breast cancer; 2% for colorectal, lung, and melanoma among whites; and 3% for prostate cancer among all races. Sixteen other sites also had delay-adjustment factors in the 1%–3% range, the only exceptions being myeloma (11%) and leukemia (13%). In this paper, I improve upon this small set of observations by considering additional cancer sites, intermediate data years, and additional data from the New York State Cancer Registry.

Methods

For this analysis, I used data from the SEER 9 registries available through version 8.1.5 of the SEER*Stat software program, along with data from the New York State Cancer Registry, where the author is based. The SEER 9 registries consist of the states of Connecticut, Hawaii, Iowa, New Mexico, and Utah, and the metropolitan areas of Atlanta, Detroit, San Francisco–Oakland, and Seattle. SEER 9 and New York collectively contain about one-sixth of the national population according to the 2010 US Census. I considered all invasive, malignant cancers diagnosed between 1999, when data submission deadlines were standardized nationally, and 2011, the most recent data available. Cases diagnosed only by death certificate or autopsy (DCOs) were excluded.

For SEER 9, delay was calculated as the ratio of case counts in later data submissions to those in the original data submissions. For example, if the number of cases of prostate...
cancer was originally reported as 20,000, and 5 years later was reported as 21,000, then the delay over this time period would be 1.05 or 5%. For New York, delay was calculated by comparing the diagnosis date of each case with the earliest date it was received by the registry. Differences greater than 23 months were considered delayed. The 2 means of calculating delay are functionally equivalent and merely reflect differently structured data. Once all cases were classified as either timely or delayed, the ratio of all cases to timely cases was calculated.

Four delay periods were chosen for analysis: up to 1 year, 1–3 years, 3–5 years, and ≤5 years (henceforth referred to as short-term, medium-term, long-term, and total delay). Of course, delays in excess of 5 years are possible, but were not considered in this analysis in order to retain a sufficiently long time series. Specifically, since the most recent
data available is for diagnosis year 2011, 5-year delay can only be assessed through 2006, leaving an 8-year time series (1999–2006). In fact, very few cases are delayed by more than 5 years; the SEER delay-adjustment factors suggest it is below one-half percent.4

Trends were calculated for all sites combined for each of the 4 delay periods. Ovarian cancer was excluded from all sites combined because of major changes in reporting requirements which began in 2001. Short-term delay trends were also examined for the following individual cancer sites: oral cavity and pharynx (oral), esophagus, stomach, colon and rectum (colorectal), liver and intrahepatic bile duct (liver), pancreas, larynx, lung and bronchus (lung), melanoma of the skin (melanoma), female breast, cervix uteri (cervix), corpus and uterus NOS (uterus), ovary (2001–2011 only), prostate, testis, urinary bladder (bladder), kidney and renal pelvis (kidney), brain and other nervous system (brain), thyroid, Hodgkin lymphoma, non-Hodgkin lymphoma, myeloma, and leukemia. These sites were chosen because they are the sites for which SEER has calculated delay-adjustment factors, and they also represent a frequently used list of the most common cancers; together these sites accounted for about 94% of all tumors diagnosed in 2011. Trend analysis consisted of ordinary least-squares linear regression using PROC REG in SAS version 9.2 (SAS Institute). Joinpoint regression using version 4.0.4 SEER’s Joinpoint Regression Program6 was also explored, but nearly all of the trends contained zero joinpoints, making them identical to ordinary least-squares regression, so these results are not reported here.

Results

Trend lines for all sites combined, by sex, for each of the 4 delay periods are shown in Figure 1. Delays generally trended downward in SEER 9, though the trends were statistically significant only for women for short-term, medium-term and total delay. Delay trended sharply and significantly downward in New York for both sexes in every delay period. In 1999, the reporting delay in New York was much higher than in SEER 9, but by 2006 it had reached comparable levels. Delays among males were consistently higher than among females, with the differences more pronounced in SEER 9.

Trend lines for short-term delay for 3 common cancer sites (prostate, breast, and colorectal) along with leukemia and myeloma are shown in Figure 2. Leukemia and myeloma were chosen for inclusion in the figure because they represent the sites with the highest delay; other sites were not included in the figure for clarity. For New York, all 8 of the depicted site-sex combinations showed statistically significant downward trends. For SEER 9, all trend lines were downward and 5 were statistically significant (leukemia among males, myeloma among males and females, and colorectal, and breast among females). Of the sites not shown, the site with the highest delay was liver, consistently about 2 percentage points above prostate cancer in both New York and SEER 9. Lung cancer had similar delay to prostate cancer, and other sites had levels below this. All of the 32 site-sex combinations not shown displayed downward trends in both New York and SEER 9, with the exception of larynx among females in SEER 9. Twenty-three of these sites showed a significant downward trend in New York while 11 showed a significant downward trend in SEER 9.

Discussion

These results demonstrate a pronounced improvement in the overall timeliness of cancer reporting between 1999 and 2010, particularly in New York State. This improvement is consistent with the adoption of more clear and comprehensive data standards, more aggressive casefinding efforts by central cancer registries, better reporting from nonhospital sources, (particularly laboratories), and more expansive interstate data exchanges.7,8 There remains much variation in delay by cancer site, with short-term delay ranging from close to zero (breast) to over 3% (myeloma and leukemia). On their own, myeloma and leukemia would fail to meet the national data standard for timeliness, which is that “within 24 months of the close of the diagnosis year, 95% of expected, unduplicated cases are available to be counted as incident cases at the central cancer registry.”9 The delay in the reporting of these 2 sites comes from their tendency to be diagnosed and treated in physician’s offices, and not reported until the patient is seen at a hospital, often for unrelated reasons. Other sites that are frequently diagnosed in physician’s offices, such as melanoma and prostate, do not share this problem to the same degree because they are typically reported in a timely manner by pathology laboratories upon receiving tissue specimens.

For most sites and particularly within SEER 9, the amount of delay among men was higher than that for women. Within the SEER registries, this was found to be mainly due to disproportionate delays in reporting from US Department of Veteran’s Affairs (VA) hospitals, which primarily serve men.10 This issue was most acute in the mid-2000s; since then, many registries have negotiated with individual facilities to restore timely reporting. A recently issued policy directive by the Veterans Health Administration could potentially offer a permanent resolution to the issue.11

Death-certificate-only (DCO) and autopsy-only cases were excluded because of their tendency to artificially lower the delay rate. By definition, these cases are not delayed because the diagnosis date is equal to the death date, but clearly they were not reported in a timely manner except in the exceedingly rare event the patient was never diagnosed at all while living. However, the DCO rates in both SEER 9 and New York State were so low that the results of this analysis were not sensitive to either their inclusion or exclusion. Specifically, DCO rates were close to 1%; for a site where about 3% of the cases were delayed, this meant the difference between 3 of 99 cases being delayed when DCOs were included (3.03%) and 3 of 100 cases being delayed when DCOs were included (3.00%).

A limitation of this analysis is that it is not possible to comment on short-term delay since 2010, medium-term delay since 2008, or long-term delay since 2006, meaning that the assessment of timeliness cannot itself be timely.
Some of the trend lines, particularly in New York, will necessarily have to flatten, as they imply an eventual negative reporting delay. There is anecdotal evidence to suggest that some of the downward trends may already be slowing. In particular, the profound complexities of the introduction of version 2 of the Collaborative Stage Data Collection System, with its many new codes and rules, for cases diagnosed in 2010 and later, may have resulted in reduced timeliness. It is also possible that staff reductions endured by some central cancer registries in recent years have limited their casefinding efforts. On the other hand, ongoing enhancements to electronic reporting systems may more than offset these obstacles. In order to remain abreast of these trends, continued monitoring of reporting delay in the manner described here should be an active part of the quality assurance activities of cancer registries.

Delay-adjusted rates have only ever been published for the SEER 13 registries (the SEER 9 registries plus the metropolitan areas of San Jose-Monterey and Los Angeles in California, Alaska Natives, and portions of rural Georgia). Cancer statistics for all other states and territories have never made such an adjustment, and so tend to portray a slightly more optimistic picture of cancer trends than is warranted. Researchers affiliated with SEER and the North American Association of Central Cancer Registries (NAACCR) are currently in the process of developing delay-adjustment models that can be applied to the remainder of the country, so it likely that this limitation will be overcome in the near future.

Acknowledgements

I would like to acknowledge Betsy Kohler of NAACCR and Kevin Henry of Temple University for their input to this project, as well as the members of the NAACCR community who took the time to answer some online questions about current reporting practices.

References

Original Article

Role of Outpatient Physicians in Reporting of Hematopoietic Malignancies for Statewide Surveillance

Sue-Min Lai, PhD, MS, MBA; John Keighley, PhD; Sarma Carimella, MBBS, MPH

Abstract: Background: The role of physicians in outpatient setting in reporting of hematopoietic malignancies is not well known. Objectives: This study described the approaches that Kansas Cancer Registry (KCR) used to ascertain completeness of hematopoietic malignancies reporting at the state level. Our study also examined the role of hematologists, oncologists and primary care physicians (PCP) in outpatient setting in reporting of hematopoietic malignancies. Methods: KCR engaged all outpatient hematologists, oncologists, and a sample of PCPs who cared for patients in geographic areas where there was limited access to hematologists/oncologists. Cases that met reportable eligibility were identified using the ICD-9-CM codes from the medical record disease index files and confirmed by reviewing patient medical records. Confirmed cases were then abstracted and sent to the registry. The study focused on 2010 diagnosed Kansan cases. Results: Of the total 2010 diagnosed cases, 18.7% were reported solely by outpatient physicians (17.0% reported by outpatient hematologists/oncologists and 1.7% reported by outpatient PCPs only). Fifty-eight percent of polycythemia vera was diagnosed and treated by outpatient hematologists, oncologists, and some PCPs. Using reportable ICD-9-CM codes only for hematopoietic malignancies causes an overestimation of the true reportable hematopoietic malignancies cases. Conclusion: Outpatient physicians are critical in the scheme of care for hematologic malignancies. Therefore collection of cancer data from these outpatient providers by a well operated statewide registry provides a far more accurate picture of what is really going on with hematopoietic malignancies.

Key words: cancer surveillance, hematopoietic malignancies, polycythemia vera

Introduction

Hematopoietic malignancies have been reportable conditions to most state central cancer registries since 1995, with a few histologies becoming reportable thereafter. Examples include polycythemia vera (PV), myelodysplastic syndromes (MDSs), and chronic myeloproliferative disorders (CMDs) that became reportable to state central registries in 2001. Hematopoietic malignancies constitute about 10% of all invasive cancers diagnosed in the United States and have not been studied for their level of complete ascertainment. Studying the level of completeness in reporting of hematopoietic malignancies is complicated because these are a collection of many diseases and have been diagnosed and treated in various settings including inpatient only, outpatient only, and a constellation of inpatient, outpatient, and other care providers. Some states have statutes that require cancer reporting from all licensed health care providers in their states while others only require cancer reporting from hospitals and pathology laboratories. The inconsistency in reporting requirements across states makes evaluation of completeness of reporting on these cancers in the United States challenging. This issue has been addressed in studies of melanoma cancer.

Due to an increasing awareness of cancers in communities, cancer clusters have become a major concern to the public, researchers, and public health practitioners. Cancer cluster investigations have continued to be carried out by local, state, and federal agencies. Based on publicly available reports on cancer cluster investigations, Goodman et al identified 428 cluster investigations evaluating 567 cancers of concern since 1990. Only 0.5% of 567 total cancer types suggested some evidence of an association between the cancer(s) of concern and hypothesized exposures. However, the level of certainty of these findings differed. Reasons for cluster investigations not having advanced our understanding of disease have been discussed. The reasons include, but are not limited to, individual cancer clusters being too small, reported clusters having vague or heterogeneous definitions of disease, poorly defined or undefined exposures, and selected populations of primary analytic interest being flawed by a posteriori reasoning. Completeness of cancer types that were being investigated has not been addressed in these reports.

In a recent effort to study the role of the environment in the origin of polycythemia vera in one area of Eastern Pennsylvania, PV cases from the Pennsylvania State Cancer Registry were used. This investigation generated several important observations including the role of physicians in outpatient setting in diagnosing these cases. Hereafter physicians in an outpatient setting are referred to as outpatient physicians. The importance of cancer case reporting from outpatient physicians for melanoma and prostate
cancers has been studied. These studies have documented underreporting of these cancers from outpatient setting providers. In 2009, a request for proposal (RFP) to improve the reporting of hematopoietic malignancies was solicited by the National Program of Cancer Registries (NPCR) with an overarching goal of identifying best practices to improve reporting of these conditions diagnosed in outpatient physicians. The Kansas Cancer Registry was one of the 3 state cancer registries awarded to address this issue, via a mechanism of subcontract with ICF–Macro.

This paper summarizes the approach the Kansas Cancer Registry (KCR) has taken to systematically ascertain the extent of hematopoietic malignancies of the blood and bone marrow that were diagnosed and treated solely by outpatient physicians. Because PV is the major focus of the RFP, this paper addresses the number of PV and also non-PV hematopoietic malignancies of the blood and bone marrow that were solely reported by outpatient hematologists, oncologists, and a sample of primary care physicians (PCPs).

### Methods

Cancer reporting in Kansas is governed by state laws and regulations. Reporting from hospitals, pathology laboratories, and freestanding surgical and radiation treatment centers has been mandated since 1997 while active reporting from outpatient physicians and other health care providers was in place since 2004. Registry operations (such as cancer reporting from providers and data quality control audits including the activities described in this paper) are part of KCR routine activities and have been approved by the Human Subjects Committee at the University of Kansas Medical Center. The approach that KCR used to examine whether outpatient providers that diagnosed and/or treated hematopoietic malignancies of the blood and bone marrow and have submitted all reportable hematopoietic cases to the registry is briefly discussed below. The reportable codes from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and International Classification of Diseases for Oncology, third edition (ICD-O-3) are shown in Table 1. This paper focused on 2010 diagnosed Kansan cases regardless of their age, sex, and race.

### Identifying and Engaging Outpatient Physicians

To supplement the physician roster that KCR has compiled since its inception, additional physicians and providers were identified using the National Provider Index (NPI) and lists of physicians with their specialty provided by the Kansas Board of Healing Arts. KCR also used internet websites (such as Administrators in Medicine and American Medical Association) to identify hematologists and oncologists.

Within each data source, physicians of interest were selected based on their self-identified specialty (hematology and/or oncology vs PCPs, including internal medicine, family medicine, pediatrics, and obstetrics and gynecology). A survey was then mailed to all identified hematologists and oncologists to obtain detailed information regarding their practice location and their affiliation with any hospital(s). Because the RFP focused on reporting of PV and non-PV hematopoietic malignancies from physicians in outpatient settings, all hematology and oncology physicians that were not affiliated with hospitals were included in this study. Rosters of physician group practices are not available, not only in Kansas, but also in many other states in the nation. As such, physicians’ addresses (if not already part of a known practice group) were mapped using Geographic Information System (GIS) software and grouped together. This effort was particularly crucial in identifying PCP practices because there were over 100 practices (or clinics).

The definition of a clinic as defined in the Kansas Administrative Regulations Article 28-70-1. In this article, clinics are defined as an establishment of 1 or more physicians who diagnose and treat patients in a permanent facility. Clinics do not provide services or accommodations for any patient to stay for more than 24 hours. With so many PCP clinics, we stratified PCP clinics into 2 large (50 or more physicians), 14 medium (10–49 physicians) and 119 small clinics (fewer than 10 physicians). Both large

<table>
<thead>
<tr>
<th>Table 1. List of Hematopoietic Histology of the Blood and Bone Marrow by ICD-9-CM and ICD-O-3 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Histology</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Plasma Cell</td>
</tr>
<tr>
<td>Immunoproliferative Diseases</td>
</tr>
<tr>
<td>Mast Cell Tumors</td>
</tr>
<tr>
<td>Histiocytosis</td>
</tr>
<tr>
<td>Leukemias</td>
</tr>
<tr>
<td>Polycythemia Vera (PV)</td>
</tr>
<tr>
<td>Chronic Myeloproliferative Disorders</td>
</tr>
<tr>
<td>Refractory Anemias, Neutropenia &amp; Thrombocytopenia</td>
</tr>
</tbody>
</table>
clinics were included in the study while 2 of the 14 medium clinics and 7 of the 119 small clinics were pseudo-randomly selected with preferences given to clinics where no hematologists and oncologists were practicing. All 14 outpatient clinics that have at least 1 hematologist or oncologist were included in this audit. One oncology clinic only diagnosed and treated solid tumors and therefore contributed no cases. Another practicing pediatric oncologist only followed up his patients in his outpatient clinic. Therefore his patients were abstracted and reported by a hospital abstractor. In summary a total of 23 clinics (11 PCPs and 12 hematology/oncology clinics) were identified and included in this study.

An introductory letter was mailed to the office administrator about the purpose of this audit, followed by phone calls to answer any questions the participating clinics might have regarding their role in this audit.

Procedures and Process

Clinics were given the freedom to submit the Medical Record Disease Index in any electronic format that their system allowed. Electronic files were first standardized at the clinic level, imported into an Access database, and combined into 1 table per clinic. Standardization processes included tasks such as removing duplicate ICD-9-CM code combinations for a patient, formatting Social Security number, date of birth, and separating patient names from 1 field into prefix, first, middle, last name, and suffix fields. Tables were then standardized across clinics, checked for the first occurrence of a patient within a clinic, and ICD-9-CM code combination across time. Only patients with reportable codes of interest (Table 1) and records with 2010 dates of service were retained. The ICD-9-CM codes were then converted to ICD-O-3 codes. All records with dates of service in 2010 were then matched with the KCR database. These records were matched based on patient identifiers and cancer type. Patients who did not match with the KCR database were used to generate a list that was then sent to participating clinics. Participating clinics were responsible for reviewing the patient’s entire medical record and abstracting cases that met the reportable criteria. KCR reviewed and abstracted cases for most of the PCP clinics since many of these small clinics were not trained to do so.

Analysis

In addition to a description of the audit process of PV cases (the focus of the RFP) from outpatient physicians, we also addressed the audit experience related to non-PV hematopoietic malignancies. PV and non-PV hematopoietic malignancies that were confirmed through this audit were described according to initiation by PV and non-PV ICD-9-CM hematopoietic malignancies codes, respectively.

In order to show the percentage of hematopoietic malignancies reported by outpatient physicians, we have included all cases that were reported by non-outpatient physicians (e.g., hospitals or path labs) in the denominator. The distribution of all 2010 confirmed diagnosed PV and
non-PV hematopoietic malignancies in Kansas was then shown according to their reporting sources. Cases were classified as outpatient setting cases if they were solely reported by outpatient radiation treatment centers, outpatient surgical centers, outpatient cancer centers, outpatient primary care clinics, or a combination of these outpatient facilities. Cases were classified as non-outpatient setting cases when they were reported solely by hospitals, a combination of hospital and outpatient facility, pathology laboratories, or a combination of hospital and pathology laboratories.

The level of completeness in ascertainment of hematopoietic malignant cases in Kansas was evaluated by calculating the expected number of hematopoietic malignancies for 2010 diagnoses year using an indirect age standardization/adjustment method by applying 2010 diagnosis-year age-specific rates from the SEER program to the corresponding age-specific population in Kansas.

**Results**

The submitted medical record disease index files from the 23 outpatient clinics contained 100,773 records (63,127 patients). Figure 1 shows records attrition from ICD-9-CM codes of reportable hematopoietic malignancies to confirmed hematopoietic malignancies using the ICD-O-3 codes (Table 1). After cleaning and removal of non-reportable codes 1.7% of the total submitted records (1,737/100,773) had reportable ICD-9-CM hematopoietic malignancies codes. Of the
Table 4. Newly Diagnosed 2010 Hematopoietic Malignancies in Kansas

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Non-Outpatient Setting</th>
<th>Outpatient Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 800</td>
<td>N = 184</td>
</tr>
<tr>
<td></td>
<td>Initiated by PCP* or HO*</td>
<td>Initiated by Inpatient Facility</td>
</tr>
<tr>
<td></td>
<td>n = 86</td>
<td>n = 714</td>
</tr>
<tr>
<td>Cancer</td>
<td>Total</td>
<td>PCP</td>
</tr>
<tr>
<td>Chronic myeloproliferative disorders</td>
<td>85</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Histiocytosis</td>
<td>7</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Immunoproliferative Diseases</td>
<td>8</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Leukemias</td>
<td>480</td>
<td>4 (0.8)</td>
</tr>
<tr>
<td>Mast cell tumors</td>
<td>2</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Plasma cell tumors</td>
<td>187</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>Polycythemia Vera</td>
<td>48</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Refractory anemias, neutropenia, thrombocytopenia</td>
<td>167</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Total cases</td>
<td>984</td>
<td>6</td>
</tr>
</tbody>
</table>

*Primary Care Physician (PCP).
†Hematologist/Oncologist (HO).
**Some cases were jointly reported by HO and PCP.

1,737 records, 544 were matched to the records that were already registered in the KCR databases. The paragraphs below summarize the results on confirmed polycythemia vera and non-polycythemia vera cases from eligible records using ICD-9 CM code of polycythemia vera and ICD-9-CM codes of non-polycythemia vera hematopoietic malignancies, separately.

One hundred and seventy-three records (164 patients) had an ICD-9-CM code of 238.4 (namely, polycythemia vera). Of these, the reporting facilities confirmed 39 polycythemia vera cases, 7 non-polycythemia vera hematopoietic malignancies, and 119 patients having no reportable conditions across all years (Figure 1). Twenty-six of the 39 polycythemia vera cases were newly diagnosed in 2010 while the remaining 13 polycythemia vera had a diagnosis year other than 2010. Four of the 7 non-polycythemia vera hematopoietic malignancies were newly diagnosed in 2010. The frequency of confirmed subtype is shown in Table 2A. Reasons for the 119 non-reportable conditions were secondary polycythemia vera (36 cases), polycythemia vera diagnosed in 2000 and earlier (7 cases), disorders of iron metabolism (7 cases), hemoglobin > 150 g/L and hematocrit > 46% (12 cases), and other miscellaneous reasons (57 cases). Among the twenty-six 2010 diagnosed polycythemia vera, Janus Kinase 2 was positive in 14 (53.8%), negative in 2 (7.7%), and not done/unknown in 10 (38.5%). Two of the 3 non-polycythemia vera hematopoietic malignancy cases that were diagnosed in 2010 were Janus Kinase 2 positive.

Of the 1,020 records (or 883 unique patients) that had non-polycythemia vera hematopoietic malignancy ICD-9-CM codes, 362 non-polycythemia vera hematopoietic malignancies and 3 polycythemia vera (from 360 patients) were confirmed by reporting facilities via chart reviews. The confirmed subtypes resulting from reviewing non-polycythemia vera hematopoietic malignancy codes are shown in Table 2B. Of the total confirmed cases, 212 non-polycythemia vera hematopoietic malignancies and 3 polycythemia vera were newly diagnosed in 2010 (Figure 1). Janus Kinase 2 positive was noted in 1 of the 3 polycythemia vera cases.

Reasons for cases that were reportable by ICD-9-CM codes of non-polycythemia vera hematopoietic malignancies, but were later confirmed by the facilities as not reportable to KCR using the ICD-O-3 are (1) not cancers/non-reportable in 523 patients and (2) having insufficient information at their facility to be abstracted in 25 patients. The 25 cases were subsequently reported by a hospital. A sample of reasons for being nonreportable is included in Table 3.

We analyzed the level of completeness in ascertaining hematopoietic malignancies before establishing the extent of such cases that were reported by outpatient clinics. There were a total of 984 hematopoietic malignancies that were newly diagnosed in 2010 and were registered in the Kansas
Cancer Registry as of April 1, 2012. The expected number of hematopoietic malignancies for the 2010 diagnosed year was calculated to be 838 using the age-specific rates from all SEER registries combined. When the age-specific rates for 2010 diagnosed cases from the Iowa SEER registry were used the expected number of hematopoietic malignancies was found to be 939 cases. We believe the registration of hematopoietic malignancies is indeed complete in Kansas.

Of the 984 hematopoietic malignancies, 48 were polycythemia vera and 936 were non-polycythemia vera hematopoietic malignancies (Table 4). One hundred and eighty-four cases (18.7%) were reported solely by outpatient physicians/clinics while 81.3% (800 cases) were reported by non-outpatient settings. Of the 800 cases, the reporting of 86 cases were initiated at an outpatient setting, followed by case completion at non-outpatient facilities. When considering the specialty of reporting outpatient physicians, 4 of the 48 polycythemia vera (8.3%) and 13 of the 936 non-polycythemia vera hematopoietic malignancies (1.4%) were reported solely by outpatient primary care physicians. The distribution of reportable polycythemia vera and non-polycythemia vera hematopoietic subtypes by non-outpatient and outpatient setting is shown in Table 4.

**Discussion**

Our statewide audit has shown the importance of outpatient hematologists, oncologists, and primary care physicians in ascertaining hematopoietic malignancies. We found 18.7% of the total reportable hematopoietic cases diagnosed in 2010 to be solely reported by outpatient physicians. This 18.7% included 17.0% being reported by outpatient hematologists and oncologists and 1.7% reported by outpatient PCPs only. Close to 19% of the hematopoietic malignancies could be missed if outpatient physicians/clinics were not required to report cancer cases to the state cancer registry. Of the 8 major groups of hematopoietic malignancies, the impact of outpatient physician reporting is most pronounced for polycythemia vera cases. Over 58% of the 2010 diagnosed PVs were reported solely by outpatient clinics (50% from the outpatient hematologists/oncologists and 8% from the PCPs) in Kansas. Outpatient diagnosed PVs were not likely to be missed in Kansas because outpatient physician reporting has been in place since 2004. In fact, the calculated expected number of hematopoietic malignancies for Kansans (838 with all SEER combined or 939 considering IOWA SEER only) relative to 984 cases registered in the KCR further supports the importance of having active reporting from outpatient providers.

Investigation of PV clusters will be problematic in states where cancer reporting from outpatient physicians is not mandated by their state law. Similarly, the degree of underreporting would also undermine the surveillance effort for non-PV hematopoietic malignancies if outpatient facilities were not reporting these subtypes to the state cancer registries.

Our observation about PVs from outpatient physicians is consistent with that observed in a recent investigation of PV in Eastern Pennsylvania.12 Our study also noted that a large proportion of hematopoietic malignancies like chronic myeloproliferative disorders, and refractory anemias, neutropenia, and thrombocytopenia were also diagnosed and treated by outpatient physicians. When it comes to cancer cluster investigations, completeness of reporting for these cancers should be of concern in states where no outpatient cancer reporting is implemented.18

The important role of outpatient physicians in diagnosing and treating patients with such disorders has long been suspected, but the extent of their contribution has not been documented until now. To our best knowledge, this paper is the first ever documentation showing their role from a statewide population-based surveillance perspective. Several reasons for not having this information readily available are described as follows. First, there is a data field Type of Reporting Source in cancer surveillance systems that captures the reporting source from where the best information (or complete information) was obtained when abstracting a case. This field has 8 possible choices using either definition. In theory, this definition is simple and clear, but it is otherwise in practice. Some cancer surveillance systems have attributed cases to the type of reporting as hospital inpatient when a case report involved a hospital regardless of the type of information that was contributed from that hospital. In our series, 86 hematopoietic malignancies (2 PVs and 84 non-PVs) that were first reported by participating outpatient clinics were reclassified as non-outpatient cases. The timing of the casefinding processes which varied across registries also played a role on how this field was coded. Second, the inconsistencies in coding this field could also be attributed to the vague definition of the term “the best information or complete information” in the coding manual. In fact, the coding manual fell short to define what “the best information or complete information” meant.

Third, the type of reporting source has undergone several revisions. For example, the North American Association of Central Cancer Registries (NAACCR) Data Standards & Data Dictionary Version 10.1 combined hospital inpatient and outpatient or clinics together21 while the NAACCR Data Dictionary Version 12, 14th edition20 has hospital inpatient as a separate code and another code to include radiation treatment centers or medical oncology centers that are either affiliated to a hospital or independent. The latest version, 16th edition22 again is not able to correctly quantify the number of cases seen in outpatient settings because radiation treatment centers or medical oncology centers that are affiliated with a hospital or being independent have been combined as a group (code=2). Lastly, outpatient physician/clinic reporting has not been actively pursued by some population-based cancer registries which can also be problematic when it comes to quantifying the role of outpatient physicians/clinics reporting of these reportable malignancies.

Another important observation from this audit was related to coding of diagnoses and treatment. The ICD-9-CM is the official system of assigning codes to diagnoses and procedures associated with hospital and outpatient clinic utilization in the United States. On the other hand, population-based cancer registries (United States and internationally all alike) currently use the International
Conclusion

Hematopoietic malignancies may be missed depending on whether state statutes require cancer reporting from outpatient physicians/clinics and the degree of proactive reporting from outpatient physicians/clinics (if state laws mandate in fact exist). Outpatient physicians and clinics are critical in the scheme of care for hematologic malignancies. Not recognizing their role in diagnosis and/or treatment of such cases can deter the effort on cancer prevention and control.

References

An Assessment of the Reliability of Race, Hispanic Ethnicity, Birthplace, and Tobacco History Data in the Massachusetts Cancer Registry, 2005–2009

R. Knowlton; S. Gershman a; A. Solis; B. Das b

Abstract: The Massachusetts Cancer Registry (MCR) reviewed the medical charts of 5,438 randomly selected breast and colorectal cancer cases diagnosed from 2005–2009 in part to assess the reliability of the race, Hispanic ethnicity, birth country, and tobacco history variables. The kappa statistic was used to assess the reliability between the originally reported variable and the reabstracted variable. There was strong agreement of kappa score for race among whites, blacks, and other (Asian, Native American), indicating a good quality of race data. The agreement for birth country was strongest among those not born in the United States with a statistically significantly higher kappa score compared to the other categories. Agreement for Hispanic ethnicity was strongest for non-Hispanics, Puerto Ricans, Central Americans, and Dominicans, groups that represent the majority of Hispanics in Massachusetts. The agreement for tobacco history was strongest among current tobacco users. This study provided useful information on the reliability of the race, Hispanic ethnicity, and birthplace variables, which are frequently used in MCR reports. It also provided heretofore unknown data on the reliability of the tobacco history variable. All categories of the race variable were very reliable as were the categories of Puerto Rican, Dominican, and Central American. Hispanic, NOS was not as reliable due to the large Portuguese population with Hispanic sounding surnames. Birth country was not as reliable due to the paucity of the data in many of the larger facilities in the state.

Key words: birthplace, Hispanic ethnicity, kappa score, race, tobacco history

Introduction

Differences in cancer incidence by racial and ethnic groups have long been documented in cancer surveillance data.1 Prominent racial/ethnic differences in the Massachusetts 2005–2009 cancer data mirrored national data with elevated incidence rates of prostate cancer among black, non-Hispanic (NH) males, liver cancer among Asian, NH males, melanoma among white, NH males and females, cervical cancer among black, NH and Hispanic females, multiple myeloma among black, NH females, and breast, lung, and ovarian cancer among white, NH females.2 Data on race, ethnicity, and birthplace are necessary for uncovering these differences and prompting further research to help explain them.3 Massachusetts has become a more racially and ethnically diverse state over the past few decades. While 3 quarters of the Massachusetts 2010 population was white, NH, Massachusetts has several minority populations that provide an opportunity to further examine health disparity outcomes.4 Given the importance of these demographic data, it is prudent to periodically review a subset of the data being reported to the central registry in order to assess its reliability. The results can be used to enhance training efforts to improve data quality.

Since 2007, hospitals and health plans in Massachusetts have been mandated by regulation 114.1 CMR 17.00 of the Massachusetts General Laws to report detailed race and ethnicity information on every patient discharge. Hospitals are required to collect data only for inpatient hospitalizations, observation unit stays, or emergency department visits, but most are collecting race/ethnicity data for patients seeking other services, typically during patient registration.5 In addition to the legislation created for improved race/ethnicity reporting, the Massachusetts Department of Public Health Office of Health Equity was created in April 2008 via An Act Eliminating Racial and Ethnic Health Disparities in the Commonwealth (H. 2234). This office coordinates and leads state work related to reduction of health disparities, including race/ethnicity data collection.

Although tobacco history was retired as a North American Association of Central Cancer Registries (NAACCR) variable in 2010, the Massachusetts Cancer Registry (MCR) continues to collect this variable. The associations between smoking and cancer have been known since the landmark study on smoking and lung cancer by Richard Doll and Bradford Hill was published in 1950.6 Several other cancers have also been linked with tobacco use: nasopharyngeal, oral, laryngeal, esophageal, pancreatic, uterine, cervical, ovarian, kidney, bladder, stomach, colorectal, and acute myeloid leukemia.7 The MCR tobacco history data have never been subjected to an audit.

*Massachusetts Cancer Registry. aWestat Corporation.

The Massachusetts Cancer Registry (MCR) was contracted to participate in a project to develop innovative uses of cancer registries for public health applications. This project was funded as part of American Recovery and Reinvestment Act (ARRA) Comparative Effectiveness Research activities through the Centers for Disease Control and Prevention. The project involved reabstracting primary payer data on breast and colorectal cancer cases diagnosed from 2005 to 2009 and abstracting primary payer data at the initiation of each treatment. In addition to the payer and treatment data, race, Hispanic ethnicity, birth country, and smoking status were also reabstracted for validity and reliability.
This study was undertaken to assess the reliability of the reporting of race, Hispanic ethnicity, birthplace and tobacco use to the MCR and to identify data items in need of improvement.

**Methods**

This quality control effort was a component of the Centers for Disease Control and Prevention/National Program of Cancer Registries (CDC/NPCR) Comparative Effectiveness Research (CER) project which awarded money to central cancer registries involved in the innovative use of registry data. The MCR was awarded money for the study of the primary payer field before and after the 2006 health care reform bill and the reliability of other demographic fields. Colorectal and breast cancers were chosen for study as they are major cancers which are more likely to be diagnosed and treated in a hospital setting.

**Study Population**

The study population for this project encompassed both men and women diagnosed with invasive colorectal cancer and women diagnosed with invasive breast cancer from 2005–2009. The cases included Massachusetts residents diagnosed at a facility within Massachusetts and reported to the MCR. Cases diagnosed at a Veterans Affairs (VA) Hospital or an out-of-state facility were excluded as were death certificate only cases, autopsy cases, cases who were diagnosed and treated at another facility other than the reporting facility, and patients who were diagnosed and received all of their first course of treatment in a physician’s office. Between 2005–2009, there were 16,840 cases of invasive colorectal cancer among male and female residents of Massachusetts and 26,449 cases of invasive breast cancer among female residents.

**Sampling Strategy**

The target sample size was planned to be 5,000 cases diagnosed from 2005–2009. Owing to the fact that the proportion of breast to colorectal cancer case was 60:40 in the MCR database and the male/female split for colorectal cancer cases was 50:50, 3,000 breast cancer cases and 2,000 colorectal cancer cases were targeted for abstraction. In addition, a reserve sample of 500 cases to compensate for the loss of sample size due to ineligibles (eg, no access to the chart) was identified. The patients were randomly selected from the list of breast and colorectal patients diagnosed during this 5-year period. As such, each patient was assigned a weight which summarized how much of the population of patients this patient represented. The weights were used to calculate statistics from the data collected, thus ensuring that the figures calculated were relevant to the population of cancer patients in Massachusetts and not merely to the sample of patients in the study. A sampling rate was set and applied to the cases. This fixed sampling rate approach attained a self-weighting sample across the 5 years. Year (2005–2009) was used as strata, and then an implicit stratification was done on the variables of interest: gender, race, ethnicity, stage, and age. The implicit stratification ensured proportional representation of cases across years and by characteristics. Sampling weights were calculated as the inverse of the probability of selection.

There were 5,348 cases randomly selected for chart review. Of the 4,981 cases selected in the original (nonreserve) sample, 4,972 were reviewed (99.8%) and of the 500 cases in the reserve sample, 376 were reviewed (75.2%). The reabstraction rate was nearly 100%. There were a total of 7,473 charts reviewed for these 5,348 cases due to many patients being seen at more than one facility.

**Analysis Plan**

The reliability of the race, Hispanic ethnicity, birthplace, and tobacco history data was examined for the study cancers for the 2005–2009 period. These fields were reabstracted by trained medical record abstractors working for the MCR. Comparisons were done between the original reported data and the reabstracted data. Owing to the sparseness of some of the data categories, these variables were categorized such that meaningful and reliable within group analyses could be produced. The Cohen’s kappa statistic (k) was used for the reliability analyses. Kappa statistics are used to measure chance corrected agreement. The agreement measurement goes from 0 (no agreement) to 1 (perfect agreement), with values greater than 0.8 considered to be excellent, values between 0.61 to 0.8 considered to be substantial, and values between 0.41 and 0.60 considered to be moderate. A 95% confidence interval was calculated for each kappa score to measure statistical significance when comparing two kappa scores.

The numbers presented in Tables 2 through 5 for the reliability analyses are based on weighted numbers derived from the randomly selected cases for this study.

**Results**

Table 1 presents the characteristics of the unweighted sample. Since this sample includes both breast and colorectal cancer cases, 80% of the sample is female. Approximately 50% of the cases were over the age of 65, over 90% were white, and 95% were non-Hispanic. Almost 90% of the cases were diagnosed in either the local or regional stage of disease. Either current or past tobacco use was reported by 43% of the cases.

Results on the reliability of reported race are presented in Table 2. There was strong agreement of kappa score for race among whites, blacks, and other (Asian, Native American), indicating a good quality of race data.

Data on birth country reliability are presented in Table 3. The agreement for birth country was strongest among those not born in the United States with a statistically significantly higher kappa score compared to the other categories. In many hospitals, the place of birth was not documented, especially for patients born either in Massachusetts or other parts of the United States. For those patients born outside the United States, the birth country was often documented in the patient history section of the hospital discharge summary or in social worker’s notes rather than in a demographic face sheet. Due to the greater frequency of foreign born patients having this birth country documentation, the kappa statistic was higher for this group.
<table>
<thead>
<tr>
<th></th>
<th>TOTAL</th>
<th>Breast</th>
<th>Colorectal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Total</td>
<td>5,348</td>
<td>100</td>
<td>3,227</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>1,047</td>
<td>19.6</td>
<td>807</td>
</tr>
<tr>
<td>51-64</td>
<td>1,577</td>
<td>29.5</td>
<td>1,063</td>
</tr>
<tr>
<td>65-74</td>
<td>1,074</td>
<td>20.1</td>
<td>619</td>
</tr>
<tr>
<td>75-84</td>
<td>1,146</td>
<td>21.4</td>
<td>538</td>
</tr>
<tr>
<td>85+</td>
<td>504</td>
<td>9.4</td>
<td>200</td>
</tr>
<tr>
<td>Sex*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,040</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4,307</td>
<td>80.5</td>
<td>3,226</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>4,968</td>
<td>92.9</td>
<td>3,003</td>
</tr>
<tr>
<td>Black</td>
<td>229</td>
<td>4.3</td>
<td>139</td>
</tr>
<tr>
<td>Other</td>
<td>135</td>
<td>2.5</td>
<td>77</td>
</tr>
<tr>
<td>Unknown</td>
<td>16</td>
<td>0.3</td>
<td>8</td>
</tr>
<tr>
<td>Birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>1,800</td>
<td>33.7</td>
<td>1,026</td>
</tr>
<tr>
<td>Other US state or Puerto Rico</td>
<td>869</td>
<td>16.2</td>
<td>526</td>
</tr>
<tr>
<td>Non-USA</td>
<td>484</td>
<td>9.1</td>
<td>268</td>
</tr>
<tr>
<td>Unknown</td>
<td>2,195</td>
<td>41.0</td>
<td>1,407</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Localized</td>
<td>3,152</td>
<td>58.9</td>
<td>2,212</td>
</tr>
<tr>
<td>Regional</td>
<td>1,607</td>
<td>30.0</td>
<td>860</td>
</tr>
<tr>
<td>Distant</td>
<td>501</td>
<td>9.4</td>
<td>138</td>
</tr>
<tr>
<td>Unstaged/unknown</td>
<td>88</td>
<td>1.6</td>
<td>17</td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>5,113</td>
<td>95.6</td>
<td>3,081</td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>48</td>
<td>0.1</td>
<td>26</td>
</tr>
<tr>
<td>Central/South American</td>
<td>36</td>
<td>0.1</td>
<td>26</td>
</tr>
<tr>
<td>Dominican</td>
<td>35</td>
<td>0.1</td>
<td>22</td>
</tr>
<tr>
<td>Hispanic, NOS</td>
<td>36</td>
<td>0.1</td>
<td>20</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>&lt;0.1</td>
<td>6</td>
</tr>
<tr>
<td>Unknown</td>
<td>72</td>
<td>0.1</td>
<td>46</td>
</tr>
<tr>
<td>Tobacco</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Used</td>
<td>2,498</td>
<td>46.7</td>
<td>1,623</td>
</tr>
<tr>
<td>Current User</td>
<td>698</td>
<td>13.1</td>
<td>413</td>
</tr>
<tr>
<td>Past User</td>
<td>1,597</td>
<td>29.9</td>
<td>926</td>
</tr>
<tr>
<td>Unknown</td>
<td>555</td>
<td>10.4</td>
<td>265</td>
</tr>
</tbody>
</table>

*Transgender case included in the total.

Table 4 shows findings on the reporting of Hispanic ethnicity. Agreement for Hispanic ethnicity was strongest for non-Hispanics, Puerto Ricans, Central Americans and Dominicans. The latter 3 groups represent the majority of Hispanics in Massachusetts. Since Cubans and Mexicans represent a very small percentage of the Massachusetts Hispanic population, any agreement measures for these groups were subject to greater data variability. The lower agreement values for Hispanic, NOS indicates that the abstractors were able to determine specific Hispanic ethnicities as a result of the chart reviews. Nearly 20% of Hispanic, NOS cases were reabstracted as non-Hispanic and 14% were reabstracted as Puerto Rican.

Data on the reliability of history of tobacco use are presented in Table 5. The agreement for tobacco history was strongest among current tobacco users. For the purposes of reabstraction, a chart needed to indicate that the patient never used tobacco in order to be classified as never used. If the chart did not clearly indicate this, the tobacco history was classified as unknown. Many cases originally reported as never used tobacco in fact had no clear statement in the chart to that effect, thus accounting for the low agreement...
among those categorized as never used and those categorized as unknown.

**Discussion**

There was a wide variety of kappa agreement values between originally reported and reabstracted values for race, Hispanic ethnicity, birthplace, and tobacco history. Among the categories for the race variable, white and black had excellent agreement according to their kappa score while other was also good. Other included Asian, NH and Native American. The agreement was very weak for cases originally reported with an unknown race, indicating that the data reabstractors were able to find an actual race for many of these cases. The strong kappa agreement for white and black cases is supported by a study which analyzed the validity of race and ethnicity data for Medicare beneficiaries. In that study, which compared self-reported race with race listed in Medicare, the sensitivities were high for white...
The categories for Hispanic ethnicity varied from excellent kappa scores for Puerto Rican, Central American, and Dominican, to moderate scores for non-Hispanic and Hispanic, not otherwise specified (NOS), to a poor kappa score for unknown. While one study found that persons classified as non-Hispanic by both surname and medical record report to the cancer registry were very likely to identify themselves as such and most self-identified non-Hispanics were classified correctly, this was not always the case in Massachusetts. This is likely due to the large Portuguese population in the state and the tendency of Portuguese surnames to be identical to Spanish surnames. Previous race audits of MCR data have discovered many Hispanic, NOS cases actually being Portuguese. As a result, these cases were reclassified as non-Hispanic. As with the race variable, the poor score for unknown indicates the ability of the data reabstractors to identify an actual Hispanic ethnicity when there was not one originally reported to the MCR. This was also the case for the Hispanic, NOS variable. The strong agreement for Puerto Ricans, Dominicans, and Central Americans was a positive finding as these ethnicities represent most of the Hispanic ethnicities in Massachusetts.

Agreement among the categories for birthplace was not as strong as the categories for race and Hispanic ethnicity. The birthplace variable was one that was not consistently collected at all hospitals in the state. Some of the largest hospitals did not collect it while some of the smallest did. At some facilities, it is not readily available in the demographic sheet of a medical record or is buried in another part of the chart such as the social worker or interpreter notes. Agreement was strongest among those cases born outside the United States. This is not surprising given the fact that, unless indicated in the demographic face sheet, a US-born patient was not likely to be documented elsewhere and was reported to the registry as having an unknown birthplace. A foreign-born patient was more likely to have his/her birthplace documented elsewhere in the chart.

The agreement for tobacco history was strongest among current tobacco users and moderate among past tobacco users. Agreements were much lower for the never used and unknown categories. A study in England which compared smoking from medical records to national survey data found that while estimates of the prevalence of current smoking were similar in both the record data and the survey data, the prevalence of former smoking was underestimated by about 5% in primary care. This same pattern may explain why the review of charts in this study yielded lower kappa scores for previous tobacco users as compared to current users.

In summary, this study provided useful information on the reliability of the race, Hispanic ethnicity, and birthplace variables, which are frequently used in MCR reports. It also provided heretofore unknown data on the reliability of the tobacco history variable. All categories of the race variable were very reliable as were the categories of Puerto Rican, Dominican, and Central American. Hispanic, NOS was not as reliable due to the large Portuguese population with Hispanic sounding surnames. Birth country was not as reliable due to the paucity of the data in many of the larger facilities in the state. While birth country data will not become more available in the future, more detailed ethnicity will be as a result of regulation 114.1 CMR 17.00 of the Massachusetts General Laws to report detailed race and ethnicity information on every patient discharge. The hospitals that did not collect birthplace at least now collect ethnicity which includes a category for American. While the NAACCR variable for race does cover many ethnicities, 2 of the predominant ones in Massachusetts (Cape Verdean and Haitian) are not in this category. Birth country, when available, is the only opportunity to collect data on these ethnicities. The tobacco history variable, retired by NAACCR in 2010, is of limited use when trying to examine the quantity of smoking as there is no pack year history recorded. Pack year history would be a better variable along with dates of smoking. The American Cancer Society reported that there are many cancers related to tobacco use and a complete tobacco use history should be collected on all cancer cases.

References
Colorectal Cancer Care at Falmouth Hospital: An Assessment of Progress Through Registry Statistics

Amanda Doodlesack

Abstract: The purpose of this study is to assess the improvement of one aspect of colorectal cancer care at Falmouth Hospital as compared to a prior quality assessment report published in 2011. This quality assessment by I. Hopewood reported the need to increase the average number of lymph nodes removed and examined per patient in order to improve compliance with national guidelines. Falmouth Hospital pathology implemented a new alcohol-formalin fixation technique in 2011 to increase the number of resected lymph nodes examined. By comparing care based on the quality of surgery (ie, number of nodes examined) and stage of diagnosis, this study aimed to assess whether the quality improvement measures implemented may have an impact. Data were obtained from the hospital’s tumor registry and the National Cancer Data Base. Results showed an increase in average number of nodes examined per patient, but no statistically significant change in the distribution of stage at diagnosis or evidence for detection of later stage cancers.

Key words: cancer registries, cancer staging, colorectal cancer

Introduction

In the United States, colorectal cancer is the third most commonly diagnosed cancer in both men and women, and is the third leading cause of cancer death. Colorectal cancer constitutes 9% of all cancer cases in both sexes, trailing breast and lung cancer in females, and prostate and lung cancer in males. The American Cancer Society estimates that in the year 2013, there will be 142,820 new cases and 50,830 deaths due to colorectal cancer in the United States. In order to minimize the number of deaths due to colorectal cancer each year, it is important that hospitals administer the highest quality of cancer care.

Resection and assessment of lymph nodes during surgery is an essential parameter of cancer care quality. The National Comprehensive Cancer Network recommends that a minimum of 12 lymph nodes be examined in order to accurately stage the disease. Accurate staging is critical to developing a proper treatment plan for the patient. Numerous studies have shown a positive correlation between the number of nodes retrieved and examined and the survival rate of patients. Detection of additional numbers of lymph nodes for pathologic examination could potentially result in increasing detection of regional or distant stage cancer. This could lead to more appropriate treatment, (eg, adjuvant chemotherapy), and eventually to improvements in survival rates. However, in a study that looked at 980 hospitals from 2004–2005, only 38% of hospitals were compliant in examining 12 or more lymph nodes in >75% of patients.

Given the clear established recommendations on surgical treatment, it could be useful for hospitals to determine whether or not they are meeting current guidelines, and how well they meet these parameters as compared to other hospitals. Performance assessment is especially practical in small community hospitals, which may have fewer resources than larger teaching and research hospitals, in order to ensure that care offered is competitive and current.

This type of analysis was conducted in 2011 for Falmouth Hospital (FH), a 95-bed community hospital located in Falmouth, Massachusetts, and the results showed that fewer than the recommended 75% of patients had 12 or more lymph nodes examined. This prompted the implementation of a new pathology procedure in 2011 to examine resected tissue for lymph nodes more thoroughly. In this procedure, colorectal tumor resection specimens are formalin fixed for 24 hours and then taken for microscopic examination using standard dissection technique. If standard lymph node dissection yields less than 12 nodes, dissected fat is placed in O-fix (Leica Biosystems), a new tissue fixative formulated with alcohol, formalin and acetic acid, for 24 hours. Samples are then reexamined for presence of additional lymph nodes turned opaque white by O-fix, which are submitted for microscopic examination (L. Max, MD, oral communication, July 2014). The present study is an additional analysis to assess the effectiveness of this improved procedure.

Methods

The study included data from years 2011–2012, which were compared to similar data from 2005–2006 included in a previous 2011 report. Data used to compare quality of surgical treatment and the distribution of stage at diagnosis at FH from 2011–2012 to 2005–2006 were obtained directly from the FH tumor registry and from the National Cancer Data Base (NCDB). The total number of cases analyzed at FH amounted to 159. Cases were first primary cases only and did not include cases of secondary metastatic tumors. The NCDB reports cases as either colon cancer or rectal cancer. To assess statistics on colorectal cancer, these 2 categories were combined.

__Falmouth Hospital, Falmouth, Massachusetts.
Address Correspondence to Amanda Doodlesack, BS. Email: amanda_doodlesack@brown.edu.
Improvement in number of lymph nodes resected and examined was assessed by analyzing 2011–2012 data from the FH tumor registry and comparing it to corresponding data from 2005–2006. Two-year spans were used in order to compile a large enough sample size for statistical significance, and to match the number of cases in each sample. The alcohol-formalin fixation technique intended to increase the number of lymph nodes examined was implemented at the beginning of 2011 after acquiring the results of the 2002–2006 data report. Therefore, by looking at 2011–2012 results as compared to 2005–2006, it could be ascertained whether or not this measure had an effect. The number of nodes examined per patient for 2011–2012 was compared to data from 2005–2006 using a Student’s t-test (calculated using Excel) and was tested at the 0.05 significance level.

Table 1. Quality of Surgery at Falmouth Hospital, Measured by Lymph Nodes Examined 2005–2006 vs 2011–2012

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N= sum of cases</td>
<td>61</td>
<td>24</td>
</tr>
<tr>
<td>Mean nodes examined</td>
<td>12.8 (11.05–14.55)</td>
<td>19.3 (15.59–22.91)</td>
</tr>
<tr>
<td>n= number surgeries with ≥ 12 nodes examined</td>
<td>35</td>
<td>19</td>
</tr>
<tr>
<td>% Surgeries with ≥12 nodes examined</td>
<td>57.4%</td>
<td>79.2%</td>
</tr>
</tbody>
</table>

*p* values were calculated using a Student’s t-test (2-sample, unpaired) to compare 2011–2012 results to 2005–2006. The *P* value was calculated to be *P* = .0009, indicating that the improvement was statistically significant beyond the .05 level.

Tumor stage at diagnosis was categorized as Stage 0, Stage I, Stage II, Stage III, and Stage IV in accordance with the AJCC sixth edition Cancer Staging Manual. Additional cases listed by the NCDB as either N/A or Unknown (2 cases) were excluded from the study. Age groups were stratified in 10-year increments, ranging from 20–29 through 90+. No cases were reported at FH in patients under 40, so results are only shown for age groups 40–49, 50–59, 60–69, 70–79, 80–89, and 90+.

Data on tumor stage at diagnosis for FH from 2011–2012 were compared to data from 2005–2006 using a chi-square test of independence (calculated using Excel). Statistical significance was tested at the 0.05 level.

### Results

#### Comparison of Quality of Surgery

The National Comprehensive Cancer Network (NCCN) guidelines recommend that at least 12 nodes be examined for each colorectal cancer surgery. In years 2011–2012, the average number of lymph nodes resected during surgery and examined by pathology at FH was 19.25 nodes per patient. This is an average of 6.45 more nodes per patient, a statistically significant improvement as compared to the 2005–2006 average of 12.8 nodes (*P* = .0009). Of all the patients undergoing colorectal surgery at FH during 2011–2012, 83% had 12 or more nodes examined compared to the 57% of surgeries that met the 12-node standard during the years 2005–2006. As stated previously, a hospital is considered to be in compliance of the NCCN guideline if more than 75% of patients have 12 or more lymph nodes examined. The improvement demonstrated since 2005–2006 has put Falmouth Hospital into NCCN guideline compliance.

#### Comparison of Stage at Diagnosis

In comparing stage of diagnosis for cases reported from 2011–2012 at Falmouth Hospital to those reported from 2005–2006 (Table 2), based on a chi-squared test of independence, the *P* value was found to be .901, greater than the critical value *P* = .05. This indicates that there was no statistically significant change in the distribution of stage at diagnosis at Falmouth Hospital from 2005–2006 vs 2011–2012.

### Discussion

After conducting a quality of care assessment for years 2002–2006, FH instituted a quality improvement process to improve quality of surgery by increasing the number of regional lymph nodes examined. This improvement process was apparently successful, as data from 2011–2012 showed a 46% improvement in the number of patients who had 12 or more lymph nodes examined compared to data collected from 2005–2006 (Table 1). As a result of this improvement, Falmouth Hospital is now considered to be one of the 38% of hospitals in compliance with the NCCN guideline, with 83% of surgeries meeting the 12-node standard.

This increase in number of lymph nodes resected and examined per patient could have resulted in an increase in the detection of metastases in lymph nodes. However, there was no increase in the proportion of cases diagnosed at later stages (Table 2). Future studies should continue to examine stage distributions.

There is significant controversy in the literature surrounding the clinical implications of improving the harvesting of lymph nodes. Several reports have supported the 12-node standard demonstrating that an increased number of nodes examined leads to more accurate staging, more appropriate treatment, and therefore potentially
longer survival rates. Other reports indicate that evidence to support the 12-node standard is limited, and that further research on the clinical implications of improving lymph node harvesting is needed. Although the statistical significance of the results presented here is limited by small sample size and due to the fact that they represent only one hospital, the lack of change in the distribution of stage at diagnosis at FH for 2011–2012 (following implementation of the lymph node harvesting improvement measure) compared to 2005–2006 does not show any clinical impact of increasing the number of lymph nodes examined.

Acknowledgements
The design of this study was guided by Peter S. Hopewood, MD, FACS and Deborah Crockett-Rice, CTR.

References
Breast Cancer Incidence Among Nebraska Women: Early- and Late-Stage Trends, 1995–2009

Kristin Yeoman, MD, MPH \(^{a,b} \); A. Blythe Ryerson, PhD, MPH \(^{a} \); Natalie Koziol, MA \(^{a} \); Bryan Buss, DVM, MPH \(^{a,b} \); Thomas Safranek, MD \(^{a} \)

**Abstract:** Differences in breast cancer incidence and stage at diagnosis have been observed among women of different socioeconomic status. We evaluated trends in invasive breast cancer incidence by stage and poverty among Nebraska women using cancer registry data. We demonstrated increasing incidence of early- and late-stage cancer among women living in the lowest poverty level tracts. Determining the underlying reasons for increases in early- and late-stage breast cancers among this population warrants further investigation.

**Key words:** breast neoplasms, censuses, incidence, poverty, social class

**Introduction**

Investigations of cancer incidence and diagnosis stage have demonstrated differences by socioeconomic class. \(^1,2\)

Among breast cancer patients in the United States, high socioeconomic status (SES) is associated with higher incidence, \(^2\) but women of low SES are more likely to be diagnosed at a later stage. \(^3,5\) Similar associations between high SES and increased breast cancer incidence have been demonstrated in other industrialized countries. \(^6,7\) In the United Kingdom, the relative risk of breast cancer was 0.84 among the most deprived women compared with the least deprived women. \(^6\) Incidence of breast cancer was more common among higher income communities in Canada. \(^7\)

The Nebraska Comprehensive Cancer Control Program and Every Woman Matters programs funded by the Centers for Disease Control and Prevention work together to promote breast cancer education, screening, follow-up, treatment, support services, and research. However, the association between breast cancer incidence, poverty level, and stage at diagnosis has not been assessed previously in Nebraska and would provide meaningful data to stakeholders engaged in developing and improving breast cancer programs and interventions within the state.

Nebraska, a largely rural Midwestern state lying in the Great Plains, has approximately 1.9 million residents, of whom approximately 81%, 5%, and 10% are non-Hispanic white, black, and Hispanic, respectively. \(^8\) During 1999–2009, a nonsignificant trend of decreasing female breast cancer incidence was noted among Nebraska women, with incidence of 131.8 and 121.0 per 100,000 persons in 1999 and 2009, respectively. \(^9\) During that period, incidence among US women declined from 135.6 per 100,000 persons in 1999 to 124.0 per 100,000 persons in 2009. \(^9\) Mortality rates from breast cancer also declined during this period among US women, with rates of 26.6 and 22.2 per 100,000 persons in 1999 and 2009, respectively; no significant declines were noted among Nebraska women, but mortality rates in 2009 were lower than the national average at 19.4 per 100,000 persons. \(^9\) To determine potential gaps in cancer programs within Nebraska, we analyzed cancer incidence data from the Nebraska Cancer Registry (NCR) from 1995–2009 by stage at diagnosis and census-tract poverty levels.

**Methods**

We included all cases of invasive breast cancer among Nebraska residents reported to the NCR during 1995–2009 among non-Hispanic white women aged ≥25 years. Records of nonwhite and Hispanic residents and women aged <50 years were excluded because of limited numbers. Partly because of missing race/ethnicity data, only 4.4% of NCR records were nonwhite women, limiting the ability to analyze data after stratification. In general, most NCR cases missing race and Hispanic ethnicity designations are in counties with high proportions of minorities. \(^10\) NCR received a Gold certification for data quality from the North American Association of Central Cancer Registries (NAACCR) for the first time in 1995; data quality prior to 1995 is unknown, and these data were therefore not included. We used International Classification of Diseases for Oncology primary site codes for breast cancer (C50.0–C50.9). We followed the NAACCR recommended geocoding guide, \(^11\) adapted to NCR, \(^12\) to code all cases in this study to 2000 US Census Bureau census tracts according to address at diagnosis; 93% of cases were coded to street address, 6% were coded to post office box addresses, and all remaining cases were coded to city center or ZIP code. We used Census 2000 rather than 2010 data to consistently align census-tract geography during all years of analysis and because of the higher margin of error for census-tract poverty estimates in 2010. We chose census-tract poverty as the key SES variable based on findings of the Public Health Disparities Geocoding Project, which concluded that the variable most useful for evaluating associations between SES and health...
outcomes was census tract poverty level because of the ability to detect expected health gradients across multiple outcomes, and the ease of geocoding, linkage to SES data, and interpretability. Following the Council of State and Territorial Epidemiologists (CSTE) guideline document recommending quartile poverty levels to allow comparison across jurisdictions, we categorized census tracts into 4 poverty levels on the basis of the 2000 US Census Bureau–estimated proportion of residents living below 100% of the federal poverty level as follows: ≥20%, 10%–19.9%, 5%–9.9%, or <5%.14

Cancer registry variables used to determine stage depended on diagnosis year. For cases diagnosed during 1995–2000, 2001–2003, and 2004 or later, we used Surveillance Epidemiology and End Results (SEER) summary stage 1997, SEER summary stage 2000, and derived SEER summary stage 2000 variables, respectively.15,16 Localized and regional/distant SEER Summary Stages were considered early- and late-stage cancers, respectively. We calculated stage-specific, direct age-adjusted (to the 2000 US standard population17) incidence rates/100,000 women by using census tract population information from the US Census for denominators and expressing results as 3-year central moving averages to decrease random variation from limited numbers. We analyzed temporal differences in incidence by calculating the average annual percentage change (AAPC) for 1995–2009 by using Joinpoint regression, defining statistically significant AAPC with a 2-sided P value <.05.18

**Results**

During 1995–2009, a total of 14,353 breast cancer cases among white women aged ≥50 years were diagnosed in Nebraska. Table 1 demonstrates the frequency and age-adjusted incidence rate of breast cancer by stage and census-tract poverty level. Women residing in <5% poverty tracts had 10.7% higher incidence overall, 11.7% higher incidence of early stage, and 8.4% higher incidence of late stage breast cancer when compared with women residing in ≥20% poverty tracts. Figure 1 illustrates age-adjusted 3-year central moving average annual incidence for early and late stages. Among women residing in <5% poverty tracts, both early- and late-stage incidence increased (AAPC, 3.1%; 95% confidence interval [CI], 1.3%–4.9%; and AAPC, 3.9%; 95% CI, 1.5%–6.4%, respectively); no significant changes in early- or late-stage incidence were demonstrated within other poverty categories.

### Discussion

Our results demonstrate an overall increase in incidence of both early- and late-stage breast cancer during 1995–2009 among white women aged ≥50 years living in the lowest poverty census tracts. Increasing incidence among affluent women has previously been demonstrated, but our observation in Nebraska of increasing late-stage incidence is unexpected.2–5 Age-adjusted incidence of breast cancer among white women using SEER data demonstrates decreasing trends; incidence in 1995 and 2009 was 137.7 and 131.2 per 100,000 persons.19 Breast cancer incidence among Nebraska women overall has decreased slightly, from 131.8/100,000 persons in 1999 to 121.0/100,000 in 2009.9 In the most recent Annual Report to the Nation on the status of cancer, breast cancer incidence rates remained stable for white women during 2001–2010, but increased among black women.20

Reasons for our observation of increasing early- and late-stage incidence among white women are unknown. The increase in early stage incidence among women in affluent areas can partly be explained by the diagnosis of cancers through screening that otherwise might never have been detected. Although富贵 areas can partly be explained by the diagnosis of cancers through screening that otherwise might never have been detected, factors such as patient access to care and social support can also play a role. Further research is needed to understand the mechanisms behind this trend and to inform public health interventions aimed at improving breast cancer outcomes among all population groups.

### Figure 1. Central Moving Average Annual Incidence of Early- and Late-stage Invasive Breast Cancer/100,000 White Women Aged ≥50 Years, Nebraska, 1995–2009

![Figure 1. Central Moving Average Annual Incidence of Early- and Late-stage Invasive Breast Cancer/100,000 White Women Aged ≥50 Years, Nebraska, 1995–2009](image)

### Table 1. Invasive Breast Cancer Cases and Average Annual Incidence† Rates among White Women Aged ≥50 Years by Stage and Poverty Level, Nebraska, 1995–2009

<table>
<thead>
<tr>
<th>Poverty Level</th>
<th>Total N</th>
<th>%</th>
<th>Incidence N</th>
<th>Early Stage Incidence N</th>
<th>%</th>
<th>Late Stage Incidence N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5</td>
<td>3,582</td>
<td>25.0</td>
<td>421.2</td>
<td>2,473</td>
<td>24.9</td>
<td>291.7</td>
<td>25.1</td>
</tr>
<tr>
<td>5–9.9</td>
<td>5,726</td>
<td>39.9</td>
<td>377.5</td>
<td>4,039</td>
<td>40.7</td>
<td>265.0</td>
<td>38.2</td>
</tr>
<tr>
<td>10–19.9</td>
<td>4,509</td>
<td>31.4</td>
<td>356.8</td>
<td>3,054</td>
<td>30.7</td>
<td>239.9</td>
<td>32.9</td>
</tr>
<tr>
<td>≥20</td>
<td>536</td>
<td>3.7</td>
<td>376.1</td>
<td>368</td>
<td>3.7</td>
<td>257.6</td>
<td>3.8</td>
</tr>
<tr>
<td>Total</td>
<td>14,353</td>
<td>100</td>
<td>9,934</td>
<td>4,419</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†Per 100,000 white women aged ≥50 years, age-adjusted to the US 2000 standard population.

A subset of cases.
presented clinically. However, it is unlikely that changes in mammography use by income level among Nebraska women aged ≥250 years during 1995–2009 would explain increasing late-stage incidence. Mammography use among Nebraska women overall slightly increased during this period, from 65.8% in 1995 to 74.9% in 2008. However, mammography rates in Nebraska were slightly lower than those for the United States overall (70.3% in 1995, 79.5% in 2008). Other potential reasons for our observation of increased early- and late-stage incidence in this population include differences in reproductive factors and hormone therapy use. Greater age at first full-term pregnancy, nulliparity, fewer full-term births, early menstruation, and late menopause are risk factors for breast cancer, but women living in high SES communities still have greater odds of breast cancer after controlling for these risk factors.

Finally, the risk of breast cancer appears to decrease as initiation and length of breastfeeding increases. Women with high family incomes and high education levels are more likely to breastfeed than those with lower incomes or education levels. We would therefore expect breastfeeding rates to be higher in high SES communities, which does not explain our findings.

A shift to more late-stage incidence within affluent areas might be explained by an influx of low-income, uninsured women without access to primary care or mammography services. Assessing the proportion of uninsured women statewide by census-tract poverty level might delineate this problem further. Because most NCR breast cancer cases diagnosed during 1995–2009 were missing insurance information, we were unable to assess insurance status. Census 2000 data were used each year, but the increasing incidence specific to low-poverty census tracts was not explained by a differential increase in population, in which a higher proportion of the population moved into low-poverty census tracts. We found that census tracts within the 2 highest poverty levels had greater proportional population increases than those within the lowest poverty level.

Although these results will be useful to breast cancer control planners and researchers within Nebraska, our study has at least 2 limitations. First, this is an ecologic analysis because census tract was used instead of individual poverty level; household income is not collected by NCR. Furthermore, census-tract poverty level was the only SES factor assessed. Education level is another aspect of SES but was not evaluated in this study because it is not systematically collected by NCR; increasing education levels are associated with increased risk of postmenopausal breast cancer, although women without education beyond high school have a higher risk of breast cancer mortality compared with those who have >12 years of education.

Second, analyses of differences in incidence and stage at diagnosis by race/ethnic group, rural–urban residence, and poverty were limited by small numbers and a relatively homogeneous population. Although the ≥20% census tract poverty level had smaller cell sizes than those of the other categories, we followed the CSTE guidelines for quartile poverty levels to allow for comparison of results with other states that choose to perform analyses using these guidelines. CSTE recommends that the cutoff of >20% be used in part because it aligns with the definition of a federally-defined poverty area. However, the >20% category can be further subdivided to allow for further detail when analyzing areas with high poverty levels. Small numbers of black and Hispanic case-patients in the NCR limited our analysis to whites. Incidence of breast cancer among black and Hispanic women in Nebraska varied substantially from year to year during 1999–2009, with incidence among black women ranging from 84.6 to 164.9 per 100,000 persons, and incidence among Hispanic women ranging from 76.4 to 163.6 per 100,000 persons. This variation makes comparisons of incidence between ethnic minorities in Nebraska with those of the United States difficult. Studies comparing incidence of breast cancer by race/ethnicity have demonstrated that white women have the highest breast cancer risk than all other minority groups, with Hispanic women having the lowest risk. However, increased breast cancer incidence with increasing SES has been demonstrated in multiple race/ethnic groups, with a strong trend among Hispanics. Furthermore, associations between census tract poverty level and breast cancer incidence have been demonstrated independent of race. Small numbers also limited our ability to evaluate differences in breast cancer incidence by rural–urban status because few rural census tracts were categorized into either the highest or lowest census tract poverty levels.

Despite these limitations, our results warrant further investigation. Similar analyses in other states with larger, more diverse populations will be useful to evaluate whether this trend is occurring elsewhere. In Nebraska, a case-control study is needed to explain our findings and evaluate changes in risk factors for early- and late-stage diagnosis among women of different census-tract poverty levels. A study that includes a more racially and ethnically diverse population in which individual-level data on risk factors for breast cancer, income, education level, and mammogram use are obtained for women with and without breast cancer and women with early- and late-stage breast cancer is important to further elucidate the underlying reasons for this increase in late-stage breast cancer incidence among Nebraska women living in high income communities. This information can then be used to engage appropriate stakeholders, leverage funding to obtain needed resources, and develop state-level plans to counter this trend.

References


In 1982, Steven Callahan, American author, architect and sailor, was crossing the Atlantic Ocean alone when his sailboat struck something in the water and sank. He was lost at sea, outside of the shipping lanes, and floating in a leaky life raft. He had very few supplies and he knew his chances of surviving were small. Seventy-six days later, the longest anyone has survived a shipwreck on a life raft alone, he was rescued by 3 fishermen. He was alive, much thinner than when he started his journey, but very much alive.

He survived this terrifying journey by his ingenuity. He learned how to catch fish and how to fix his broken solar still so that he could evaporate salt water for drinking.

But here’s a part of his story I find fascinating. He learned how to keep himself going when hope had vanished and there seemed no point in continuing to struggle just to stay alive. He could have easily given up. His life raft was punctured and he spent over a week struggling to fix it. Even when his body began to fail and he was weak and dehydrated, he did not give up. The leaky raft was wearing him out. He was hungry and exhausted and could easily have laid down and let his body slip into the abyss.

When people find themselves in these types of situations, they can do something with their minds that gives them the courage to keep going, or they can go crazy or just give up. Survivors, however, have a different way of managing their thoughts. They find the guts to carry on in spite of overwhelming odds. We see this every day in our cancer centers and hospitals around the world. Cancer survivors are among some of the toughest people, mentally and physically, you will find anywhere.

Callahan wrote about his experience and said, “I tell myself I can handle it. Compared to what others have been through, I’m fortunate. I tell myself these things over and over, building up fortitude…”

His words bear repeating until our own minds understand and compel us to never give up. When you feel like your work, or life, is using you for a punching bag or your problems are too overwhelming, say his words aloud. Keep saying them until you come back to your senses.

Here’s the truth: our circumstances are only bad when we compare them to something better. But, without much effort, you can always find someone who has gone through something much worse than you have. Next time you walk through your hospital or treatment center, look around. If you are observant, you will realize that you are lucky to be where you are, who you are, and to have the means to make a living. No matter how bad it may seem, you are blessed and fortunate. This is not bootstrap thinking. It is full-on, eyes-wide-open thinking that will turn your life around and fill you with a sense of gratitude, love, and compassion that you must have to succeed at anything in life.

So, the next time you feel tired, beaten down, and exhausted, tell yourself that you can handle it. Keep repeating this until your thinking is repaired. Compared to what others have been through, you are blessed! Step away from whatever it is that is dragging you down. Use Callahan’s mantra or find your own and tell it to yourself over and over. It will help you get through the rough spots and leave with you a stronger purpose and a sense of overwhelming gratitude and purpose.

“Greatness is not in where we stand, but in what direction we are moving. We must sail sometimes with the wind and sometimes against it—

but sail we must, and not drift, nor lie at anchor.”

~ Oliver Wendell Holmes

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.

Reference
A new version of the NAACCR edits metafile, NAACCR v14A (version 14A), was posted on July 15, 2014 and can be downloaded from the NAACCR website. To download the new metafile, go to http://www.naaccr.org/StandardsandRegistryOperations/VolumeIV.aspx. Under Version 14, you will find the NAACCR v14A metafile, a spreadsheet listing the changes, an Edit Detail Report (a PDF file with descriptions of all edits), and a spreadsheet of edits/edit sets.

**Highlights of edit changes:**

**Edit Name:** CS Lymph Nodes Eval, Nodes Ex (CS)
(previous name: CS Lymph Nodes Eval, RX Summ—Scope, Nodes Ex [CS])

**Change:** When “CS Lymph Nodes Eval [2840]” indicates that regional lymph nodes were removed for examination (codes 3, 5, or 6), this edit does not allow “Regional Nodes Examined [830]” to be coded ‘00’ (none). The edit no longer checks “Scope of Regional Lymph Node Surgery [1292]”.

**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)
CS Mets at DX-Lung, CS Mets at DX (CS)

**Change:** When “CS Mets at DX [2850]” is coded ‘98’ (not applicable), the edits allowed “CS Mets at DX-Bone, Brain, Liver, and Lung [2851, 2852, 2853, and 2854]” to be coded ‘8’ (not applicable) only when the primary site code was not ‘C809’ (unknown primary). The edits now allow ‘8’ for these variables when the primary site is coded to the ’IllDefinedOther’ schema, which includes ‘C809’.

**Edit Name:** CS SSF 9, Head and Neck Schemas (CS)

**Change:** Code ‘00’ (none) was added to the list of codes allowed for “Regional Nodes Positive [820]” when “CS Site-Specific Factor 9 [2863]” is coded ‘998’ (no histopathologic examination of regional lymph nodes).

**Edit Names:**
Lymph-vascular Invasion, Histology ICDO3 (COC)
Lymph-vascular Invasion, Histology ICDO3 (CS)

**Change:** When “Behavior [523]” is coded ‘0’ (benign), ‘1’ (borderline), or ‘2’ (in situ), this edit does not allow “Lymph-vascular Invasion [1182]” to be coded ‘1’ (lymph-vascular invasion present/identified). The CoC version expects all solid tumors to be coded, while the CS version limits the requirement to the penis and testis schemas that use this information to assign stage group. Cases that pass the CoC edit will pass the CS edit; however, passing the CS edit does not guarantee the case will pass the CoC edit.

**Note:** This issue was submitted to the CAnswer Forum with the following response from the AJCC Expert Panel Member/CAP Cancer Committee Member. Occasionally, a pathologist may report a tumor of in-situ behavior with lymph-vascular invasion (LVI). If the tumor truly has LVI, then it is not truly an in-situ tumor. This case should be coded as an invasive carcinoma, present in lymphatic spaces (no stromal invasion), associated with in-situ. This is an unusual situation that is rarely seen in breast specimens, and is extremely unlikely to occur in other organs.

**Edit Name:** Primary Site, Morphology-Imposs ICDO3 (SEER IF38)

**Change:** The following site/histology combinations are no longer considered impossible:
C383 (Mediastinum, NOS) 8240 (Carcinoid)
C710-C719 (Brain) 8070 (Squamous Cell)

Rationale: Intracranial squamous cell carcinoma is very rare with most cases arising from a preexisting benign epidermoid cyst. These cases will still fail the edit “Primary Site, Morphology-Type,Beh ICDO3(SEER IF25)” requiring that the case be reviewed and the “Over-ride Site/Type [2030]” flag set, if the coded information is found to be correct, in order to pass.

**Edit Name:** Census Tract 2000, State, County, 2000-2009 (NPCR)

**New edit:** Just like the existing edit ‘Census Tract 2000, State, County at DX (NPCR),’ this edit verifies that the “Census Tract 2000 [130]” code is valid for the coded “Addr at DX—State [80]” and “County at DX [90],” except that it applies only to cases diagnosed 2000–2009.

**Edit Name:** Census Tract 2010, State, County, 2010-2019 (NPCR)

**New edit:** Just like the existing edit ‘Census Tract 2010, State, County at DX (NPCR),’ this edit verifies that the “Census Tract 2010 [135]” code is valid for the coded “Addr at DX—State [80]” and “County at DX [90],” except that it applies only to cases diagnosed 2010–2019.

NAACCR v14A includes several additional small modifications, mostly to the edit descriptions. Please see the spreadsheet of changes for additional information.
Introduction

Quality of data is something we all take seriously and registrars work diligently to abstract correctly every time. Review of one’s work is necessary to ensure accuracy of the abstract and accuracy of the registrars’ perception of the rules and guidelines. In addition to accuracy, it is also important to provide timely feedback and a way to track trends.

A little bit of history:

Three years ago our office went through tremendous change. Retirements, registrars leaving, and an increase in the number of reportable cases (hospital mergers) seemed to unfold simultaneously. Today, we are an office of 8 registrars with very diverse work schedules.

- 2 registrars—full time and in the office full time
- 3 registrars—less than full time
- 3 registrars—full time but work remotely part time

My challenge: Each day the number of registrars “in house” is varied. Timely feedback of quality reviews became difficult. I also needed a better way to track quality review percentages and monthly productivity.

Method and Discussion

Our registry already had a paper-based quality review tool that worked well. As a department, we reviewed and discussed this tool for changes and updates. All of the registrars had input regarding which data items to list and the number of points attached for each item. Decisions, such as allowing for partial points, were made. Example: Date of diagnosis is 6 points, but if month and year are correct (and the day is incorrect), then the registrar would have 2 points deducted, not 6 points. Using this updated paper-based tool as my template, I created an Excel spreadsheet, password protected, for each registrar.

At the top of the quality review spreadsheet (Table 1), the registrar will find their cumulative percentage score as well as current month productivity. Also listed are the primary site, MRN numbers, and the score for each. Two reviews of 1 primary site are completed each month per registrar.

The body of the quality review spreadsheet (Table 2) lists the data items, point value for each data item and an area for comments.

- The **Point Value** column shows the point value for each data item.
- The **Points Received** column auto-calculates the total number of points in the last (bottom) cell. If a data item is found to be incorrect I will deduct the applicable point.
Table 2. Body of the Quality Review Spreadsheet

<table>
<thead>
<tr>
<th>Point Value</th>
<th>Points Received</th>
<th>Comments — Registry (optional)</th>
<th>Point Value</th>
<th>Points Received</th>
<th>Comments — Registry (optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Information Tab</td>
<td>Case #1</td>
<td>Patient Information Tab</td>
<td>Case #2</td>
<td>Patient Information Tab</td>
<td>Case #2</td>
</tr>
<tr>
<td>1</td>
<td>Patient full name</td>
<td>1</td>
<td>1</td>
<td>Patient full name</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Social Security number</td>
<td>2</td>
<td>2</td>
<td>Social Security number</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>Sex</td>
<td>1</td>
<td>1</td>
<td>Sex</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Race</td>
<td>1</td>
<td>1</td>
<td>Race</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Date of birth</td>
<td>1</td>
<td>1</td>
<td>Date of birth</td>
<td>1</td>
</tr>
<tr>
<td>Facility Specific Tab</td>
<td>Facility Specific Tab</td>
<td>Facility Specific Tab</td>
<td>Facility Specific Tab</td>
<td>Facility Specific Tab</td>
<td>Facility Specific Tab</td>
</tr>
<tr>
<td>2</td>
<td>Medical record number (11 digits)</td>
<td>2</td>
<td>2</td>
<td>Medical record number (11 digits)</td>
<td>2</td>
</tr>
<tr>
<td>Summary Tab</td>
<td>Summary Tab</td>
<td>Summary Tab</td>
<td>Summary Tab</td>
<td>Summary Tab</td>
<td>Summary Tab</td>
</tr>
<tr>
<td>5</td>
<td>Primary site (CXX)</td>
<td>5</td>
<td>5</td>
<td>Primary site (CXX)</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>Histology/Behavior</td>
<td>5</td>
<td>5</td>
<td>Histology/Behavior</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>Sequence number</td>
<td>1</td>
<td>1</td>
<td>Sequence number</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Laterality</td>
<td>1</td>
<td>1</td>
<td>Laterality</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Date of diagnosis (m/d/y)</td>
<td>6</td>
<td>6</td>
<td>Date of diagnosis (m/d/y)</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Grade</td>
<td>2</td>
<td>2</td>
<td>Grade</td>
<td>2</td>
</tr>
<tr>
<td>Facility Tab</td>
<td>Facility Tab</td>
<td>Facility Tab</td>
<td>Facility Tab</td>
<td>Facility Tab</td>
<td>Facility Tab</td>
</tr>
<tr>
<td>2</td>
<td>Class of case</td>
<td>2</td>
<td>2</td>
<td>Class of case</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Date of first contact (m/d/y)</td>
<td>6</td>
<td>6</td>
<td>Date of first contact (m/d/y)</td>
<td>4</td>
</tr>
<tr>
<td>Collaborative Tab</td>
<td>Collaborative Tab</td>
<td>Collaborative Tab</td>
<td>Collaborative Tab</td>
<td>Collaborative Tab</td>
<td>Collaborative Tab</td>
</tr>
<tr>
<td>1</td>
<td>CS Tumor size</td>
<td>1</td>
<td>1</td>
<td>CS Tumor size</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS Extension</td>
<td>1</td>
<td>1</td>
<td>CS Extension</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS TS/Ext eval</td>
<td>1</td>
<td>1</td>
<td>CS TS/Ext eval</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS Lymph nodes eval</td>
<td>1</td>
<td>1</td>
<td>CS Lymph nodes eval</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS Reg LN positive</td>
<td>1</td>
<td>1</td>
<td>CS Reg LN positive</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS Reg LN examined</td>
<td>1</td>
<td>1</td>
<td>CS Reg LN examined</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS Mets at diagnosis</td>
<td>1</td>
<td>1</td>
<td>CS Mets at diagnosis</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS Mets eval</td>
<td>1</td>
<td>1</td>
<td>CS Mets eval</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS SSF 1</td>
<td>1</td>
<td>1</td>
<td>CS SSF 1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS SSF 2</td>
<td>1</td>
<td>1</td>
<td>CS SSF 2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS SSF 3</td>
<td>1</td>
<td>1</td>
<td>CS SSF 3</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS SSF 4</td>
<td>1</td>
<td>1</td>
<td>CS SSF 4</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS SSF 5</td>
<td>1</td>
<td>1</td>
<td>CS SSF 5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>CS SSF 6</td>
<td>1</td>
<td>1</td>
<td>CS SSF 6</td>
<td>1</td>
</tr>
<tr>
<td>AJCC Staging Tab</td>
<td>AJCC Staging Tab</td>
<td>AJCC Staging Tab</td>
<td>AJCC Staging Tab</td>
<td>AJCC Staging Tab</td>
<td>AJCC Staging Tab</td>
</tr>
<tr>
<td>6</td>
<td>Clinical T N M</td>
<td>6</td>
<td>6</td>
<td>Clinical T N M</td>
<td>6</td>
</tr>
<tr>
<td>Treatment — Summary Tab</td>
<td>Treatment — Summary Tab</td>
<td>Treatment — Summary Tab</td>
<td>Treatment — Summary Tab</td>
<td>Treatment — Summary Tab</td>
<td>Treatment — Summary Tab</td>
</tr>
</tbody>
</table>
### Table 2, cont. Body of the Quality Review Spreadsheet

<table>
<thead>
<tr>
<th>Point Value</th>
<th>Case #1</th>
<th>Comments — Quality review</th>
<th>Comments — Registrar (optional)</th>
<th>Case #2</th>
<th>Comments — Quality review</th>
<th>Comments — Registrar (optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Dx/Staging/ Palliative Procedure code</td>
<td>1</td>
<td>Dx/Staging/ Palliative Procedure code</td>
<td>1</td>
<td>Dx/Staging/ Palliative Procedure code</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Dx/Staging/ Palliative Procedure date</td>
<td>6</td>
<td>Dx/Staging/ Palliative Procedure date</td>
<td>6</td>
<td>Dx/Staging/ Palliative Procedure date</td>
</tr>
<tr>
<td>Treatment — Surgery Tab</td>
<td></td>
<td>Surgery code (definitive surgery)</td>
<td>1</td>
<td>Surgery code (definitive surgery)</td>
<td>1</td>
<td>Surgery code (definitive surgery)</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Surgery date (m/d/y)</td>
<td>6</td>
<td>Surgery date (m/d/y)</td>
<td>6</td>
<td>Surgery date (m/d/y)</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>Scope of regional LN surgery</td>
<td>1</td>
<td>Scope of regional LN surgery</td>
<td>1</td>
<td>Scope of regional LN surgery</td>
</tr>
<tr>
<td>Treatment — Radiation Tab</td>
<td></td>
<td>Radiation start date (m/d/y)</td>
<td>6</td>
<td>Radiation start date (m/d/y)</td>
<td>6</td>
<td>Radiation start date (m/d/y)</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>Radiation modality</td>
<td>1</td>
<td>Radiation modality</td>
<td>1</td>
<td>Radiation modality</td>
</tr>
<tr>
<td>Treatment — Chemotherapy Tab</td>
<td></td>
<td>Chemotherapy (drop-down menu)</td>
<td>1</td>
<td>Chemotherapy (drop-down menu)</td>
<td>1</td>
<td>Chemotherapy (drop-down menu)</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Chemo course started (m/d/y)</td>
<td>6</td>
<td>Chemo course started (m/d/y)</td>
<td>6</td>
<td>Chemo course started (m/d/y)</td>
</tr>
<tr>
<td>Treatment — Hormone/ Immunotherapy Tab</td>
<td></td>
<td>Hormone Code</td>
<td>1</td>
<td>Hormone Code</td>
<td>1</td>
<td>Hormone Code</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Hormone date (m/d/y)</td>
<td>6</td>
<td>Hormone date (m/d/y)</td>
<td>6</td>
<td>Hormone date (m/d/y)</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Immunotherapy date (m/d/y)</td>
<td>6</td>
<td>Immunotherapy date (m/d/y)</td>
<td>6</td>
<td>Immunotherapy date (m/d/y)</td>
</tr>
<tr>
<td>Treatment — Other/ Recurrence Tab</td>
<td></td>
<td>Hematologic Tran/ Endocrine Proc</td>
<td>1</td>
<td>Hematologic Tran/ Endocrine Proc</td>
<td>1</td>
<td>Hematologic Tran/ Endocrine Proc</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>General comments — 0 point value items that may require registrar review.</td>
<td></td>
<td></td>
<td>General comments — 0 point value items that may require registrar review.</td>
<td></td>
</tr>
<tr>
<td>General comments — 0 point value items that may require registrar review.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Your text boxes were very thorough and complete! :-)</td>
</tr>
</tbody>
</table>
from the Points Received box. (See Date of First Contact in the example below. Four points were received of a 6-point value)

- **Comments – Quality review** is where I will site my source for any points deducted.
- **Comments – Registrar** is where a reply can be stated by the registrar.
- The 2 comment boxes are where the dialogue begins and where the true value of a quality review rests.
- At the bottom of the spreadsheet is a third area for **General Comments**. This area is where I will list nonpoint value items that I observed while reviewing the case. For example, if text did not include tumor size, I will add a note for the registrar to please add tumor size in the Pathologic text box.
- General Comments are also where I will add positive feedback such as “Excellent text!” or “This was a tough case…nice work!!”
- At the bottom of the quality review spreadsheet (Figure 1), you will see the 12 monthly worksheets for the abstract year.
- Each month lists a different site for review with the cumulative score percentages at the top.
- Once a quality review is complete, I will send an email to the registrar that “July’s QR is ready for review.”
- Once reviewed by the registrar, if there are any questions or concerns, we will set a day and time to meet and discuss.
- For ease of tracking, I created an Abstracting Quality Review Snapshot spreadsheet.
- This tool is accessed by me and the departmental supervisor only, and is password protected.
- Once all 8 quality reviews are complete, I will send a note to the supervisor that the results are ready for review.
- Columns G and J have built in formulas to auto calculate each month’s totals.

Results

The Quality Review spreadsheet and Abstract Quality Review Snapshot has been implemented for 3 years and are working very well.

Each registrar’s quality review spreadsheet is password protected using their 3 initials plus a 4-digit number of their choice. I keep all passwords on file. The registrars, whether in-house or working remotely, have immediate access to their reviews and ongoing score.

The Abstract Quality Review snapshot is an easy way to monitor trends and to take action if someone’s score or productivity drops. It is also useful for end-of-year assessments.

Each year it is easy to create new spreadsheets from a blank template.

In addition to quality reviews, we also do monthly Peer Reviews using the same data items. Peer reviews are one registrar reviewing an abstract of another registrar. Peer reviews are educational reviews only (no points attached). One peer review per registrar is completed each month. Peer reviews have proven to be of tremendous value as a platform for open discussion of accuracy as well as review of text.
SURVEILLANCE OF THE FREQUENCY AND RESULTS OF TESTING OF INCIDENT OROPHARYNGEAL CANCERS FOR HUMAN PAPILLOMAVIRUS: THE POTENTIAL ROLE OF POPULATION-BASED CANCER REGISTRIES

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:
• Discuss the relationship between human papillomavirus (HPV) infection and certain cancers of the oropharynx
• Explain the importance of accurately coding HPV status in the cancer registry abstract for oropharyngeal cancers
• Compare and contrast various methods for HPV testing

1. Oral cavity–pharynx cancers in the US population have been:
   a) declining
   b) increasing
   c) stable
   d) eradicated

2. Certain oropharyngeal cancers are associated with:
   a) HIV (human immunodeficiency virus)
   b) HPV (human papillomavirus)
   c) HepB (hepatitis B)
   d) HepC (hepatitis C)

3. The group of HPV-associated oral cavity-pharynx cancers includes:
   a) adenocarcinomas
   b) cholangiocarcinomas
   c) squamous cell carcinomas
   d) small cell carcinomas

4. Incidence data on HPV-related cancers are of special interest because:
   a) there is a potential for primary prevention with dietary changes
   b) there is a potential for primary prevention with smoking cessation
   c) incidence rates are projected to decrease
   d) incidence rates are projected to increase

5. The HPV status of oral cavity–pharynx tumors is a site-specific factor required by:
   a) Commission on Cancer–accredited programs
   b) National Program of Cancer Registries
   c) National Cancer Registrars Association
   d) North American Association of Central Cancer Registries

6. The most common HPV-related sites in the oral cavity-pharynx are:
   a) lateral wall and tonsil
   b) lateral wall and lip
   c) base of tongue and tonsil
   d) base of tongue and lip

7. According to Table 2, Distribution of Data on AJCC Collaborative Stage Site-Specific Factor 10 (SSF 10) on Human Papillomavirus (HPV) Testing of Tumor, for Incident Invasive BTT Cancers Diagnosed in 2010–2012 in the Connecticut Tumor Registry, site-specific factor (SSF) 10 was most often coded as:
   a) negative
   b) positive
   c) not done
   d) unknown

8. According to Table 3, Human Papillomavirus (HPV) Test Result for BTT Cases in the Connecticut Tumor Registry Coded as Either HPV Positive or HPV Negative Using Site-Specific Factor 10 (SSF 10):
   a) a higher percentage of males had positive HPV results
   b) a higher percentage of females had positive HPV results
   c) HPV-positive test results showed an increase from 2010 to 2011
   d) HPV-positive test results showed an increase from 2010 to 2012

9. Limitations of the coding rules for SSF 10 on HPV status of oropharyngeal cancers include:
   a) no information on the specific test used
   b) use of HPV detected by serology alone
   c) p16 IHC markers are coded as positive
   d) various tests are equivalent in analytic performance

10. This study concluded that data from the Connecticut SEER Registry show:
    a) no utility for the collection of SSF 10
    b) no utility for monitoring HPV testing rates
    c) decreased incidence rates for base of tongue and tonsil cancers
    d) increased incidence rates for base of tongue and tonsil cancers

The JRM Quiz and answers are now available through NCRA’s Center for Cancer Registry Education (CCRE). For your convenience, the JRM article and quiz can be accessed online at www.CancerRegistryEducation.org/jrm-quizzes. Download the article, complete the quiz and claim CE credit all online.
Instructions: Mark your answers clearly by filling in the correct answer, like this ■ not like this x. Passing score of 70% entitles one (1) CE clock hour per quiz.

Please use black ballpoint pen.

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

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

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: 4103.01

Quiz Identification Number: December 31, 2016

JRM Quiz Article: SURVEILLANCE OF THE FREQUENCY AND RESULTS OF TESTING OF INCIDENT OROPHARYNGEAL CANCERS FOR HUMAN PAPILLOMAVIRUS: THE POTENTIAL ROLE OF POPULATION-BASED CANCER REGISTRIES

☐ 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

Print clearly in black ballpoint pen.

First Name ________________________________ M.I. ________________________________ Last Name ________________________________

Address

Address

City ________________________________ State/Province ________________________________ ZIP Code/Postal Code ________________________________

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

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

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

Manuscript submission requirements are given in “Information for Authors” found on the inside back cover of each Journal and on the NCRA Web site at [http://www.ncra-usa.org/jrm](http://www.ncra-usa.org/jrm).
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 (CTR); 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, RHT, CTR at: izoo4@cdc.gov, or at: CDC/NCCDHP/DPCP/CSB, 4770 Buford Drive, MS K-53, Atlanta, GA 30341-3717, 770-488-6400 (office), 770-488-6759 (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 JRM, and to facilitate technical processing of submissions. Failure to follow the guidelines may delay consideration of your manuscript. Authors who are unfamiliar with preparation and submission of manuscripts for publication are encouraged to contact the Editor for clarification or additional assistance.

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

Journal of Registry Management 2014 Volume 41 Number 3
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](http://www.ncra-usa.org/membership)*  
*Phone: 703-299-6640 Ext. 319*  
*E-mail: member@ncra-usa.org.*
Finding Your Dream Job is Just a Click Away

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

Need a Clinical Placement Site?
NCRA’s Job Bank also includes sites where students can complete their clinical hours and search for mentors.

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