Visioning Timeliness, Improving Accuracy, and Enhancing Efficiency: Evaluation of Incident Data and Cancer Reporting to Central Registries

UNIVERSITY OF UTAH HEALTH SCIENCES

Antoinette Stroup, PhD; Rosemary Dibble, CTR; Kimberly Herget, BA;
C. Janna Harrell, MS; SuAnn McFadden, CTR; Sharon Bair, CTR

Utah Cancer Registry, University of Utah

ABSTRACT

In 2008, the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute considered requiring a set of ***core incidence data*** to be reported to central registries within 6 months of diagnosis. However, it is unknown whether central registries receive these data within this time. The completeness and accuracy of the initial data are unknown. This study monitored timeliness and accuracy of records transmitted to the Utah Cancer Registry (UCR) for 2009 diagnoses. Mean lag time for E-path records was six days with 63% complete. Mean lag time for hospital abstracts was 147 days with 89% complete. Edits to core data were required for 50% of sampled records, of which 30% were due to coding errors. When compared to case counts for 2009 diagnoses captured in May 2011, the UCR was approximately 85% complete at the time data was extracted for this study (March 1, 2010).

INTRODUCTION

- In 2008, SEER investigators considered a two-tiered reporting system, with Tier 1 being a core set of incidence data reported to central registries within six months of diagnosis.
- However, two essential questions must be addressed before this system can be implemented:
  - Do central registries receive these data within six months after diagnosis?
  - To what extent are the initial data complete and accurate?

Core Incidence Data (Tier 1)

Demographics
- Name
- Social Security Number
- Address
- Date of Birth
- Sex
- Race
- Spanish Origin

Tumor Information
- Diagnosis Date
- Primary Site
- Histology
- Behavior
- Laterality

AIMS

1. Monitor the timeliness of receiving electronic and paper pathology reports and hospital abstracts submitted to the central cancer registry. Assess completeness of case finding as of March 1, 2010.
2. Document completeness of core incident data from pathology reports and hospital abstracts and evaluate the timeliness of specific data elements.
3. Evaluate the quality of core incident data from initial pathology reports and hospital abstracts by examining changes and edits made by coders for a sample of cases.

ACKNOWLEDGMENTS

Research was supported by the National Cancer Institute's SEER Program, Contract #HSN261201000026C, with additional support from the Utah State Department of Health and the University of Utah. Special thanks to Huntsman Cancer Institute's Communications Department for support in designing this poster.

METHODS

AIM 1:
- Collect original pathology and hospital abstract data for 2009 diagnoses before record consolidation.
- Calculate lag times between diagnosis and initial report to the registry, and examine differences by record type.
- Ascertain completeness for 2009 diagnoses by comparing study data extracted on March 1, 2010 to incident data as of May 2011.

AIM 2:
- Calculate percentage of 2009 cases with complete core incident data by record type.
- Compare initial source data to data from subsequent sources for 500 randomly selected records within site-specific strata of colorectal, lung and bronchus, melanoma, prostate, female breast, and all other cancers combined.
- Apply Cox Proportional Hazards modeling to evaluate when specific core data elements were populated, controlling for county at diagnosis, sex, race and ethnicity, site, behavior, and reporting hospital.

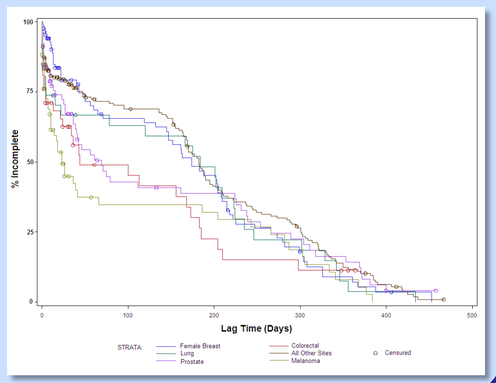
AIM 3:
- Use logistic regression to evaluate factors (for example, site, gender, reporting facility, and others) that may contribute to coding changes as a result of visual editing.

Completion Rates and Mean Lag Times by Record Type

Source Data	No. Records	No. Cases	Cases with Complete Core Data†	Mean Lag Time (Days)
E-Path 2009	14,462	10,105	63%	6
Paper Path Only 2009	1,092	1,082	53%	24
Abstract Only 2009	5,420	4,492	89%	147

† Excludes middle name, county, and Spanish origin

Core Data Completion Curve by Site



RESULTS

- **Lag Time**
  - E-Path had the shortest mean lag time (6 days), but hospital abstracts were most complete in terms of core incidence data (89%).
- **Incomplete Data**
  - 13% of E-Path cases lacked address data; 23% lacked middle name and race.
  - Less than 10% of abstracts lacked zip code, county, or diagnosis day.
- **Quality of Core Data in Initial Reports**
  - The mean time between diagnosis and data extraction was 117 days (4 months) for the 498 cases reviewed by trained CTRs (N=2 were recurrences).
  - 46% of the cases required at least one change to a core data item; 28% of those were errors, and 26% were "censored" because a hospital abstract was not available to assess data quality.
  - Prostate (HR=2.0, p<0.001), colorectal (HR=1.89, p<0.01), and lung cancer (HR=1.88, p<0.001) cases were more likely to be complete than all other cancers. Incident records for males were less likely to be complete (HR=0.75, p<0.05) than those for females, even after controlling for primary site.
  - Compared to 2009 case counts in May 2011, we found that our original files with combined e-path, paper path, and abstracts source records were 85% complete on March 1, 2010 (3 months after closing the 2009 diagnosis calendar year).

DISCUSSION

Implementing a two-tiered cancer surveillance system will require central registries to continue (1) improving completeness and quality of electronically transmitted core incidence data, (2) working closely with reporting facilities to improve timeliness of abstract submissions, and (3) considering visual editing efficiency measures to focus on sites more likely to have missing core data and/or errors in auto-coded data.