REGISTRAR PIP

Visit SEER*Educate: A comprehensive training platform for registry professionals

How to Use Audit Results to Improve Data Quality

Introduction

Hospital registry data submitted to a population-based registry is an essential part of cancer surveillance in the region. Central registry data is used to monitor incidence of, and survival from, all types of cancer. It is critical in evaluating cancer control programs focused on reducing the cancer burden. Complete, timely, and accurate data is also necessary to estimate variations in population subgroups and changes among population subgroups over time.

Ensuring the quality of registry data is considered a fundamental element in maintaining a registry. Without an adequate level of quality and some way of assuring that this level will be maintained, there will be no users of the data. Quality control efforts consider both the needs of the users and the available registry resources (i.e., time, money, and personnel) to determine the level of quality able to be maintained.

There is a possibility inaccurate data can exist in any registry database when there is limited data monitoring. The result is potentially compromising the quality and impact of research done using our data. How do we know if our data is accurate? We have to look at it! The more familiar we are with the data, the easier it is to identify something that seems odd. The oddities in our findings will be quality issues that need to be improved.

"Looking at our data" translates to "auditing it"!

Background

In early 2023, CSS staff began conducting audits targeting the coding and supporting text on abstracts submitted by hospital registrars. For the purposes of this audit, only cases abstracted at a single facility were included. An Excel spreadsheet was prepared for each facility which included patient and abstractor identifiers for the cases being audited. Each auditor was required to record if any casefinding, abstracting, or coding issue was identified for each case, the corrective action taken, and who performed the audit for the facility.

This audit was not intended to review the accuracy of every data item collected by hospital registrars. We chose to audit the following most critical data items involved in determining the accuracy of reportability, incidence, stage, and treatment:

- Race
- Dx Date
- Residence
- Site (includes SSDIs related to Schema selection)
- Histology
- Behavior
- Stage (includes EOD and Summary Stage)
- Treatment
- Text support for coding

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I

The audit results for each hospital conducted January through March 2023 have been returned to the submitting hospital registry manager to share with staff members. In this edition, we will be providing regional results so one can compare how staff members at your facility performed against all the registrars in the 13-county CSS region.

FYI: We will be conducting another round of audits later this year. Individual hospital results will again be reported to the registry manager and the regional summary results will be reported in an upcoming edition of the Registrar PIP.

Summary Results

There were two broad categories of errors evaluated which include 1) lack of documentation supporting the codes selected and 2) codes not reflecting the documentation provided.

Documentation Omissions

Documentation is needed to support all codes selected. Poor documentation lacks the sufficient clarity, specificity, and completeness for others to be able to consistently confirm the accuracy of selected codes when using the data for investigative purposes, consolidating information involving multiple sources or simply auditing procedures. The most common cause of poor documentation seems to be a lack of understanding regarding the specific information that needs to be included for coding purposes.

The lack of documentation identified during this audit for race and EOD/staging is problematic. We need to improve our abstract documentation in these two areas because of the importance of these data items.

- **Race** information is important for public health monitoring, cancer research and clinical practice because it can affect and help explain different health beliefs, risks, behaviors, and outcomes. Use a full text description of the race rather than an abbreviation. Some abbreviations are problematic such as AA. Does the abstractor mean African American or Asian American?

- **EOD/Staging** information is important at both population and patient levels. At the population level, staging data is used to develop and evaluate health promotion and treatment programs, facilitate resource allocation appropriate to address the needs of cancer patients diagnosed at different stages of disease involvement, and to stratify survival analyses to compare the results of different patient groups.

The documentation needs to be complete and accurate to facilitate effective communication between the hospital and central registry staff and various users (e.g., clinicians, researchers, etc.) of the registry data. While documentation serves many purposes, we need to remember its primary focus is to provide effective patient care. We need to recognize that ultimately patients will rely on information collected in registries because researchers and clinicians use our data to suggest and promote the best health care options for their patients.

Coding Errors

Table 1 indicates the number of cases involved in this first audit, the number of cases with errors and the total number of errors identified.

Table 1 Audit Summary Cases Reviewed 1/1/23 - 3/31/23	
Number of cases audited	499
Number of cases with errors	182
Total Number of Errors	236
NOTE: 36.5% of the audited cases had errors in the fields selected for review, with 1.3 errors in each of these cases.	

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I

Table 2 shows the number of errors by data item and the percentage of the 499 audited cases with that error type. For this article we will focus on the seven most common errors observed during this audit, the coding missteps made, and include reminders of documented coding guidelines we need to follow.

Table 2
Summary of Coding Error Types
Cases Reviewed 1/1/23 - 3/31/23

Type of Coding Error	Number of Errors	Percentage of the 499 audited cases with error type
EOD Primary Tumor	**56**	**11.2%**
Date of Diagnosis	**51**	**10.2%**
Primary Site	**28**	**5.6%**
Summary Stage	**25**	**5.0%**
Histology	**23**	**4.6%**
Surgery of Primary Site	**15**	**3.0%**
Incidence - 4 resulted in overreporting - 3 resulted in underreporting	**7**	**1.4%**
Behavior	**4**	**0.8%**
EOD Mets	**4**	**0.8%**
Chemotherapy	**3**	**0.6%**
Date of Birth	**2**	**0.4%**
Date of First Surgical Procedure	**2**	**0.4%**
Scope of Regional Lymph Node Surgery	**2**	**0.4%**
SSDI Lymph Node Size of Mets	**2**	**0.4%**
Tumor Size Clinical	**2**	**0.4%**
Grade Clinical	**1**	**0.2%**
Grade Pathological	**1**	**0.2%**
Hematopoietic Transplant/Endocrine Procedures	**1**	**0.2%**
Hormone	**1**	**0.2%**
Laterality	**1**	**0.2%**
Mets at Dx -- Bone	**1**	**0.2%**
Mets at Dx- - Other	**1**	**0.2%**
Other Therapy	**1**	**0.2%**
Schema Discriminator 1	**1**	**0.2%**
Systemic Surgery Sequence	**1**	**0.2%**
Total number of errors	**236**	

- **Incidence Reporting** - While not the most common type of error observed during this audit, the most critical one involves case reportability. There were a total of seven errors resulting in either over- or underreporting incidence. The reasons for these errors are associated with:

  ✓ Applying the current Solid Tumor Rules incorrectly (multiple errors)
  ✓ Recording a current in-area address as the patient's residence at diagnosis when they were actually a resident of another state at the time of diagnosis
  ✓ Submitting a nonreportable histology as reportable
  ✓ Missing a clinical diagnosis of a second primary noted on a scan
  ✓ Using a nonreportable ambiguous term to incorrectly report a primary

When we have a lot to do, sometimes we rush our coding decisions because we opt to trust our memory rather than checking our manuals for help. This can lead to errors because we do not apply the latest new coding rules, recognize recent changes in reportability for certain histologies, forget which is a reportable versus nonreportable ambiguous term, or overlook a second primary revealed in the abstracted text we correctly recorded but didn't

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I

report. With the release of updated manuals, be sure to check for an associated Change Log to quickly identify many changes, especially those associated with reportability.

For those of us who use abstracting software that auto-populates the **Address at Diagnosis** field when we enter the **Current Address** field, we need to remember it is necessary to doublecheck the accuracy of the Address at Diagnosis field because the auto-populating programming "help" will result in incorrect regional reporting if we do not manually update the field when appropriate.

- **EOD/Summary Stage** - When coding errors are made in one of the EOD fields, it is not uncommon to also miscode the Summary Stage field too . . . definitely, a one-two punch we need to try to avoid. The reasons for these errors include:

  ✓ Interpreting available text documentation incorrectly
  ✓ Recording as metastatic involvement when involvement by direct extension actually applies
  ✓ Overlooking the site-specific notes associated with a schema
  ✓ Using the lesser rather than the greater extension code when a ***range*** of involvement is provided in the documentation
  ✓ Ignoring direct extension when it was described using ambiguously terminology considered synonymous with involvement

Some of the most problematic primary sites when coding this field include CNS, prostate, melanoma, and lung. Here are some helpful hints to improve our coding:

We need to improve our understanding of the ***CNS*** anatomy. Being able to identify infra- and supratentorial involvement is important to accurately distinguish between localized versus regional disease involvement.

For ***melanomas***, when Clark level is stated to be IV-V, code the greater rather than the lesser stated extension (i.e., Clark level V).

Given that many ***lung*** primaries are not surgically removed, careful documentation of lobe, node, and tumor extension from scans is critical for staging. Make sure to check the body of the scan if the final impression does not provide the necessary tumor involvement details. Also, before coding 100 (Minimally invasive adenocarcinoma) for lung primaries, be sure to check the specific criteria described in Note 3 to determine whether code 30 (Any size tumor that is localized) more appropriately applies.

A ***prostate*** case needs a digital rectal exam recorded to support a code selection that confirms whether the tumor was clinically apparent or inapparent.

- **Diagnosis Date** - Most errors in this field are the result of ignoring the ***"clinically recognized medical practitioner"*** portion of the guideline that indicates "the date of diagnosis should reflect the date the reportable neoplasm was first identified, ***clinically*** or microscopically, ***by a recognized medical practitioner.***" Don't be too quick to code the date of diagnosis to the date of a positive biopsy before checking to see whether there is a prior clinical reportable description of the disease in a scan or physician note. Do not change the date of diagnosis when a clinical diagnosis is subsequently confirmed by positive histology or cytology.

- **Primary Site/Histology** - Major errors in coding one or both of these fields are those that result in a change in the SEER site recode. The SEER site recodes define the major cancer primary site/histology groups commonly used in SEER's reporting of cancer incidence data and published statistics. The site recode variables are added to the SEER databases as a convenience for researchers. Thankfully, the majority of errors were not major errors so the impact on the site recode variable was minimal.

The most common errors in these fields include:

- ✓ Incorrectly distinguishing between nodal and extranodal lymphoma primaries
- ✓ Not using the Hematopoietic and Lymphoid Neoplasm Coding Manual to help code primary site for hematopoietic and lymphoid neoplasms
- ✓ Using incorrect subsite codes for breast, lung, and CNS tumors
- ✓ Ignoring (or not recognizing) modifiers to the histology NOS descriptions that would have resulted in a more specific histology being coding (Perhaps we opted to code histology to a value that would allow AJCC staging to be performed.)

Per Serban Negoita, Data Quality, Analysis, and Interpretation Branch Chief, use the following instructions to code histology independent of whether we are able to assign AJCC staging:

- ✓ Always code the histology strictly based on the terms used by the pathologist (or managing physician if a pathology report is not available), not based on whether a case is eligible for AJCC staging.
- ✓ Assign the histology code independent of and before assessing eligibility to stage the case.
- ✓ Do not assess eligibility for AJCC staging until you have assigned the histology code.

- **Surgery of Primary Site** - The following were the most common coding issues identified in this audit:

  - ✓ Bilateral nipple sparing mastectomies were incorrectly coded to 42 (total mastectomy w/ removal of uninvolved contralateral breast) rather than 30
  - ✓ Spinal cord "resection" is coded to 22 and not 30 unless the resection is explicitly stated and documented to be a "radical, total, or gross total" resection
  - ✓ Choosing the wrong modifier associated with the removal of the primary site (e.g., a ***total*** nephrectomy has different surgery code than a ***radical*** nephrectomy)
  - ✓ Coding a more extensive surgery occurring ***after*** the completion of first course of therapy for which a less extensive surgical procedure appeared more appropriate because no supporting documentation or initial treatment plan was provided to confirm the reason for the delay associated with performing the more extensive surgery (e.g., pt had lumpectomy with negative margins followed by letrozole followed by bilateral nipple-sparing double mastectomy almost a year later)
  - ✓ Coding surgery ***planned*** as part of first course of therapy that ***occurs after progression*** of disease (Procedures performed after progression of disease are considered part of second course treatment, not first course.)
  - ✓ Incorrectly inferring the type of surgery based on pathology findings rather than using the surgeon's title and corresponding description of the procedure performed

The operative report is the go-to document when coding Surgery of Primary Site. As indicated in the SEER Manual, "Use the ***entire operative report*** as the primary source document to determine the best surgery of primary site code. The body of the operative report will designate the surgeon's planned procedure as well as a description of the procedure that was actually performed. The ***pathology report may be used to complement*** the information appearing in the operative report, but the operative report takes precedence."

Conclusion

Training data collection personnel is essential to collecting high quality data. Part of training includes the need to not only consider how to identify, quantify, and correct coding issues, but also committing to providing opportunities to discuss those issues with the staff in order to improve future performance. Using audit results is one effective way to identify the training needs of the registry staff.

Given the critical errors identified during this first audit of nearly 500 hospital registrar abstracts, hospital registry managers might consider using these audit results to identify coding modules in SEER*Educate to help train their new and experienced registrars. SEER*Educate training modules are updated annually to reflect the changes to

coding and casefinding guidelines. There are two primary advantages for using the prepared cases in SEER*Educate to train staff on the coding guidelines required by the standard setters:

- The ease in comparing each abstractor's performance against the latest standards because everyone is considering the same case scenarios, expected codes, and rationales supporting the required coding

- The ability to collect summary information on coding results and identify in-person training needs for an individual staff member or for the entire registry staff

Ongoing training is crucial to manage the continuous change in coding and casefinding requirements mandated by the standard setters and to achieve the goal of maintaining high quality of registry data. By annually taking advantage of training opportunities offered at your facility or through WSTRA, NAACCR, NCRA, and SEER*Educate, we can each reach the goal of improving our data quality and the databases to which we contribute.