SEER*DMS Auto-Consolidation Work Group
Teleconference Summary
November 8, 2018
2:00 to 3:00 p.m. EDT

Representatives from the NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 11 cancer registries participated in the SEER*DMS Auto-Consolidation Workgroup (WG) conference call on November 8, 2018. Participants included:

REGISTRIES:
Alaska
California Central
Detroit
Hawaii
Iowa
Kentucky
Minnesota
New Jersey
New York
Seattle
Utah

NCI: Peggy Adamo, Melissa Bruno, Marina Matatova, Serban Negoita, Kai Wong

IMS: Suzanne Adams, Linda Coyle, Fabian Depry

SCG: Kathryn Brown-Huamani, rapporteur

Westat: Laura Lourenco

Action Items

Participants agreed to the following action items:

- Linda and Marina agreed to set up and distribute the invite to an administrative meeting to discuss validation of source records.
- Linda agreed to post the list of priority fields for auto-consolidation in a Squish issue.
- Suzanne and Linda agreed to draft documentation of basic "known over unknown" logic for each priority field. The Work Group (WG) will need to define specific logic for each field.
- IMS will examine the list of fields to determine final field groupings.
- Frances Ross agreed to post the revised Lymph Vascular Invasion (LVI) documentation/instructions on Squish and request feedback.
- Linda, Bobbi, and Frances agreed to meet to discuss the revision of steps 5 and 6 in the LVI documentation. Revisions should include the option of manual review.
- WG members should review and provide comments on the revised priority fields and LVI documents by November 20, 2018. For the LVI document, members also should describe the processes they use for LVI and respond to questions at the bottom of the LVI document.

Two Workgroup Initiatives

Auto consolidation is performed with the assumption that all records are accurate. The WG had recommended two initiatives: 1) developing algorithms to validate source records and 2) developing auto-consolidation rules. Participants agreed to discuss these topics during regular WG calls and to alternate between the two topics. For example, if the December call agenda includes a discussion of algorithms for source records, the January call agenda should include a discussion of auto-consolidation rules.

Cheryl Moody indicated that she and the staff at the California Central Cancer Registry are willing to work on both initiatives. Cheryl will lead the validation efforts.

Discussion

Participants discussed the possibility of the New York and New Jersey registries participating in this effort. These registries have a system that checks incoming cases when they pass through Edits and rejects cases with serious problems such as missing critical fields. This process, however, is not the same as validating source data.

Participants raised the possibility of using SEER Edits to validate interfield data in the source records. Linda agreed that the existing edits should be used when possible. New edits might need to be added but the base would be standard-setter edits. Some cases might pass edits and still have inaccurate data. The workgroup would need to determine whether there are additional fields not currently in an edit that could support or invalidate the code in that particular field.

The WG needs to further evaluate what might be required for validation now and in the future. Questions that need to be answered are: 1) What factors play a role in supporting a specific code? and 2) How can a code be validated without relying on text? It might be useful to determine whether a single code is supported by multiple sources because it is best to validate a code using more than one source. Validation against multiple sources is an intermediate step that could be automated before going to text mining.

Participants discussed steps required to confirm that a record is valid for auto-consolidation. One option is to check the record item by item. The WG needs to consider how this task can be automated. For example, can registrars run specific edits to produce valid codes? The group should first determine objectives for facilitating review of records for auto-consolidation and then identify near-, mid-, and long-term steps for accomplishing those objectives.

Participants agreed to discuss validation of source data at an administrative meeting before the next Auto consolidation WG meeting focusing on validation of source data. The administrative meeting will focus on developing an effective structure and identifying goals and focus areas.

IMS List of Priority Fields

Linda presented a spreadsheet of SEER required data items prioritized for auto-consolidation. These data items fall into the following categories: 1) field categories that are calculated/summarized; 2) items obtained through linkage; 3) items being added or updated in the SEER manual (these will not be included until the manual is updated); 4) rules in development; and 5) rules in production, which include summarized treatment fields and historic data items. Fields were also identified for which "known over unknown" auto-consolidation rules could be developed. IMS has started to group fields for which values must stay together during the auto-consolidation process. For example, laterality should be taken from the facility/record that provided primary site. Participants agreed that the spreadsheet is a useful working document that WG members can reference.

Discussion

In response to a question, Linda explained that most Site-Specific Data Items (SSDIs) were deferred until coding rules are updated. All SEER required data items are included in the spreadsheet. Participants should review this spreadsheet and let Linda know if any item is misclassified.

Participants suggested that IMS evaluate the frequency at which a data item needs to be consolidated relative to other data items. This information would help guide prioritization of fields. Linda agreed to

follow this suggestion and test some simple data items in December. Participants added that fields should be grouped before Linda undertakes this task because some fields affect other fields.

Summary stage is not calculated from EOD fields at most, but not all, registries. Participants agreed to group EOD metastases with Metastases at diagnosis (bone, brain, distant lymph nodes, liver, lung, and other) because a code in any of the latter fields requires an EOD metastases code and vice versa. Data elements for lymph nodes also should be grouped.

The California Central registry staff perform grouping by examining all the fields that are associated with edits. Changing the site code can take the coder to a different and incorrect schema, creating a series of errors. Cheryl Moody recommended examining the edits associated with each field to determine downstream effects of different codes in that field, to avoid conflicting automation logic.

In response to a question, Linda noted that Tumor Size Summary is not a required field for SEER but is required in the NAACCR submission when available. SEER requires Tumor Size Clinical and Tumor Size Pathologic, which are needed to derive Tumor Size Summary.

Logic for Lymph Vascular Invasion (LVI)

Participants reviewed, discussed, and updated the draft documentation/instructions on LVI.
LVI will have more complex logic beyond "unknown" versus "known." They recommended first assigning codes based on schemas and behavior as documented in the SEER Manual and the STORE manual. Next, the hierarchy of valid codes should be defined based on WG input. Participants discussed revisions to the document, which were implemented during the call.

Discussion

Participants agreed that standards should be set to adhere to the Commission on Cancer (CoC) coding instructions because those are the most complex among standard setters. SEER Manual and STORE Manual instructions regarding codes 8 and 9 are different. SEER requires LVI to be coded for penis and testis cases only with all other sites coded as 8. However, the CoC requires LVI to be coded for many more schemas. Participants using CoC instructions were concerned that applying the auto-consolidation logic could lead to 9's (unknown) being changed to 8 (Not applicable). One option would be to place codes 8 (not applicable) and 9 in the current hierarchy of codes, or change step 5 in the draft conversion instructions to read "code as an 8 if 8, code as a 9 if 9, and then use defaults for a blank field." Participants agreed that code 9 should remain available on the source and only the CTC field automatically would be coded 8 if none of the previous conditions applied. Some registries wanted to code LVI sites as 8 if the field truly is not applicable for a particular cancer site as defined by the STORE Manual. When there is no value and "not applicable" does not apply for that cancer site, participants agreed to use code 9. Participants noted that these rules could be included in source record coding logic rather than in auto-consolidation logic.

Instructions must distinguish between source and consolidation logic. At the California Central registry, staff run source and consolidation logic almost simultaneously to identify flaws in the records. The logic should not be directed at improving source information from admissions, but selecting the best admissions information. Registries also should be able to trace all codes back to a source document or have a valid rule for why the source code was overwritten.

Linda, Bobbi, and Frances agreed to implement revisions to steps 5 and 6 in the draft LVI logic based on discussion during this call. They will post the revised draft for review and feedback from WG members. Linda noted that all steps do not need to be automated and sending the user to manual review should

remain an option for steps 5 and 6. When reviewing the revised LVI logic/instructions, WG members should consider how neoadjuvant therapy will affect the coding logic.

Next Steps

Participants agreed to discuss auto-consolidation rules during the next WG call, which would allow more time to schedule an administrative meeting on the validation of source documents.

Next Auto-Consolidation Work Group Call

The next Auto-Consolidation WG call is scheduled for December 13, 2018.