SEER*DMS Auto-Consolidation and Validation Work Group
Teleconference Summary
July 27, 2021
1:00 to 2:30 p.m. ET

Representatives from the NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 13 cancer registries participated in the SEER*DMS Auto-Consolidation and Validation Work Group (WG) conference call on July 27, 2021. Participants included:

REGISTRIES:

Alaska
California Central
Connecticut
Detroit
Idaho
Iowa (Bobbi Matt, WG co-chair)
Kentucky
Louisiana
Minnesota
New Jersey
New Mexico
Seattle
Utah

NCI: Peggy Adamo, Serban Negoita

IMS: Suzanne Adams, Linda Coyle, Fabian Depry, Nicki Schussler, Alex Song

SCG: Kathryn Brown-Huamani, rapporteur

Action Items

Participants agreed to the following action items:

- Registries should review the percentages of lymphovascular invasion (LVI) brain and central nervous system (CNS) cases in Squish issue 9581 and note if they agree to coding all such cases as 8.
- Linda agreed to post a data search showing cases where laterality for nonpaired sites is not equal to 0 (Squish 9246).
- Linda agreed to share and discuss the list of laterality and nonpaired site codes with Jenna Mazreku.
- Linda agreed to have IMS create a data search that demonstrates how the new American Joint Committee on Cancer (AJCC) Tumor, Nodes, Metastasis (TNM) Staging Known over Unknown logic works and distribute this search to the registries.
- Linda agreed to begin with an evaluation of the SSDI guidelines in the SEER Manual (see https://www.naaccr.org/wp-content/uploads/2021/03/SSDI-Manual_v-2.0.pdf?v=1627750277) and then reach out to the SSDI WG.
- Serban agreed to speak with Jennifer Ruhl and Jim Hofferkamp about scheduling a meeting with the SSDI WG. Linda agreed to notify Serban when she is ready to meet with the SSDI WG.
- Participants should send any questions for the SSDI WG to Linda and Suzanne.
- Linda and Suzanne agreed to discuss Suzanne's proposals for including Regional Nodes Positive and Examined and the six Mets at Diagnosis fields in autoconsolidation and develop a proposal to present to the full Autoconsolidation WG.

IMS Update

IMS continues to work on action items from the last Autoconsolidation and Validation WG meeting. Specifically, IMS is working on implementing "Known over Unknown" rules for any field that does not currently have an auto-consolidation rule. As part of this process, Linda will be creating Squish issues to ask the WG members whether IMS can remove certain auto-build logic, which will be replaced by new known over unknown rules. SEER*DMS is moving away from setting default values in an auto-build rule in favor of handling this process in autoconsolidation.

LVI Benign Brain Defaults

Prior to 2018, the SEER Manual had no consistent guidance for coding LVI for Benign and Borderline Brain and CNS cases. The revised SEER Manual now has clear guidance for coding this field. During the last meeting of this WG, participants decided that code 8 would be an appropriate default code for LVI Benign Brain, regardless of the code on the incoming record (which would remain unchanged). Participants agreed to this decision pending review of Benign Brain cases at their registries.

Linda delivered a presentation on LVI Percentages for Brain, CNS (see also Squish issue 9581) based on NPCR or SEER reportable cases diagnosed in 2010 or later across all SEER*DMS registries. She reviewed cases for which LVI was not coded 8 and one of the following was true:

- Schema ID was 00721, 00722, or 00723.
- Collaborative Stage Schema ID was brain, CNS_other, or intracranial_gland
- Tumor/Node/Metastasis (TNM) Schema ID was brain, CNS_other, or intracranial_gland.

Linda presented the results for analyses using the Iowa registry data and combined data from all registries. Bobbi noted that in both analyses, similar proportions of cases were coded 8 and 9. She recommended recoding all LVI Benign Brain, CNS cases diagnosed since 2010 as 8 and creating an edit to ensure consistent coding in the future. Linda requested that individual registries review and comment on results of the analyses and share their final recommendations for coding in Squish issue 9581.

Discussion

- Jenna Mazreku (California Central) noted that no SEER edits enforce the code of 8 for LVI Benign Brain. Peggy suggested raising this concern with the Edits WG. The WG should first ensure, however, that such an edit has not been planned for 2021 cases. Jennifer Stevens of IMS could confirm whether such an edit was in the 2021 edits package. Linda asked that registries provide a comment in the Squish issue indicating whether using code 8 is acceptable.
- Edit N6437 (in 00721, 0722 and 00723 Schemas) does not include the code 8.
- Participants clarified that the default code 8 is not appropriate for invasive Brain, CNS cases because those cases could have LVI.
- Bobbi asked whether the new LVI Benign Brain, CNS coding rule should be a NAACCR edit, a SEER patient set edit, or a CTC edit. Any change to NAACCR edits would need to be addressed by the NAACCR Edits WG. Two members of the Autoconsolidation and Validation WG serve on the NAACCR Edits WG. At this point, the edit would be a SEER guideline only.
- Participants asked about the research value of Benign Brain, CNS codes other than 8. Other codes received from hospitals would be conserved and remain available. Only during the autoconsolidation process would LVI Benign Brain, CNS cases be coded 8.
- LVI Percentages for Brain, CNS will be available for review in the Squish issue. Registry staff should contact Linda if they have difficulty accessing the data search results.

- Suzanne pointed out that the version of the SEER Manual revised in 2018 instructed registries to use code 8 for LVI Benign/Borderline Brain, CNS but this instruction was removed in the 2021 manual. Per Peggy Adamo, the upcoming SEER Manual 2022 will include that instruction.

Laterality and Non-Paired Sites

IMS worked with NCI to define a list of paired and nonpaired sites and sites that could be categorized as either. The SEER*DMS table that stores the primary site codes includes the paired site codes. The list will be used in SEER*Edits, autoconsolidation, and other SEER*DMS algorithms.

WG members agreed that laterality should be set to 0 on the CTC for nonpaired sites. Linda agreed to post the results of a search showing cases where laterality for nonpaired sites does not equal 0. Participants agreed with forcing the laterality code of 0 for nonpaired sites after checking to ensure the site was nonpaired. Linda agreed to update the Squish issue for laterality and nonpaired sites.

Discussion

Jenna asked about sharing queries for different registries and possibly discussing this issue offline with IMS. Linda agreed to speak with Jenna offline and share the list of site codes with her.

Tumor Size Clinical and Tumor Size Pathological

Based on the discussion during the last Autoconsolidation and Validation WG webinar, IMS developed proposed logic for Tumor Size Clinical and Tumor Size Pathological that can be tested. The SEER Manual provides detailed instructions for when to assign codes 000 and 999 for Tumor Size Clinical. The problem is that the sites and schemas that should be coded 999 could also have an EOD Primary Tumor code 800, which should be coded 000 per the currently published instructions. These coding instructions, however, included contradictions that led to conflicts being identified that actually were not conflicts. NCI and IMS are attempting to resolve this problem by adding exceptions to the instructions to code Tumor Site Clinical as 000 when Extent of Disease (EOD) Primary Tumor is coded 800 (no evidence of primary tumor). These added instructions read “for any schema except 00830 (excluding C422), 00458, 00790, 00795, 00672, 00671, 00822, 00821.” Schemas that would be coded 999 therefore need to be excluded when executing the EOD Primary Tumor 800 coding.

Linda pushed two data searches out to the registries: (1) Tumor Size Clinical conflict, and (2) Tumor Size Pathological conflict. Participants should review Squish issue 9547 and post their feedback.

Discussion

Nancy Lozon at the Detroit registry noted that she has cases in SEER*DMS for which there is no evidence of a primary tumor with a Tumor Size Clinical code of 999 when is should be 000. A conflict should occur in these cases but is difficult to see in SEER*DMS. Suzanne suggested that registries export data to an Excel spreadsheet to examine the codes for various combinations of sites and schemas. In Excel, filters can be used to select cases that should be coded 999. This approach will allow registrars to systematically identify cases that are incorrectly coded. Participants tried this approach using a sample of Detroit data. During this test, a participant raised a concern about soft tissue sarcoma primary site being coded 999. Linda agreed to re-examine that site.

AJCC TNM Known over Unknown

Suzanne presented proposed logic for the AJCC TNM Staging Known over Unknown rule. The logic would apply to all four sets of TNM fields. The logic would keep sets together in a record. On the CTC, however, each set could come from a different source. The new logic will take all of the fields from the same record if the CTC values are unknown or the values in the record do not conflict with those in the CTC and some of the fields can be completed.

Discussion

Participants discussed the proposed AJCC TNM Staging Known over Unknown logic. Linda clarified that the T, T Suffix; N, N Suffix; M, and Stage Group fields would come from a single record when a record can be identified. If all records have the same values for those fields, those values would be taken. If one record has a complete set of values and there is no conflict with any known values across the record, the complete set of values will be taken. If there are conflicts in any known values, a manual review would be forced.

Serban expressed concern about taking all of the fields from only one record because two records might come from different hospitals for the same case and only one might have information such as an imaging of a distant site. Serban requested examples to demonstrate how the proposed logic would work. Linda agreed to have IMS create a data search that demonstrates how the new AJCC TNM Staging Known over Unknown logic works and provides examples of cases that conflict with the logic.

In response to a query from Suzanne, participants indicated that TNM sets should not be broken apart. The WG agreed to discuss the proposed logic at a future meeting after WG members have reviewed some examples.

Upcoming Efforts

Evaluate Progress of Rules in Development

New Tumor Size Clinical and Pathological rules are ready for evaluation. The evaluation should begin by defining ways that known values conflict.

Grouping SSDIs

The WG has proposed organizing site-specific data items (SSDIs) into groups of similar items (e.g., laboratory values, interpretation fields) to simplify the process of creating logic for autoconsolidation.

Discussion

- Jenna agreed to obtain feedback from other colleagues at the Central California registry on the grouping of the SSDIs.
- Linda suggested that participants consider specific fields that would always be taken from the same record.
- Nicki noted that SSDIs are complicated and therefore should be organized and reviewed by schema.
- As a start, similar fields such as laboratory values and ranges and Gleason patterns and scores could be grouped. Fields that must come from the same source also could be grouped. Once groupings are established and logic is developed, range values between specific fields will need to be considered. Participants suggested reviewing the schema for each SSDI in SEER*Edits which might provide a map of what needs to be grouped.

- Linda emphasized the importance of keeping the logic for consolidating SSDIs simple. Whenever possible, a record that completes an unknown value for a field should be used. With SSDIs, IMS needs to identify values that rely upon multiple fields.
- Serban suggested asking the SSDI WG for input on grouping and prioritizing SSDIs for autoconsolidation. Linda agreed with this suggestion but still would like to begin developing an approach for grouping SSDIs using a single schema. Nicki suggested beginning with breast, prostate, colorectal, or lung schemas because these sites have the most SSDIs. Linda agreed that prostate might be a good place to start and other participants proposed using the prostate PSA schema. The SSDI WG will need examples.
- A participant raised a concern about different interpretations of lab values, which could complicate the autoconsolidation logic. A standard approach for addressing two known values in autoconsolidation would be needed with exceptions.
- Linda pointed out that the autoconsolidation logic should mimic what is done manually. The various guidelines for coding SSDI fields therefore should be considered (AJCC, SEER, American Colleges). Linda agreed to begin with an evaluation of the SSDI guidelines in the SEER Manual, and then reach out to the SSDI WG (see https://www.naaccr.org/wp-content/uploads/2021/03/SSDI-Manual_v-2.0.pdf?v=1627750277). Linda asked Serban and Nicki to arrange a meeting with members of the SSDI WG to discuss priorities and approaches for autoconsolidating the SSDI fields. Serban agreed to speak with Jennifer Ruhl and Jim Hofferkamp about scheduling a meeting. Linda agreed to notify Serban when she is ready to meet with the SSDI WG. Serban asked that participants send any questions for the SSDI WG to Linda and Suzanne.

Other Fields

Suzanne asked for ideas for other fields that should be included in autoconsolidation. She suggested Regional Nodes Positive and Examined. These fields do not need to be combined with EOD Nodes. Suzanne also suggested considering the six Mets at Diagnosis fields. The consistency of current coding of 8 versus 9 in these fields across sites should be examined. Linda pointed out that the six Mets at Diagnosis fields could not be included in autoconsolidation without including the EOD fields. Linda and Suzanne agreed to discuss Suzanne's proposals in a smaller group meeting and then present a proposal to the full Autoconsolidation WG.

Suzanne suggested first reviewing the relatively simple Sentinel Nodes field. Collection for this field only began in 2018 and it includes only two schemas. Linda agreed to this proposal.

Announcements

- IMS and NCI are working together to develop COVID autoconsolidation logic.

Discussion

A participant inquired about a Squish issue she submitted regarding Date of Diagnosis and linkage of HL7 and other records received. She found some cases in which a later Date of Diagnosis on a new record replaced the original Date of Diagnosis in SEER*DMS. Linda agreed to investigate this problem.

Participants asked when the EOD Consolidation Manual would be used to develop autoconsolidation logic. IMS is examining this Manual to determine the best approach for developing the logic. Once this task is complete, Linda will discuss the planned approach for developing autoconsolidation logic with this WG.