SEER*DMS Change Control Advisory Board (CCAB) Users Group
Teleconference
September 25, 2023
3:00 p.m. to 4:00 p.m. EDT

Representatives from the NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 26 cancer registries participated in the SEER*DMS Webinar on September 25, 2023. Participants included:

REGISTRIES:

Alaska
Arkansas
California Cancer Registry
Cherokee Nation
Connecticut
Detroit
Florida
Georgia
Greater Bay Area
Greater California
Hawaii
Idaho
Illinois
Iowa
Kentucky
Los Angeles
Louisiana
Massachusetts
Minnesota
New Jersey
New Mexico
New York
Seattle
Texas
Utah
Wisconsin

NCI: Kathy Cronin, Betsy Hsu, Marina Matatova, Serban Negoita

IMS: Suzanne Adams, Linda Coyle, Nicki Schussler, Jennifer Stevens

SCG: Kathy Brown-Huamani, rapporteur

Action Items

- Registries should submit to IMS their questions about the changes in the data submission process.
- Linda agreed to create a Squish issue to collect registries' feedback on the Cancer PathCHART "impossible" or "unlikely" tumor site morphology combinations list.
- IMS needs to ensure that the Michigan registry is on the CCAB's meeting invitation list.

Overview of Meeting **Linda Coyle, Marina Matatova**

Linda welcomed participants and noted that today's agenda includes an update on the annual SEER submissions related to SEER*DMS, status of changes on those submissions, real time reporting, and ongoing projects.

She announced that SEER*DMS went live at the Michigan registry on September 11, 2023. The Detroit and Michigan registries are sharing a SEER*DMS instance.

Annual SEER Submissions **Linda Coyle**

The submission requirements can be found at SEER Data Submission Requirements website. Final updates to these requirements were posted on September 22, 2023.

Status of SEER*DMS Changes to Support Annual Submissions

Most of the 2023 changes to SEER*Edits were released on August 30, 2023 and some remaining updates were scheduled for release in late September. The changes include new edits for Census 2020 fields and minor updates to existing edits.

SEER*DMS also is being updated with SEER, NPCR, and NAACCR requirements for 2023. EXT-02 (now called Submission Extract) has been updated to support all submission files required by the different standard setters, including NPCR, NAACCR, National Childhood Cancer Registry (NCCR), Virtual Pooled Registry (VPR), Indian Health Services (IHS) linkage and SEER-Medicare linkage. This will enable registries to create all necessary submission files from the same cut of data.

Date suppression options for the NAACCR Call for Data were expected to be deployed in EXT-02 as part of the September 25, 2023 update . These changes will also be implemented in NAACCR*Prep. IMS received a request for a date suppression option for the NPCR extract and will be making those changes in SEER-DMS accordingly (see Squish #12544). The goal is to equip EXT-02 with the necessary options to eliminate or minimize the need for post-processing.

Linda led a demonstration of SEER*DMS EXT-02 and noted the following: 1) the SEER transmission data file is set to "yes" by default; 2) all other data files are set to "no" by default; 3) IMS has included suppression options to support the preferences and policies of specific registries; 4) dates always are masked for the SEER submissions; and 5) IMS options for DOD training packages soon will be removed to avoid inadvertent use.

Questions and Answers

A participant asked whether the NAACCR submission file extract can be run without running the SEER extract. IMS confirmed that it can and explained that the option for SEER Transmission Data File would be set to "no". Registries should include the linkage files when selecting the SEER transmission data file in the final submission, but if the extract is being run for testing or edits purposes then the linkage options can be turned off,

The Seattle registry representative noted that the SEER-Medicare Finders file requirements will be deployed after the data submission is due. Jennifer Stevens has been working to ensure that those requirements are released earlier and has provided a preliminary draft of the file to CMS. She noted that the 2023 changes to the SEER-Medicare Finders file involve providing a tumor-level rather than patient-level file. In the past, linking a patient-level file inadvertently included some CTCs that were "restricted-

release". EXT-02 now generates a tumor-level file and there are various restricted release options: 1) exclude all patient CTCs if at least one is restricted, 2) ignore all restrictions and include all CTCs 3) include only CTCs that are not restricted.

Marina asked whether a Squish issue was created for registry staff to ask questions about submission changes. Linda asked participants whether they preferred to use Squish or submit individual questions to IMS and note concerns affecting multiple registries. Registries preferred submitting their questions individually.

Submission Procedures/Announcements

In terms of edits, all standards setters (e.g., NAACCR) have released their meta-files. The recommendation is to run SEER*Edits and GenEDITS to get an accurate sense of the status of edits in the data. Registries should check for edits in SEER*DMS that may need to be activated and check for a edits with a high number of failures in SEER*Edits and/or GenEDITS. Registries should contact IMS via Squish for assistance.

Genomic data (from Castle and Decipher) have been moved from the linkage file to the main transmission file, which also contains standard data items. Pharmacy and medical claims data are in the linkage file. The Central California registry will create its own transmission file and add the genomic data; IMS can provide any needed support.

Best Practices for Identifying Cases to Submit

Linda suggested identifying best practices related to submission processes and asked registries about how they currently focus on cleaning data for submission. She indicated that some registries may "freeze" new imports, and other registries may use an ID file to focus on cases to clean and submit. This may be a topic of discussion between NCI and IMS to standardize in 2024.

A Seattle registry participant noted that the size of a registry affects whether data need to be frozen at some point. The Seattle registry, which is a smaller registry, can continue to process records longer than the larger registries (e.g., California, New York, Texas). The Utah registry pauses imports of hospital abstracts about a month ahead of the data submission and sometimes pauses pathology (path) report imports within a few weeks of submission. The New York registry will first clean the CFD cases, then start working on path screens from the next year so as not to affect CFD cases.

Serban asked if other registries stopped importing abstracts and/or path reports before the SEER data submission, and then restarted imports, for example, between the SEER, NPCR, and NAACCR submissions.

The Georgia registry creates submission files for all standard setters (SEER, NPCR and NAACCR) from the same cut of data. The registry typically stops imports about a week before the SEER submission is due. The Illinois, Idaho, and Kentucky registries employ a similar procedure. The Kentucky registry discontinues imports 1 week before the submission is due and freezes database activities while preparing extracts for submission. Serban asked that registries contact NCI if they are creating extracts for the SEER, NPCR, and NAACCR submissions at different times without freezing.

The Minnesota registry uses a process involving testing, cleaning, retesting in GenEDITS. This process ensures that corrections are made in both the production and test servers.

The Idaho registry pauses imports at a certain point to work on edits and other data refinements. When ready to submit, registry staff cut the data files and run them through edits for NPCR, NAACCR, and SEER. Cleaning is performed as needed. All files are cut once they are cleaned.

The New Jersey registry pauses imports, which is time dependent. The Seattle registry does not stop imports until the final days before the submission. This process is connected to their real-time reporting workflow.

The Texas registry has developed an identification (ID) file using CTC date to limit cases in a file. The Texas registry will not stop imports this year due to difficulties in the past caused by frezing.

The Iowa registry will stop imports closer to November 1 of any given year and limits which tasks staff can do during the last 2 weeks before submission.

To better support future submissions, Linda proposed a tool that would flag all CTCs created for submission so that any new cases would not be added in the last period right before November 1. The only work required for registries would be if new data for existing cases was added and consolidated automatically. This proposal raises the question about what exactly 'freezing' the data means and whether "freezing" includes stopping automated consolidation and linkage to the identified CTC. IMS is starting to explore ways to standardize these processes, and ideas are welcomed.

Real-time Reporting Workflow **Linda Coyle**

Real-time reporting refers to creating CTCs from path reports and not waiting for the abstract, thus making a larger proportion of the data from the CTCs available to public more quickly.

In 2023, IMS worked with pilot registries to test a real-time reporting workflow. A manual review conducted as part of this pilot study revealed that many prostate cancer CTCs (C619) created from path reports were non-reportable. The Path Extraction API that auto-codes path reports using natural language processing (NLP) was overidentifying malignant prostate path reports. Until the Path Extraction API predictions can be improved, SEER will require manual screening of these prostate path reports (unless they can be matched to an existing abstract based on dates). The pilot study also revealed that some of the CTCs created were built with unusual or invalid codes. A registry recommended using the Cancer Pathology Coding Histology And Registration Terminology (Cancer PathCHART) impossible site/histology combination list to force a manual review.

Marina wondered whether registries should provide feedback on real-time reporting before implementing any changes. Linda will create a Squish issue to collect registry comments.

Discussion

Betsy noted that NCI is planning to meet with the DOE team and Cancer PathCHART leadership to discuss how to use lists in the context of the Path Extraction API NLP predictions. In response to a question from Marina about use cases, Betsy noted that the NCI-DOE team had discussed incorporating the "impossible" lists, but none had been published by Cancer PathCHART at the time. Now that lists for the first set of cancer sites are being released, a meeting with NCI, DOE, and Cancer PathCHART is being planned. Linda asked registries whether they would want the abstractor to see a prompt or warning if they are attempting to set an impossible site/histology combination in a manual path screening task.

Bobbi Matt (Iowa registry) commented that an impossible site/histology combination would need to be corrected. Reports with an "unlikely" but not impossible combination could move forward in the workflow and be corrected or overridden at a later time. When the codes are unlikely, Serban asked Bobbi whether it would be best to the path screener review the path report and correct the combination, rather than move it forward in the workflow to the next abstractor. Bobbi responded that if the path screener is the user who created the unlikely or impossible combination, it should go to a second reviewer.

Linda clarified that if the API generates an impossible or unlikely combination, the task would move to manual review in path screening. If a person performing the path screening task coded an impossible combination, then the system should post a warning. If an unlikely combination is returned, then that should be coded on the path report, but without automatically building a CTC. The next step would be to wait for the casefinding and the abstract to build the CTC.

Colleen Sherman (New York registry) reported reviewing potential tumors with unlikely combination codes that were generated by the API. She prefers not having a warning, which would slow the workflow at her registry. She would rather see errors or warnings when manually building the CTCs. Linda clarified that it would only be the impossible combinations that would generate a warning. IMS will look at numbers to see how often this is happening.

Serban called attention to the recently published Cancer PathCHART list of impossible and unlikely combinations (2024 Cancer PathCHART ICD-O-3 Site Morphology Validation List) and reiterated that the SEER will work with the DOE team to update the API algorithm to ensure that unlikely and impossible combinations are not created.

Nicki provided a brief overview of Cancer PathCHART. When standard setters release new ICD codes, it can take time to reach the U.S. cancer registry community. The purpose of Cancer PathCHART is to communicate new standards more efficiently. A Site Morphology Validation List was released first. This list includes all solid tumor site histologies, stratified by disease. Pathologists specializing in specific diseases review and classify various types of combinations as unlikely or impossible and work together to reach consensus.

When asked whether cases diagnosed in 2023 and 2023 would need to be validated, Suzanne clarified that cases from 2024 and forward will be evaluated. For the purposes of flagging unusual cases in the SEER*DMS workflow, Linda asked the registries to submit a chat comment about whether to use the impossible and unlikely combinations for cases pre-2024.

Linda noted that registries should check their real-time reporting data to determine if they had increases in the number of cases for 2022 to 2023 and provide any updates by the end of the calendar year. The period between the November 2023 and February 2024 submissions would be an opportune time to focus on improving the real-time reporting workflow. Any registries wanting more information on Cancer PathCHART can contact the NCI. IMS will update SEER*DMS to include warnings for reports that were coded as an impossible or unlikely combination.

Next Steps **Linda Coyle**

Marina would like feedback from the registries prior to implementing any changes in SEER*DMS based on Cancer PathCHART lists.

The next SEER*DMS workshop will be held on November 6, 2023.