SEER*DMS Change Control Advisory Board (CCAB) Users Group
Teleconference Summary
January 10, 2019
3:00 to 4:30 p.m. EDT

Representatives from NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 21 cancer registries participated in the SEER*DMS Users Group conference call on January 10, 2019. Participants included:

REGISTRIES:
Alaska
Cherokee Nation
Connecticut
Detroit
Georgia
Greater California
Greater Bay Area
Hawaii
Idaho
Iowa
Kentucky
Los Angeles
Louisiana
Massachusetts
Minnesota
New Jersey
New Mexico
New York
Seattle
Utah
Wisconsin

NCI: Jessica Boten, Melissa Bruno, Kathy Cronin, Lois Dickie, Marina Matatova, Serban Negoita, Alyssa Wang, Kai Wong

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

Westat: Laura Lourenco

SCG: Kathy Brown-Huamani, rapporteur

Action Items

- Serban agreed to post his proposal for pathology metrics as a Squish issue.
- IMS and NCI will modify the language in Serban's description of the proposed metrics to clarify the approach and incorporate recommendations made by registry participants.
- Linda, Marina, and Serban agreed to discuss the metrics with each registry separately to ensure that registry staff understand how the metrics will be implemented at their location and to make modifications for processes unique to each registry.
- Linda agreed to create Squish issues on new Data Standards changes to allow registries to provide feedback and ask questions. Participants requested two Squish issues, one on SSDIs (SSFs) and the other on broad levels of editing. Registries should provide feedback in Squish.
- IMS will include the metrics in a Squish issue to obtain additional feedback.
- IMS will contact registries in the near future to discuss approaches to automating sequence number.

Announcements

- NCI and IMS are working toward deploying SEER*DMS at the SEER registries that still do not have the system.

- NCI and IMS are planning to work with registries on the classification and new metrics.
- Future CCAB meetings will include proposals to obtain input from registries. This information will be used to drive Working Group (WG) and NCI/IMS efforts.

Pathology Metrics Review **Serban Negoita**

Serban presented a draft proposal for the development of pathology metrics that can be used at all SEER registries and requested feedback on this proposal. He also requested input on next steps toward developing metrics and best practices for acquisition, processing, and use of ePath reports. In addition, NCI/IMS are working on categorization of pathology reports in SEER*DMS.

NCI/IMS began work on ePath metrics that could be calculated before the reports are uploaded to SEER*DMS. Once the CCB determines the types of metrics that would be useful, IMS can move forward with the development of those metrics as well as a brief glossary of terms to help identify reportable and non-reportable cases and define an image, structured, and unstructured report.

NCI wants information on registry criteria for case reportability. NCI also wants to develop pathology metrics that consider the nonreportable pathology reports that reside in SEER*DMS.

Serban proposed six metrics but noted that the exact implementation of each metric might vary by registry because each registry might have a slightly different definition of structured and unstructured records. He noted that ePath reports are comprised of image, structured, and unstructured information. The reason for developing the metrics was to learn how registries are processing pathology reports and the volume of pathology reports being processed. NCI would like to explore questions such as 1) What time lag do registries experience between receiving pathology reports and abstracts? 2) What tools and technologies are registries employing to process pathology reports? This information will guide NCI quality initiatives such those focusing on AIM technology and NLP as well as relevant future initiatives.

The first metric would measure the proportion of cases histologically confirmed through pathology reports generated within 60 days of the date of diagnosis. Serban asked about the appropriate time period after date of diagnosis for identifying these cases (e.g., 60 days as stated in the solid tumor manual, 30 days, less?). Participants recommended examining how much the proportion of cases with pathology reports would increase from 60 days to 90 to 120 days after diagnosis, and so on. IMS has examined pathology reports for various projects for factors such as the date a specimen was collected compared to the date of diagnosis and the surgery date. IMS found that most pathology reports are received within 60 days of diagnosis. Linda noted that operations and availability of data could be examined over the next year.

Registry participants pointed out that many abstracts are received long after the associated pathology reports are received so information on the date of diagnosis might not be available for several months after the pathology report is received. Linda explained that the denominator is the number of CTCs and the metric would examine the proportion of CTCs linked to pathology reports. All CTCs have a date of diagnosis. Participants agreed that the metric would need to be used about a year after pathology reports were received to allow time for them to be linked with an abstract.

The second metric was similar to the first but focused on the proportion of pathology reports received more than 12 months after the date of diagnosis by cancer site and year of diagnosis. The assumption is that these pathology reports reflect a recurrence. NCI is interested in compiling and analyzing data on recurrence and possibly developing an NLP algorithm to automatically collect data on recurrence. Serban clarified that the pathology report date used in the first and second metrics would be the date of specimen collection, not the date the pathology report was received by the registry.

The third metric measures the total number rather than the proportion of pathology reports by type (structured, unstructured, and image), year, registry, and possibly other categories such as cancer type. Linda suggested also examining the number of manually coded pathology reports to see patterns over time. For the third metric, the date pathology reports were uploaded to SEER*DMS will be used rather than the date of specimen collection.

The fourth metric would measure the proportion of CTCs linked to structured, unstructured, or image pathology reports by year of diagnosis. The goal is for the proportion of structured reports to increase over time and the proportion of image reports to decrease to 0. Participants asked about the approach to CTCs linked to more than one type of pathology report and suggested a hierarchical system to evaluate these CTCs.

The fifth metric will examine the number of death reports linked to a patient set but not a CTC over time. The sixth metric will examine the number of CTCs with no pathology report. Serban clarified that this metric would reference only confirmation via pathology report and would not include radiologic or clinical confirmation. Participants suggested creating another metric to compare cases that are histologically confirmed and those confirmed through other means to evaluate the effectiveness of ePath data streams.

In the DoE NLP project, IMS found that 60 to > 85 percent of CTCs have pathology reports, depending on registry. Among cases with pathology reports, most had more than one report indicating that multiple pathology reports often are received for the same case. Many histologically confirmed cases also are missing a pathology report, which needs to be investigated. Some registries do not collect pathology reports for a case that has a NAACCR abstract. Linda explained that this approach was fine when pathology reports primarily were used to determine whether a registry had a case. NCI's goal is to obtain data items from the pathology reports themselves.

Serban asked participants for feedback on the classification of pathology reports in the glossary he presented. Classification is based on whether the report is an image report and whether it is structured or unstructured. Some registries are starting to receive medical images via HL7 files, so the term "image" might need to be redefined. Participants asked if SEER*DMS would flag unique reports. Although unique reports are not specifically flagged, a report created from an HL7 file is distinguished from a PDF file that is loaded or entered. HL7 reports currently are classified as unstructured, but participants noted that HL7 reports are heterogeneous with regard to degree of structure. Participants asked about the rationale for classifying reports differently if they all are treated the same. The rationale for classifying reports as structured versus unstructured is to be able to monitor the proportion of each type of report received. NCI wants the proportion of structured reports to increase over time. In addition, image reports affect the workflow because SEER*DMS cannot read these reports, so the text needs to be entered. On the other hand, free text reports can be read by NLP algorithms. Although the goal is to increase the number of reports with coded data items, to date, IMS has seen few such reports. Definitions 4 and 5 in the glossary presented by Serban address this concern.

Discussion

Reportability varies according state laws. Facilities that use AIM have reportability criteria programmed into their ePath Reporter. Several registries that receive records via AIM perform additional screening once cases are uploaded to confirm that they are reportable. Some registries also upload nonreportable cases to SEER*DMS. For example, in Minnesota, treatment for dermatology cases is considered non-reportable, so the registry staff abstract information on those cases. The registry also receives and uploads cases from neighboring states, which would be considered nonreportable. Minnesota has a list of specific

types of nonreportable pathology reports the registry wants to retain. In New York, Georgia, and Louisiana, AIM sends cases to SEER*DMS and registry staff screen for reportability. Cases loaded into SEER*DMS are not deleted after they are deemed nonreportable. A reportability flag is available at the record level for registries that upload nonreportable cases. The Utah registry retains nonreportable pathology reports to update records at follow up.

Some representatives from registries that do not yet use SEER*DMS asked for clarification about the process for receiving and using the metrics with pathology reports. They also asked for clarification about how to use the metrics with unlinked pathology reports that might represent a new diagnosis. Histology would not be known until the pathology report has been processed. NCI and IMS will need to determine an approach for handling unlinked pathology reports that represent a new diagnosis. Participants added that the records used for fifth metric do not include a date of diagnosis, therefore, the date of the pathology report might be used.

The New York registry reported difficulty linking pathology reports to CTCs, especially those from independent pathology laboratories. This point highlights the limited resources of some registries, which might impede their ability to use the proposed metrics. The metrics will help NCI understand the needs of each registry, for example, based on the number of reports that are not automatically matched.

Marina asked if registries other than New York were sometimes unable to match pathology reports after reviewing them. She would like to know how registries handle and classify those reports to determine whether a metric is needed for unmatched pathology reports. Reportability sometimes cannot be determined because geographic and other information are missing. This is particularly true for reports from hematopoietic laboratories. The New York registry classifies reports that cannot be confirmed as reportable as auditable, and they are not counted until the registry can confirm reportability through the physician who requested the pathology report. The Minnesota registry employs a similar approach. California registries received screening pathology reports with specimen dates in November 2018 that have not yet been linked to a CTC because they were screened before receiving the abstract for the cases. These registries wait to see if the report will be linked to a tumor because, in California, they have a 6-month reporting delay. Participants agreed that the fifth metric should examine reportability and use a date other than the diagnosis date.

During casefinding, the New York registry found that certain types of cancers, such as melanomas, were missing because they were not treated in hospitals. As a result, the registry decided to obtain most pathology reports from independent laboratories to ensure complete casefinding. Participants suggested examining the proportion of pathology reports received from independent versus hospital laboratories. Reports received from hospital laboratories are easier to link because they usually include more information.

SEER*DMS Updates **Linda Coyle and Marina Matatova**

Linda discussed NCI's efforts to inform the CCB and the plan to change the CCB to the Change Control Advisory Board (CCAB). Linda asked for registry input to identify topics for upcoming CCAB meetings and to share challenges at individual registries. Registry input will drive changes to SEER*DMS functionality. Information about specific changes to SEER*DMS functionality are posted in Squish.

Linda provided the following updates:

- Since September 2018, IMS made many changes to SEER*DMS in response to NAACCR 2018. These changes are expected to decrease in 2019 but work on data standards related to NAACCR 2018 will continue.
- New SEER*DMS migrations are underway at the Idaho and Kentucky registries and are expected to be completed in the first quarter of 2019.
- IMS is building a migration framework that will be used at the Massachusetts and California registries. The version history will include items related to this new framework.
- Much work over the past year focused on submission requirements. IMS worked closely with New York on SEER submissions.
- IMS put substantial effort into the API for autocoding pathology reports using NLP.
- In the first quarter of 2019, IMS will focus on usability testing with the goal of developing and releasing a new prototype of SEER*DMS. The prototype will be reviewed with partners and testing will begin in 2019.
- IMS is testing version 5 and implementing recommendations from work groups. The new version will include changes to display and logic.

Data Standards Changes

Linda discussed changes related to Data Standards. As an example, she displayed the staging page for a 2018 case. SSF (SSDI) data items appear near the bottom of the screen and required fields are bolded. In response to registry requests, the SSDI section was divided into two sections for clarity (required items in Standard Editing and others in Minimal Editing). Items that always are required will be separated and placed at the top of the screen. Lower SSDIs might be bolded because these are required by some groups. The impact of these changes is an expected agenda item for the next SEER Managers meeting. Manager discussions should focus on ways that CTR efforts can inform the usability project. In addition to screen changes, IMS will create known over unknown rules for SSDIs.

Discussion

In response to participant questions, Linda clarified that IMS only is working on SSDIs at this point. Many other categories of data items are required, when available, therefore, the language associated with these categories will need to be discussed at the Managers meeting.