SEER*DMS Change Control Advisory Board (CCAB)
Teleconference Summary
May 9, 2019
3:00 to 4:30 p.m. EDT

Representatives from NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 17 cancer registries participated in the SEER*DMS CCAB conference call on May 9, 2019. Participants included:

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

NCI: Peggy Adamo, Kathy Cronin, Lois Dickie, Steve Friedman, Marina Matatova, Serban Negoita

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

Westat: Laura Lourenco

SCG: Kathy Brown-Huamani, rapporteur

Action Items

- Linda agreed to discuss the evaluation of the new algorithm to link pathology reports to patient sets and CTCs at a future CCAB meeting, probably in September.
- NCI staff will contact the SSA to clarify the accuracy of SSA vital status information.
- Marina and Melissa Bruno will be contacting each registry to schedule discussions regarding the registry's results from the ePath metrics.
- IMS will work on creating an ePath metrics data report for each registry so that registries can review the results in more depth. IMS will add documentation to clarify which numbers represent paper versus electronic pathology reports.
- Marina and Linda agreed to ensure that all comments regarding the metrics have been addressed. Registry representatives should submit any additional questions as soon as possible via the Squish issue.
- NCI will work with IMS to develop a format for a report of ePath metrics results once all input is received from registries.
- Linda agreed to add logic from the queries to the metrics reports for each registry.
- IMS will include the ePath metrics in a Squish ticket to obtain ongoing feedback.
- Registries that received an email notification from IMS about updates to the Data Confidentiality Agreement should respond with any comments before next week.
- Registry representatives should submit proposed CCAB call topics to Linda and Marina.

Announcements **Linda Coyle**

Linda provided the following updates:

- IMS is releasing the SEER Solid Tumor Rules this week. The multiple primary (MPH) rules were turned off for the CTC matching algorithm for 2018 cases, but now will be turned on, allowing incoming 2018 records to be matched. IMS can work with registries to evaluate the matching of records already in the system. IMS plans to implement quarterly updates and collaborate with Lois to determine the best way to implement the updates.
- The Idaho registry now is using SEER*DMS.
- The source code implementation of the Autoconsolidation module has been updated to support more complex rules being defined by the Autoconsolidation Work Group.
- IMS continues to implement new Autoconsolidation rules.
- IMS added new, more specific fields to SEER*DMS to indicate the source of the data in a CTC. The new fields will indicate reporting by the Department of Defense (DoD), by other facilities that might have fewer data restrictions (Other), and whether the information came from a data exchange. These fields can be used for data restrictions when needed and might replace the current CTC restricted field. These fields currently are being implemented at the Kentucky and Seattle registries. IMS will work with each registry on the conversion to these fields. IMS will create a Squish issue to allow registries to communicate their specific needs in implementing these fields. If the DoD is noted as a source, the U.S. Department of Veterans Affairs (VA) facility will need to be recorded.
- IMS has implemented changes to Data Standards, including updates to Edits, staging algorithms, libraries, polishers, defaults, and data conversions for 2018 fields.
- IMS conducted several security tests, including a formal risk assessment, and made related adjustments to SEER*DMS. IMS currently is responding to a system audit.
- IMS is working on algorithms to automate sequence number so that all sequences will be accounted for in the patient set.
- Three large registries expressed interest in the ability to upload image files and have those files automatically linked to existing records. IMS has been working on this functionality, which it expects to release in the next few weeks.
- IMS is making progress on the Natural Language Processing (NLP) algorithm. Linda expects changes to be made to SEER*DMS in late 2019 related to this algorithm.
- A new prototype of SEER*DMS was developed in response to usability testing and is being reviewed for release later this year. NCI plans to deploy new functionalities in SEER*DMS through an iterative process.
- The Department of Energy (DoE) and the Usability team will conduct webinars and other activities to educate registry staff about changes that will be made to SEER*DMS as a result of the usability testing and allow them to provide feedback on those changes. Registry staff likely will begin to receive notifications about the SEER*DMS changes and webinars on those changes in August 2019.
- IMS is developing a standard template for data exchanges that should be available soon.

Discussion

A participant asked if the new source of data category, Other, would include cause of death (CoD) as reported by the National Death Index (NDI). If CoD only is obtained from the NDI, registries are not allowed to release that information and should not use the Other field to report NDI as a source for CoD. Registries should flag CoD information that is obtained only from the NDI.

ePath Workflow Review **Linda Coyle**

A few registries have requested changes to their ePath workflow. Linda has submitted Squish issues about the review of the ePath workflow in preparation for these changes. This review might serve as a model for periodic review of other SEER*DMS workflows. Linda recommended performing detailed reviews of any workflow that is undergoing significant change. Registries can benefit from the workflow ideas and analyses of other registries. The sharing of information across registries should help to standardize workflows. IMS will review workflows to reduce maintenance and prepare for future changes. IMS staff want to understand the variations in workflows among registries to more efficiently implement changes and updates in SEER*DMS. To date, IMS has identified two main areas in which registries differ in their ePath workflows:

1. The sequence of manual pathology screening versus matching. In some registries, the pathology report is matched against the patient set prior to manual pathology screening. In other registries, no matching occurs until the manual pathology screening task is complete. Performing the match prior to manual pathology screening might be the best approach because the registry would have more information about the patient at the time of pathology screening.
2. The timing for building the CTC. Some registries build the CTC immediately for a subset of pathology reports, usually those cancers for which the registry does not expect to receive an abstract. Other registries do not build any pathology-only CTCs until casefinding steps are completed. Linda encouraged participants to consider other possible approaches to building the CTC.

IMS created an option to match pathology reports to the patient set before performing pathology screening. The algorithm developed to perform this match also can match pathology reports to CTCs when they are available. Pathology reports would not undergo screening when they matched a CTC. The pathology report normally is received before the CTC is created, so few reports will match to a CTC. Utah and Georgia registries currently use this new algorithm to match pathology reports to both patient sets and a few CTCs. The Louisiana registry uses the algorithm to match pathology reports at the patient level and is considering matching to CTCs, when available. IMS is working on implementing the algorithm at the Iowa registry and Linda is speaking with staff at the Detroit and Connecticut registries about implementing the algorithm. IMS plans to use this algorithm as the default. The Kentucky, Idaho, Massachusetts, and California registries also will use and test this algorithm to determine if IMS needs to make changes to meet their specific needs. These registries should carefully examine their data to determine whether they want to use the algorithm to link pathology reports to CTCs. IMS will work with individual registries to automate this process. Linda has not yet contacted all registries regarding the algorithm.

IMS will continue to evaluate the algorithm at the various registries to determine appropriate levels of standardization and customization of the ePath workflow. Linda would like to develop a plan for keeping customized workflows current.

Discussion

Participants asked about SEER*DMS autocoding capabilities for distinguishing between reportable and nonreportable cases. The autocoding has not been implemented yet but will focus on coding physicians and facilities. The autocoding feature might be used to code other variables, such as site or histology, in the future.

With the new algorithm, the record is matched to a CTC when the date that the specimen was collected matches the date of diagnosis or surgery. Linda clarified that the algorithm notes case reportability when linking to a CTC.

ePath Metrics Project **Marina Matatova**

NCI/IMS began work on ePath metrics that could be implemented before pathology reports are uploaded to SEER*DMS. Once the CCAB finalizes the metrics, IMS will develop a brief glossary of terms to help identify reportable and non-reportable cases and define an image, structured, and unstructured report.

The discussion of ePath metrics began at the September 2018 Face-to-Face meeting. At that meeting, participants expressed interest in continued conversations regarding the ePath metrics. At the recent SEER Managers meeting, Linda and Serban presented updated metrics and proposed approaches for collecting pathology information. NCI plans to conduct a review of registry ePath activities regularly (either yearly or quarterly).

The original plan was to develop similar ePath metrics for all registries, but registry feedback showed substantial logistical differences between registries due to factors such as state policy. NCI asked registries to complete an ePath pathology landscape survey in 2018.

The first metric focuses on pathology report coverage. The question asks if the registry had at least one pathology report with a specimen date within 60 days of the CTC date of diagnosis that was linked to a SEER reportable, microscopically confirmed CTC. Linda presented data on the percentage of SEER-reportable CTCs with a pathology report by registry for diagnosis years 2016 and 2017.

The second metric examines the proportion of pathology reports received more than 12 months after the date of diagnosis by cancer site and year of diagnosis. The question asks how many SEER reportable pathology reports are received after first-course treatment. Few registries had at least one pathology report with a specimen date at least 1 year after the date of diagnosis.

The third metric examines the types of pathology reports linked to CTCs (structured, unstructured, or image pathology reports) by year of diagnosis. Structured is the preferred type of report, which is received with a coded site, histology, and behavior. Currently, a small proportion of pathology reports received by registries are structured. The typical HL7 pathology report is unstructured and must undergo pathology screening to be coded. Image pathology reports are scanned documents. One purpose of this metric is to identify the proportion of CTCs for which a pathology report never will be received. Cases that will never have pathology reports can be excluded from certain studies. Linda presented all possibilities with regard to pathology reports, including no pathology report and paper versus electronic reports along with report type. She recommended tracking the number and type of pathology reports linked to CTCs over time.

A fourth metric examines the number of CTCs with no pathology report. Linda presented the proportion of CTCs indicating the cancer was microscopically confirmed with and without pathology reports. Registries might want to examine CTCs that indicate microscopically confirmed disease but that have no pathology report. Data suggest that, depending on the registry, up to nine percent of cases will have no information from pathology reports. Participants noted that facilities that do not submit pathology reports electronically but have a cancer registry will send the cancer registry rather than the pathology report. An abstractor also could have had access to the pathology report when they were coding the abstract without the pathology report existing in SEER*DMS. In general, if a facility does not use ePath, its pathology reports would not be in SEER*DMS.

A fifth metric examines pathology reports rather than CTCs and the types of pathology report (e.g., paper vs. electronic, structured vs. unstructured). At most registries, the great majority of pathology reports are electronic, but the proportion varies widely. It might be useful to identify cases that were diagnostically confirmed using a pathology report.

A sixth metric examines the number of months between the receipt of the pathology report and the receipt of the first abstract for a case (based on dates loaded in SEER*DMS). The purpose of this metric is to examine timeframes for receiving pathology reports compared to abstracts. The time lapse between the receipt of these two items could affect the usefulness of SEER data for some studies.

A seventh metric examines how many pathology reports are linked to CTCs. Linda presented the results of this metric for each registry for diagnosis years 2015 to 2017. Serban added that some registries have policies and procedures that increase the proportion of pathology reports not linked to CTCs. NCI and IMS will want to discuss these policies/practices with those registries.

The final metric examines changes in the proportion of linked pathology reports from year to year.

Next steps will include calls between NCI, IMS, and staff at each registry to review the registry's metric results and discuss possible reasons for those results. NCI wants registry-specific use cases.

Linda concluded that the effort to develop metrics relates to other SEER efforts, including efforts to standardize the workflow, improve usability, and increase capacity to link to reports and eliminate duplication. The Patient Set Editor has not been updated since 2005 but is expected to be improved through the usability project. The goals for the next few years will be to standardize and declutter SEER*DMS.

Discussion

Seattle continues to link pathology reports at the patient level. If the registry begins to link pathology reports to CTCs, the linkage rate might be higher because of the registry's large number of pathology-only records. CTC-pathology report linkages with less visual clutter would facilitate linkages at the Seattle registry.

Participants wondered why some cases that are not microscopically confirmed had pathology reports. These likely represent errors that could be corrected. An inconclusive or non-diagnostic pathology report could be linked to a CTC for which the cancer was diagnosed clinically.

In response to a question, Linda indicated that the third metric does not consider whether a pathology report is reportable. Reportability depends on the registry linkage rules. The purpose of this metric is to identify CTCs that might require further review.

Participants suggested considering the source of the reports. At some registries, the state might require pathology reporting, but hospitals run by the Department of Veteran's Affairs do not have to comply with this requirement. In addition, out-of-state data exchange files would not necessarily include pathology reports.

Linda clarified that all numbers she presented include both paper and electronic pathology reports unless the number/proportion of paper and electronic reports was measured specifically.

Social Security Administration (SSA) Linkage

New processes will be used for the SSA linkage in 2019 with the hope that data files will be returned to the registries more quickly. This year, IMS created a single file with masked identifiers for registries and will need a few days to divide up the files for each registry and return them to the registries. Linda encouraged registries to ask questions during the call if they were new to SEER*DMS and will be performing the SSA linkage for the first time.

Discussion

Two registries that are using SEER*DMS for the first time will be comparing followup rates before and after the SSA linkage and will report their findings to the CCAB. Other registry participants noted that the SSA linkage has been their top source of followup data.

One registry is waiting until processing of SSA results is complete before submitting data to the NDI. The registry representative wanted to know if other registries were actively correcting erroneous Social Security numbers (SSNs) identified through the SSA linkage. The Georgia registry corrects SSNs based on the linkage. Linda explained that the SSA will return files indicating whether the record matches a record in their file with the date of death if the patient is deceased. The SSA no longer provides specific reasons for nonmatches. Some registries used the linkage for follow back when SSA provided more information on the reasons for nonlinkage. The SSA also does not provide a date of last contact for patients coded as alive. Some registries impute a missing date of last contact based on when they receive the SSA file for a living patient (usually 2-3 months after), but provide caveats that the patient might have died in the past 2 to 3 months. SSA data include most individuals who are working and who pay Social Security taxes.

Participants asked about the level of evidence that the SSA would be able to provide regarding an individual's vital status. They suggested posing this question to SSA contacts.