SEER*DMS Change Control Advisory Board (CCAB) Users Group
Teleconference
January 30, 2024
2:30 p.m. to 4:00 p.m. EST

Representatives from the NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 26 cancer registries participated in the SEER*DMS Users Group conference call on January 30, 2024. 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: Peggy Adamo, Kathy Cronin, Lois Dickie, Marina Matatova, Serban Negoita

IMS: Suzanne Adams, David Angelaszek, Gary Beverungen, Emily Carver, Linda Coyle, Christopher (Chris) Johnson, Chuck May, Nikki Schussler, Jenifer Stevens

SCG: Kathy Brown-Huamani, rapporteur

Action Items

- Linda agreed to follow-up with the Seattle registry to resolve data reporting issues.
- IMS plans to run preliminary pathology counts per the CLIA numbers and forward them to the registries the week of February 5 2024.
- IMS and NCI will prioritize a meeting with the Connecticut registry to discuss ePath reporting solutions.
- IMS will schedule a meeting with the Hawaii registry to discuss technical issues with ePath reporting.
- Registries participating in the comparison analysis should complete their work by March 1, 2024.
- Linda agreed check schedules to schedule the next CCAB meeting.

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

This meeting focused on an update on the Data Management Report (DMR), ePath reporting, and geocoding in SEER*DMS.

2024 SEER DMR (Squish #12982)

DMRs are submitted by SEER-funded registries. A Squish issue (#12982) was created to collect feedback from the registries about the Facility List Template. Marina emphasized that the purpose of this discussion was to address questions and provide clarification before the February 28, 2024 deadline.

The majority of the questions were related to the combination of data items in the DMR, and Marina reviewed the template along with the DMR from the New York registry. Columns A to G list the SEER*DMS facilities (e.g., hospitals) with path reports in either 2022 or 2023. Column H has the Clinical Laboratory Improvement Amendments (CLIA) number, to which Centers for Medicare & Medicaid Services (CMS) data are linked. Columns H through P is based on the sending laboratory.

IMS collects CLIA information from the MSH4 binary HL7, which is populated in the sending laboratory as well as a field in path reports in SEER*DMS. David, Linda, and Marina described the various DMR fields and workflow. Multiple facility fields exist on a path report. Similar to 2022, the facility ID (FAC-ID) fields in the record are populated in a manner such that they can be used to request an abstract and can be linked to the Abstract Facility Lead (AFL) workflow. Facility fields reported in a registry's 2022 DMR have been expanded to include CLIA information. IMS will provide the new template for collecting CLIA information to registries to review as well as a new version of the report with the counts by month. Column E, if not already filled in, should be left blank.

Discussion

The New York registry receives path reports from the Albany Medical Center. Because this facility also is a reference laboratory, the CLIA number and the facility name are confusing. The Albany Medical Center is where the specimens reside and where the examination occurred. This Center also sends samples to other laboratories, such as Genomic Health, Inc. for additional testing. The New York registry receives direct feeds from other laboratories, which also is confusing. April was expecting the CLIA number and the facility name to be located in the left columns of the spreadsheet.

The Albany Medical Center is considered the parent facility that sends specimens for testing in a separate laboratory. NCI anticipates that Column C (CLIA Laboratories) will include numerous underlying laboratories that provide data to the main facility in Column I (CLIA facility). The purpose of the list is to fully understand operational infrastructure at the registries. The NCI is increasing investments in different tools and technologies to help support registry operations and wants to ensure that this investment meets the needs of individual reporting facilities. April noted that laboratories listed on the spreadsheet are independent companies and not part of the Albany Medical Center system. Therefore, these laboratories would need to be included in the list.

NCI would like to know how reporting facilities such as Albany Medical Center access specimens. In New York, facilities are required to send a reference test and incorporate those results into the original report. April expressed concern that Albany Medical Center examines thousands of specimens annually, but does not have its name, address, and CLIA number shown in one row. Serban shared April's concern. He noted that the FAC ID and different CMS CLIA numbers are useful for the NCI to understand how to develop a solution for Albany Medical Center. Linda explained that the spreadsheet shows how the path

reports are coded in SEER*DMS; the default is "unknown." Serban added that placing the combination of external laboratories at the beginning of the table for the New York registry is not ideal.

Jovanka Harrison, also with the New York registry, noted the need for a mapping or a link between the CLIA number in the MSH4 that links on import and the FAC ID for a sending and receiving facility. Marina noted that the goal is to understand and clean up those mapping issues prior to the 2025 submissions. IMS will recheck the sending laboratory (Column E) in the New York registry's spreadsheet and make corrections, accordingly.

The Utah registry participant was unclear on the purpose of the list regarding Column C and multiple samples and path reports from a single laboratory. Marina explained that this information helps the NCI to determine the best solution for the reporting facilities. Serban added that, in 2024, NCI and IMS will focus on ensuring that the counts are accurate for Column I. Marina emphasized NCI's need to fully understand the data flow through the full system at the registries.

The Seattle registry received the spreadsheet, but typically processes their path reports outside of SEER*DMS. These reports are sent without the HL7 format and discussions sometimes are needed about what data to include. Linda agreed to discuss the needs of the Seattle registry in a separate call.

Next Steps

Registries can email remaining questions directly to IMS or continue to post in Squish issue #12982. IMS will run preliminary pathology counts based on CLIA numbers and forward to the registries in the coming week. Registries can begin to review and identify any issues. IMS planned to provide the final counts on February 15, 2024.

ePath Processing in SEER*DMS Updates

In terms of information technology (IT) support to registries and reporting facilities, some facilities have entered into their own contracts with Inspirata, Inc. At four registries, reporting facilities have agreed to send unfiltered path reports and IMS is working on the filtering in SEER*DMS. A significant number of facilities are using the Electronic Mapping, Reporting, and Coding (eMaRC Plus Lite) software.

Regarding report filtering, IMS has developed the first version of a simple, keyword-based reportability algorithm using the Seattle registry data. The next steps will be to deploy this algorithm (currently running in test mode) at a facility that has no other option. Any registry that could serve as a "use case" for this model can contact Linda. IMS has been testing the Reportability API deployed as part of the NCI-Department of Energy (DOE) Modeling Outcomes Using Surveillance Data and Scalable Artificial Intelligence for Cancer (MOSSAIC) project. Registries that are receiving unfiltered path reports are using the NCI-DOE Reportability API to prioritize reports for manual path screening.

In addition, the IMS team completed the development of SEER*Transfer and is providing analytical support to the NCI to monitor the impact of changes in path processing. In response to registries receiving a nonstandard HL7 format, IMS implemented lab-specific HL7 imports to allow these registries to receive unfiltered path reports: nine in the Louisiana, two in Utah, and three in New Mexico. The IMS team is participating in calls with eight registries and their facilities most affected by the Inspirata change: California, Georgia, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, and Utah. IMS planned to reach out to three other registries in February: Connecticut, Hawaii, and Seattle. Four registries (Arkansas, Detroit, Michigan, and Minnesota) were not affected by the change in the contract with Inspirata. Other registries IMS has not met with that are experiencing issues with ePath reporting can contact Linda to schedule a call. IMS also has identified critical priorities and is testing filtering in

SEER*DMS. The Kentucky registry is performing similar tests on their own systems. A high priority is testing the Reportability API for facilities.

SEER*Transfer System

SEER*Transfer is a new software package that was completed at the end of December 2023. IMS is starting to implement SEER*Transfer at the SEER*DMS registries. SEER*Transfer is a web-based interface for manual or automatic data transfers using a REST (representational state transfer) API. This secure system allows login access via the open internet, which reduces the need for IMS technical support for account maintenance. Data are never stored in this system. Phase 1 of SEER*Transfer is focused on the transmission of hospital or laboratory data from reporting facilities. The community feed Island used by Walgreens and CVS will continue. In future Phases, IMS anticipates adding the reportability algorithms, follow-back modules, and other functionalities to SEER*Transfer. This system will support all existing inputs, with the exception of mass changes, and will include file validation, parsing to split it into registry-specific files, and transfer of files to SEER*DMS.

Registries interested in using SEER*Transfer will need an amendment to the Interconnection Security Agreement (ISA) between IMS and the registry. Nikki will be reaching out to all the registries in the coming weeks with further details. Data use agreements will be established between the user/facility and IMS. The Idaho registry, which rapidly completed their ISA amendment, will be the first to implement SEER*Transfer in production. This registry allows hospital users to login to SEER*DMS to import files and database access in minimal. Those users will be moved to SEER*Transfer. Linda noted the priorities for onboarding other registries: any registry that needs SEER*Transfer prevent interruptions in E-path feeds or any registry that can rapidly move the amendment process.

Discussion

The Connecticut registry has a one facility composed of three hospitals that is continuing with Inspirata and E-Path. Another health system, consisting of seven hospitals is trying to create its own solution, but likely will need to sign a 1-year agreement with Inspirata. Several other facilities are urgently seeking an alternative, such as eMaRC Plus Lite. Several hospitals are uncertain on their next steps. Marina and Linda agreed that the Connecticut registry should be prioritized for a meeting with NCI and IMS.

The Alaska registry has no alternative to Inspirata. The feed is still in place, but without a contract and no concrete guidance on how to proceed. This registry is unclear as to whether to cease operation of the Inspirata server or continue to use the path reports that the server is providing.

The Connecticut registry still has the receiver for Inspirata, has a continuance that ends in mid-June, and is trying to execute a no-cost contract extension.

Linda asked whether any registry receiving Inspirata data that is going through a facility but is not coming through a TransMed server. The Iowa registry does not have the Inspirata receiver, is receiving data, and is testing this setup. The Hawaii registry has one facility that extended its Inspirata contract but has not follow-up about receiving that feed. Marina suggested a call with the NCI and IMS to further discuss any technical issues. The Alaska registry has 636 processed path reports and is unsure of the legal ramifications of using them. The Alaska registry representative, Garrett Zimpelman, agreed to contact IMS for further discussion. Because of case load issues, the Seattle registry will need to discuss with IMS, installing keyword algorithm at a facility. They will need to do pre-filtering to meet concerns at some facilities.

Status of Geocoding Analysis (Squish #12882) **IMS**

IMS has met with a few registries as they were conducting the comparison analysis and is willing to meet with others. Time has been set aside on Mondays at 4 p.m., EST for a 30-minute meeting with the Idaho registry. Any registry interested in attending these meetings can contact Linda. Results from the Idaho, Minnesota, and New Jersey registries were similar. Linda invited representatives of any of those registries that have completed their analysis to summarize their data.

The Minnesota registry geocoded the uncertain values with a Minnesota-specific geocoder and subjected them to Tableau filtering, producing countless decisions. Their results support Chris's findings with the Idaho data, with both Automated Geospatial Geocoding Interface Environment (AGGIE) Geocoder and Geocodio, finding 98 percent of records. The Minnesota registry thinks that Geocodio would be a viable alternative for future geocoding.

Chris noted that he examined the differences between urban and rural areas because Geocodio and AGGIE use different reference datasets. The Geocodio company confirmed that they can support registry-specific reference data and also has alias city name tables, a feature which has been built into their system. Geocodio is able to accommodate additions for a particular registry. Additional SAS code is available to registries if they chose to make the rural versus urban comparison.

Discussion

Inaf Tabassum (principal investigator, New York Department of Health) asked whether the purpose of this exercise is to determine which geocoder would be implemented within the SEER*DMS environment, especially since the New York registry has its own geocoder. Linda clarified that the purpose is to identify a viable alternative as an integrated geocoder and called attention to Squish # 12882 that provides the details for conducting the comparison analysis. April confirmed that she is performing the comparison analysis for the New York registry.

In response to a question from the Louisiana registry representative, Chris clarified that the intent is to populate the Excel file that is provided and evaluate and makes decisions based on the combinations of the Geocodio quality metrics as well as AGGIE quality metrics.

Linda and Chris discussed the idea of using a hybrid geocoding model, where high-quality matches (97%) from Geocodia are used for direct matching and lower-quality matches (2%) are sent to AGGIE for Batch processing. Chris also mentioned that upstream issues with spelling errors and street names were more common in the Idaho data and should be corrected using LexisNexis before geocoding. Although one registry expressed concern about continuity and quality of geocoding, Chris noted that the results have been similar in the analysis. Mariana asked about a timeline for completing the initial analysis and then convening a group meeting.

Linda would like to hear registries feedback about not having the AGGIE Geocoder, the process to adopt an alternative integration, and thoughts on Batch processing. Comments can be posted in Squish #12882.

The Texas registry representatives noted having comparable results on its data analysis and expressed the need for a geocoder. They think Geocodio would be a good fit but plan to run the urban versus rural comparison prior to making a decision. In addition, the representatives added that they previously performed geocoding by linking with a different group in their department, which created manual work for staff. Having an integrated model would be this registry's choice.

The Seattle and Massachusetts registries miss having the integrated geocoder. Linda clarified that IMS developed an export for creating the Batch file that is sent to AGGIE and that there is a system task to load the results. Problems identified by the Louisiana registry, regarding AGGIE Batch have been corrected and changes should have been implemented. The Illinois registry thinks geocoding is acceptable and mentioned the back-and-forth work required. When asked about decisions on uncertain values, Chris recommended a third-party arbiter for tiebreaker decisions, such as Bing or Google and suggested a sampling strategy for manual review of those records.

Emily is streamlining the SAS code for rural versus urban analysis and will post the code to Squish #12882. Linda assessed that registries would want a geocoding decision finalized by the end of April and the viable alternative made available. She asked about putting into production Geocodio for registries already decided. Marina suggested first meeting with the Lynne Penberthy (Associate Director, NCI's Surveillance Research Program [SRP]) and geocoding point staff to review the initial data analysis by the end of February.

From today's discussion, Linda anticipates having a decision about using Geocodio in registries ready for production by the end of February, while others can continue their analysis. Chris prefers a group meeting to review the assessments across registries, particularly the urban versus rural differences. Linda will check calendars and plan for a group meeting in early March. Mariana agreed with reviewing the initial analyses and confirmations from the registries in a group meeting and then having an internal SRP/NCI meeting.

Next Steps **CCAB**

The deadline for the DMRs was February 28. Registries aimed to complete their comparison analysis and conclusions by March 1. The next regular CCAB meeting is scheduled for May 30, 2024.