SEER*DMS Change Control Advisory Board (CCAB) Users Group
Teleconference Summary
October 31, 2019
3:00 to 5:00 p.m. EDT

Representatives from NCI, IMS, the Scientific Consulting Group, Inc. (SCG), and 20 cancer registries participated in the SEER*DMS Users Group conference call on October 31, 2019. Participants included:

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

NCI: Melissa Bruno, Marina Matatova, Serban Negoita, Donna Rivera, Winny Roshala

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

SCG: Kathy Brown-Huamani, rapporteur

Action Items

- Fawn Vigneau of the Detroit registry agreed to work with IMS to send a poll to the SEER registries to gauge the interest of registry researchers in the adding the Middle Eastern surname algorithm to SEER*DMS.
- Linda agreed to generate surname case counts for each registry for Fawn to distribute with the poll. These counts would give registry staff an idea of the algorithm's potential utility to them.
- The Minnesota registry representative agreed to send IMS information about the registry's plans for automating the pathology reporting process.
- NCI and IMS will consider registry procedures for handling pathology reports and propose ePath metrics that are likely to work for all SEER*DMS registries.
- Linda agreed to create a Squish issue where registries can report the feasibility of obtaining pathology report counts by transport mechanism.
- Serban agreed to investigate what kind of information on pathology reporting facilities is available from AIM and the National Program of Cancer Registries (NPCR), and then contact each registry to determine whether staff can access information about characteristics of pathology reporting facilities.
- IMS will create a Squish issue that registry staff can use to propose other ePath metrics that would be feasible and valuable to the SEER registries.
- Linda agreed to contact registries to discuss options for improving workflow.

Middle Eastern Surname Algorithm   **Kendra Schwartz and Julie Ruterbusch**

Dr. Kendra Schwartz of the Detroit registry discussed a Middle Eastern surname algorithm developed at Detroit and California registries. The algorithm has been tested, validated, and used in multiple studies. Kendra proposed including the algorithm in SEER*DMS to generate a Middle Eastern ethnicity field that would help investigators study cancer in this population. Dr. Julie Ruterbusch would like to poll other registries to gauge their interest in the Middle Eastern surname algorithm. Linda recommended that IMS implement the algorithm for all interested registries. Linda also recommended discussing this algorithm during a SEER Research call led by Dr. Kathy Cronin.

Discussion

Linda asked if all SEER*DMS users should be able to see the Middle Eastern surname algorithm logic or last names used in the algorithm, or whether users would only be able to view the algorithm result. The presenters suggested protecting the list of names used in the algorithm in a table in SEER*DMS that is not viewable to most users. The Middle Eastern surname field with the result of the algorithm would be viewable to SEER*DMS users.

In California, the Middle Eastern ethnicity indicator is part of the standard data dictionary. The California definition distinguishes between Arab Americans and immigrants from Middle Eastern countries as a whole. Participants discussed whether it was worthwhile to examine differences and overlap between Arab and other Middle Eastern and North African surnames.

ePath Metrics Brainstorming   **Melissa Bruno, Linda Coyle, and Serban Negoita**

NCI SEER has been examining how each SEER*DMS registry processes pathology reports and their ePath metrics. NCI now wants to understand pathology report metrics and processes across registries that do not use SEER*DMS. Participants discussed the feasibility and value of the different pathology metrics used by the registries. Linda asked registry participants to share information about metrics and processes they already are using to manage the pathology reports they receive. Three proposals were submitted for ePath metrics from registries and NCI.

The Connecticut registry initially examines counts of incoming pathology reports in AIM. Pathology reports are moved to the SEER*DMS autoloader on a weekly basis. Timing problems led to fewer pathology reports being identified by the registry's AIM ePath monitor than appeared in SEER*DMS. To resolve this issue, registry staff developed a program to examine what was being sent to the autoloader. The registry found that, because ePath reports are received in real time, reports received from a hospital after the batches are run will appear in the monitor but not in the imported file. The Connecticut registry continues to work on this timing problem. The registry also receives paper pathology reports. These reports are counted for each facility and compared with the number of NAACCR abstracts received from that facility. These numbers should be similar because the state of Connecticut mandates that all hospitals and private laboratories send the registry a pathology report for every histologically confirmed diagnosis. The metrics used by the Connecticut registry are volume of pathology reports by system (ePath monitor and the autoloader) and counts for each facility by report type (paper and HL7) compared with abstracts. A problem is that addenda make volumes appear higher than they are.

Fist Proposed Metrics

NCI proposed some new metrics and is seeking feedback on these metrics from the registries. The first set of metrics proposed by NCI include volume of reports and number of reports transmitted by vendor, pathway, and facility. Marina asked if other registries were using metrics similar to those used in

Connecticut. The Utah registry uses the same metrics to track ePath reports. The registry implemented these metrics because of transmission interruptions from certain reporting laboratories. A drop in the volume of reports would indicate a transmission interruption somewhere. Louisiana, Central California, and Minnesota also track pathology report volume by facility to identify problems in transmission. The Minnesota registry checks reports against abstracts to ensure the correct number of reportable and nonreportable cases are being transmitted. Registry staff perform manual processing but are working with Mayo Clinic to automate the counting of pathology reports. The Detroit registry tracks the volume of pathology reports transmitted daily to identify transmission problems quickly. The Kentucky registry also has a process for tracking facility reporting to identify changes in the number of pathology reports received. The Alaska registry is setting up its workflow and would like metrics to track the volume of pathology reports. Currently, the registry is attempting to resolve problems with its ePath monitoring system. At the Seattle registry, staff track pathology reporting daily so that they can contact a facility if a file is not received on the scheduled date. Most ePath reporting goes directly to the Washington state registry. This registry sends a subset of ePath files to the Seattle registry. The registry created a logging system that records the name of the file received, the number of reports in it, and the date and time stamp to facilitate follow back with the state registry and the CDC when potential transmission lapses are identified. The Seattle registry staff currently are investigating the cause of underreporting from one large laboratory. The Seattle metric is volume by pathway and facility.

Marina asked if any registries did not have a method for tracking changes in pathology reporting and if registries would be interested in metrics to assist them in this process. The New York registry periodically imports large batches of pathology reports into SEER*DMS but does not import them regularly. The registry representative indicated that metrics would be useful once the New York registry begins to regularly import pathology reports.

Participants discussed ways that ePath processes could be automated. Linda asked if registries received a log of filenames from reporting facilities to compare with the filenames in SEER*DMS. Registry participants indicated that they can easily obtain information on the number of files received from a reporting facility, but not the number of files sent from a facility.

Second Proposed Metrics

A second set of metrics proposed by NCI include reportability and filtering. Participants discussed the proportion of nonreportable pathology records filtered at various points in the reporting pathway. The Alaska registry representative indicated that his registry obtains counts of the number of HL7 reports received and the number of those reports that are reportable. The Process Summary notes the number of reports received and processed/not processed, the number transmitted (because they were considered reportable based on the algorithm) and not transmitted (because they were deemed nonreportable); as well as the number of warnings. The Report Type Summary notes the type of each report processed.

AIM should be able to identify reportable and nonreportable records, but some registry representatives noted that AIM does not filter out all nonreportable records. Serban asked about the ability of other applications used by the registries to transmit and count reportable pathology reports. Utah obtains these counts through Python batch groups, but does not use AIM EMC. The selection criteria used by the various applications that filter HL7 records might determine the number of nonreportable records received by registries.

The New York registry creates quality reports of incoming pathology records. The registry uses SEER*DMS to identify reportable and nonreportable records. Laboratories that recently began to report to the New York registry also have been trained to screen for reportability. The Kentucky registry receives the reports that are marked reportable through AIM and other systems. This registry, however,

3

also reviews the reports in-house for reportability, and then counts the final number reportable pathology cases. The Utah registry has a similar process. Serban recommended developing standard criteria for reportable pathology records based on the criteria used by registries. Linda recommended a flag indicating a reportable disease.

Ideally, the reportability of a pathology report should be determined before it goes to SEER*DMS. Seattle and California registries determine reportability of pathology reports prior to importing them into SEER*DMS. The Seattle registry has its own system for filtering and reportability checks, similar to AIM. Potentially reportable cases are identified prior to pathology screening. After pathology screening, some reports undergo manual review to confirm reportability. Linda recommended that registries distinguish between reportable and potentially reportable pathology reports. In California, AIM criteria for filtering pathology reports initially was liberal to ensure that no reportable cases were missed. Over time, California registries have worked with AIM to refine these criteria. As with the Seattle registry, California registries review reports received to confirm that they are reportable. As many as 45 percent of the pathology reports determined by AIM to be reportable are ultimately determined to be nonreportable in California. Eureka now includes a Natural Language Processing (NLP) algorithm for reportability to further identify nonreportable cases received from AIM, which might improve this percentage.

Participants asked if ePath was designed to filter for reportability before screening at the registry level. NCI and IMS are seeking ePath metrics to incorporate into SEER*DMS to screen pathology reports in a way that aligns with processes used by all registries. Marina noted three types of reportability screening: 1) pre-screening that occurs at the laboratories based on training conducted by registry staff or pre-screening at the registry, 2) screening that occurs at the registry using in-house software or performed manually, and 3) ePath screening within SEER*DMS to confirm reportability through checks of various fields. Manual screening might be used either to screen incoming messages from pathology laboratories, or to screen PDF reports after they have been received by the registry. At some registries, pathology reports that are not linked to a patient set in SEER*DMS are reviewed manually to determine whether they represent a new primary tumor. Seattle has remote access to most pathology facilities reporting in their area, which allows the registry to perform a brief chart review to confirm reportability. The Connecticut registry performs manual screening of paper reports before they are scanned and loaded in SEER*DMS to prevent the inclusion of non-reportable cases. The Louisiana registry staff engage in the same process, but the registry receives few paper reports. The New Jersey, New Mexico, and New York registries have the same process. The Minnesota registry receives paper reports but reporting facilities are responsible for ensuring that only reportable cases are sent to the registry. The registry also receives PDF versions of pathology reports on a CD from one large dermatology facility. These reports are first imported into SEER*DMS and then reviewed, and the non-reportable cases deleted. The Minnesota registry is working on implementing ePath reporting at this facility, which sends approximately 80,000 reports each year.

Serban asked about the feasibility of generating counts of pathology reports before they are loaded into SEER*DMS by reporting pathway (e.g., AIM, other reporting systems, paper, etc.) and by initial reportability and nonreportability. Some registries, such as Louisiana, New Jersey, Detroit, Georgia, Minnesota, and New York do not perform any processing of electronic pathology reports before they enter SEER*DMS. In New York, however, facilities perform a prescreening process that eliminates many nonreportable cases because, by law, they cannot send cases that are deemed nonreportable to the state registry. Laboratories in Utah and New Mexico also screen pathology reports for reportability prior to sending them to the registry. These registries would not have information on how many pathology reports were withheld by the reporting laboratory because they were determined to be nonreportable. AIM facilities can provide counts of reports that they determined to reportable and nonreportable; but this information cannot always be obtained from independent laboratories. The Idaho registry is continuing to refine its pathology report processing procedures. The registry still reviews a substantial number of paper

reports for reportability. Seattle and California registries screen for reportability prior to loading pathology reports into SEER*DMS.

Third Proposed Set of Metrics

The third proposal for ePath metrics was a count by the transmitting laboratory. NCI also would like to collect demographic information on reporting laboratories to determine the types of laboratories that have good ePath coverage and the support needed to improve ePath reporting for other types of laboratories. Most registries collect some demographic information about reporting laboratories. Laboratories can be classified based on the number of services they provide; setting (e.g., academic, hospital, commercial, small pathology group); and Lab Information System (LIS). NCI will investigate the types of demographic information collected by the registries to identify some standard characteristics that describe pathology laboratories reporting to registries.

NCI also would like to examine counts by transport mechanism. The Louisiana registry staff can determine what reports are transmitted through each transport mechanism based on information in the messages received. The registry representative recommended counting the number of HL7 messages with filenames beginning with TMG (for AIM). Most registries only receive pathology reports through no more than two transport mechanisms. SEER*DMS now stores messages in a way that allows them to be queried, which should facilitate counting by filename prefix.

The New York registry staff would not be able determine the transport mechanism used by each reporting laboratory, because some laboratories used multiple transport mechanisms. Other registries, however, can obtain counts by transport mechanism.

Discussion

The first few characters of a pathology report indicate the reporting facility and sometimes the type of facility. The Minnesota registry performs prescreening based on these characters. The New York registry collects information on services, setting, and LIS for reporting laboratories, but the number of vendors is limited (about 6-7). The state of New York conducts a survey of reporting laboratories.

The Utah registry has difficulty collecting information about reporting facilities other than large hospitals. The Seattle registry anticipates similar difficulties. In California, a survey of reporting laboratories revealed the difficulty of obtaining information about these laboratories.

National associations have information about large laboratories and some groups of independent laboratories. AIM should have information about LIS vendors. The NPCR works directly with pathology laboratories and likely has some information on many smaller independent national laboratories. A first step would be for NCI to work with NPCR and AIM to determine what information could be collected on reporting laboratories. The next step would be to determine what information each registry has on its pathology reporting facilities, and what information can be shared with NCI.

Participants agreed that the types of pathology laboratories reporting to registries are worth exploring. For example, Physician Office-Based Laboratories (POLs) is a type of facility that reports to some registries but information about these laboratories is not always stored in a readily accessible system.

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

- IMS is working on the NAACCR duplicate algorithm, making updates to National Death Index processing, and improving the HL7 workflow.

- A Workflow presentation will be added to the agenda for the January SEER*DMS CCAB call.
- The next CCAB meeting is scheduled for January 2020.