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

Representatives from NCI, IMS, NAACCR, the Centers for Disease Control and Prevention (CDC), the Scientific Consulting Group, Inc. (SCG), and 23 cancer registries participated in the SEER*DMS Users Group conference call on May 30, 2024. Participants included:

REGISTRIES:

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

NCI: Marina Matatova, Serban Negoita

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

CDC: Sanjeev Baral, Vicki Bernard, Sandy Jones

NAACCR: Stephanie Hill

SCG: Carolyn Fisher, rapporteur

Action Items

- Sandy (CDC) agreed to contact the Minnesota registry to discuss their examples of synoptic report issues.
- Registries should decide whether to review both reportable and unreported messages that have unclear site information and contact the CDC.
- Stephanie (NAACCR) agreed to provide the New York and Texas registries with the names of facilities that have established relationships with the Ped SSDI Work Group (WG).
- Suzanne agreed to distribute the presentation slides for this meeting to the attendees.

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

The agenda included an update on CDC's National Program of Cancer Registries (NPCR) Electronic Mapping, Reporting, and Coding (eMaRC)-Light version (eMaRC Plus Lite) software and NCI's Pediatric Data Collection System (PDCS).

Marina noted that some registries already are working through implementation of eMaRC Plus Lite and that NCI has been meeting with CDC to better understand the application and the software. They provided an overview of eMaRC Plus Lite, addressed questions, and highlighted use in the registries. The PDCS and pediatric staging also was discussed.

Demonstration of eMaRC Plus Lite **Sandy Jones**

Sandy Jones, Lead Public Health Advisor, Cancer Surveillance Branch (CSB), Division of Cancer Prevention and Control, CDC, explained that eMaRC Plus Lite is an easy to use tool developed for laboratories to filter their narrative reports as a first pass for reporting to the central registries. The goal of this tool is to identify reportable and non-reportable cases from narrative reports. Hospitals can use the software to filter laboratory reports they receive. The reportability feature of the eMaRC software has existed for more than 20 years. Sanjeev Baral, Technical Lead, Bitwise Apps & Analytics Corp, CDC contractor, developed the software and made it easy to train.

Two versions of eMaRC Plus Lite are in operation. Version 1 (v1) is for pathology laboratories. Registries can select cases to test the software and post to the API launched upon installation. JSON code is created once eMaRC Plus Lite is applied to the data. The JSON response can be stored into the laboratory system to identify cases to send to the central registry. Both central and hospital registries can use this version. The v2 is designed for the cancer registries to repost nonreportable cases. Sandy reviewed and demonstrated the model parameter terms for Histology/Reportability, Primary Site, Brain Site, Mask, Pre-negation, Post-negation, Cytology Report, and natural language processing (NLP) for several cancer types, associated files, and messages. Sandy also reviewed examples of auto-coded cases, noting that this is an added benefit.

The CDC-NPCR default model is included in the tool as a start. The ultimate goal is to have a core and expanded terms list to accommodate laboratories reporting to more than one state and those state registries that collect additional data. The default model does not allow changes. Registries will need to use their own data to create their registry-specific model. Sandy demonstrated the creation of such a model and where it links to the NPCR. The search terms have been built over the years and the CSB has been improving this list with the eMaRC Plus Lite release, specifically for skin melanoma. An effective start and end year as well as site histology criteria are necessary to run this model. The algorithm for primary site terms was recently updated to manage skin cases slightly differently. Previously, a skin histology table and a skin nonreportable sites table were used, but were causing conflicts in reportability. The CDC found it easier to manage these tables together and is testing this new approach.

The eMaRC Plus Lite software only processes HL7 v2 messages. Laboratories not using the recommended Logical Observation Identifiers, Names, and Codes (LOINC) codes when setting up their reporting for pathology causes the information to be entered into a text diagnosis field for analysis. This tool will be challenged to exclude the clinical history and structured data and improves the reportability filtering. eMaRC Plus Lite can create a map based on a term a laboratory uses for a final diagnosis or a LONIC code. The pre and post-negation terms are flexible and can be changed.

eMaRC Plus Lite can export the entire table to an Excel spreadsheet. Users can make edits and imports to the default model. With the SEER model, users can import data, but will receive a warning that all data in

the table will be replaced. Users can process batches and batch size can be defined. The reports are received and entered into a source folder and the JSON code creates a reportables folder and a non-reportables folder. eMaRC Plus Lite remembers the settings when messages move back and forth. It exports HL7 messages and a NAACCR XML.

Sandy posed a question to the participants about their registry preference for reviewing both reportable and nonreportable messages with unclear information on the ill-defined sites, such as abdomen or prostate. Several states have boiler plate-like paragraphs on their path reports that can affect the NLP and its ability to flag a case as reportable or nonreportable. If eMaRC Plus Lite is unable to determine the reportable status, it will err on the side of caution by flagging a case as reportable. Models can be shared across registries and best practices leveraged. Several facilities already are using the eMaRC Plus Lite API, including the Cleveland Clinic, MD Anderson Cancer Center, and Mayo Clinic. The Hartford Health System is planning to use this tool.

Discussion

Amy Casey from the Minnesota registry noted that cases without a final diagnosis include a synoptic report but labels are negated. Sandy explained that the NLP options can mask the Observation/Results segment (OBX).5.1 text. Sanjeev added that the software can be set to ignore an OBX segment that starts with a specific word such as synoptic. The application will not read a synoptic report. Amy expressed concern that that a negated term, which is found near the diagnosis, would be deleted. Sandy requested an example of the Minnesota registry's file for review, noting that this registry is adding the information via a text file not via a code. Sandy requested input on addressing ill-defined sites from all registries.

When building the pipelines for registries, Mariana asked whether transfer is supported in eMaRC Plus Lite. Sandy highlighted that the Association of Public Health Laboratories (APHL) is moving towards hospitals reporting their data across the APHL Informatics Messaging Services (AIMS) platform. This approach aligns with the infectious disease data reporting and APHL is planning to add a cancer feed at little or no cost to the hospitals. Marina noted that NCI and IMS have invested time in building SEER*Transfer and are anticipating aligning it with existing tools to bring data into the IMS enclaves.

Sandy asked whether SEER*Transfer will restructure the data according to the NAACCR 5.5. Marina noted all forms are accepted and that transformation occurs in the IMS enclaves. Linda added that formatting will occur in SEER*DMS and clarified that SEER*Transfer is a secure transfer mechanism. Sandy explained that the CDC has not addressed restructuring and making compliant HL7 v2 messages to NAACCR 5.5. She discussed future plans, including expanding the core terms which is expected to be a surveillance community activity.

Registry Discussion Related to eMaRC Plus Lite **Kevin Ward**

Kevin described a project to validate eMaRC Plus Lite using the Georgia Cancer Registry (GCR) data. The registry processed a subset of the 2021 reports for reportable cases that have used the Artificial Intelligence in Medicine, Inc. (AIM) E-Path tool from a single hospital for a single month, process reports from all hospitals in single month, and process all reports from all hospitals excluding free standing laboratories over 12 months. The results were reviewed and the model refined after each step.

The GCR processed their reports in batch format, which involved adding all pathology reports to a single input directory, setting up and running the model, and selecting cases as reportable or nonreportable. The results were shared with the CDC NPCR teams to update the base model. The GCR processed 201,597 Inspirata pathology reports for 2021 cases and 184,473 (92%) were reportable, 17,124 (8%) nonreportable. Approximately 20 percent of the reportable GCR cases historically have been false

positives. Of the 17,124 nonreportable cases, 97 percent were confirmed, which will save registry staff time reviewing these cases. The remaining 571 nonreportable cases that were classified as reportable by the GCR were likely were missed cases and were reduced to 123 post manual review. Some cases were misclassified as auditable (clinical history only findings). The next steps for the GCR will be to determine, in one facility, the number of additional false positive and true positive reports that eMaRC Plus Lite will identify that were not sent through the E-Path tool. The GCR will reprocess all reports using the updated CDC model.

Kevin encouraged other registries to perform this validation of their AIM reports using the updated CDC model. A major advantage is that the eMaRC Plus Lite tool is available at no cost to the user.

Marina noted that registries can email any followup questions or submit a Squish to IMS about eMaRC Plus Lite. Sandy expressed appreciation to NCI for the invitation and to Suzanne Schwartz and the New Jersey registry for contributing to this project.

PDCS: Requirements and Modifications **Serban Negoita, Stephanie Hill**

Overview

The PDCS is a system developed by NCI and IMS to support the implementation of Toronto Conceptual Principles and Guidelines for Pediatric Cancer Stage in Population-Based Cancer Registries (PBCRs). These guidelines have been endorsed by the International Association of Cancer Registries and used in the International Benchmarking of Childhood Cancer Survival by Stage Project (BENCHISTA), which examines survival by stage at diagnosis for childhood cancers. BENCHISTA has been implemented in multiple PBCRs in Europe, Australia, Japan, and other countries that collect the data elements necessary to calculate or recalculate the stage according to the Toronto guidelines. NCI is proposing the PDCS system for the U.S. cancer registries. The NAACCR Pediatric Site-Specific Data Items Work Group (Ped SSDI WG), NCI, and IMS staff have supported this system.

The PDCS (or Toronto Childhood Cancer Staging System) is a Union for International Cancer Control (UICC) TNM-based system, consisting of 3 core data items (Pediatric Tumor, Pediatric Regional Nodes, and Pediatric Mets); 9 new prognostic factors; 4 revised prognostic factors; 8 derived items; and 33 schemas. The PDCS has been sponsored by NCI SEER program to add a new data item to the NAACCR standards recommended by the Mid-Level Tactical Group (MLTG) for implementation with NAACCR v24 and reconfirmed for v25; and approved for implementation by the High-Level Strategic Group (HLSG) with NAACCR v25. The PDCS will be included in the NAACCR Data Standards and Data Dictionary and will be listed in the Required Status (RS*) Table for the SEER program. Early adopters among vendors and registries already have implemented the PDCS in 2024, particularly the Kentucky and Seattle cancer registries and the Louisiana Tumor Registry.

Eligibility of data items for PDCS will be based on age. Although the National Childhood Cancer Registry (NCCR) covers cases diagnosed at ages 0 to 39 for certain schemas, the MLTG/HLSG recommended and approved new data items for cases diagnosed at ages 0 to 19, which SEER will implement. NCI is considered the best type of facility to abstract for PDCS and is not necessarily the diagnosing facility. A tertiary facility that provides the first course of treatment (FCOT) is being considered. Principal investigators (PIs) can use their discretion regarding age-based eligibility and a facility for abstracting in the PDCS.

Serban briefly reviewed the PDCS 2024–2025 implementation plan for SEER*DMS composed of two main steps. Test using the SEER API and dynamic link libraries (DLLs) not the AJCC DLLs. Access data collections, including the feasibility by facility type (e.g., caseloads, services), registrars' training in the

PDCS code, and ability to increase awareness about the Toronto Childhood Cancer Staging System. In addition, the PIs of core SEER*DMS registries will be expected to select facilities for the implementation; exceptions can be approved by the SEER Program Coordinator, Mr. Steve Friedman. Each registry is expected to select at least two facilities to participate, and report the ID to IMS. Registries are encouraged to consider additional facilities, especially those that routinely provide FCOT. Serban clarified that RS* is referring to required size-specific (when available) and that site-specific is referring to the 33 site-specific schemas available in the PDCS. He emphasized that the pediatric schemas in the PDCS are different from the Extent of Disease (EOD) or AJCC schemas or chapters and reviewed examples of the RS* requirements. Further details can be found on the SEER Registrar Staging Assistant (SEER*RSA) website.

Implementation Strategy: Plans

Stephanie Hill, Associate Director, NAACCR, detailed the PDCS implementation plans and acknowledged the PDCS Implementation Plan team: NCI SEER, Jennifer Ruhl and Serban; IMS, Nikki Schussler; and NAACCR, Stephanie. She noted that NAACCR also is coordinating data elements for the NCCR. The New PDCS API will be stand alone, was designed to avoid dual coding by registries, and will not include AJCC DLLs, which are captured by the current SEER API. Stephanie highlighted the anticipated milestones by diagnosis year:

- 2023–2024—Beta tested in the Kentucky Cancer Registry (KCR)
- 2024—Pilot testing in select hospitals in SEER states
- 2025—Limited implementation in all core SEER registries
- 2026—Full implementation in all core SEER registries
- TBD—Implementation in other NCCR registries

With the 2023–2024 beta test, the KCR reported that implementation of the PDCS API statewide (beginning with 2023 cases) was similar to a small-scale annual NAACCR change and easy to integrate into their cancer data management system, with minimal disruptions in the abstracting workflow of the two main, large facilities. Stephanie highlighted that the KCR provides the software for all of their hospitals.

The 2024 Pilot testing is in progress in three hospitals in the Louisiana and Seattle registries. The Implementation Team has engaged three hospital registry software vendors (Elekta, C/NET, and ONCOlog) and has provided these vendors a special data dictionary XML extension. NAACCR has developed a separate Edits Metafile and will provide targeted training to registries in the summer.

In the 2025 implementations, PDCS data elements have been approved and are included in the v25 NAACCR data dictionary. PIs will select two facilities. The Implementation Team will coordinate with central registries, vendors, and facilities. The plan follows the MTLG's implementation timeline and October 2024 is the deadline for any changes. Suggested criteria for facility selection include pediatric case workload; software vendors (or registry contacts) already participating in the pilot testing; and existing relationships in the states and their involvement with the Ped SSDI WG, particularly New York and Texas.

Stephanie next highlighted the specific implementation activities:

- July–August 2024—C/NET testing begins
- August 1, 2024—PIs notify SEER/NAACCR of selected facilities

- Mid-August 2024—NAACCR v25 vendor meeting (annual)
- Mid-September 2024—Kickoff meeting for selected hospital registries
- October 1, 2024—Notify hospital registry vendors
- November–December 2024—NAACCR training webinars, including pediatric facilities
- January 1–June 30, 2025—API included in hospital software updates

Products available to assist with implementations include XML Data Dictionary (by request, 2024 only), NAACCR v25 Data Standards and Dictionary, API/DLL, and Edits Metafile. Registries can access these tool from the PEDIATRIC Data website. In terms of training, the topic will be included in the discussion of the SEER coding workshop in September 2024, the webinars will be recorded, the topic will be included in the National Cancer Registrars Association (NCRA) 2025 Annual Meeting, and an Ask a SEER Registrar subject area specific to pediatric data collection will be established.

Discussion

Serban clarified that NCI will work with SEER research support registries interested selecting facilities for implementations.

The Utah registry representative asked whether the requirement is to select individual hospitals or healthcare organizations as facilities. Serban noted that the observation is that a number individual hospitals where pediatric patients receive FCOT are limited and include tertiary care facilities but noted that the selection is up to the registry to decide in this first phase of the implementations.

In response to requests, Stephanie will provide the New York and Texas registries the names of facilities in their areas who already have established relationships with the Ped SSDI WG.

Registries who want to know more about the PDCS in general can reach out to NCI (Serban or Steve), or for implementations, they can contact Stephanie or Nikki. Email or Squish can be used for the communications. Suzanne will distribute the presentation slides for this meeting to the attendees.

Next Steps **CCAB**

The next regular CCAB meeting is scheduled for December 9, 2024.