



Creating a Global Updates Script Library to Use for Data Quality Control: Gateway to Process Improvement for the Missouri Cancer Registry and Research Center (MCR-ARC)



Saba Yemane, BS, BA^{1,2}; Iris Zachary, PhD, MS, CTR^{1,2,3}; Jeannette Jackson-Thompson, MSPH, PhD^{1,2,3}; Nancy Rold, BA, CTR^{1,2}
1) University of Missouri (MU) School of Medicine, Dept. of Health Management and Informatics, Columbia, MO, United States
2) Missouri Cancer Registry and Research Center, Columbia, MO, United States
3) MU Informatics Institute, Columbia, MO, United States

Background

Central Cancer Registries (CCRs) are important for cancer reporting, cancer control, evaluation and research. However, only complete, accurate and timely data are valuable for these activities. MCR-ARC has a SQL Server 2008 database that contains more than one million records. All CCRs' data must pass standard edits and required national edits to be submitted for certification and meet program requirements. This ongoing process takes time and effort and can be more efficient if appropriate global changes are carefully applied to the database.

Purpose

To make the process of quality control more efficient with less effort through global updates.

Methods

As a general practice, we analyze the CDC National Program of Cancer Registries (NPCR) Advanced Edits Report to identify the most common errors. After running queries to retrieve data that triggers errors, MCR Operations staff work together to determine the fixes for the errors. The Database Administrator writes scripts to correct the errors. For the errors that are common, the scripts are run against the database on a regular basis. Over the years, we have developed a library of scripts to run against the database to clear edits on a regular basis.

Results

Running the NPCR advanced edits for the first time this year, we found a total of 100,466 errors in 605,284 records for diagnosis years 1996 to 2014. Five errors accounted for 4,749 to 41,312 counts (Table 1). Together, these five errors accounted for 98% of the total errors. Thirteen errors (not shown) accounted for 1 to 451 counts. Most of the smaller counts were cleaned manually by QA staff. Many of the larger counts were cleaned by writing and running scripts using global updates (see Table 2 for examples).

Table 1: Errors and Global Updates

| Error Message | Before Global Changes | After Global Changes |
|--|-----------------------|----------------------|
| ICD Revision Number, Cause of Death | 41,312 | 2,815 |
| Census Tract 2000, State, County, 2000-2009 * | 22,592 | 17,006 |
| ICD Revision, Vital Stat, Date of Last Contact | 20,228 | 0 |
| Follow-up Source Central, Date of DX | 9,555 | 0 |
| Census Tract 2010, State, County, 2010-2019 * | 4,749 | 823 |

* For these edits, the county code and census tract combination tables need to be reviewed. Cases with these errors were recoded in 2016 by the GIS Director, Office of Geospatial Information, State of Missouri, and came back with the same census tract codes.

Discussion/Conclusions

By using global updates, we were able to clear about 77,792 error counts. Through the use of standardized global update scripts, we are able to make the process of data cleaning and quality control more efficient and, therefore, more cost effective. This script library is becoming more comprehensive over time as we add additional scripts for data quality control.

Script Samples

Table 2: Global Update Script Library

| Partial List of Scripts | |
|-------------------------|---|
| 1 | Update Follow-up Information from Death Certificate |
| 2 | Update Information from BCCCP Linkage |
| 3 | Update Census and Geocoding Information |
| 4 | Update Database to Clear Various NAACCR and NPCR Core Edits |
| 5 | Update Database to Clear NPCR Advanced Edits |
| 6 | Update Database Information from NDI Linkage |
| 7 | Scripts to Generate Data Requests |