2022 NPCR NEW HAMPSHIRE SUCCESS STORY

New Hampshire State Cancer Registry; Judy Rees and Adriane Burke

A Quality Improvement Proof of Concept Exercise Using Pseudonymization to Identify Duplicate Cancer Cases Across State Boundaries

National Program of Cancer Registries
SUCCESS STORY

SUMMARY

This collaborative, four-state pilot quality improvement project sought to identify duplicate cancer cases across state boundaries - Maine: New Hampshire, Rhode Island, and Vermont. Using a publicly available tool to encrypt patient identifiers[1], linking cases across state boundaries, we identified and reconciled possible duplicate cases among neighboring states.

CHALLENGE

Cancer treatment may involve travel across state lines. This movement can create duplicate cancer reporting in two or more states. Duplicate cases could distort our understanding of the burden of cancer in a state, region, or even nationally. A project reported previously by Chris Johnson[2] assessed duplicates in Idaho and Utah and found that approximately 0.2% of cases had also been registered in the neighboring state. Mr. Johnson concluded that up to 1.6% of Idaho's Death Clearance Only (DCO) cases could disappear with an effort to identify cancers registered in other states.

While data exchange agreements allow states to send data to a patient's home state, there's no easy way for states to exchange patient identifiers unless there is evidence that a case "belongs" to the other state. This pilot quality improvement project used a publicly available tool to encrypt patient identifiers and then link cases across state boundaries to identify and reconcile possible duplicate cases among a group of neighboring states.

It is important to understand how many potential duplicates exist, and for registry data quality improvement, remove duplicate reports. In this quality assurance initiative, we tested whether pseudonymization techniques can help us de-duplicate cases between states to improve cancer surveillance. This project was carried out using pre-XML data, 1995-2018 from four participating state's registry. We used a publicly available, standalone, open-source tool designed to replace identifiable fields with a pseudonymized file, which could then be linked with neighboring states data to identify duplicates.

SOLUTION

The goal of this multi-state pilot was to identify duplicate cancers across state boundaries using pseudonymization to encrypt patient identifiers. We believe the study design will be of interest to other state cancer registries seeking a way to examine and reconcile duplicate cancer cases in neighboring states.

The advantages of this proposal are that a) it is easy to implement, b) no personal identifiers are exchanged, and c) the required software is available and free. The disadvantages are that a) the linkage method is deterministic - meaning that there is no mechanism to assess uncertainty when it is unclear if 2 cases from 2 states are or are not the same person, and b) the assumption is made that the data are clean.

Challenges in the pilot, included the time needed to create the files, link the files and discuss the matched cases. In the future, a more complex series of deterministic linkages could be developed to increase the accuracy of matching.

RESULTS

New Hampshire received the encrypted file from each state as the broker. Linking the digests between the states resulted in the results displayed in Table 1. The results were returned to sending states.

TABLE 1. RESULTS OF THE HASHED STATE-TO-STATE LINKAGES: COUNTS OF TOTAL AND POTENTIAL DUPLICATERECORDS BY STATE (DIAGNOSIS YEARS: 1995-2018). PERCENTAGES REPRESENT THE PROPORTION OF A STATE'S TOTAL CASES THAT ARE POTENTIAL DUPLICATES.

State	Total Cases	Link to NH	Link to ME	Link to RI	Link to VT
New Hampshire (NH)	200,753		531 (0.26%)	90 (0.04%)	190 (0.09%)
Maine (ME)	208,401	531 (0.25%)		82 (0.04%)	78 (0.04%)
Rhode Island (RI)	160,830	90 (0.06%)	82 (0.05%)		17 (0.01%)
Vermont (VT)	91,411	190 (0.21%)	78 (0.09%)	17 (0.02%)	

The reconciliation process was very labor-intensive, requiring two registrars to independently review the case in their own registries. We learned a few important lessons. First, there are duplicate cases for the same primary as patients move between states. Second, reconciliation requires the registrars to open the cases on their respective computers to see all the elements and discuss the issues. Third, reconciliation is very time intensive.

SUSTAINING SUCCESS

Through this quality improvement proof-of-concept exercise, we demonstrated the value of using pseudonymization to identify duplicate cancer cases reported to more than one state andto resolve a proportion of DCO cases. We've created protocols and provided the SAS[2] code used in this process on our website (https://geiselmed.dartmouth.edu/nhscr/projects/). Our protocol provides an affordable and accessible method to estimate the number of duplicate cases between two states and to examine and reconcile the duplicates. This exercise suggests it is worthwhile and the practice increases the confidence in the validity of the data. The major obstacle is the time commitment to do the reconciliation. The reconciliation process could become more efficient by development of a software tool to import a select set of variables that could be displayed to make the review faster. In future attempts, it may be possible to build a more sophisticated linkage algorithm using multiple pseudonymized digests of variable subsets, to increase the efficiency of the process.

At the end of the exercise, important questions arose to consider: how many duplicate reports exist between the states; what is the potential impact of those duplicates on regional and national rates; and how many Death Clearance Only (DCOs) could be eliminated through a multi-state effort to identify duplicates?

1. University of Nottingham UK. OpenPseudonymiser (OP) https://www.openpseudonymiser.org/. Accessed 07 05 2022
2. Johnson C, Herget KA, Dibble R, Carson S, Stroup AM. Assessment of Duplicate Cancer Cases in Utah and Idaho: Improving Interstate Cancer Surveillance NAACCR 2013 Conference, 2013.
3. SAS/ACCESS® 9.2 Interface to ADABAS: Reference. Cary NSII. SAS Institute Inc.; 2013.
4. Ansolabehere S, Hersh ED. ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name. Statistics and Public Policy 2017;4 (1) :1-10.

STORY QUOTE

"Where could this go in the future? Protocols and SAS codes are freely available on the NHSCR website. We invite other states to take up the challenge and try it out - improve on the process, improve on the linkage algorithm, and make the reconciliation process more efficient." – Judy Rees, NHSCR Director

REGISTRY CONTACT INFORMATION

603-653-6265
New Hampshire State Cancer Registry Website

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention