



Deep Dive into Data Quality - California's Approach



Cheryl Moody, BA, CTR; Marilyn Scocozza, CTR
California Cancer Reporting and Epidemiologic Surveillance (CalCARES) Program*
Institute for Population Health Improvement, UC Davis Health System

Background

The National Program of Cancer Registries (NPCR) routinely provides a data quality indicator (DQI) report which illustrates California's percentages in specific data fields as compared to the median percentage across all NPCR registries. California Cancer Registry (CCR) reviews the report to determine data fields that meet the following criteria:

- California's percent for a data field is higher than the NPCR Median Range
- California's percent for a data field is within the NPCR Median range, but has been increasing over time.

The most recent report evaluated cases diagnosed 2011-2015. SQL queries were run on two data fields identified meeting the above criteria.

- Spanish Origin
- Diagnostic Confirmation

Project Objective

Our poster will illustrate the results of our analysis focusing specifically on any patterns or trends identified; corrective actions implemented; and outline our plan for eliminating the root cause of any coding issues identified.

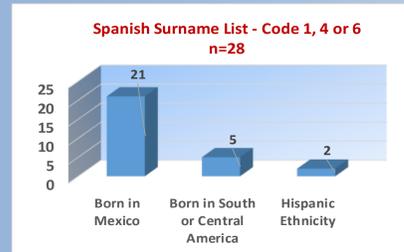
Spanish Origin

Methodology

California's data quality percent of cases coded to unknown for Spanish Origin has been rising over time and is now higher than the NPCR Median range.

- 2011 - CA: 1.74; NPCR Median Range 1.74
- 2015 - CA: 2.30; NPCR Median Range 1.88

The database was queried and identified 24,719 patients diagnosed between 2011 and 2015 where Spanish/Hispanic Origin was coded to 9 - unknown. A second query was run on the 24,719 patients to determine if any were on the Spanish Surname list. There were 1,359 (5.5%) identified that should have been coded to 7 at a minimum.



Further review of the 1,359 patients on the Spanish Surname list revealed that 28 had a known birthplace or Hispanic documentation in the text and should have been coded to 1, 4 or 6.

- 21 born in Mexico
- 5 born in South or Central America
- 2 with Hispanic text documentation

In addition, there were 171 patients with a known country of birth and were NOT on the Spanish Surname List. Of those, 5 patients were born in the Philippines.

Issues Identified

- Married female, no maiden name, born in Mexico, married last name NOT on Spanish Surname list and no text regarding ethnicity
 - SEER's RESPONSE: Code 1 (Mexican)
- Married female, no maiden name, born in the Philippines, married last name NOT on Spanish Surname list, text states "Hispanic"
 - SEER's RESPONSE: Text says Hispanic, code 6 (Spanish, NOS)
- Married Female, no maiden name, born in Peru, married last name IS on Spanish Surname list, no text regarding ethnicity in abstract
 - SEER's RESPONSE: No text regarding ethnicity but born in Peru – code to 4 (South American)
- Patient has two last names. One of the last names is on Spanish Surname list
 - SEER's RESPONSE: Code to 7 (Spanish Surname Only)

Note: As a result of these questions, SEER's Program Manual was revised to provide clarification on how to code these scenarios and now includes the above examples.

Recommendations

- Analyze feasibility of auto-checking last names against Spanish Surname list and defaulting those on the list to code 7 for Spanish Origin
- Create a Q-Tip educational reminder regarding Spanish Origin
- Update California's Abstractor Coding Manual with revisions that provide clarity for coding Spanish Origin
- Continue to evaluate California's data quality percentage for Spanish Origin coded to 9 on the NPCR DQI Report

Diagnostic Confirmation

Methodology

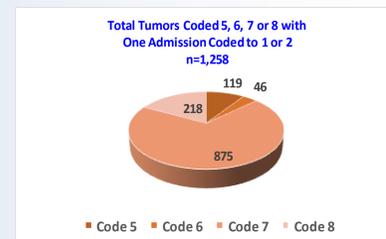
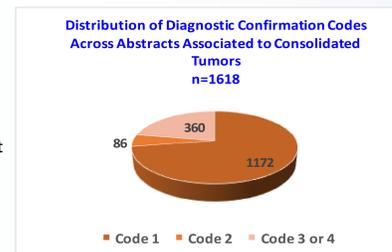
California's data quality percent of cases coded to Diagnostic Confirmation 5, 6, 7 or 8 - not microscopically confirmed, is higher than NPCR median range.

- 2011 - CA: 4.25; NPCR Median Range 3.85
- 2015 - CA: 4.26; NPCR Median Range 4.09

The database was queried and identified 1,618 consolidated tumors with at least one abstract coded to a lower number than the consolidated tumor's value.

1,258 (78%) of the tumors had at least one abstract coded to 1 or 2 and were reviewed for this presentation. 360 tumors had at least one abstract coded to 3 or 4 and were not reviewed.

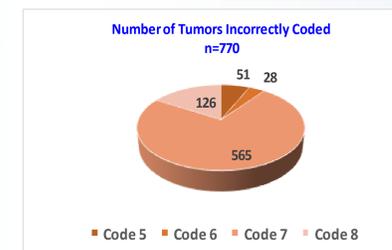
- 1,172 coded to 1
- 86 coded to 2
- 360 coded to 3 or 4



This chart shows the total number of tumors coded 5, 6, 7 or 8 (488 coded correctly and 770 coded incorrectly)

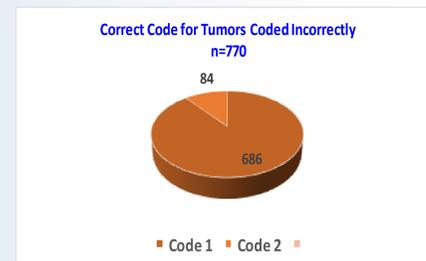
- 119 coded to 5
- 46 coded to 6
- 875 coded to 7
- 218 coded to 8

- 488 (39%) of the tumors were coded correctly
- 770 (61%) of the tumors were coded incorrectly at the consolidated tumor level, however, an associated abstract was correctly coded.
 - * 51 incorrectly coded to 5
 - * 28 incorrectly coded to 6
 - * 565 incorrectly coded to 7
 - * 126 incorrectly coded to 8



Of the 770 tumors coded incorrectly

- 686 (89%) of the tumors should have been coded to 1
- 84 (11%) of the tumors should have been coded to 2



Issues Identified



- Subsequent treatment of surgery or biopsy would change the 5, 6, 7 or 8 code to positive histology (code 1)
- Frequently, an associated admission was coded appropriately
- Information collected over the course of the disease was not used to revise diagnostic confirmation
 - It is not limited to first course treatment
 - This instruction appeared to be missed by regional staff performing consolidation tasks

Recommendations

- An auto-consolidation rule is being drafted to auto-check cases coded to 5, 6, 7 or 8 and if the case meets specific criteria, the diagnostic confirmation code will be revised.
- An educational Q-Tip or reminder will be distributed for regional staff to remind them that diagnostic confirmation should be revised appropriately based on information received over the course of the disease and is not limited to first course treatment.

Combined Results

The number of cases coded to unknown for Spanish Origin has risen from 1.74% to 2.30% over this time period and the number of cases coded to 5, 6, 7 or 8 for Diagnostic Confirmation is higher in California than the NPCR median for 2011-2015.



Central Registries perform multiple data query checks on a routine basis. Taking the time to evaluate reports provided from national standard setters and analyze registry data by taking a deep dive into the results allows the opportunity to identify coding discrepancies that may otherwise be overlooked during the course of routine edit error correction and/or visual editing. Implementing corrective actions and providing education and training on these issues will result in enhanced data quality.

Overall Conclusions

- Periodic analysis of the NPCR DQI Report should be performed
- Review the report for data fields where the state registry average is higher than the NPCR average
- Review the report for data fields within NPCR's average, however steadily increasing over time
- Use this opportunity of NPCE DQI Report review to identify areas for correction and/or education in order to improve the registry's data quality



*The CalCARES Program partners with the California Department of Public Health (CDPH) to manage the operations of the state mandated California Cancer Registry program

The following staff contributed to the design and presentation of this poster: Ghenadie Ciornni, BS, CPBI