Deep Dive into Data Quality - California’s Approach

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

Background
The National Program of Cancer Registrars (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
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.

- 2013 - CA: 2.14, NPCR Median Range 1.74
- 2015 - CA: 2.36, 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. The database was queried and identified 24,719 patients diagnosed between 2011 and 2015 where Spanish/Hispanic Origin was coded to 9.

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.

Recommendations
- An auto-consolidation rule is being drafted to automatically check cases coded to 5, 6, 7 or 8 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.

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.

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 NPCR DQI Report review to identify areas for correction and/or education in order to improve the registry’s data quality.

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 the NPCR median range.

- 2013 - CA: 4.21, NPCR Median Range 3.85
- 2015 - CA: 4.26, NPCR Median Range 4.09

The database was queried and identified 1,618 consolidat-ed 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
- 380 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

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.

Deep Dive into Data Quality - California’s Approach

General Information
- 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.
- 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
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.

- 2013 - CA: 2.14, NPCR Median Range 1.74
- 2015 - CA: 2.36, 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.

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 7 at a minimum.

In addition, there were 171 patients with a known country of birth and were NOT on the Spanish Surname list. Of the 171 patients, 5 patients were born in the Philippines. In addition, there were 171 patients with a known country of birth and were NOT on the Spanish Surname list. Of the 171 patients, 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.
- Married female, no maiden name, born in the Philippines, married last name NOT on Spanish Surname list, text states “Hispanic”.
- Married Male, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Married Female, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Patient has two last names. One of the last names is on Spanish Surname list.

Issues Identified
- Married female, no maiden name, born in Mexico, married last name NOT on Spanish Surname list and no text regarding ethnicity.
- Married female, no maiden name, born in the Philippines, married last name NOT on Spanish Surname list, text states “Hispanic”.
- Married male, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Married Female, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Patient has two last names. One of the last names is on Spanish Surname list.

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 Abstracter 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.

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.

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 NPCR DQI Report review to identify areas for correction and/or education in order to improve the registry’s data quality.

Deep Dive into Data Quality - California’s Approach

General Information
- 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.
- 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
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.

- 2013 - CA: 2.14, NPCR Median Range 1.74
- 2015 - CA: 2.36, 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.

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 7 at a minimum.

In addition, there were 171 patients with a known country of birth and were NOT on the Spanish Surname list. Of the 5 patients, 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.
- Married female, no maiden name, born in the Philippines, married last name NOT on Spanish Surname list, text states “Hispanic”.
- Married Male, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Married Female, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Patient has two last names. One of the last names is on Spanish Surname list.

Issues Identified
- Married female, no maiden name, born in Mexico, married last name NOT on Spanish Surname list and no text regarding ethnicity.
- Married female, no maiden name, born in the Philippines, married last name NOT on Spanish Surname list, text states “Hispanic”.
- Married Male, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Married Female, no maiden name, born in Peru, married last name is on Spanish Surname list, no text regarding ethnicity in abstract.
- Patient has two last names. One of the last names is on Spanish Surname list.

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 Abstracter 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.

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.

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 NPCR DQI Report review to identify areas for correction and/or education in order to improve the registry’s data quality.