Using Changes Made in the Visual Editing Process to Create More Targeted Quality Control

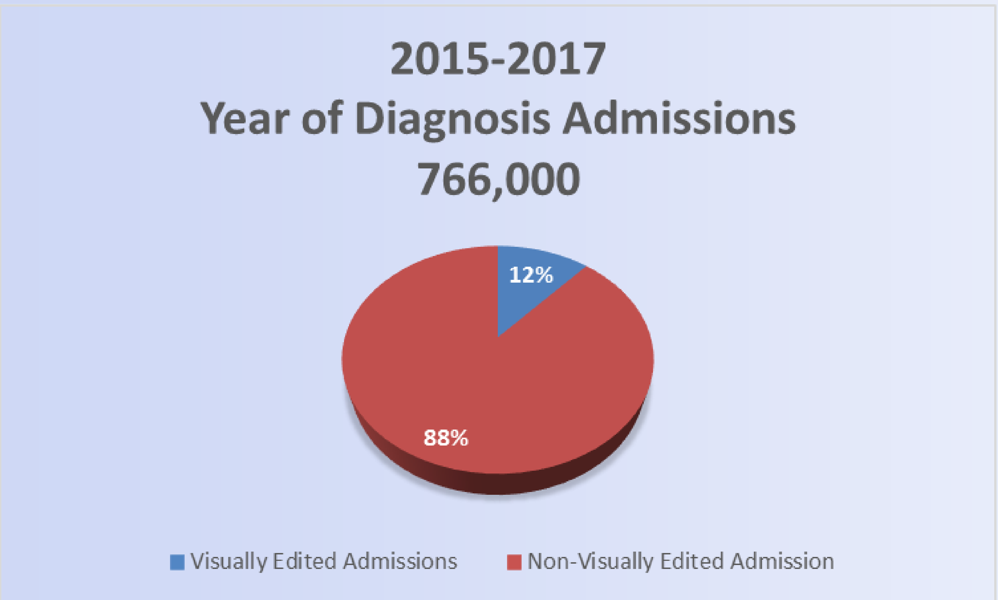
S. Wood, B.A

California Cancer Reporting and Epidemiologic Surveillance (CalCARES) Program*

Institute for Population Health Improvement, UC Davis Health System

Background

As the California Cancer Registry (CCR) moves more and more towards automatic processes, the organization needs to identify new methods for quality control activities. Previously all incoming hospital abstracts uploaded into Eureka (CCR's Database Management System) were "visually edited" to ensure that all coded values were supported by the incoming text related to the diagnosis. The sheer volume of data makes it impossible to manually perform this method of quality control on all cases. Currently the criteria for requiring this quality control step is limited to specified abstractors and other randomly selected cases. In total only around 10%-20% of incoming hospital abstracts require this text to code review.

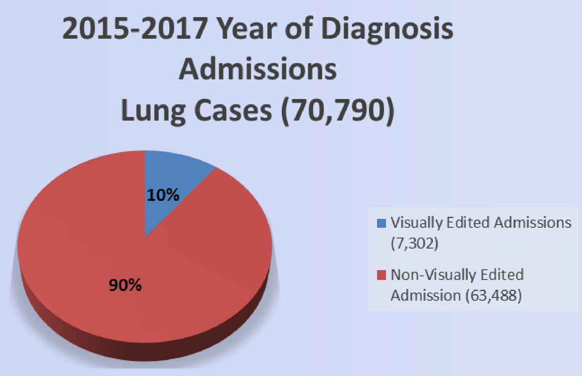


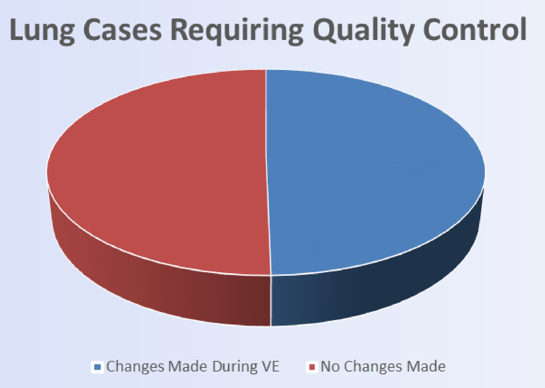
Objective

As we continue to strive for high quality data as well as efficient processes and procedures to ensure high quality data, it is worth exploring new selection criteria for additional quality control processes.

Of the over 766,000 2015-2017 date of diagnosis admissions already uploaded to the CCR, just over 88,600 require visual editing (12%). Some of this 12% is selected based on specified abstractors, but the majority of the cases are chosen at random for the additional quality control process. At the conclusion of the visual editing process many of the cases do not require significant corrections. With such a large manual effort involved in the visual editing process, whether changes are required or not, changing the selection criteria for cases requiring additional quality control would serve to improve data quality. Targeting this quality control activity to cases that are more likely to require quality control will be a more efficient use of manual work and help identify patterns of potential data quality issues.

With just over 70,000 lung cases with a 2015-2017 date of diagnosis reported to the CCR the lung cases will serve as a good pool of cases for analysis . Of those 70,000 cases uploaded, 7302 required additional quality control activities, and of those cases only around 3,500 cases required changes that could be counted as discrepancies and reported back to the abstractor.





Most Common Fields Changed in Visual Editing for Lung Cases

- CS Site-Specific Factor 2
- CS Extension
- CS Site-Specific Factor 1
- CS Mets at DX
- Date of Diagnosis
- Site
- Grade
- TNM Staging
- CS Tumor Size
- TNM Clinical Stage Group
- TNM Clinical M
- Histology

Methodology

Step 1: Find a pattern to the primary site code changes

In order to find repeatable patterns with relatively simple decision points, I took a close look at all of the instances that the primary site of a lung case was changed during the visual editing process. Cursory analysis of changes made to the primary site showed that keywords in the text fields (below) informed the quality control specialist that the primary site be coded differently from how it was abstracted.

Text Diagnostic Procedure Physical Exam
Text Diagnostic Procedure Scopes
Text Diagnostic Procedure X-ray/Scan

In cases where the quality control process determined that the primary site should be coded to **C340** analysis showed the following terms in at least one of the above text fields:

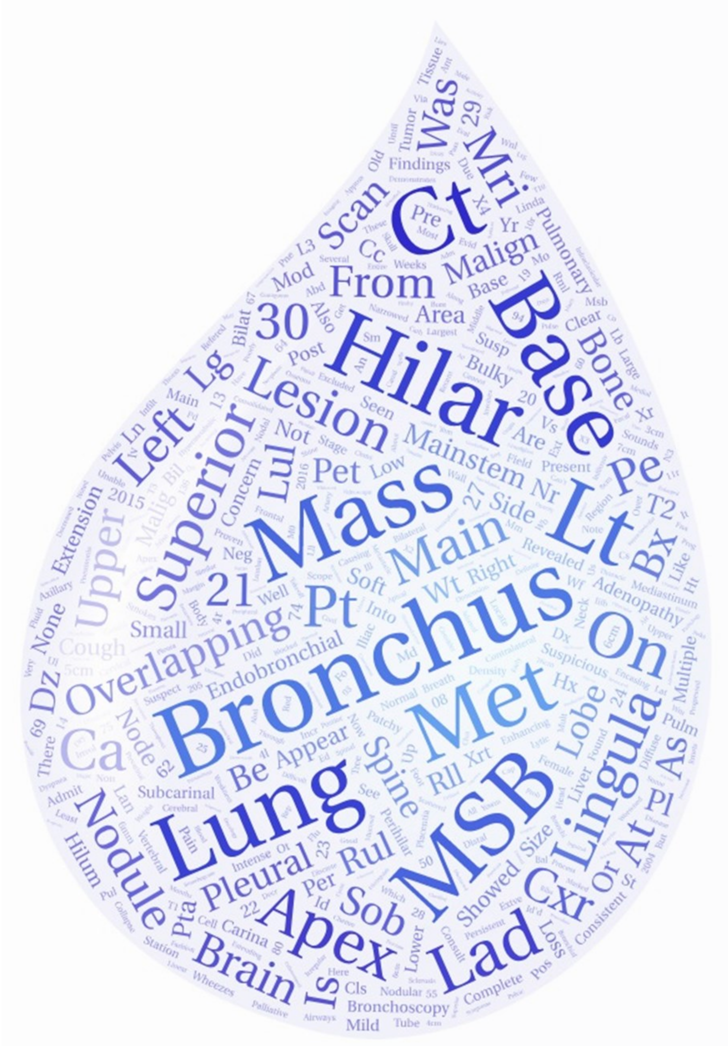
Main (Stem) Bronchus
Hilar
Carina

In cases where the quality control process determined that the primary site to be **C341** analysis showed the following terms:

Upper Lung (RUL, LUL, UL)
Lingula
Apex
Pancoast
Superior

In cases where the quality control process determined that the primary site to be **C343** analysis showed the following terms:

Lower Lung (RLL, LLL)
Base



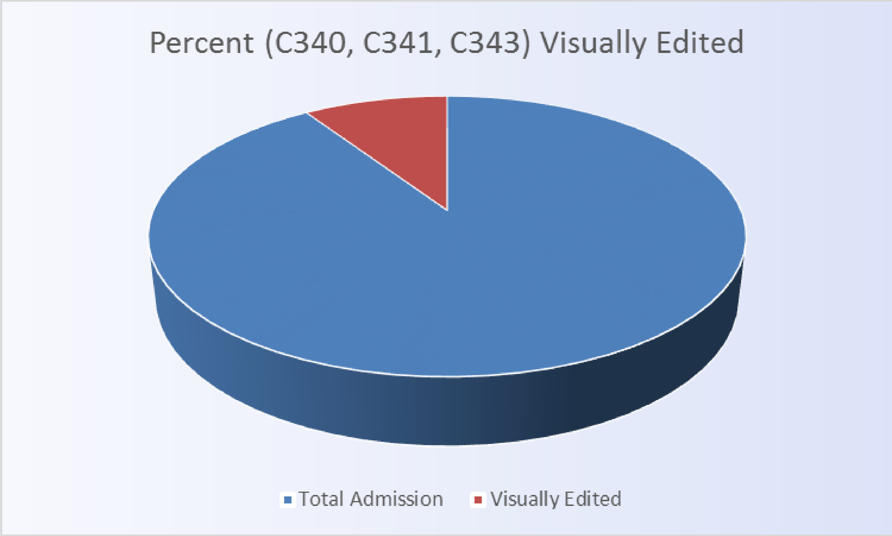
Step 2: Test for patterns identified in Step 1 in the larger data set

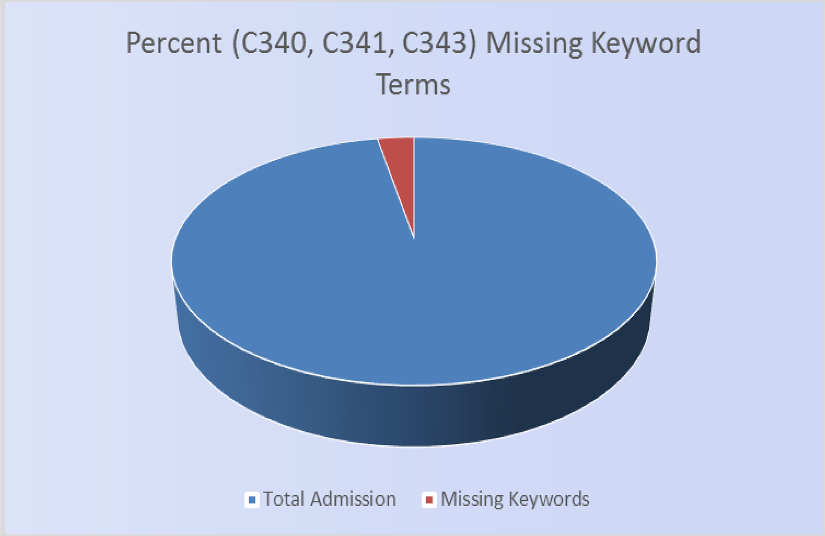
By searching for the terms expected to indicate each site in the text we can determine how indicative of each site those terms are. Looking at the hit rate for each term we can look at the cases that fall out and either refine the search in order to see if any additional terms/ patterns emerge.

Site	Total Cases	Cases With Terms Present	Cases With Missing Terms	Percent Missing Terms
C340	2466	2394	72	2.92%
C341	31850	31353	497	1.56%
C343	17859	17012	847	4.74%

Overall Results

For the three lung sites involved in this analysis, main bronchus (C340), upper lung (C341), and lower lung (C343), the overall success rate for finding relevant terms in at least one of three text fields was 97%. Deeper analysis of the cases that do not include the search terms could reduce the percentage even more as there may be different expectations about the quality of the text depending on class of case or type of reporting source.





Conclusion

Although the scope of the analysis only served to confirm that the text in the case supported the code for the primary site and not the breadth of the entire dataset, this was done in a very basic manner by looking for keywords in the text and could quickly confirm that 97% of the cases appear be coded correctly. Further refinement of the queries could search for negative terms or terms that might suggest a neighboring site. Analysis of the text can also find patterns to help determine if a case is properly coded if there appears to be a mention of metastatic disease.

Starting with the 2018 Data Item Changes the CCR has 21 fields/field groupings that are considered part of the visual editing process wherein a quality control specialist analyzes the data to ensure proper coding and provides feedback to the abstractor if there are any discrepancies. While this text matching effort was 97% successful in identifying the primary site code based on keywords in the text, this logic only applies to these three primary sites, and while the logic for each and every primary site may not be overly complicated, it will be an intense work effort to work through all of the various sites as well as the visually edited data items.

Leveraging the changes already made to cases in our system seems to be a great starting point for analysis of what fields and conditions contribute to a "supported" code. What is a quality control specialist looking for in order to verify a code? Can the CCR leverage each of those conditions in order to flag cases for additional quality control when the known quality conditions do not appear to be met?

The current process for visual editing is for a quality control specialist to review the text and codes while viewing a (tabbed) page that looks that the entire data set. In the future, using target queries, the quality control specialist could view a more limited and custom selection of the data to help determine, at a glance, whether additional quality control activities may be required. Although this is additional step in the process, a more targeted selection of cases to perform quality control will only help to identify cases that may need to be checked, which will also serve to help identify additional patterns and create a continuous process of refining the selection criteria for cases that may require additional quality control.

UC DAVIS
INSTITUTE FOR POPULATION HEALTH IMPROVEMENT

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

I would like to thank the following contributors: Mary Brant, B.A., CTR, Marilyn Scocozza, CTR.