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.

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 closer 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.

<table>
<thead>
<tr>
<th>Site</th>
<th>Total Cases</th>
<th>Cases With Terms Present</th>
<th>Cases With Missing Terms</th>
<th>Percent Missing Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>C340</td>
<td>2466</td>
<td>2394</td>
<td>72</td>
<td>2.92%</td>
</tr>
<tr>
<td>C341</td>
<td>31850</td>
<td>31353</td>
<td>497</td>
<td>1.56%</td>
</tr>
<tr>
<td>C343</td>
<td>17859</td>
<td>17012</td>
<td>847</td>
<td>4.74%</td>
</tr>
</tbody>
</table>

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 to 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.

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.