Automated Versus Manual ICD-10 Coding of Neuroimaging Reports ("Radnosis" Project)

**Primary Category:** Informatics Innovations

**Secondary Category:** Quality and Safety

**Area of Focus:** Diagnostic Radiology
Author:
Jessica Moore, BS; MPH Student

Co-Authors:
Alexander M. McKinney, MD
Jeffrey Rykken, MD
Bharathidasan Jagadeesan, MD
Kevin D. Campbell, MHI
Zeke J. McKinney, MD, MHI, MPH

Disclosures:
AM, ZM, KC are primary owners of Veev, Inc.
KC is the principal owner of Campbell Analytics, LLC.
No other relevant financial disclosures for the author or co-authors.

Affiliations:
1. University of Minnesota School of Public Health, Minneapolis, MN
2. University of Minnesota Department of Radiology, Minneapolis, MN
4. Campbell Analytics, LLC, Minneapolis, MN
5. HealthPartners Occupational and Environmental Medicine Residency, St. Paul, MN
6. HealthPartners Institute, Bloomington, MN
Background

• **2015: Implementation of ICD-10 in United States**
  – Significant impact on US billing and coding

• **Imaging Coding Requirements**
  – Include diagnostic codes affecting reimbursement
  – Typically radiologists do not aid in coding!

• **US Shift to Value-Based Care**
  – Billing and reimbursement to be linked to quality measures
  – Limited or no discrete data generated from radiology reports
  – How can value in radiology be measured?
Objective

This study aimed to test an automated algorithm’s accuracy compared to manual coding for routine neuroimaging reports.
Methods: Demographics

• Study Population
  – Study date range: Jan. – Jun., 2016
  – 200 head, neck, and spine CT or MRI examination reports
  – Randomly-selected from 10,488 studies
  – 10,488 studies randomly-selected from larger set of studies performed during study time period
  – Equally-distributed from inpatient, outpatient, and ED

• Study Site
  – Academic health center serving most of Minnesota
  – Performs approx. 1 million imaging examinations per year
Methods: Coding

• **Automated Coding**
  – Experimental, proprietary natural language processing (NLP) engine
  – Extracted 645 “phrases” from 200 examination reports
    • 3.28 phrases per report
  – Coded phrases with multiple ICD-10 codes
  – Codes ranked by a score indicating likelihood of correctness
Methods: Coding

• **Manual Coding**
  – Simple web-based interface for phrase review
  – Select single “best” ICD-10 code per extracted phrase
  – Alternatively check “no applicable ICD-10 code exists” box

– Four (4) physicians:
  • 3 neuroradiologists
  • 2 with content expertise in ICD-10 coding
– Physicians blinded to NLP engine results
### Updated Phrase 4935 with Best Diagnosis 'Mass of left temporal lobe' (R22.0).

<table>
<thead>
<tr>
<th>Impression Line</th>
<th>Phrase</th>
<th>Phrase ID</th>
<th>Best Dx Text</th>
<th>Best Dx ICD-10</th>
<th>Update/Delete</th>
<th>ICD-10 N/A</th>
<th>ICD-10</th>
<th>Diagnosis Name</th>
<th>ICD-10</th>
<th>Diagnosis ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3 CM HYPERDENSE MASS OF THE LEFT TEMPORAL LOBE WITH EXTENSIVE SURROUNDING</td>
<td>hyperdense neoplasm left temporal lobe surrounding vasogenic edema</td>
<td>4935</td>
<td>Mass of left temporal lobe</td>
<td>R22.0</td>
<td>Delete</td>
<td></td>
<td></td>
<td>Mass of petros temporal bone</td>
<td>M85.9</td>
<td>198279</td>
</tr>
<tr>
<td>VASOGENIC EDEMA. MILD ADJACENT LOSS OF THE GRAY/WHITE DISTINCTION IS QUESTIONED</td>
<td>mild adjacent loss gray white distinction questioned which could</td>
<td>4936</td>
<td>Sclerosis changes on head CT</td>
<td>67.82</td>
<td>Delete</td>
<td></td>
<td></td>
<td>Right temporal lobe mass</td>
<td>R22.0</td>
<td>200208</td>
</tr>
<tr>
<td>WHICH COULD REFLECT SOME LOCAL PRESSURE-INDUCED ISCHEMIC CHANGE. DIFFERENTIAL</td>
<td>local pressure induced ischemia change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mass of right temporal lobe</td>
<td>R22.0</td>
<td>200209</td>
</tr>
<tr>
<td>CONSIDERATIONS OVERWHELMINGLY FAVOR TUMOR, EITHER METASTATIC OR PRIMARY</td>
<td>differential considerations overwhelmingly favor neoplasm either</td>
<td>4937</td>
<td>Primary brain neoplasm (H)</td>
<td>649.6</td>
<td>Delete</td>
<td></td>
<td></td>
<td>Left parietotemporal mass</td>
<td>G93.9</td>
<td>215312</td>
</tr>
<tr>
<td>BRAIN. MASS EFFECT FROM THE TUMOR AND ADJACENT VASOGENIC EDEMA IS RELATIVELY</td>
<td>metastasis primary brain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mass of left parietotemporal region</td>
<td>G93.9</td>
<td>632488</td>
</tr>
<tr>
<td>PRONOUNCED, WITH 8 MM OF LEFT-TO-RIGHT MIDLINE SHIFT AT THE LEVEL OF FORAMEN</td>
<td>compression neoplasm adjacent vasogenic edema relative pronounced</td>
<td>4938</td>
<td>Uncal herniation (H)</td>
<td>G93.5</td>
<td>Delete</td>
<td></td>
<td></td>
<td>Mass of left temporal lobe</td>
<td>R22.0</td>
<td>635807</td>
</tr>
<tr>
<td>MONRO AND SLIGHT UNCAL HERNIATION.</td>
<td>left right herniation level foramen monro slight uncal hernation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mass of right parietotemporal region</td>
<td>G93.9</td>
<td>635810</td>
</tr>
</tbody>
</table>
Methods: Manual Coding Analysis

• Agreement of Applicable ICD-10 Codes
  – Fleiss’ $\kappa$ (kappa): agreement between $\geq 3$ raters

• Agreement of “Best” ICD-10 Code Among Reviewers
  – % agreement between combinations of 2-4 reviewers
  – Phrases where combination of reviewers assigned ICD-10 code
  – Varied for length of ICD-10 code (3, 4, 7 characters)
Methods: Coding Length

• ICD-10 Code Length
  – Maximum 7 characters (excluding period)

  – “Cerebral infarction due to thrombosis of left MCA” (I63.312)

<table>
<thead>
<tr>
<th>ICD-10 Character Length</th>
<th>Character Specification</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>General pathology</td>
<td>“I63”</td>
</tr>
<tr>
<td>4</td>
<td>Specific pathology</td>
<td>“I63.3”</td>
</tr>
<tr>
<td>5</td>
<td>Anatomic localization</td>
<td>“I63.31”</td>
</tr>
<tr>
<td>6</td>
<td>Detailed specification, e.g., laterality</td>
<td>“I63.612”</td>
</tr>
<tr>
<td>7</td>
<td>Initial or subsequent encounter, sequela</td>
<td>“I63.612A”</td>
</tr>
</tbody>
</table>
Methods: Automated Coding Analysis

• Agreement of “Best” ICD-10 Code with NLP Engine
  – Evaluation for engine matches with all 4 reviewers
  – Varied for length of ICD-10 code (3, 4, 7 characters)
  – Varied for number of ranked engine ICD-10 matches
    • e.g., top 1, top 5, top 10

  – Threshold engine score:
    • Average likelihood score for the top ranked matches where reviewers and engine agreed at a specified ICD-10 code length

<table>
<thead>
<tr>
<th>Classification</th>
<th>Reviewers</th>
<th>Engine</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive (TP)</td>
<td>All agreed</td>
<td>Agreed with reviewers</td>
</tr>
<tr>
<td>False Negative (FN)</td>
<td>All agreed</td>
<td>Did not agree with reviewers</td>
</tr>
<tr>
<td>True Negative (TN)</td>
<td>All not agreed</td>
<td>Score &lt; threshold</td>
</tr>
<tr>
<td>False Positive</td>
<td>All not agreed</td>
<td>Score ≥ threshold</td>
</tr>
</tbody>
</table>
Results: Reviewer Agreement

- Agreement of an Applicable ICD-10 Code
  - Fleiss’ $\kappa$ (kappa) = 0.546 (moderate agreement, $p << 0.05$)

- Agreement of “Best” ICD-10 Code Among Reviewers

![Bar chart showing percent reviewer agreement for different combinations of four reviewers.](chart.png)
Results: Engine Performance

Performance metrics:
- Sensitivity (recall): \( \frac{TP}{TP + FN} \)
- Specificity: \( \frac{TN}{TN + FP} \)
- Accuracy: \( \frac{TP + FP}{n} \)
- Odds Ratio (OR): \( \frac{TP / FP}{FN / TN} \)

n = number of phrases (645)
Conclusions

• Agreement in Manual Coding
  – Limited agreement on whether ICD-10 codes can be applied
  – Limited agreement on assignment of ICD-10 codes

• Agreement with Automated Coding
  – NLP-based coding algorithm highly sensitive for top 5 codes
  – Increasing to top 20 codes decreases performance

Note: Having multiple codes attached to a report is relevant
  • as up to 4 can be attached to billing
Conclusions

• Implications
  – Coding variability among expert reviewers
  – Need to improve coding homogeneity
  – Automated coding-based decision support for radiologists during dictation ("real-time") can aid in improving coding homogeneity