Rule-Based Mapping Between
ICD-10 and ICD-10-AM

Jeewani Anupama Ginige

H. Shafruddin
Aaron O’Donnell
Michael Tran
Robert Smith
Chris Boulamatsis
Introduction

• Mapping Carried out from ICD-10 to ICD-10-AM
• Originally maps were created manually and updated biennially
• An algorithmic approach is explored: Rule-based mapping with a customised search engine
• Can achieve high sensitivity with zero false positive.
• Support longitudinal morbidity statistics and time serious analysis for epidemiological reporting
Motivation

• ICD-10 contains 12,187 codes, whereas ICD-10-AM contains 16,953 codes
• Manual mapping requires a huge amount of resources and could be prone to errors
• System generated mappings are compared with mapped ground-truth data
What the Maps are?

- Linkage from an international classification source to target
- Source: International Classification of Diseases (ICD-10)
- Target: International Classification of Diseases Australian Modification (ICD-10-AM)

**INPUTS**

- ICD-10 codes
  - Its full description along with inclusions
  - Optional fifth-digit level codes
- ICD-10-AM 10th edition
  - Complete code, descriptions and exclusions

**OUTPUTS**

- Mapping from ICD-10 codes and description to ICD-10-AM codes and descriptions as well as comments
Types of Maps

SIMPLE MAP
• Links a single ICD-10 to a single ICD-10-AM code
• Cardinality of simple map is one-to-one

ICD-10-AM code

A00.9 ← A00.9

ICD-10 code

COMPLEX MAP
• Links a single ICD-10 code to multiple ICD-10-AM codes
• Cardinality is one-to-many

C88.0

C88.00

C88.01

Waldenstrom macroglobulinaemia without mention of remission
Waldenstrom macroglobulinaemia with mention of remission
### One-to-One Mapping

#### 1. Direct map & description

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>ICD-10-AM Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>A00.9</td>
<td>A00.9</td>
</tr>
<tr>
<td>Cholera, unspecified</td>
<td>Cholera, unspecified</td>
</tr>
</tbody>
</table>

#### 2. Slight change in ICD-10-AM description include dagger code

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>ICD-10-AM Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>F00.9</td>
<td>F00.9</td>
</tr>
<tr>
<td>Dementia in Alzheimer disease, unspecified</td>
<td>Dementia in Alzheimer’s disease, unspecified (G30.9†)</td>
</tr>
</tbody>
</table>

#### 3. Different code

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>ICD-10-AM Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>U84.8</td>
<td>Z06.78</td>
</tr>
<tr>
<td>Resistance to other specified antimicrobial drug</td>
<td>Resistance to other specified antimicrobial drug</td>
</tr>
</tbody>
</table>

#### 4. Spelt Differently

<table>
<thead>
<tr>
<th>ICD-10 Code</th>
<th>ICD-10-AM Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>T36.1</td>
<td>T36.1</td>
</tr>
<tr>
<td>Poisoning: Cefalosporins and other beta-lactam antibiotics</td>
<td>Cephalosporins and other beta-lactam antibiotics</td>
</tr>
</tbody>
</table>
One-to-Many Mapping

ICD-10 code

A41.5
Sepsis due to other Gram-negative organisms

Expanded ICD-10-AM

A41.50 → Sepsis due to unspecified Gram-negative organisms
A41.51 → Sepsis due to Escherichia coli [E. Coli]
A41.52 → Sepsis due to Pseudomonas
A41.58 → Sepsis due to other Gram-negative organisms
List of Techniques

• Customised search engine: Elastic search with optimisation
  ➢ Vector Space Model technique: To find vectors
    ▶ Term frequency
    ▶ Inverted document index
  ➢ Similarity Index: comparing resulting vector and search term vector

• Apply pre-processing techniques before indexing
  ➢ Indices and search terms are converted to lowercase
  ➢ Remove: Apostrophes, dashes and stop words

• Synonyms and language localisation normalisation
  ➢ Difference in spelling ICD-10-AM and the International version

• N-grams technique: To get more fine-grained text similarity search
Material

List of

- ICD-10 ECLs (CSV)
- ICD-10-AM Tenth edition ECLs (CSV)
- 4,635 human mapped files (CSV)
Methods

Stage 1

1. Obtain search terms from ICD-10 parent descriptions
2. Search for each ICD-10 code against ICD-10-AM
3. Top result is checked if it has same code
4. Code with similar description match has high confidence of mapping correctly
5. If it is different code, a series of passes are used to find correct mapping
   • Use combination of parent descriptions to give context to the child code
   • Remaining difference checked using 3-letter n-gram technique
6. Unmatched mappings provided with list of 10 possible results
7. Expert clinical coders choose manually the most accurate code if any
Methods

Stage 2

Additional logic is used

- Check parent code of ICD-10 maps equivalent to ICD-10-AM parent
- If Yes, all children of ICD-10-AM map automatically to parent code
- In some instance, ICD-10 codes map directly to child code of ICD-10-AM
Results

- Out of 12,187 ICD-10 codes, 12,121 mappings were found
- Sixty-One codes: No ICD-10-AM codes exists
- Remaining false negatives includes:
  - Entries with vastly different descriptions
  - Sentence structure make text similarity approach challenging
- No inaccuracies found in generated maps when compared with 4635 human created maps
- Remaining 7552 maps do not require human intervention

<table>
<thead>
<tr>
<th></th>
<th>Ground Truth</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>POSITIVE</td>
<td>NEGATIVE</td>
</tr>
<tr>
<td>True Positive (TP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positive (FP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Negative (FN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Negative (TN)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>POSITIVE</th>
<th>NEGATIVE</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITIVE</td>
<td>12,121</td>
<td>0</td>
<td>12,121</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>5</td>
<td>61</td>
<td>66</td>
</tr>
<tr>
<td>Total</td>
<td>12,126</td>
<td>61</td>
<td>12,187</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.96 %</td>
<td>100 %</td>
<td>99.96%</td>
</tr>
</tbody>
</table>
Conclusion

• Rule-based mapping from ICD-10 to ICD-10-AM using computer-aided algorithm
• Unsolved mapping is small enough for human intervention to achieve 100% mapping
• Reverse mapping (ICD-10-AM to ICD-10) can be achieved with same method
• Need to explore on the possibility of using the same method to map totally different classification systems such as ICHI and ACHI
Thanks to

- Marla Tun
- Filippa Pretty
- Vera Dimitropoulos
- Dr. Qingsheng Zhou
- Professor Richard Madden
Rule-Based Mapping Between
ICD-10 and ICD-10-AM

Jeewani Anupama Ginige

H. Shafruddin
Aaron O’Donnell
Michael Tran
Robert Smith
Chris Boulamatsis