DERIVING PHYSICIANS’ EXPERTISE PROFILES BASED ON ICD9-CODED ENCOUNTER NOTE LOGS

Victor Brodsky, MD
Tru V. Tran
Stephen E. Kessler II
Curtis Cole, MD
Weill Cornell Medical College
Why? Can’t we just ask?

- Medical scientists are erratic at self profiling
  - Accuracy
  - Completeness
  - Future Goals vs. Past Expertise
  - Regular updates
  - Time consuming process
- We know their specialties and billing activities
- Publication history & other sources still useful
Can someone else do it?

- The Challenge
  - To accurately and efficiently determine a particular physician’s areas of expertise, by disease/condition

- The Approach
  - Automate the analysis of the ICD9-coded diagnostic information in the electronic medical patient records signed by physicians to determine areas of expertise
A Branch of ICD9 Hierarchical Tree

**Infectious and Parasitic Diseases (001-139)**  
- **Intestinal Infectious Diseases (001-009)**  
  - 001 Cholera  
  - 002 Typhoid and Paratyphoid Fevers  
  - **003 Other Salmonella infections**  
    - 003.0 Salmonella gastroenteritis  
    - 003.1 Salmonella septicemia  
    - **003.2 Localized salmonella infections**  
      - 003.20 Localized salmonella infection, unspecified  
      - 003.21 Salmonella meningitis  
      - **003.22 Salmonella pneumonia**
A Branch of ICD9 Hierarchical Tree

Level 0

Level 1

Level 2

Level 3

Level 4

Level 5

Root

Infectious and Parasitic Diseases (001-139)

Intestinal Infectious Diseases (001-009)

Etc.

Etc.

Etc.

001 Cholera

002 Typhoid and Paratyphoid Fevers

003 Other Salmonella Infections

003.0 Salmonella gastroenteritis

003.1 Salmonella septicemia

003.2 Localized salmonella infections

003.20 Localized salmonella infection, unspecified

003.21 Salmonella meningitis

003.22 Salmonella pneumonia
ICD9 Tree Statistics: 17,414 Items

- **8,562** • Level 5 - Disease; Most Specific
- **7,438** • Level 4 - Disease; Specific
- **1,231** • Level 3 – Disease
- **164** • Level 2 – Subcategories
- **19** • Level 1 - Categories
Physician Population

- 2,443 physicians who saw > 5 patients

- Of 17,231 ICD9 Codes, each physician encountered from 1 to 1,747
Ranking ICD9s

• Only some of logged ICD9 codes are pertinent to the physician
  – A psychiatrist may log two ICD9 codes in the EMR for a patient visit:
    • 300.3 Obsessive-compulsive disorders
    • 278.0 Overweight and obesity

• For each pair (a physician and an ICD9), we produce a score with a magnitude predicting how pertinent the given ICD9 code is to the given physician’s expertise

• In a successful ranking system, ICD9 codes that are directly applicable to a physician’s expertise are ranked higher than those that are incidental
Measuring Ranking Success

- Sample physicians selected
- Each ICD9 that they had logged was subjectively marked as either relevant or irrelevant to their area of expertise based on their specialty
- Per-Physician ICD9 lists sorted by score
- Hamming distance to nearest target list found
- Will need to confirm scores by presenting profiles to the “owners” for grading
Hamming Distance

Hamming Distance $= 2$
Hamming Distance, Nearest Target

- 3 Relevant, 1 is not in top 3 $\Rightarrow$ HD = 1
First Efforts – Sort by # of Patients

If all relevant ICD9 Codes (in blue) had many patients, division between relevant & irrelevant codes would be here.
Incidence Ratios

- A physician may see one ICD9 more often than other ICD9s
- A physician may diagnose an ICD9 more often than other physicians diagnose that ICD9
- A disease may occur more often than another disease in our patient population

Core Score = \[
\frac{\text{# of patients seen for an ICD9}}{\text{total # of patients seen}}
\]

One provider

All providers
Utilizing the Tree Structure of the ICD9

- Diagnoses by a particular physician tend to cluster within specific branches of the tree
- These clusters tend to be related to the physician’s specialty
- We can utilize this info by giving these clusters greater weight
Climbing up the Tree

0 & (8+6)/3 & (8+6+4+8+1+27+9+39)/10

Branch 1

Branch 2
Accounting for coverage

- Branch 1
  - 6
  - 16

- Branch 2
  - 6
  - 3
  - 5
  - 4
  - 1
  - 3
### Average with specialty average

<table>
<thead>
<tr>
<th>100% Personal</th>
<th>50% Specialty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transient Tic disorder 43</td>
<td>Screening for lipoid disorders (cholesterol level) 23</td>
</tr>
<tr>
<td>T7-T12 level w/ unspecified spinal cord injury 101</td>
<td>Generalized Anxiety Disorder 39</td>
</tr>
<tr>
<td>Circadian rhythm sleep disorder, jet lag type 112</td>
<td>Screening examination for viral diseases 8</td>
</tr>
<tr>
<td>Pedal cycle accident injuring unspecified person</td>
<td>Bipolar II Disorder</td>
</tr>
<tr>
<td>Lethal midline granuloma</td>
<td>Mammographic microcalcification of breast</td>
</tr>
<tr>
<td>Pertussis alone (vaccination)</td>
<td>Schizophrenia, Undifferentiated Type</td>
</tr>
<tr>
<td>DTP with typhoid-paratyphoid (vaccination)</td>
<td>Shortness of Breath</td>
</tr>
<tr>
<td>Screening examination for viral diseases</td>
<td>Diabetes, Type II, Adult Onset</td>
</tr>
<tr>
<td>Counseling on substance use and abuse</td>
<td>Panic Disorder without Agoraphobia</td>
</tr>
<tr>
<td>Other abnormality of red blood cells</td>
<td>Anxiety state, unspecified</td>
</tr>
</tbody>
</table>
Clustering to extract profile

39 are beyond 2 standard deviations; k-means adds 63

ICD9 Codes

Score

By Patients

By Scores

VIVO
Resulting profile statistics

% of Relevant ICD9 codes in the profile

Physician Profiles

- ER
- Hematology 1
- Hematology 2
- IM 1
- IM 2

Legend:
- % Relevant in Top Cluster (By Patients)
- % Relevant in Top Cluster (By Scores)
Sample profile

- Immune thrombocytopenic purpura
- Primary thrombocytopenia, unspecified
- Other sickle-cell disease without crisis
- Constitutional red blood cell aplasia
- Sickle-cell thalassemia without crisis
- Other thalassemia
- Evans' syndrome
- Constitutional aplastic anemia
- Sickle-cell/Hb-C disease without crisis
### Physician ID list for a sample ICD9

**ICD9 Code: 157.3 – Pancreatic Cancer**

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.91309</td>
<td>• 129219 HEMATOLOGY/ONCOLOGY</td>
</tr>
<tr>
<td>9.308263</td>
<td>• 113716 ONCOLOGY (MEDICAL); HEMATOLOGY; INTERNAL MEDICINE; ONCOLOGY; HEMATOLOGY; PEDIATRIC HEMATOLOGY/ONCOLOGY;</td>
</tr>
<tr>
<td>9.137091</td>
<td>• 118712 COLON AND RECTAL SURGERY; SURGERY</td>
</tr>
<tr>
<td>6.915659</td>
<td>• 119084 GASTROENTEROLOGY</td>
</tr>
<tr>
<td>6.33550</td>
<td>• 114357 SURGERY; GASTROENTEROLOGY</td>
</tr>
<tr>
<td>6.034863</td>
<td>• 115212 MEDICINE HEMATOLOGY/ONCOLOGY - SOLID TUMOR; ONCOLOGY; HEMATOLOGY; INTERNAL</td>
</tr>
<tr>
<td>4.562387</td>
<td>• 115362 ONCOLOGY; HEMATOLOGY</td>
</tr>
<tr>
<td>4.494329</td>
<td>• 129763 ONCOLOGY; HEMATOLOGY</td>
</tr>
<tr>
<td>4.248922</td>
<td>• 130479 HEMATOLOGY/ONCOLOGY - SOLID TUMOR; ONCOLOGY; HEMATOLOGY</td>
</tr>
</tbody>
</table>
Hierarchically Clustering Specialists

Euclidian Distance in n-dimensional space
Work in progress

- For a physician with 5 ICD9s (& patients), does each of the ICD9’s constitute “20% of what she sees”?
- Does the only physician who has seen one specific ICD9 become “institution’s top specialist” in it?
- Are the physician’s diagnosis and note correct?
- Calculating “intrinsic” ICD9 scores to separate “rare diseases” from “bicycle accidents”?
- Normalizing and making scores “meaningful”
More data sources & applications

- **Sources**
  - Orders placed per diagnosis
  - Referrals

- **Applications**
  - Profiling based on medical procedures
  - Climbing meshes of terms instead of rooted trees