Expressing Observations from Electronic Medical Record Flowsheets in an i2b2 based Clinical Data Repository to Support Research and Quality Improvement

Russ Waitman¹, Judith Warren², LaVerne Manos², Dan Connolly¹
¹Department of Biostatistics, ²School of Nursing
The University of Kansas Medical Center
October 26, 2011
Overview

- **Background:** i2b2, data warehouse for clinical research
- **Background:** “Nursing” flowsheets
- **Goal:** Exploit vast data for clinical research and understand clinical documentation investment
- **Challenge:** Reorganizing the local unit ontologies to support clinical research and analysis
i2b2: Informatics for Integrating Bench to Bedside

- Star schema based data warehouse software
  - Developed by National Center for Biomedical Computing at Partners Healthcare (Harvard University, Boston, MA)
  - Cohort discovery and hypothesis generation
  - Initial domain: phenome/genome integration
- i2b2 novelty: Entity/Attribute/Value (EAV) model combined with hierarchical concept ontologies

<table>
<thead>
<tr>
<th>Concept Path</th>
<th>Concept Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>i2b2\Diagnoses\Circulatory system (390-459)\</td>
<td></td>
</tr>
<tr>
<td>i2b2\Diagnoses\Circulatory system (390-459)(448) Disease of capillaries\</td>
<td>ICD9:448</td>
</tr>
<tr>
<td>i2b2\Diagnoses\Circulatory system (390-459)(448) Disease of capillaries(448-1) Nevus, non-neoplastic\</td>
<td>ICD9:448.1</td>
</tr>
</tbody>
</table>

Flowsheets: Untapped Data Source

- Typically the data existed on paper; rarely used beyond the bedside.
  - Exchange has lagged other observations and billing information because there’s no regulatory requirement
    - Changes with Phase 2 meaningful use (pressure ulcers)?
- Fundamental, structured data about the patient
  - Weight, temperature, pain, fluid intake/output
- Key process measures
  - Regulatory: Restraints documentation
- Perhaps the largest IT investment in clinician time.
  - Mobility for physical therapy research
  - Homeless discharge status for public health research
How are “nursing” flowsheets organized?
Method: System Architecture

- Data from Epic Clarity database (> 7,000 tables & 60,000 columns)
- Transformed into an i2b2 compatible schema. Then, de-identified, and loaded on a separate database server to be accessed by i2b2.
- De-identified data used by i2b2 is deemed non-human subjects research by our institutional review board
Results: Flowsheets Dwarf Diagnoses, Labs and Medications

- 411,000,000 flowsheet observations loaded (35 months).
  - 2,200,000 diagnoses, 18,000,000 meds and 25,000,000 labs
- 147,000 distinct patients (majority inpatient observations)
  - 56,000 encountered the Emergency Room,
  - 8,000 in the outpatient cancer center,
  - 46,000 in perioperative, procedural and ambulatory settings.
- 6,742 distinct measures used
- most frequent: pulse 6,000,000 facts
- Approximately 100 measures are half the observations.
- Data density corresponds to clinical practice.
  - Facts/Patient Density: Surgical Intensive Care 14,150, Emergency Department 63
  - Measure/Encounter: Central venous pressure 111; pain goal 8; height 1.3
Method: Exploring Ontology Usability
ETL decision: “Lazy” Loading of Terms

- Support alternative views of reality
  - Load the data with the local ontology. Map concepts to standards secondarily in the concept space.
  - Allows multiple ontologies for observations and works around mapping challenges with contributing organizations

Further technical details described at: http://informatics.kumc.edu/work/wiki/HERON
Results: KU Hospital Flowsheets in i2b2

- “KU IP” inpatient flowsheets
- “Airway” Template
- “Tracheostomy Tube” Group
- “Trach Site Appearance” Flowsheet Measure
- Choices for the Measure (“Dry and Intact”, “ Bloody”, etc.)
Clinical Research and Quality Perspective

- Researcher just wants all Temperatures, Respiratory Rate
- Cancer Center’s Vital Signs Template (#920553)
- Vitals and Extended Vitals Groups (#9209769, #9209770)
- Obstetrics Maternal Vitals Template (#12001)
  - Maternal Vitals Group (#12000)
- Note Issue: #1190 Respirations versus #9 Resp
Method: Cluster Similar Groups or Templates
e.g. Vital Signs

- Ob/Labor Vitals #1481
- Critical Care Vitals Adult #1481
- Vitals #2091
Method: Pruning and Clustering

- Written in SQL with Python for the iterative clustering algorithm.
- Acts against the i2b2 metadata schema to promote reusability for other institutions adopting i2b2 with different EMRs.
- Removed vendor and test builds from the ontology
- Agglomerative Hierarchical Clustering of groups and then templates
- **Crude Naming convention**: the most frequent words included in the names of the merged groups or templates.
- Graphs are generated that diagram the clustering process.
- Ontology size calculated for thresholds of 0.9, 0.5, 0.25, and 0.1 and compared with ICD9 Diagnoses ontology in i2b2.
Results: Clustering Visualization 0.5 Similarity Illustrating Merger of Pediatric Templates

Newborn VS and Newborn Flowsheet: 0.909

Merged Newborn with Merged Assessment: 0.537
Results: Pruning and Clustering Impact on Ontology Size

<table>
<thead>
<tr>
<th>Pruning/Clustering Step</th>
<th>Total concepts</th>
<th>Templates</th>
<th>Groups (distinct)</th>
<th>Measures</th>
<th>Measure choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>270,641</td>
<td>720</td>
<td>5,379 (2,292)</td>
<td>37,656</td>
<td>226,666</td>
</tr>
<tr>
<td>Actual Build</td>
<td>221,531</td>
<td>542</td>
<td>4,070 (1,655)</td>
<td>30,604</td>
<td>186,151</td>
</tr>
<tr>
<td>Remove &lt; 35</td>
<td>121,841</td>
<td>428</td>
<td>3,350 (1,325)</td>
<td>21,983</td>
<td>95,916</td>
</tr>
<tr>
<td>Threshold = 0.9 (G and T)</td>
<td>102,285</td>
<td>390</td>
<td>3,008 (1,026)</td>
<td>19,937</td>
<td>78,818</td>
</tr>
<tr>
<td>0.5 (G and T)</td>
<td>50,146</td>
<td>266</td>
<td>1,359 (799)</td>
<td>10,699</td>
<td>37,751</td>
</tr>
<tr>
<td>0.25 (G and T)</td>
<td>36,594</td>
<td>193</td>
<td>953 (687)</td>
<td>8,228</td>
<td>27,169</td>
</tr>
<tr>
<td>0.1 (G and T)</td>
<td>30,371</td>
<td>150</td>
<td>743 (615)</td>
<td>6,950</td>
<td>22,497</td>
</tr>
</tbody>
</table>

- Removing spurious builds and undocumented/infrequent measures reduced the number of distinct measures from 13,659 to 4,066
  - Before pruning, measures occurred in 5.4 places in the ontology
  - After pruning, 1.7 places
- For comparison, ICD9-based diagnosis ontology from i2b2
  - 44,577 total concepts, 158 high level concepts, and 20,116 ICD9 codes (8,231 were used locally)
  - ICD9 codes occur in one location in this ontology
Toward a Usable Flowsheet Ontology

- **Encouraging**: moderately pruned ontology is smaller than ICD9-based diagnosis ontology
- **However**, this pruning approach doesn’t create a usable ontology.
  - Names assigned to merged groups requires human curation.
- **Future Work**: Automated approach should improve when coupled with expert guided terminology development, feedback from users, and evaluation.
Flowsheets: the Wild West of Terminology

- Hospitals and clinics customize flowsheets to optimize clinical workflow
  - Resulting challenges:
    - maintaining the large flowsheet ontology
    - validating proper utilization of terminology
    - measuring compliance in nursing documentation
- Implementation Science Opportunity:
  - Measure variability of a local ontology from the model ontology
  - Develop tools to assist organizations when their local ontologies need to meet external requirements (e.g. meaningful use, SNOMED encoding).
Limitations of this Research

- One medical center; primarily inpatient focused.
  - EMR ambulatory deployment: Fall 2010 through 2012.
- For some observations (lines, drains, airways) unsolved ETL challenge limits analytical utility as explained in the paper.
Flowsheets Provide a Foundation for Clinical and Informatics Research

- Meeting the needs of Clinical Research
  - Our Flowsheet data (ex. BMI) is being used for prospective clinical trial recruitment and retrospective data analysis.
- Vast New Data Source Opens New Areas of Informatics Research
  - How does flowsheet documentation measure consumption of clinical resources?
    - Example: ICU practice change to intensive insulin therapy impacts nursing workload to provide more care and requires additional documentation.
  - What data is most critical for clinical decision making versus regulatory compliance and non-clinical activities?

Questions?