Chart Abstraction on VINCI: tools and services

Olga Patterson
Dan Denhalter
• Overview of chart abstraction
• Annotation project workflow
• Demonstration of tools:
  – eHost
  – Chart Review
• VINCI Annotation Services
• Questions and discussion
• Chart Abstraction
  – Chart review
  – Medical record review
  – Chart annotation
• A research methodology of data collection for retrospective investigation
How is EMR Created?

Creating patient documentation
An Electronic Medical Record contains information stored in both structured and unstructured formats.

- Structured data tables most often include objective information like labs and vital signs.
- Unstructured notes most often contain the patient’s experience, the providers work-up and diagnosis, the treatment plan, and outcomes.
Structured Data

- Stored in database tables
- Include labs, meds, vitals, demographics, visit information, Codes, etc.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>DateTime</th>
<th>LabID</th>
<th>TestName</th>
<th>Specimen</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>101010</td>
<td>2012-10-06 1600</td>
<td>0409101</td>
<td>Potassium</td>
<td>Serum</td>
<td>4.9</td>
</tr>
<tr>
<td>101010</td>
<td>2012-10-06 1600</td>
<td>71368168</td>
<td>Sodium</td>
<td>Serum</td>
<td>136</td>
</tr>
<tr>
<td>202020</td>
<td>2012-10-06 0900</td>
<td>0409101</td>
<td>Potassium</td>
<td>Serum</td>
<td>3.6</td>
</tr>
<tr>
<td>202020</td>
<td>2012-10-06 0900</td>
<td>71368168</td>
<td>Sodium</td>
<td>Serum</td>
<td>145</td>
</tr>
</tbody>
</table>
Unstructured Data

• Stored in the database, but with text fields.

• Includes written or dictated text notes like progress notes, discharge summaries, and radiology notes.

• Also includes semi-structured information like templates and comment fields.
SUBJECTIVE: The patient is in with several medical problems. He complains his mouth being sore since last week and also some "trouble with my eyes." He states that they feel "funny" but he is seeing okay. He denies any more diarrhea or abdominal pain. Bowels are working okay. He denies nausea or diarrhea. Eating is okay. He is emptying his bladder okay. He denies dysuria. His back is hurting worse. He complains of right shoulder pain and neck pain over the last week but denies any injury. He reports that his cough is about the same.

CURRENT MEDICATIONS: Metronidazole 250 mg q.i.d., Lortab 5/500 b.i.d., Allegra 180 mg daily, Levothroid 100 mcg daily, Lasix 20 mg daily, Flomax 0.4 mg at h.s., aspirin 81 mg daily, Celexa 40 mg daily, verapamil SR 180 mg one and a half tablet daily, Zetia 10 mg daily, Feosol b.i.d.

ALLERGIES: Lamisil, Equagesic, Bactrim, Dilatrate, cyclobenzaprine.


HEENT: Head was normocephalic. Examination of the throat reveals it to be clear. He does have a few slight red patches on his upper inner lip consistent with yeast dermatitis.

Neck: Supple without adenopathy or thyromegaly.

Extremities: He has full range of motion of his shoulders but some tenderness to the trapezius over the right shoulder. Back has limited range of motion. He is nontender to his back. Deep tendon reflexes are 2+ bilaterally in lower extremities. Straight leg raising is positive for back pain on the right side at 90 degrees.

Abdomen: Soft, nontender without hepatosplenomegaly or mass. He has normal bowel sounds.

ASSESSMENT:
1. Clostridium difficile enteritis, improved.
2. Right shoulder pain.
3. Chronic low back pain.
4. Yeast thrush.
5. Coronary artery disease.
6. Urinary retention, which is doing better.

PLAN: I put him on Diflucan 200 mg daily for seven days. We will have him stop his metronidazole little earlier at his request. He can drop it down to t.i.d. until Friday of this week and then finish Friday's dose and then stop the metronidazole and that will be more than a 10-day course. I ordered physical therapy to evaluate and treat his right shoulder and neck as indicated x 6 visits and he may see Dr. XYZ p.r.n. for his eye discomfort and his left eye pterygium which is noted on exam (minimal redness is noted to the conjunctiva on the left side but no mattering was seen.) Recheck with me in two to three weeks.
21.6 million patients

6.9 billion lab tests

4.0 billion orders

2.9 billion procedures

2.7 billion clinical notes

2.1 billion diagnoses

2.0 billion medication fills

2.1 billion outpatient visits

14 million inpatient visits
• Retrospective study
  – Utilizing clinical or administrative historical data that was originally collected for reasons other than research

• Chart abstraction focuses on unstructured data sources but can be combined with structured data
Text is where the majority of clinical information is stored.

• The patient experience is in text.
  – “Patient reports his knee hurts so bad he cannot sleep. He is also at risk of losing his job because he cannot work without sitting down”

• The type of illness, symptoms, and severity are in text.
  – “Diagnosed with relapsing remitting MS, currently mild tingling and weak grip.”

• The timing of the episode is in text.
  – “The patient saw ENT last week and surg was scheduled. She was cleared by cardiology last Monday, labs yesterday were normal. Pt taken to OR at 3:00 PM for tonsillectomy, she was taken to PACU in good condition, returned to same day, and discharged at 8 pm.”

• The disease course is detailed in text.
  – “The chest pain started at while at the gym running, but resolved with rest. The following day the patient again had CP while walking. Today the pain was constant and he presented to the ER. The pain resolved with aspirin, oxygen, nitro, MS.”
The treatment course is detailed in text.

“The patient was started on albuterol, then changed to xop and spiriva after pulm consult and cardiology saw her.”

The outcomes are in text.

“The patient was started on 48 weeks of peg interferon and ribavirin, but tx was stopped due to constant fatigue, anxiety, and concerns of his wife that he may harm himself.”

Even structured elements that are missing from the database are in text.

“Patient is transferring care from university hospital. He is genotype 1, VI 391,000, hep B immune, HIV negative.”

The only thing not in text is what the provider failed to write.
EMR is written by providers for other providers

Difficulty of document interpretation
- Photocopies
- Misspellings and grammar errors are pervasive
- Terminology differs from non-clinical text
  - "patient endorses being verbally abused"
  - "patient status post spinal fusion"
  - "Angina, r/o MI"
- Abbreviations and acronyms are common
  - "50 y/o pt c DM2, HTN, c/o SOB & CP. R/O MI"

Missing data

Incomplete and inconsistent documentation
• Retrospective clinical research
  – Case-control studies, case series

• Quality reporting

• Compliance auditing

• Guideline development

• Reference standard for computerized text processing
• Annotation = label that assigns meaning to data
  – Contain a pointer to start and stop points in a text
  – Can have class or attribute information with them
  – Generated by human, machine or human+machine.

“The CXR shows **LLL consolidation**.”

Span: 15:31
Class: Finding
Assertion: Present
1. Define concepts and variables
2. Select annotation tool
3. Document selection
4. Develop annotation guideline
5. Identify annotator qualifications
6. Train and manage annotators
7. Adjudication or Annotation quality measurement
• Concept definition
  – Diagnosis, lab test, action, event…

• Variable definition – Values that the concepts can have:
  – Diagnosis: explicitly mentioned or inferred
  – Lab test: exact numeric value or range or direction
  – Action: planned or occurred
  – Event: explicitly mentioned or inferred
Example of concept and variable definition for a study on quality of colonoscopy procedures

<table>
<thead>
<tr>
<th>Concept</th>
<th>Variable definition</th>
<th>Source</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowel preparation</td>
<td>Explicitly stated quality of bowel</td>
<td>Colonoscopy report</td>
<td>Excellent, Good, Fair,</td>
</tr>
<tr>
<td></td>
<td>preparation.</td>
<td></td>
<td>Poor, Inadequate</td>
</tr>
<tr>
<td>Procedure indication</td>
<td>Explicitly stated indication for</td>
<td>Colonoscopy report</td>
<td>Screening, High Risk/</td>
</tr>
<tr>
<td></td>
<td>procedure.</td>
<td></td>
<td>diagnostic Treatment</td>
</tr>
</tbody>
</table>
### Concept and variable definition

<table>
<thead>
<tr>
<th>Concept</th>
<th>Variable definition</th>
<th>Source</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anemia</td>
<td>Any evidence of anemia</td>
<td>Any clinical</td>
<td>Affirmed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>note</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept</th>
<th>Variable definition</th>
<th>Source</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anemia</td>
<td>Explicitly stated diagnosis of aplastic anemia or pancytopenia consistent with ICD9 code 284, excluding non-specified anemia</td>
<td>TIU documents</td>
<td>Affirmed</td>
</tr>
</tbody>
</table>
Annotation tools

- Chart Review – on VINCI
- eHost - https://code.google.com/p/ehost/
- Knowtator – http://knowtator.sourceforge.net/
- RapTAT – https://code.google.com/p/raptat/
- BRAT - http://brat.nlplab.org/
- .....
• Data sources in CDW
  – TIU Documents
  – Radiology notes
  – Comment and text fields in various tables

• Other sources
Sample size for a descriptive study of a dichotomous variable 95-percent confidence interval

<table>
<thead>
<tr>
<th>WIDTH OF THE CONFIDENCE INTERVAL (W)</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected proportion (P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>138</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>196</td>
<td>87</td>
<td>49</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>246</td>
<td>109</td>
<td>61</td>
<td>39</td>
<td>27</td>
</tr>
<tr>
<td>0.25</td>
<td>288</td>
<td>128</td>
<td>72</td>
<td>46</td>
<td>32</td>
</tr>
<tr>
<td>0.30</td>
<td>323</td>
<td>143</td>
<td>81</td>
<td>52</td>
<td>36</td>
</tr>
<tr>
<td>0.40</td>
<td>369</td>
<td>164</td>
<td>92</td>
<td>59</td>
<td>41</td>
</tr>
<tr>
<td>0.50</td>
<td>384</td>
<td>171</td>
<td>96</td>
<td>61</td>
<td>43</td>
</tr>
</tbody>
</table>

Annotation guideline

- Annotation schema
  - Annotation types = classes
  - Attributes
  - Relationships

- Formal step-by-step description of the annotation process

- Examples
Who Should Annotate?

- **Who**: depends on the clinical use case and information targets

- Domain expertise (physicians, nurses, nurse practitioners, pharmacists, physician assistants, coders, etc).

- Expensive domain experts are often not needed for specific tasks or portions of the task.

Human chart reviewer = annotator
Annotation Workflow

Pre-annotate

Human A Annotates

Post-Process

Human A Annotates

Human C Adjudicates

Post-Process

Pre-annotate

Human B Annotates

Human A Annotates
– Inter-rater agreement
  • Proportion of matched annotations to the number of all annotations performed by two people

\[
\frac{\text{Number matched annotations}}{\text{Total number of annotations}}
\]

– Cohen’s Kappa
  • \( \rho_o \) - observed proportionate agreement
  • \( \rho_e \) - probability of random agreement

\[
k = \frac{\rho_o - \rho_e}{1 - \rho_e}
\]
VINCI Services Team

Jeff Scehnet
Kevin Malohi

VINCI@va.gov
VINCIServices@va.gov

- Concierge
- Data Provisioning
- Compliance
- Feasibility
- Recruitment
- **Annotation / Chart Review**
- Natural Language Processing
- Analytics and Data Services
- Application Development
• Range of services
  – Education and training
  – Project definition and guideline development
  – Contracting annotators
  – Full chart abstraction project
VINCI Services Team

Jeff Scehnet
Kevin Malohi

VINCI@va.gov
VINCIServices@va.gov

• Concierge
• Data Provisioning
• Compliance
• Feasibility
• Recruitment
• **Annotation / Chart Review**
• Natural Language Processing
• Analytics and Data Services
• Application Development

10/01/2015
eHost

Chart-Review
CAROTID SERIES COMPLETE

Reason: TIA

UNDERLYING MEDICAL CONDITION:
72 year old man with TIA
REASON FOR THIS EXAMINATION:
Eval for carotid stenosis
CAROTID STUDY

HISTORY: TIA.

FINDINGS: No prior studies for comparison. There is complete occlusion of the left and associated internalization of the external carotid artery on the left. There is absent flow involving the left vertebral artery. There is a significant focal hypoechoic plaque involving the right ICA. Similar plaque involving the right ECA. The peak systolic velocities on the right are 264, 56, and 15 cm per second for the ICA, CCA, and ECA respectively. The ICA to CCA ratio is 5.3. There is antegrade flow involving the right vertebral artery.

IMPRESSION:
1. Occluded left-sided ICA and left-sided vertebral artery.
2. 70-79% right ICA stenosis.
3. Near occlusion of the right external carotid artery.
# Chart-Review: Configuration

## Show Clinical Element Configuration

<table>
<thead>
<tr>
<th>Name</th>
<th>Lab Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Simple lab element based on example database.</td>
</tr>
<tr>
<td>Active</td>
<td>true</td>
</tr>
<tr>
<td>Title Field</td>
<td>LAB_NAME</td>
</tr>
<tr>
<td>Description Field</td>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>Type</td>
<td>LIST</td>
</tr>
<tr>
<td>All Elements By Patient Id Query</td>
<td><code>select lab.lab_date, lab_test_lookup.lab_name, lab.result, concat(lab_test_lookup.lab_name, ', ', cast(lab.result as char)) as description, lab.id, lab.lab_performed_by from lab, lab_test_lookup, patient where lab.lab_performed_id = lab_test_lookup.id and lab.patient_id = patient.id and patient.id = ?</code></td>
</tr>
<tr>
<td>Single Element Query</td>
<td><code>select lab.lab_date, lab_test_lookup.lab_name, lab.result, concat(lab_test_lookup.lab_name, ', ', cast(lab.result as char)) as description, lab.id, lab.lab_performed_by from lab, lab_test_lookup, patient where lab.lab_performed_id = lab_test_lookup.id and lab.patient_id = patient.id and lab.id = ?</code></td>
</tr>
<tr>
<td>Created By</td>
<td>admin</td>
</tr>
<tr>
<td>Created Date</td>
<td>2014-03-05 10:23:16.0</td>
</tr>
</tbody>
</table>
Step 3 - Users and Clinical Elements In This Process

Task Creation Query

Enter the query to create process tasks.

select id from patient;

Users

- admin
- project1
- project2

Other

Assignment style: By Process

Create Tasks and Finish
Dale McClanahan, 60 Year old Male

previous admissions, Routine susceptibility report demonstrated susceptibility to Daptomycin but after two days of sustained bacteremia and worsening picture, gentamicin was added and his PICC line was discontinued. The patient remains on the floor, but has been persistently febrile. Transthoracic echocardiogram shows new tricuspid valve regurgitation and a 3 cm vegetation. He endorses subjective fevers and chills but does not otherwise localize his symptoms. He reports feeling depressed about his outlook.

Hematocrit (%)

24 / 41.50%
Acknowledgements

• Resources and Facilities
  – Veterans Affairs Salt Lake City Health Care System
  – Department of Epidemiology, University of Utah

• Funding Support
  – VA Informatics and Computing Infrastructure VA HSR HIR 08-204
  – VA Consortium for Health Informatics Research VA HSR HIR 08-374


