Developing an “Academic” EHR form the Vanderbilt Synthetic Derivative

Cynthia S. Gadd
Christoph U Lehmann
Academic EHR

- Trainees need practice with EHRs to learn to
  - Extract important data
  - Differentiate between valuable and useless information
  - Develop new content
  - Understand pitfalls of EHRs
    - Duplication
    - Noise
    - Hidden information
Synthetic Derivative

- Database with Clinical Information
  - Derived from Vanderbilt’s EHR
    - Starpanel
    - Enterprise Data Warehouse
      - EPIC (ADT)
      - Medipac (ADT, billing)
      - WIZ order
  - Uncoupled from identified medical records
Synthetic Derivative

• How?
  • Continuously updated from clinical data
  • De-identified

• What
  • Image of clinical data

• Why
  • For research purposes
Synthetic Derivative

• Data from 2.2 million individuals
• De-identification
  • primarily commercial software
    • applied & assessed for effective scrubbing of identifiers
    • Example:
      • “John Smith” appeared in the original medical record
      • SD corresponding record replaced with a tag [NAMEAAA, BBB]
      • Post processing – replaced with random name
  • Uses US Census name list, regional phone books, street directories, voter lists, Vanderbilt staff/faculty/student lists
  • maintains the semantic integrity of the text
  • Dates replaced with a randomly generated date
    • All other dates shifted with identical delta
## Examples under-marked

<table>
<thead>
<tr>
<th>Pre-scrub</th>
<th>After scrub</th>
<th>Error Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rx for Lortab 10, #60 w/ one refill 12/8/4</td>
<td>Rx for Lortab 10, #60 w/ one refill 12/8/4</td>
<td>Date (Complete but malformed)</td>
</tr>
<tr>
<td>SOCIAL HISTORY: He currently lives at 77 Spruce Loop; Crossville, Tennessee</td>
<td>SOCIAL HISTORY: He currently lives at 77 Spruce Loop; **PLACE, Tennessee</td>
<td>Street Address</td>
</tr>
<tr>
<td>DATE OF BIRTH: 02/22/1912</td>
<td>DATE OF BIRTH: <strong>DATE[Jun 22 1912].</strong></td>
<td>Age Over 90</td>
</tr>
<tr>
<td>number of the ventilator is 98141 Patient being monitored with oximetry The</td>
<td>number of the ventilator is 98141 Patient being monitored with oximetry The</td>
<td>Device ID</td>
</tr>
<tr>
<td>Severe Left Thigh Hematoma (Traumatic) 6/00</td>
<td>Severe Left Thigh Hematoma (Traumatic) 6/00</td>
<td>Partial date</td>
</tr>
</tbody>
</table>
# Examples over-marked

<table>
<thead>
<tr>
<th>Pre-scrub</th>
<th>After scrub</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI: soft, ND, normal bowel sounds, non tender, no hepatomegaly, no splenomegaly</td>
<td>GI: **PLACE, ND, normal bowel sounds, non tender, no hepatomegaly, no splenomegaly</td>
</tr>
<tr>
<td>with iron, 40 gm protein daily, and <strong>PLACE</strong> as a calorie source</td>
<td>with iron, 40 gm protein daily, and <strong>ID-PAT</strong> calories daily.</td>
</tr>
<tr>
<td>Standardized Balance Tests: BERG Total score: 34 Pt required frequent rest</td>
<td>Standardized Balance Tests: <strong>NAME[XXX: WWW]</strong> score: 34 Pt required frequent rest</td>
</tr>
<tr>
<td>An attending Cardiologist was present throughout the diagnostic study.</td>
<td>An attending <strong>NAME[SSS]</strong> was present throughout the diagnostic study.</td>
</tr>
<tr>
<td>filled through the Easter Seals. The patient is also requesting additional</td>
<td>filled through the The patient is also requesting additional</td>
</tr>
</tbody>
</table>
Synthetic Derivative

- Search interface
  - Input
    - basic clinical and demographic information
      - age and gender
    - ICD 9 codes
    - CPT procedure codes
    - Medications
    - lab values
  - Return
    - de-identified data for review and selection
BioVU

- Biobank
  - contains de-identified DNA
  - extracted from leftover blood after clinically-indicated testing of Vanderbilt patients
    - who have not opted out
    - Linked to Synthetic Derivative
  - Future expectation of other tissue types:
    - serum proteomics, possibly surgical tissues
Academic EHR

- HIT training is important in the training process for future clinicians
- What is needed:
  - Safe training environment
    - No accidental deletion
    - No accidental addition to actual patient records
  - Protection of privacy
  - Verisimilitude (thank you, Harold)
Feedback SD content to Starpanel

SMITH, HELLEN (02/01/1949 - 56YO F) <999-99-9999> (555) 556-5555 Alert PCP: Mary, Johanson

Oncology Clinic Note 2004/09/28 14:09 By: Maredith Carter-Grant, M.D. Signed by: ********

DIAGNOSIS: Stage II invasive mammary breast cancer “T2 N0 M0.”

ONCOLOGIC HISTORY: Ms. Smith is a 55-year-old female who is post menopausal who was found to have an abnormality on her mammogram. She subsequently had an ultrasound-guided FNA which showed malignant cells. She was referred to the breast Center where she underwent a core biopsy on August 15, 2004, which showed infiltrating mammary carcinoma. She was subsequently seen by Dr. Owens who, on August 15, 2004, did a left modified radical mastectomy. Pathology from this revealed an invasive mammary carcinoma, no special type, with lobular features, 2.0 cm in greatest dimension, which was intermediate combined histologic grade with low proliferative rate tumor, extending to 1.9 mm in the lower, lateral, deep margin. There was no evidence of lymphovascular invasion present. Thirteen lymph nodes were negative for malignancy. Her tumor was ER positive, PR 1% positive, HER2/neu negative. She, at the time of surgery, had placement of a tissue expander, for immediate first stage reconstruction of her left breast, by Dr. McDonald. It was decided, since her final pathology showed tumor extending to 1.8 mm from the lower, lateral deep margin, that she be referred to Wilbur Slouse who was planning on doing radiation therapy after she received chemotherapy. She had a MUGA scan done on September 28, 2004, which showed a normal ejection fraction with a left ventricular ejection fraction of 68%. She is here to receive her first cycle of Adriamycin and Cytotoxan. We discussed the risks and benefits of chemotherapy and she has decided to proceed with chemotherapy.
Future Goal

• Sandbox in Starpanel
• Populated with Synthetic Derivative content
• Allows for training and experimentation
Thank you