Health Informatics

Lecture 3

Samantha Kleinberg
samantha.kleinberg@stevens.edu
Who accesses/creates data?

- Nurses
- Clinicians
- Patients

EHR

?
EHR components

• Patient data (view/entry)
• Ordering
• Decision support
• Communication
Surgery Service, Dr. Jones

S: No further Chest Pain or Shortness of Breath. "Feeling better today." Patient reports headache.

O: Afebrile, P 84, R 16, BP 130/82. No acute distress.

   Neck no JVD, Lungs clear
   Cor RRR
   Abd Bowel sounds present, mild RLQ tenderness, less than yesterday. Wounds look clean.
   Ext without edema

A: Patient is a 37 year old man on post-operative day 2 for laparoscopic appendectomy.

P: Recovering well. Advance diet. Continue to monitor labs. Follow-up with Cardiology within three days of discharge for stress testing as an out-patient. Prepare for discharge home tomorrow morning.

Associated signs and symptoms: «*»

Review of Symptoms:

Constitutional: negative ROSNEGGEN «ROSGener...»

Eyes: negative ROSNEGEYES «ROSEyes...»

Ears: negative ROSNEGears

Nose/Mouth/Throat: negative ROSNEGnmt «ROSENMT...»

Cardiovascular: negative ROSNEGcardio «ROSCvs...»

Respiratory: negative ROSNEGresp «ROSResp...»
Order entry
Types of data

• Narrative text
  – Medical history, description of current illness, family history
  – Format may be semi-structured template or free text

• Numerical values
  – Laboratory tests, vital signs, some aggregate measures

• Image
  – X-ray, ultrasound

• Other
  – Genomic
  – Medications
Structuring data

• Controlled vocabularies
  – Term list: set of concepts, no overlap
  – Ontologies: concepts and structure

• Why do we need controlled vocabularies?
Some major vocabularies

• ICD9/10 – diagnosis codes
• SNOMED-CT (Systematized NOmenclature of MEDicine)
• RxNorm – drug names
• MeSH (medical subject heading)
• LOINC (Logical Observation Identifier Names and Codes)
• DSM – mental disorders
ICD9

428 Heart failure
- 428.0 Congestive heart failure, unspecified convert 428.0 to ICD-10-CM
- 428.1 Left heart failure convert 428.1 to ICD-10-CM
- 428.2 Systolic heart failure
  - 428.20 Systolic heart failure, unspecified convert 428.20 to ICD-10-CM
  - 428.21 Acute systolic heart failure convert 428.21 to ICD-10-CM
  - 428.22 Chronic systolic heart failure convert 428.22 to ICD-10-CM
  - 428.23 Acute on chronic systolic heart failure convert 428.23 to ICD-10-CM
- 428.3 Diastolic heart failure
  - 428.30 Diastolic heart failure, unspecified convert 428.30 to ICD-10-CM
  - 428.31 Acute diastolic heart failure convert 428.31 to ICD-10-CM
  - 428.32 Chronic diastolic heart failure convert 428.32 to ICD-10-CM
  - 428.33 Acute on chronic diastolic heart failure convert 428.33 to ICD-10-CM
- 428.4 Combined systolic and diastolic heart failure
  - 428.40 Combined systolic and diastolic heart failure, unspecified convert 428.40 to ICD-10-CM
  - 428.41 Acute combined systolic and diastolic heart failure convert 428.41 to ICD-10-CM
  - 428.42 Chronic combined systolic and diastolic heart failure convert 428.42 to ICD-10-CM
  - 428.43 Acute on chronic combined systolic and diastolic heart failure convert 428.43 to ICD-10-CM
- 428.9 Heart failure, unspecified convert 428.9 to ICD-10-CM

ICD10

Codes
- I50  Heart failure
  - I50.1  Left ventricular failure, unspecified
  - I50.2  Systolic (congestive) heart failure
    - I50.20  Unspecified systolic (congestive) heart failure
    - I50.21  Acute systolic (congestive) heart failure
    - I50.22  Chronic systolic (congestive) heart failure
    - I50.23  Acute on chronic systolic (congestive) heart failure
  - I50.3  Diastolic (congestive) heart failure
    - I50.30  Unspecified diastolic (congestive) heart failure
    - I50.31  Acute diastolic (congestive) heart failure
    - I50.32  Chronic diastolic (congestive) heart failure
    - I50.33  Acute on chronic diastolic (congestive) heart failure
  - I50.4  Combined systolic (congestive) and diastolic (congestive) heart failure
    - I50.40  Unspecified combined systolic (congestive) and diastolic (congestive) heart failure
    - I50.41  Acute combined systolic (congestive) and diastolic (congestive) heart failure
    - I50.42  Chronic combined systolic (congestive) and diastolic (congestive) heart failure
    - I50.43  Acute on chronic combined systolic (congestive) and diastolic (congestive) heart failure
  - I50.8  Other heart failure
    - I50.81  Right heart failure
      - I50.810  ...... unspecified
      - I50.811  Acute right heart failure
      - I50.812  Chronic right heart failure
      - I50.813  Acute on chronic right heart failure
      - I50.814  ...... due to left heart failure
    - I50.82  Biventricular heart failure
    - I50.83  High output heart failure
    - I50.84  End stage heart failure
    - I50.89  Other heart failure
  - I50.9  Heart failure, unspecified
ICD10

http://www.healthcareitnews.com
Claims data

- ICD9/10 codes

- Coder assigns billing codes for visit based on info in EHR

- Many studies use only claims data.
  - What’s the problem with that?
Measuring Diagnoses: ICD Code Accuracy
What does it mean for a code to be accurate?

• Compare expert assessments of same record
  – how reliable is code

• Compare EHR coder to physician assessment
  – How well do clinical assessments match billing
Are ICD10 codes more accurate?

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In Charlson Index</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>12.8</td>
<td>9.6</td>
<td>8.4</td>
<td>3.2</td>
<td>4.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>8.1</td>
<td>4.6</td>
<td>4.5</td>
<td>3.5</td>
<td>3.6</td>
<td>.642</td>
</tr>
<tr>
<td>Rheumatic disease</td>
<td>2.6</td>
<td>1.5</td>
<td>1.4</td>
<td>1.1</td>
<td>1.2</td>
<td>.683</td>
</tr>
<tr>
<td>Dementia</td>
<td>3.3</td>
<td>1.1</td>
<td>2.4</td>
<td>2.2</td>
<td>0.9</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>In Elixhauser Index</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>21.8</td>
<td>9.4</td>
<td>9.1</td>
<td>12.4</td>
<td>12.7</td>
<td>.241</td>
</tr>
<tr>
<td>Pulmonary circulation disorders</td>
<td>2.7</td>
<td>1.6</td>
<td>1.6</td>
<td>1.1</td>
<td>1.1</td>
<td>.578</td>
</tr>
<tr>
<td>Valvular disease</td>
<td>7.0</td>
<td>3.2</td>
<td>3.5</td>
<td>3.8</td>
<td>3.5</td>
<td>.134</td>
</tr>
<tr>
<td>Hypertension</td>
<td>30.2</td>
<td>25.2</td>
<td>22.2</td>
<td>5.0</td>
<td>8.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>8.8</td>
<td>6.2</td>
<td>3.7</td>
<td>2.6</td>
<td>5.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
<td>0.1</td>
<td>0.2</td>
<td>.157</td>
</tr>
<tr>
<td>Solid tumor without metastasis</td>
<td>9.5</td>
<td>7.4</td>
<td>7.4</td>
<td>2.1</td>
<td>2.1</td>
<td>.736</td>
</tr>
<tr>
<td>Renal failure</td>
<td>4.0</td>
<td>4.6</td>
<td>4.9</td>
<td>-0.6</td>
<td>-0.9</td>
<td>.180</td>
</tr>
<tr>
<td>Blood loss anemia</td>
<td>1.1</td>
<td>0.7</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
<td>.858</td>
</tr>
<tr>
<td>Deficiency anemia</td>
<td>1.9</td>
<td>1.9</td>
<td>1.4</td>
<td>0</td>
<td>0.5</td>
<td>.011</td>
</tr>
<tr>
<td>Coagulopathy</td>
<td>7.7</td>
<td>1.8</td>
<td>1.8</td>
<td>5.9</td>
<td>5.9</td>
<td>1.000</td>
</tr>
<tr>
<td>Fluid and electrolyte disorders</td>
<td>11.1</td>
<td>6.1</td>
<td>5.6</td>
<td>5.0</td>
<td>5.5</td>
<td>.089</td>
</tr>
</tbody>
</table>

- Calculate how often code mentioned given opportunities (dates with some ICD9 codes recorded)

- Use distribution of proportion to calculate differential entropy

- Entropy captures variability/uncertainty

- Had physicians create gold standard (categorizing as chronic/acute), used majority
Fig. 5. Documentation probability over time for all diabetes patients, those with greater than 20 positive mentions, and those with less than or equal to 20 positive mentions.
Search Results

- Arterial bypass graft (procedure)
- Operative procedure on coronary artery (procedure)
- Repair of heart (procedure)
- Thoracic artery repair (procedure)

Name: Coronary artery bypass grafting (procedure)
Concept ID: 232717009

- Allograft bypass of coronary artery (procedure)
- Aortocoronary artery bypass graft with saphenous vein graft (procedure)
- Aortocoronary bypass grafting (procedure)
- Connection of mammary artery to coronary artery (procedure)
- Coronary artery bypass graft x 1 (procedure)
- Coronary artery bypass graft, anastomosis of artery of thorax to coronary artery (procedure)
- Coronary artery bypass grafts greater than 5 (procedure)
- Coronary artery bypass grafts x 2 (procedure)
- Coronary artery bypass grafts x 3 (procedure)
- Coronary artery bypass grafts x 4 (procedure)
- Coronary artery bypass grafts x 5 (procedure)
- Emergency coronary artery bypass graft (procedure)
- Prosthetic bypass of coronary artery (procedure)
SNOMED CT Components, Hierarchies and Outputs

SNOMED CT DESIGN

<table>
<thead>
<tr>
<th>Concept Id</th>
<th>Descriptions</th>
<th>Description type</th>
<th>US English Language Reference Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>385627004</td>
<td>Cellulitis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118932009</td>
<td>Disorder of foot</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **385627004 Cellulitis**
- **118932009 Disorder of foot**

- **62837005 Cellulitis of hand**
- **128276007 Cellulitis of foot**
- **95345008 Ulcer of foot**

MI – Myocardial infarction

Synonym

High granularity
DISCHARGE SUMMARY

Name: [**Known patient lastname**], [**Known patient firstname**]

[**Unit Number 626**]

Admission Date: [**2016-11-07**]

Discharge Date: [**2016-11-22**]

Date of Birth: [**1972-09-20**]

Sex: F

HISTORY OF PRESENT ILLNESS: Patient is a 44-year-old lady status post living related kidney transplant on [**2016-10-19**], who presented at [**Hospital 36**] for end-stage renal disease secondary to type 1 diabetes mellitus.

She presented to [**Hospital1**] on [**2016-11-07**] with increased drainage from her surgical wound and JP, increased abdominal pain, and anuria x4 days. The patient reported constipation for a week. She denies flatus. She was complaining of nausea and vomiting. Her abdominal pain had become progressively worse left lower quadrant most notable. There is no radiation to the back or elsewhere. She denied any fevers, chills. She noted decreased p.o. intake recently. Her drainage from her wound incision and JP was notable for yellowish clear urine smelling fluid.
10/18 7AM-11PM

S/V: Pt has had a very eventful day. At 6:45 AM he was noted to have SBP 40's by NBP, with HR 60's. Initially responsive, but rapidly decreasing responsiveness followed by respiratory arrest. Pt was intubated with 100% F102, then **Month/Day**. An A-line was placed; we have consistently been able to easily draw blood from the line, but it appears dampened and reads quite a bit lower than the NBP, so we have been using the NBP all day. He soon required pressors for SBP 70's. He was started initially on Neo, but was titrated up to a max of 120 mcg/min with little if any effect. He was then started on Levo. Over several hours, with some difficulty, the Neo was weaned off to with the Levo as high as 40 mcg/min. He was transiently on Dopa, as high as 10 mcg/kg/min, but it was soon D/C'd 8/c HR into the 140's. Around 1PM his BP again began to fall into the 50's. His extremities were cold, and HR dropped into the 60's again. He was given 250cc fluid bolus, and Dopa was again attempted, at a lower dose. This time, however, he began to have lots of ventricular ectopy, including short runs of VT. Dopa was again D/C'd. Levo increased more, and he again stabilized for a few hours. About 7:45 he suddenly went into sustained VT. A-line tracing was flat (though is has never been reliable). In the interest of saving time, a cuff pressure was not checked. He was unresponsive, and was defibrillated once with 200J. He converted initially to ST with lots of ectopy, then settled down into NSR after a few minutes. He has remained in NSR since. BP is borderline on high-dose Levo. EKG shows ST depressions, but not much changed from yesterday. CK's, Troponin added to earlier labs.

F/E: Pt is dialysis-dependant. He has had >2.5L fluid since MN, and will be dialyzed tomorrow. Lytes have been followed closely. Mg repleted after episode of VT, and he has been given 15g Kaexolate for borderline hyperkalemia.

NEURO: Pt initially unresponsive this AM. Over the day he has been agitated with ANY intervention. Initially well sedated on **Month/Day 15***, but he was changed to Fentanyl gtt w/prn Ativan to try to avoid hypotension from the **Month/Day 15***. Fentanyl has increased a couple of times. He is OK when left alone, but easily agitated.

**Month/Day**: Hct 30-32, stable. Coags greatly elevated with INR 5.1 this AM. He was given 2mg Vit K SQ, but coags worse afterwards. No further intervention at present.

GI: Vomitted brown O+ material both before and after intubation. Belly soft, obese, obviously tender. Too unstable to go to CT. Plan was for U/S, but he was hypotensive to 50's when they came, so it was deferred. Medium loose brown, foul-smelling stool this AM (sent for C-diff). On Protonix.

ID: Temp rising to max of 101.7 this evening. He has been fully cultured and is on multiple abx. Amphotericin was up when he arrested this AM and was stopped with half of it infused. He did not receive the rest.....HN aware. WBC 30-40K, Lactate has risen to 7.9. He has a worsening metabolic acidosis, with bicarb now down to 12.

RESP: Intubated, vented. Current settings A/C .5/750/24/PEEP 5. ABG's show adequate oxygenation, compensated metabolic acidosis. LS diminished. He has minimal secretions, but he was found to have green beans in the back of his throat on intubation, and we suctioned a few pieces out...none since this AM.

SKIN: He has 2 small decubiti on buttocks, covered with Duoderm. Also has open area in left groin.

ACCESS: A-line as described above. He has a right femoral tunnelled **Male First Name Last Name** catheter. A cuffed left EJ line was removed this AM. Multiple attempts at other access have been made by many people without success.

SOCIAL: Pt has a sister **Last Name (un) 140** who was in. He also has a very involved home care nurse named **First Name 6 141** **Last Name 142** who was extremely upset about his condition. She was in to visit this evening, and was here for the 6th episode. The pt's lawyer also came briefly. He does not have a proxy; SW notified by case manager of his admission, serious condition, and need for proxy determination.

A: septic shock with multiple potential sources.

P: continue abx, follow cx results. Support BP and resp as needed. Follow labs closely. Anticipate possible need for CVP/HD is does not tolerate HD. SW consult for proxy.
Co-codamol was prescribed. This markedly reduced the pain condition.

CLEF Corpus: Semantic Annotation of Clinical Text
NLP task examples

• Which patients met Framingham CHF criteria?
• What drugs and symptoms are often mentioned together?

• Question Answering
• Summarization
NLP approaches

• Bag of words
• Linguistic rules
• Data driven
Available data

• MIMIC II deidentified EHR data + notes
  – ShARe/CLEF
  – https://sites.google.com/site/shareclefehealth/data
  – http://clefehealth2014.dcu.ie

• i2b2 challenges
  – Relations, coreference, temporal statements
  – https://www.i2b2.org/NLP/DataSets/Main.php
Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies

Raphael Cohen¹, Michael Elhadad¹ and Noémie Elhadad²

Abstract

Background: The increasing availability of Electronic Health Record (EHR) data and specifically free-text patient notes presents opportunities for phenotype extraction. Text-mining methods in particular can help disease modeling by mapping named-entities mentions to terminologies and clustering semantically related terms. EHR corpora, however, exhibit specific statistical and linguistic characteristics when compared with corpora in the biomedical literature domain. We focus on copy-and-paste redundancy: clinicians typically copy and paste information from previous notes when documenting a current patient encounter. Thus, within a longitudinal patient record, one expects to observe heavy redundancy. In this paper, we ask three research questions: (i) How can redundancy be quantified in large-scale text corpora? (ii) Conventional wisdom is that larger corpora yield better results in text mining. But how does the observed EHR redundancy affect text mining? Does such redundancy introduce a bias that distorts learned models? Or does the redundancy introduce benefits by highlighting stable and important subsets of the corpus? (iii) How can one mitigate the impact of redundancy on text mining?
NLP challenges

• Time
• Negations, Uncertainty
• Context
  – Obese vs. morbidly obese
  – Distinguishing semantics (MI = myocardial infarction, mitral insufficiency)
• Resolving references
Levels of granularity

A hybrid knowledge-based and data-driven approach to identifying semantically similar concepts

Rimma Pivovarov, Noémie Elhadad *

Department of Biomedical Informatics, Columbia University, 622 W. 168th Street, VC-5, New York, NY 10032, USA
Research and applications

Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis

Swapna Abhyankar, Dina Demner-Fushman, Fiona M Callaghan, Clement J McDonald

+ Author Affiliations

Correspondence to
Dr Swapna Abhyankar, National Library of Medicine, Lister Hill National Center for Biomedical Communications, 8600 Rockville Pike, Building 38A/9N917, Bethesda, MD 20894, USA; swapna.abhyankar@nih.gov

Received 12 April 2013
Revised 30 October 2013
Accepted 13 December 2013
Published Online First 2 January 2014
Personal health record

Basic idea: patient managed, can be shared with others

Types:

- Patient access to part of EHR (institution-based)
- Self-maintained (patient creates their history/updates all info)
- Export EHR to 3\textsuperscript{rd} party site, adds data (e.g. Microsoft Health Vault)
Now, with a few years of experience, we’ve observed that Google Health is not having the broad impact that we hoped it would. There has been adoption among certain groups of users like tech-savvy patients and their caregivers, and more recently fitness and wellness enthusiasts. But we haven’t found a way to translate that limited usage into widespread adoption in the daily health routines of millions of people. That’s why we’ve made the difficult decision to discontinue the Google Health service. We’ll continue to operate the Google Health site as usual through January 1, 2012, and we’ll provide an ongoing way for people to download their health data for an additional year beyond that, through January 1, 2013. Any data that remains in Google Health after that point will be permanently deleted.
Online communities

www.patientslikeme.com
PHRs and privacy

• Because they’re not medical providers, external/linked PHR not subject to HIPAA
• Provider PHRs ARE covered by HIPAA
Who’s using PHRs?

HEALTH CARE REFORM

The Digital Divide in Adoption and Use of a Personal Health Record

Cyrus K. Yamin, BS; Srinivas Emani, PhD; Deborah H. Williams, MHA; Stuart R. Lipsitz, ScD; Andrew S. Karson, MD, MPH; Jonathan S. Wald, MD, MPH; David W. Bates, MD, MSc
Intensity of use

• 51% low use (0-1 log ins over 2 years)
• 27% high use (10+ log ins)
  – 41% of these were age 51-65

• Common message types
  – Question about care, medication refills, referrals/appointments

• More comorbidities = more messages
PHR pros/cons

- Maintain and reuse information (HIE)
- Patients can correct errors, fill in gaps
  - Centralized, can deal with multiple providers
- Better understanding of health

- Barriers to use/access
- Data quality
  - Patients introducing errors, removing information
- Security
- Patient concern/ not understanding data
For next week

• For this Friday, reading as usual
• For Monday, complete human subjects training, submit to MIMIC to gain access to data
  – See Canvas, follow MIMIC instructions EXACTLY. If you take the wrong course, you won’t be able to access the data.