Summarizing the Patient Record & Modeling Diseases from EHR Observations

Noémie Elhadad
noemie.elhadad@columbia.edu
Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy

CHAPTER SEVEN

Electronic Health Records ................................................................. 33
Overview of Findings .................................................................... 33
Qualitative Findings ...................................................................... 33
Improved Professional Satisfaction: EHRs Facilitate Better Access to Patient Data ............................... 34
Improved Professional Satisfaction: EHRs Improve Some Aspects of Quality of Care ....................... 34
Improved Professional Satisfaction: Better Communication with Patients and Between Providers ................................................................. 35
Worsened Professional Satisfaction: Time-Consuming Data Entry ................................................. 35
Worsened Professional Satisfaction: User Interfaces That Do Not Match Clinical Workflow ............ 37
Worsened Professional Satisfaction: Interference with Face-to-Face Care ..................................... 38
Worsened Professional Satisfaction: Insufficient Health Information Exchange ........................... 39
Worsened Professional Satisfaction: Information Overload ............................................................. 40
Worsened Professional Satisfaction: Mismatch Between Meaningful-Use Criteria and Clinical Practice .................................................................................................................. 40
Worsened Professional Satisfaction: EHRs Threaten Practice Finances ........................................ 41
Worsened Professional Satisfaction: EHRs Require Physicians to Perform Lower-Skilled Work ........ 42
A revolt is brewing among doctors and hospital administrators over electronic medical records systems mandated by one of President Obama’s early health care reforms.

The American Medical Association called for a “design overhaul” of the entire electronic health records system in September because, said AMA president-elect Steven Stack, electronic records “fail to support efficient and effective clinical work.”

That has “resulted in physicians feeling increasingly demoralized by technology that interferes with their ability to provide first-rate medical care to their patients,” Stack said.
Today we’ll be talking about

- What’s the story of the patient I am taking care of? What is my patient at risk for?

- Information overload
- (Disease Progression Prediction)
- Summarization of patient record
- Modeling diseases from EHR observations
Information overload

• Present at all levels of care
  – Primary / inpatient / emergency care
  – Health information exchange

• EHR data is cognitively taxing to navigate
  – Lots of it
  – Heterogeneous data
  – Primarily organized chronologically

What’s my patient at risk for?

• Disease progression prediction
  – Chronic kidney disease (CKD)
  – Difficult for clinicians, because of uncertainty, but also information overload

• State of the art for risk prediction models for CKD
  – Varying model type (Logistic, Cox)
  – Varying features (demographics, eGFR, diagnoses, laboratory tests)
  – Varying outcomes (creatinine, eGFR, complications, kidney failure)
Our goal

• Use *longitudinal*, *heterogeneous* data sources to predict risk of a near-term CKD outcome that should be sensitive to short-term medical decisions.

• In contrast to previous studies, we:
  – Use EHR data
  – Use longitudinal data (up to 20 years back)
  – Use heterogeneous data (demographics, labs, notes)
  – Use stage III CKD as a trigger for prediction and stage IV as the outcome.

Data + Models

• Data
  – ~20k patients visiting primary care clinic
  – ~3k with stage III CKD and ~307 with stage IV CKD

• 5 predictive models compared – all incorporated into a basic Cox
  – eGFR – Estimated glomerular filtration rate
  – RLT – Recent Laboratory tests
  – TKF – Text Kalman filter
  – LKF – Laboratory test Kalman filter
  – LTKF – Laboratory test and Text Kalman filter
eGFR and RLT (recent lab tests) models
TKF (Text Kalman Filter)
LKF (Lab Kalman Filter)

Stage III CKD

Stage IV CKD

Prediction

Time
LTKF (Lab & Test Kalman Filter)
Methods

• Component models
  – Model of text (latent Dirichlet allocation (LDA))
    • K=50
  – Model of the past (Kalman Filter)
    • Discrete time, binned by month, observations included 19 laboratory values and notes represented as log transformation of topic proportions
  – Model of the future (Cox proportional hazards)
    • Covariates include Kalman filter latent values at stage III onset, Kalman filter offsets, and demographics.
    • Dependent variable is time to stage IV
## Results

<table>
<thead>
<tr>
<th></th>
<th>Δ LTKF</th>
<th>Δ LKF</th>
<th>Δ TKF</th>
<th>Δ RLT</th>
<th>Δ eGFR</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTKF</td>
<td></td>
<td></td>
<td>***</td>
<td>*</td>
<td>***</td>
<td>0.849</td>
</tr>
<tr>
<td>LKF</td>
<td></td>
<td></td>
<td>***</td>
<td></td>
<td>**</td>
<td>0.836</td>
</tr>
<tr>
<td>TKF</td>
<td></td>
<td></td>
<td></td>
<td>***</td>
<td></td>
<td>0.733</td>
</tr>
<tr>
<td>RLT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>0.819</td>
</tr>
<tr>
<td>eGFR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.779</td>
</tr>
</tbody>
</table>

*=p<0.05, **=p<0.01, ***=p<0.001
## Results – risk factors

<table>
<thead>
<tr>
<th>Topic 3 (heart failure)</th>
<th>Topic 32 (diabetes)</th>
<th>Topic 29 (dialysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lasix</td>
<td>units</td>
<td>q15</td>
</tr>
<tr>
<td>volume</td>
<td>insulin</td>
<td>Dialysis</td>
</tr>
<tr>
<td>edema</td>
<td>subcutaneous</td>
<td>Fistula</td>
</tr>
<tr>
<td>heart</td>
<td>lantus</td>
<td>Volume</td>
</tr>
<tr>
<td>failure</td>
<td>glucose</td>
<td>Bid</td>
</tr>
<tr>
<td>worsening</td>
<td>diabetes</td>
<td>Lasix</td>
</tr>
<tr>
<td>diuresis</td>
<td>times</td>
<td>Placement</td>
</tr>
<tr>
<td>severe</td>
<td>70/30</td>
<td>Improved</td>
</tr>
<tr>
<td>diastolic</td>
<td>diabetic</td>
<td>Heparin</td>
</tr>
<tr>
<td>overload</td>
<td>days</td>
<td>Examined</td>
</tr>
</tbody>
</table>
## Results – protective factors

<table>
<thead>
<tr>
<th>Topic 33 (family history)</th>
<th>Topic 35 (health maintenance)</th>
<th>Topic 41 (non-specific)</th>
<th>Topic 43 (gynecological)</th>
<th>Topic 45 (asthma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>died</td>
<td>died</td>
<td>history</td>
<td>breast</td>
<td>Albuterol</td>
</tr>
<tr>
<td>age</td>
<td>flu</td>
<td>pressure</td>
<td>vaginal</td>
<td>Asthma</td>
</tr>
<tr>
<td>years</td>
<td>visit</td>
<td>rate</td>
<td>mammo</td>
<td>Inhaled</td>
</tr>
<tr>
<td>mother</td>
<td>fasting</td>
<td>count</td>
<td>cancer</td>
<td>Lung</td>
</tr>
<tr>
<td>father</td>
<td>colonoscopy</td>
<td>three</td>
<td>hx</td>
<td>obstructive</td>
</tr>
<tr>
<td>brother</td>
<td>year</td>
<td>revealed</td>
<td>pap</td>
<td>Wheezing</td>
</tr>
<tr>
<td>sister</td>
<td>shot</td>
<td>times</td>
<td>nl</td>
<td>Advair</td>
</tr>
<tr>
<td>worked</td>
<td>vaccine</td>
<td>shortness</td>
<td>age</td>
<td>Pulm</td>
</tr>
<tr>
<td>children</td>
<td>wnl</td>
<td>discharged</td>
<td>will</td>
<td>restrictive</td>
</tr>
<tr>
<td>deceased</td>
<td>check</td>
<td>creatinine</td>
<td>endometrial</td>
<td>Puffs</td>
</tr>
</tbody>
</table>
What about many diseases at once?

**Figure 1:** A comparison of standard survival analysis (top frame) and the survival filter (bottom frame). A filled circle represents an observed event, while an empty circle represents a censored one. In the case of standard survival analysis, patients in a cohort are aligned by an event. In the survival filter, patients are not aligned and unlike standard survival analysis, many conditions are considered simultaneously.

Predictive modeling on EHR data

• Incorporating longitudinal information helps
• Incorporating types of evidence (text, labs) helps

• Meaningful data science + EHR:
  – How to make this type of predictions useful for clinicians?
  – How to make them useful within their workflow?
Back to information overload

• Present at all levels of care
  – Primary / inpatient / emergency care
  – Health information exchange

• EHR data is cognitively taxing to navigate
  – Lots of it
  – Heterogeneous data
  – Primarily organized chronologically

Patient record summarization

“The act of collecting, distilling, and synthesizing patient information for the purpose of facilitating any of a wide range of clinical tasks”

Previous approaches

- focus on specific disease
- focus on specific care setting (ICU)
- largely ignore EHR text
- deployment and study of impact is lacking

How clinicians summarize patient information

92yo woman

How clinicians summarize patient information

Average time in each section of the EHR

On average, physicians spent
• 50% of their time in the Notes section
• 25% of their time in the Laboratory section

How clinicians summarize patient information

- All physicians visited the “Notes” section first
- No established ordering of summary content
  - Problem-oriented view of the patient

Functionality wish list for an EHR summarizer

• Aggregate information from the whole record
• But allow for zooming in and out of particular parts of the record
• Use notes as primary content selection source
• Facilitate finding supporting evidence in documentation
• Be problem oriented
• Be interactive
• Update in “real time”
HARVEST

- Extracts content from a patient’s longitudinal documentation
- Aggregates information from multiple care settings
- Visualizes content through a timeline of a patient’s problem documentation and clinical encounters
- Distributed computing infrastructure
- Deployed at New York-Presbyterian hospital

local harvest

The HARVEST system processes data from visits and HL7 message feeds through a map-reduce indexing process. The resulting data are stored in an HBase database, which contains tables for visits, notes, and problems:

- **Visits** (visit_id, MRN, dates, attending, visit_type, primary_ICD9)
- **Notes** (MRN, visit_id, note_id, note_type, date, author, text)
- **Problems** (MRN, note_id, CUI, lexical_item, char_offsets, salience)

The data is then made available for online web visualization through query summary content and Javascript visualization, which requires authentication with a physician's MRN.
Natural language processing of clinical documentation

• Extract problems mentioned in all the notes of a record
  – Conditions, as well as signs and symptoms
• Compute salience of problem documentation for a given time frame in a patient record

• Challenges
  – Robust processing across all note types
  – Identify and merge problems that are semantically similar
  – Handle redundancy within longitudinal record

Natural language processing of clinical documentation

• Distributed infrastructure
  – 650,000 notes/month avg. are authored at NYP
  – 20,000 notes/second parsing and indexing
    (compared to 500 notes/second in a non-distributed infrastructure)
Use cases

• “What’s the story?”
  – ED visit
  – Hospital admission
  – Walk-in at clinic

• Quality indicators
  – 2-hour on average per patient
  – HARVEST use shortens chart review by 20 mins on average (log analysis) and increases confidence of abstraction (survey)

• Researchers and trial coordinators
• Education
Next steps

• How to handle
  – Not mentions in documentation, but actual presence of a condition, based on all EHR observations
  – Conditions not diagnosed yet, but documentation supports their presence

• Need a mechanism to describe the presence of a disease for any time slice of a patient record
Disease modeling

- Isn’t there a list of problems somewhere in the EHR we can look up?
  - There are manually curated problem lists, but not guarantee they are filled or maintained by clinicians
  - Doesn’t handle yet-to-be diagnosed conditions
- Couldn’t we ask clinicians to describe each disease as a set of patient characteristics and go from there?
  - eMERGE PheKB
  - 42 diseases phenotyped so far
Phenotyping wish list

• Portable across institutions
• Data-Driven
• Not expert intensive
• Probabilistic
• Robust to very large datasets
• Robust to many diseases (up to 1000)
• Robust to many patients
Large-scale, probabilistic phenotyping

Patient Records
Structured and Unstructured

Learned Probabilistic phenotypes

Diabetes Mellitus
Congestive Heart Failure
Depressive Disorder
Joint Disorder
Lupus
Chronic Kidney Disease
Breast Cancer
Colon Cancer
Asthma
Hyperlipidemia
...

Inference mechanism
Uphenome (unsupervised)
Uphenome (unsupervised)
Uphenome (unsupervised)
Can we learn phenotypes across institutions and care settings?

<table>
<thead>
<tr>
<th></th>
<th>MIMIC - ICU Total / Unique</th>
<th>NYPH - Outpatient Total / Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Patients</em></td>
<td>18,697 / 18,697</td>
<td>9,828 / 9,828</td>
</tr>
<tr>
<td><em>Words</em></td>
<td>13,086,278 / 12,919</td>
<td>13,494,149 / 13,158</td>
</tr>
<tr>
<td><em>Medications</em></td>
<td>1,044,541 / 855</td>
<td>9,978 / 273</td>
</tr>
<tr>
<td><em>Lab Tests</em></td>
<td>7,499,446 / 309</td>
<td>351,992 / 300</td>
</tr>
<tr>
<td><em>Diagnoses</em></td>
<td>159,740 / 985</td>
<td>177,420 / 931</td>
</tr>
</tbody>
</table>
Experiments

• How good are the learned phenotypes
  – Physician-rated phenotype coherence
  – Physician-rated phenotype granularity
  – Physician-rated phenotype comparison to baseline (LDA-all)

• How well does the model infer phenotypes for unseen patients
  – Compare learned phenotypes to gold-standard annotations in notes
Experimental setup

80% of data used for training set, 20% for test set

Parameters

- \( P = 250 \)
- \( \alpha = 0.1 \)
- \( \mu, \nu, \xi, \pi = 0.1 \)
- \# training Gibbs sampling iterations = 7,000
- \# testing Gibbs sampling iterations = 1,000
Phenotype example

Words from notes written by clinicians:
- lupus
- ana
- sle
- complement
- rheum
- anti
- mg
- ab
- rash
- absent
- esr
- ulcers
- igg
- plaquenil
- dna
- alopecia
- wt
- antibody
- urine
- systematic
- dsdna
- neg
- rheumatology
- crp
- positive
- antimalarials
- metamucil
- prednisone
- c4_complement
- c3_complement
- esr
- rbc urine
- total_hemolytic_complement
- dna antibody igg
- crphi
- random_urea_protein
- antidna_antibodies
- urine_protein_random
- urine_creatinine
- random_urea_creatinine

Laboratory Tests Performed:
- 710.0_systemic_lupus_erythematosus

Medications prescribed:
- antimalarials
- metamucil
- prednisone
- c4_complement
- c3_complement

ICD9 codes billed:
- 710.0_systemic_lupus_erythematosus
Results – coherence

Distribution of Coherence Scores for Each Model

- Phenome Model
- LDA_all Model

Un-interpretable → Great
Results – granularity

Phenome Model

- Single disease: large portion
- Mix of diseases: small portion
- Non-disease: negligible portion

LDA-all

- Single disease: large portion
- Mix of diseases: small portion
- Non-disease: negligible portion
Results – comparison to baseline

80.4% of the time, the Phenome model was preferred.
Results – inference on unseen patients

Disorders that appear in a patient record
(assigned to patient, not negated, not generic)

vs.

Phenotypes that are inferred for that patient
Phenotype 7

Phenotype 7

mitral valve regurgitation repair severe
replacement mvr moderate tricuspid furosemide
potassium-chloride warfarin heparin-sodium docusate-sodium acetaminophen epinephrine
magnesium-sulfate milrinone potassium hct
hgb glucose sodium inr-pt plt-count creat mch magnesium ptt rdw mchc pt urea-n mcv rbc total-co2
wbc chloride 424.0-mitral-valve-disorders
398.91-rheumatic-heart-failure-congestive
397.0-diseases_of_tricuspid-valve
Conclusions

• Disease modeling
  – Leveraging heterogeneous data helps, but need for appropriate models

• EHR summarization
  – Robust NLP of underlying data
  – Information visualization
  – Computing infrastructure to enable operational summarization

• Virtuous circle
Thank you!

people.dbmi.columbia.edu/noemie/phenosum

National Library of Medicine R01 LM010027
National Science Foundation award #1344668