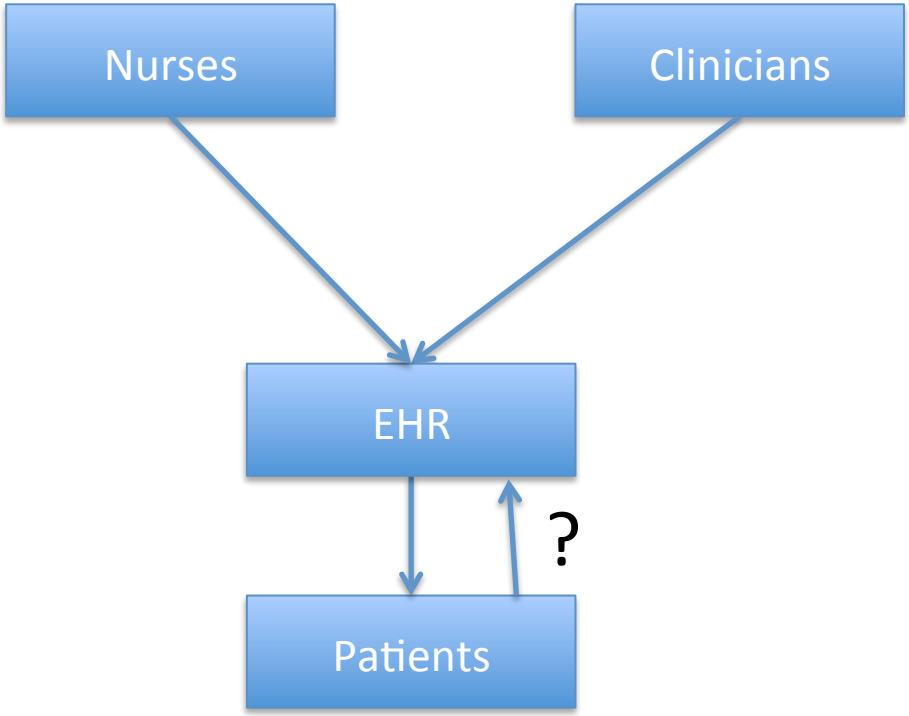
Health Informatics

Lecture 3

Samantha Kleinberg
samantha.kleinberg@stevens.edu

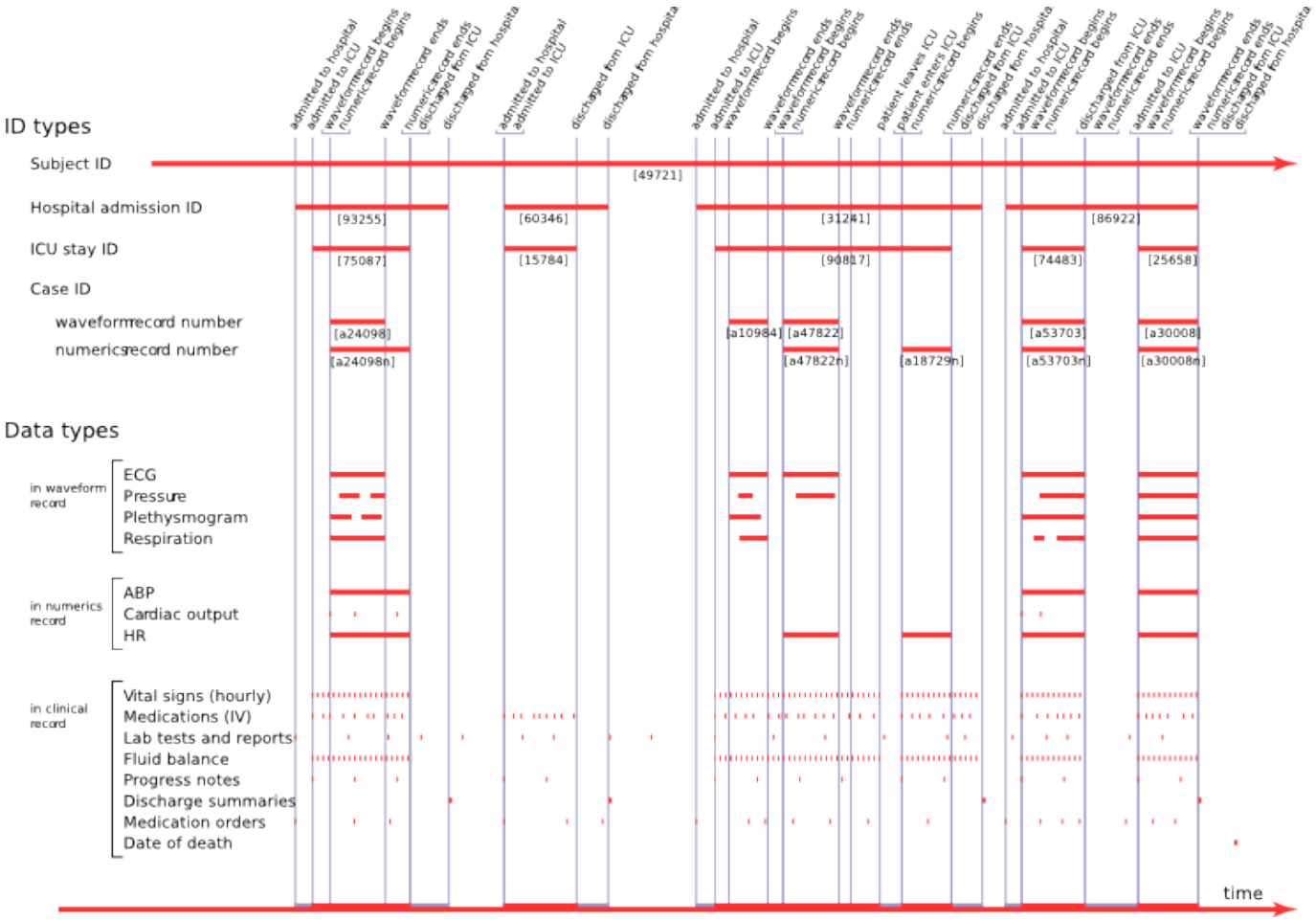
Who accesses/creates data?



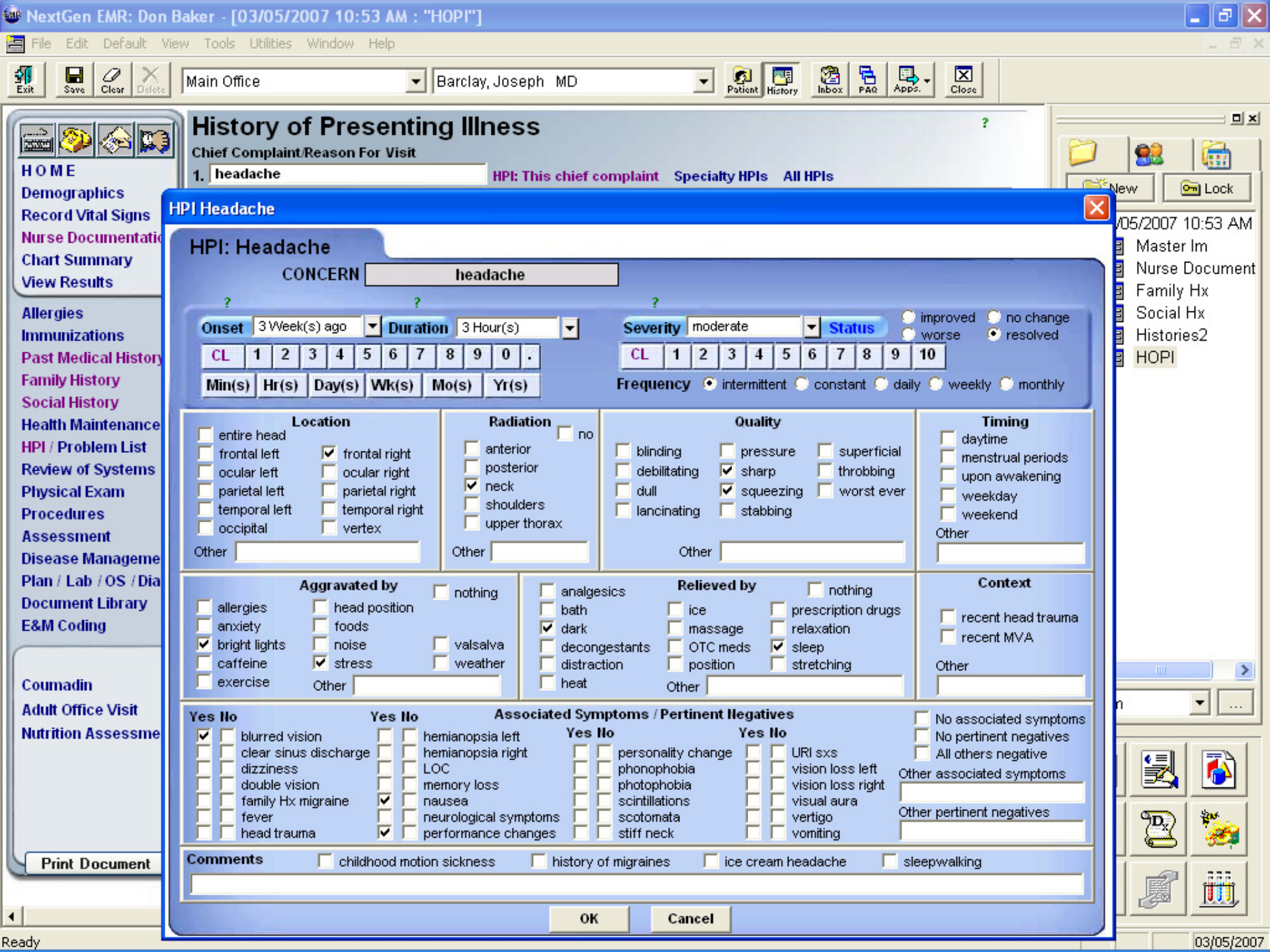
EHR components

- Patient data (view/entry)
- Ordering
- Decision support
- Communication

MIMIC – patient example



https://physionet.org/mimic2/UserGuide/node12.html


NextGen EMR: Don Baker - [03/05/2007 10:53 AM : "HOPI"]
File Edit Default View Tools Utilities Window Help
Exit Save Clear Delete
Main Office
Barclay, Joseph MD
Patient History Inbox PAQ Apps. Close
HOME
Demographics
Record Vital Signs
Nurse Documentatio
Chart Summary
View Results
Allergies
Immunizations
Past Medical History
Family History
Social History
Health Maintenance
HPI / Problem List
Review of Systems
Physical Exam
Procedures
Assessment
Disease Manageme
Plan / Lab / OS / Dia
Document Library
E&M Coding
Coumadin
Adult Office Visit
Nutrition Assessme
Print Document
History of Presenting Illness
Chief Complaint/Reason For Visit
1. headache
HPI: This chief complaint
Specialty HPIs
All HPIs
New
Lock
05/2007 10:53 AM
Master Im
Nurse Document
Family Hx
Social Hx
Histories2
HOPI
HPI Headache
HPI: Headache
CONCERN
headache
Onset
3 Week(s) ago
Duration
3 Hour(s)
Severity
moderate
Status
improved
no change
worse
resolved
CL 1 2 3 4 5 6 7 8 9 0 .
CL 1 2 3 4 5 6 7 8 9 10
Min(s) Hr(s) Day(s) Wk(s) Mo(s) Yr(s)
Frequency
intermittent
constant
daily
weekly
monthly
Location
entire head
frontal left
frontal right
ocular left
ocular right
parietal left
parietal right
temporal left
temporal right
occipital
vertex
Other
Radiation
no
anterior
posterior
neck
shoulders
upper thorax
Other
Quality
blinding
pressure
superficial
debilitating
sharp
throbbing
dull
squeezing
worst ever
lancinating
stabbing
Other
Timing
daytime
menstrual periods
upon awakening
weekday
weekend
Other
Aggravated by
nothing
allergies
head position
anxiety
foods
bright lights
noise
valsalva
caffeine
stress
weather
exercise
Other
Relieved by
nothing
analgesics
bath
ice
prescription drugs
dark
massage
relaxation
decongestants
OTC meds
sleep
distraction
position
stretching
heat
Other
Context
recent head trauma
recent MVA
Other
Associated Symptoms / Pertinent Negatives
Yes No
blurred vision
clear sinus discharge
dizziness
double vision
family Hx migraine
fever
head trauma
hemianopsia left
hemianopsia right
LOC
memory loss
nausea
neurological symptoms
performance changes
personality change
phonophobia
photophobia
scintillations
scotomata
stiff neck
URI sxs
vision loss left
vision loss right
visual aura
vertigo
vomiting
No associated symptoms
No pertinent negatives
All others negative
Other associated symptoms
Other pertinent negatives
Comments
childhood motion sickness
history of migraines
ice cream headache
sleepwalking
OK
Cancel
Ready
03/05/2007



SOAP

Vital Signs:

Vitals as of: 11/9/2012: Systolic 110: Diastolic 84: Heart Rate 65: Temp 98.3F: Respiratory Rate 16: O2 Saturation 98

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...»

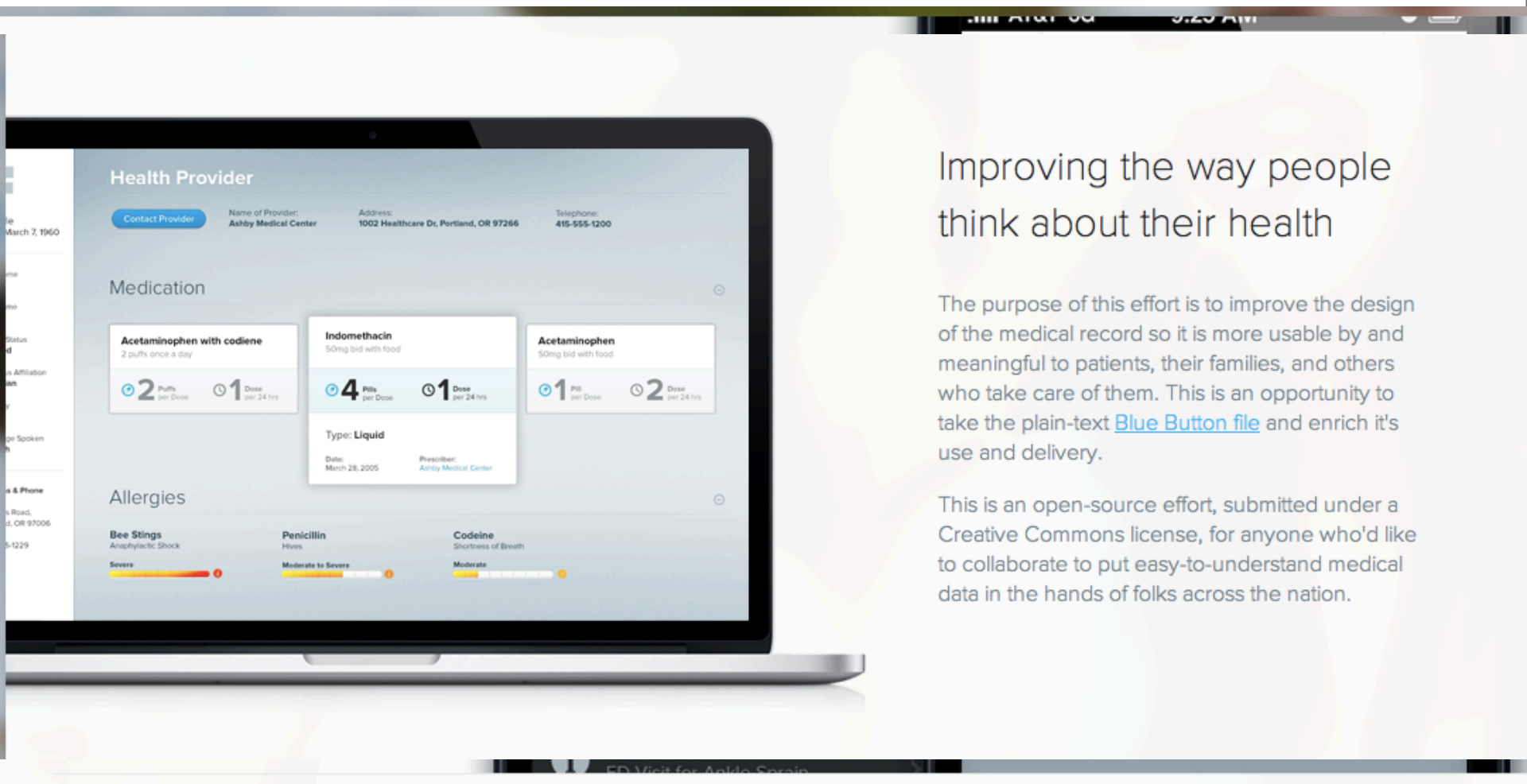
Eyes: negative ROSNEGEYES «ROSEyes...»

Ears: negative ROSNEGEARS

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

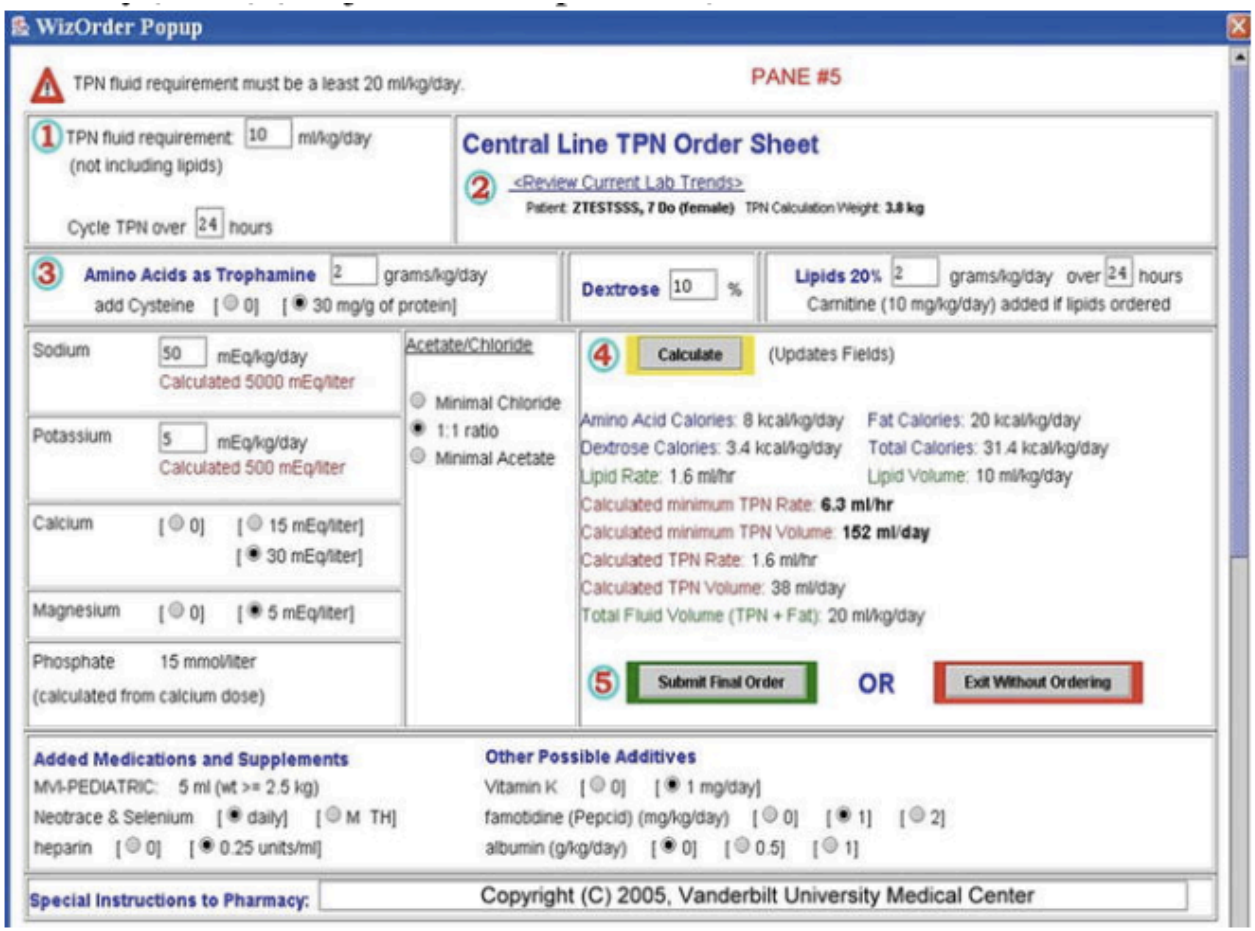
Cardiovascular: negative ROSNEGCARDIO«ROSCVS...»

Respiratory: negative ROSNEGRESP«ROSResp...»

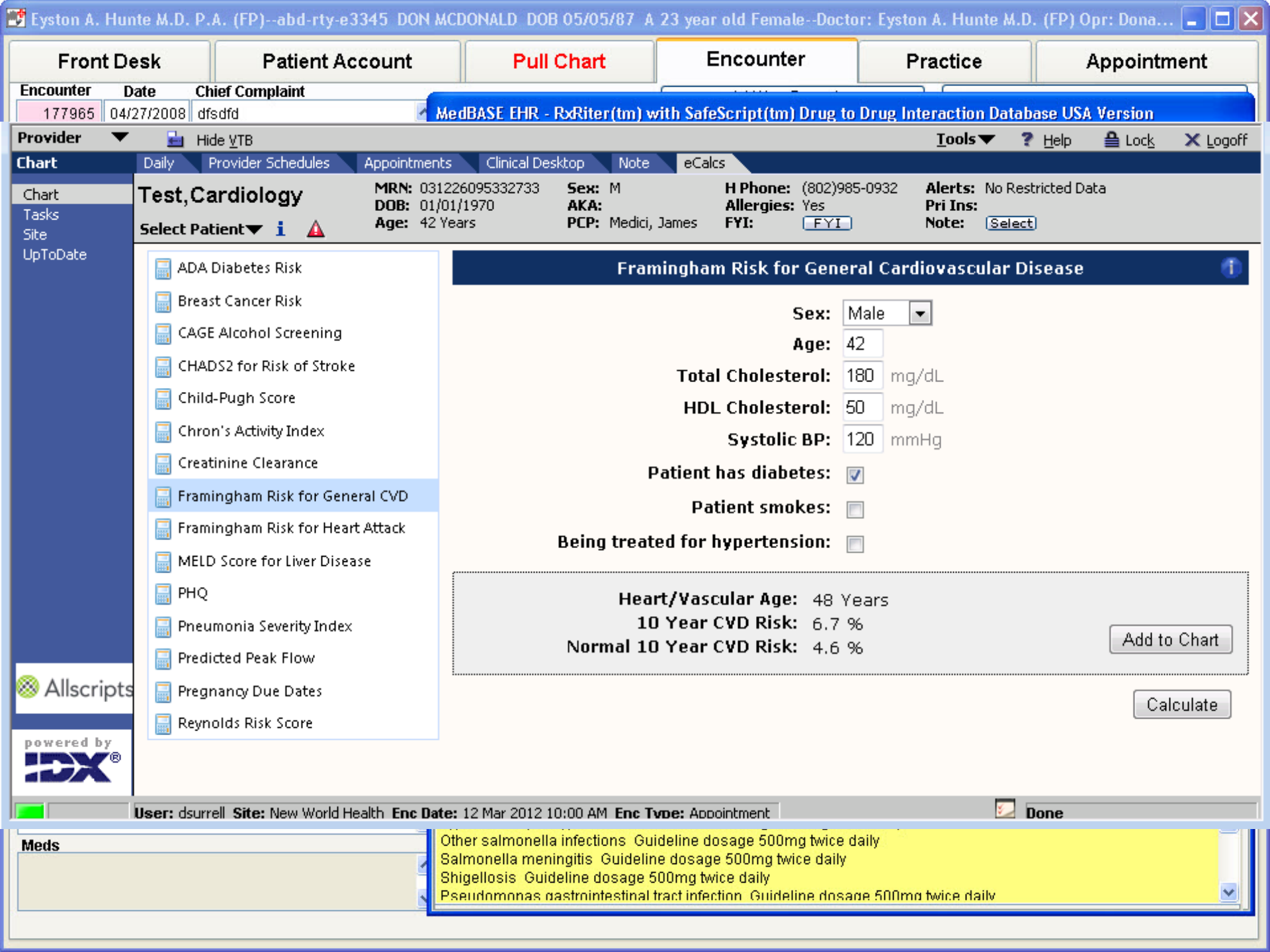


http://healthdesignchallenge.com

Order entry



Shortliffe EH, Cimino JJ (2013) Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Springer Verlag


Eyston A. Hunte M.D. P.A. (FP)--abd-rty-e3345 DON MCDONALD DOB 05/05/87 A 23 year old Female--Doctor: Eyston A. Hunte M.D. (FP) Opr: Dona...
Front Desk
Patient Account
Pull Chart
Encounter
Practice
Appointment
Encounter
Date
Chief Complaint
177965
04/27/2008
dfsdfd
MedBASE EHR - RxRiter(tm) with SafeScript(tm) Drug to Drug Interaction Database USA Version
Provider
Hide VTB
Tools
Help
Lock
Logoff
Chart
Daily
Provider Schedules
Appointments
Clinical Desktop
Note
eCalcs
Chart
Tasks
Site
UpToDate
Test,Cardiology
Select Patient
MRN: 031226095332733
DOB: 01/01/1970
Age: 42 Years
Sex: M
AKA:
PCP: Medici, James
H Phone: (802)985-0932
Allergies: Yes
FYI
Alerts: No Restricted Data
Pri Ins:
Note: Select
ADA Diabetes Risk
Breast Cancer Risk
CAGE Alcohol Screening
CHADS2 for Risk of Stroke
Child-Pugh Score
Chron's Activity Index
Creatinine Clearance
Framingham Risk for General CVD
Framingham Risk for Heart Attack
MELD Score for Liver Disease
PHQ
Pneumonia Severity Index
Predicted Peak Flow
Pregnancy Due Dates
Reynolds Risk Score
Framingham Risk for General Cardiovascular Disease
Sex: Male
Age: 42
Total Cholesterol: 180 mg/dL
HDL Cholesterol: 50 mg/dL
Systolic BP: 120 mmHg
Patient has diabetes:
Patient smokes:
Being treated for hypertension:
Heart/Vascular Age: 48 Years
10 Year CVD Risk: 6.7 %
Normal 10 Year CVD Risk: 4.6 %
Add to Chart
Calculate
Allscripts
powered by IDX
User: dsurrell Site: New World Health Enc Date: 12 Mar 2012 10:00 AM Enc Type: Appointment
Done
Meds
Other salmonella infections Guideline dosage 500mg twice daily
Salmonella meningitis Guideline dosage 500mg twice daily
Shigellosis Guideline dosage 500mg twice daily
Pseudomonas gastrointestinal tract infection Guideline dosage 500mg twice daily

Types of data

- Narrative text
  - Medical history, description of current illness, family history
  - Template, 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)

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

http://www.icd9data.com/2013/Volume1/390-459/420-429/428/default.htm

ICD10

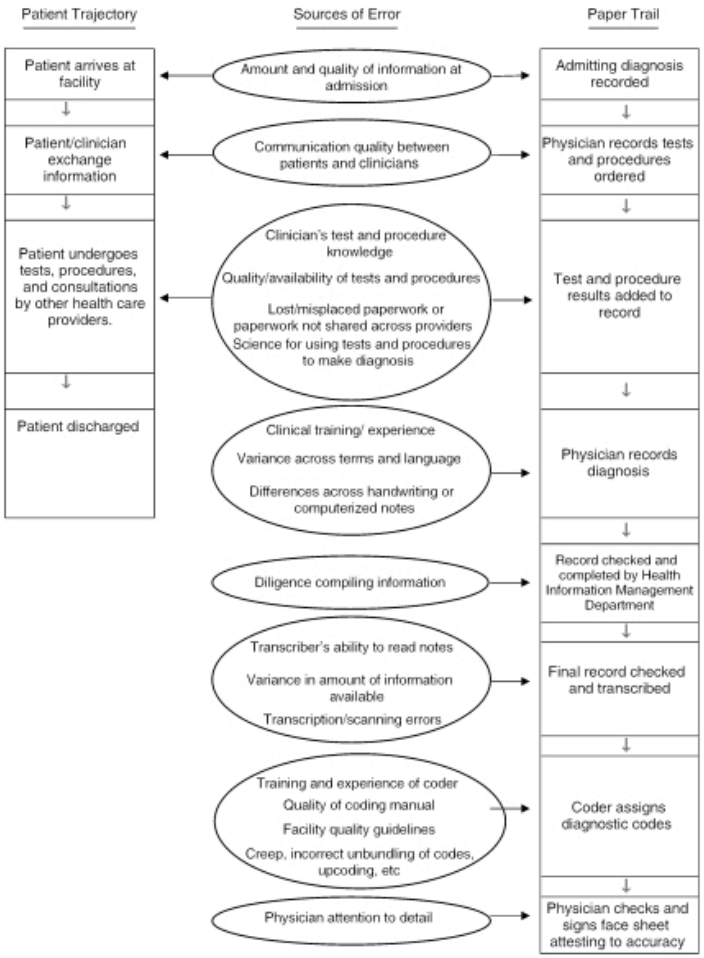


http://www.healthcareitnews.com

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?

ICD9 for research?



Health Serv Res. 2008 August; 43(4): 1424–1441.
doi: 10.1111/j.1475-6773.2007.00822.x
Health Serv Res. 2005 October; 40(5 Pt 2): 1620–1639.
doi: 10.1111/j.1475-6773.2005.00444.x

IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 17, NO. 2, MARCH 2013

Temporal Properties of Diagnosis Code Time Series in Aggregate

Adler Perotte and George Hripcsak

-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

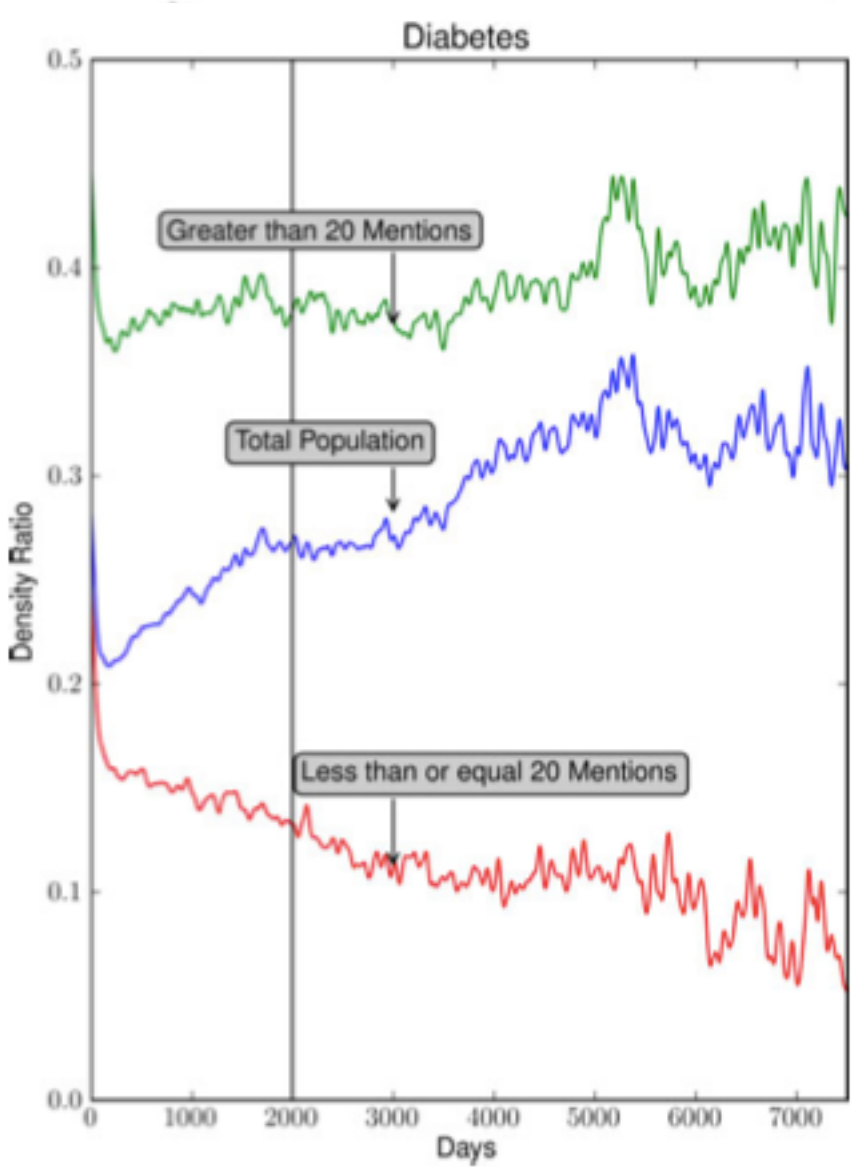


Fig. 5. Documentation probability over time for all diabetes patients, those with greater than 20 positive mentions, and those with less than or equal to 20 positive mentions.

1600

SNOMED

Search Results

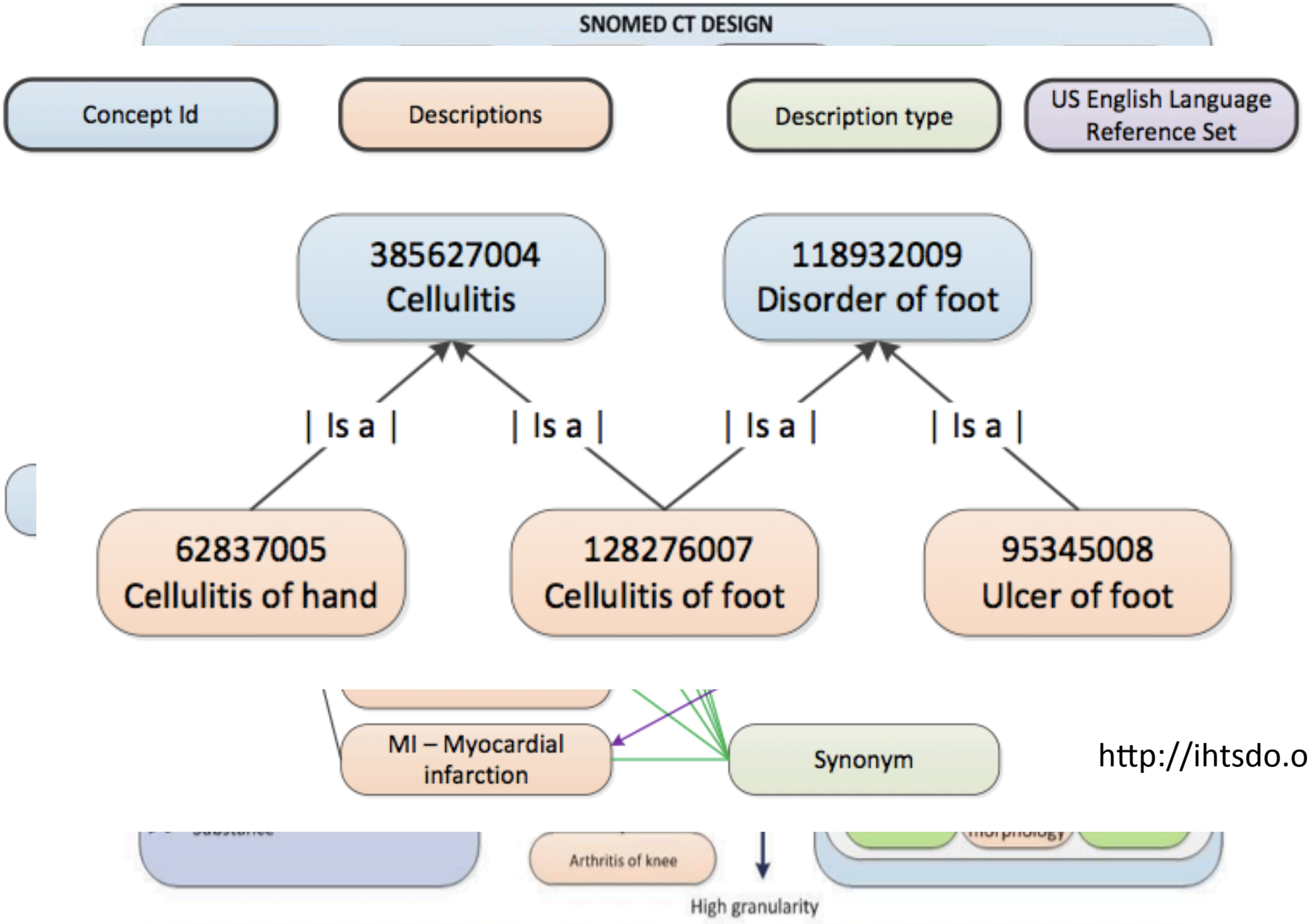
- Arterial bypass graft (procedure)
- Operative procedure on coronary artery (procedure)
- Repair of heart (procedure)
- Thoracic artery repair (procedure)

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)

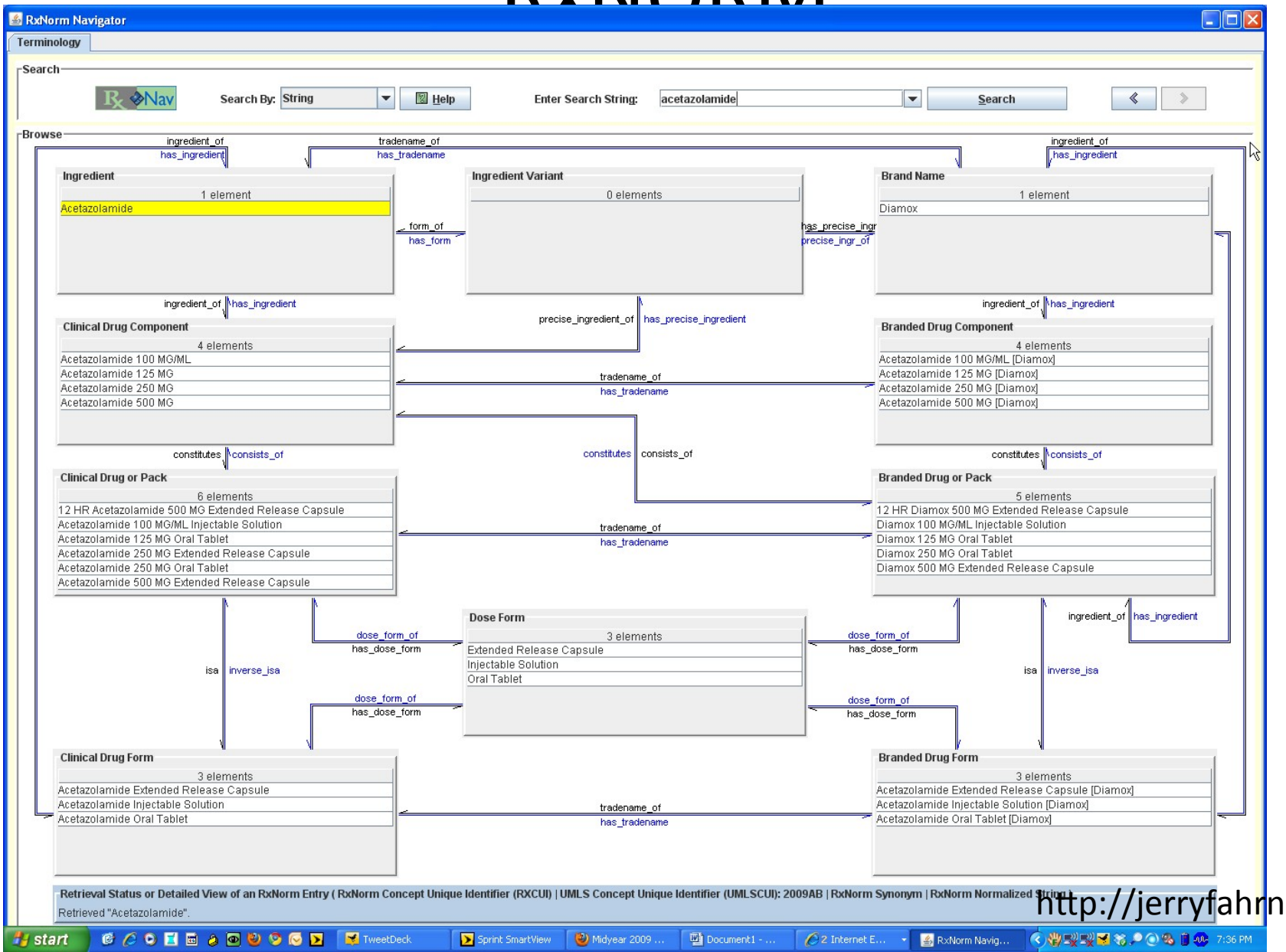
Staging Scale
Substance

SNOMED CT Components, Hierarchies and Outputs



http://ihtsdo.org/

RxNORM



http://jerryfahrni.com/

==========

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.

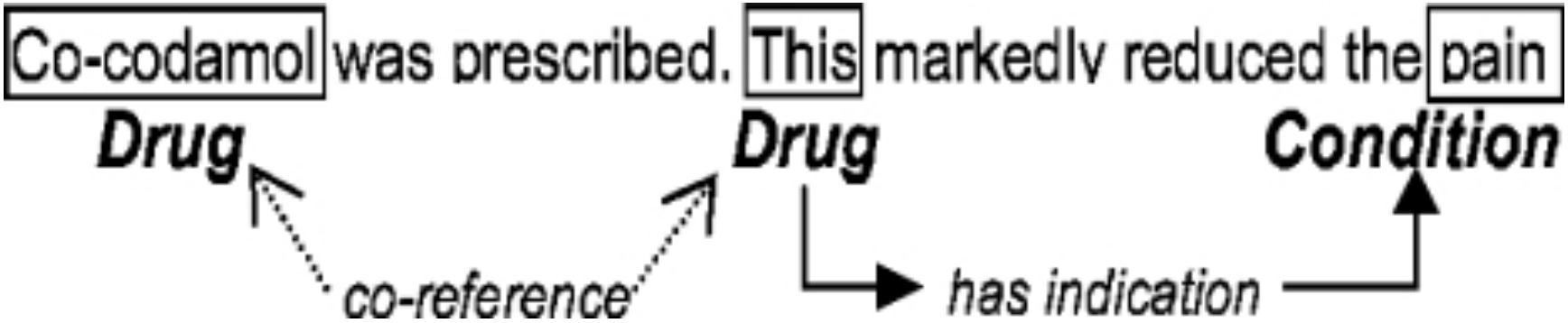
==========

http://mimic.physionet.org

2013-10-18 22:41:00
NPN 7AM-11PM:
S/O: 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 ambued with 100% FiO2, 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, which 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 to off 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 d/t 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 15gm 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 with prn Ativan to try to avoid hypotension from the [**Month/Day 15**]. Fentanyl has been 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 OB+ 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. Ampho dose which was up when he arrested this AM was stopped with ~half of it infused. He did not recieve the rest.....HO 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 have suctioned a few pieces out...none since this AM.
SKIN: He has 2 small decubs on buttocks, covered with Duoderm. Also has open area in left groin.
ACCESS: A-line as described above. He has a right femoral tunneled [**Male First Name (un) 139**] catheter. A clotted 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 Name8 141**] [**Last Name 142**] who was extremely upset about his condition. She was in to visit this evening, and was here for the VT episode. The pt's lawyer also came in 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 CVVHD is does not tolerate HD. SW consult for proxy.

https://physionet.org/mimic2/UserGuide/node14.html

NLP example



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

RESEARCH ARTICLE Open Access

Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies

Raphael Cohen[1*], Michael Elhadad[1] and Noémie Elhadad[2]

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

Journal of Biomedical Informatics 45 (2012) 471–481

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin



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

Today's paper

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

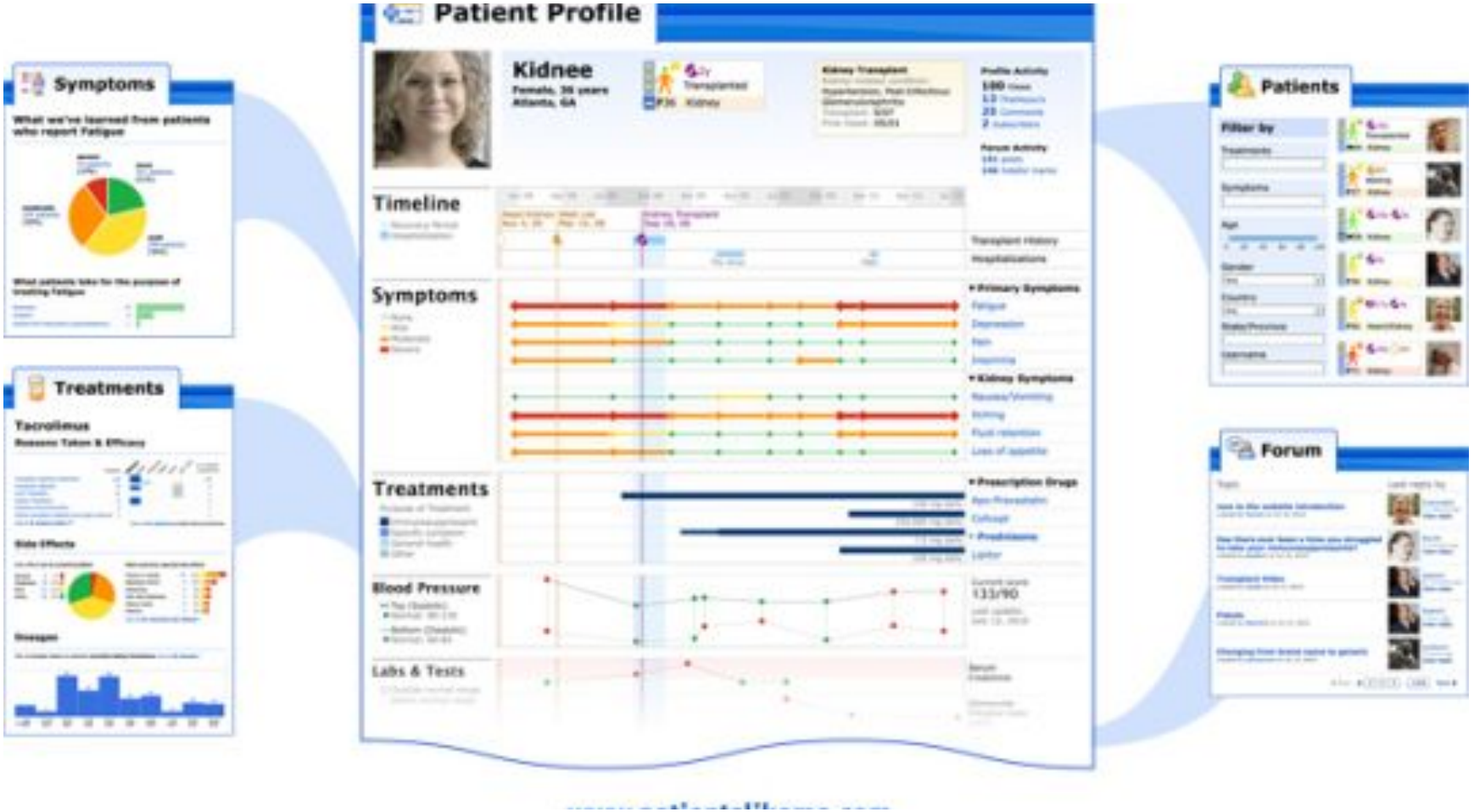
Basic idea: patient managed, can be shared with others

Types:

- Patient access to part of EHR (institution-based)
- Self-maintained (patient creates their history/ updates all info)
- Export EHR to 3rd party site, adds data (e.g. Microsoft Health Vault)


Google health
Search the Web
Jane
Options
Print
Share
Private
Age: 42 years old
Sex: Female
Race / Ethnicity: White
Blood type: O+
Edit
Summary
All records
Wellness
Add
Hide wellness
Blood pressure (15)
118 / 74 mmHg - Aug 15, 2010
Cholesterol, total (3)
194 mg/dL - Jul 12, 2010
Coffee consumption (15)
2 cups/day - Aug 15, 2010
Steps taken (22)
10014 steps - Sep 7, 2010
Weight (with BMI) (15)
144.5 lb, 22 BMI - Sep 12, 2010
Remove from summary
Delete forever
Problems
Add
Show problems (2)
Medications
Interaction warnings
Add
Hide medications
Atorvastatin (1)
Current
20 mg, By mouth - 1 tablet, 1 time per day
Fosinopril (3)
Current
10 mg, By mouth - 1 tablet, 1 time per day
Vitamin C (1)
Past
500 mg, By mouth - 1 tablet, 1 time per day
Remove from summary
Delete forever

Online communities



patientslikeme.com

PHRs and privacy

- Because they're not medical providers, external/linked PHR not subject to HIPAA
- Provider PHRs ARE covered by HIPAA
  - Any problems?

Who's using PHRs?

ORIGINAL INVESTIGATION

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

sers

Table 2. Adjusted Odds Ratios of PHR Adoption[a]

Characteristic	Odds Ratio (95% CI)
Sex	
Male	1 [Reference]
Female	1.15 (1.08-1.21)
Age, y	
18-35	0.79 (0.73-0.86)
36-50	0.94 (0.89-0.99)
51-65	1 [Reference]
>65	0.84 (0.79-0.90)
Race/ethnicity	
White	1 [Reference]
Black	0.50 (0.45-0.55)
Hispanic	0.64 (0.57-0.73)
Asian	0.74 (0.68-0.80)
Insurance status	
Commercial	1 [Reference]
Medicare	0.59 (0.52-0.74)
Medicaid	0.66 (0.58-0.74)
Self-pay	1.21 (1.09-1.32)
Comorbidities, total	
0	1 [Reference]
1	1.23 (1.22-1.30)
2-4	1.27 (1.17-1.30)
Marketing intensity	
Normal	1 [Reference]
Aggressive	2.92 (1.58-5.40)
Annual household income, quartile	
1, High SES	1 [Reference]
2	0.98 (0.93-1.02)
3	0.96 (0.91-1.00)
4, Low SES	0.86 (0.82-0.92)
Visits	
0-1	1 [Reference]
2-4	0.67 (0.63-0.73)
≥5	0.49 (0.44-0.55)

Abbreviations: CI, confidence interval; PHR, personal health record; SES, socioeconomic status.

obreviations: CHF, congestive heart failure;
HR, personal health record;
ES, socioeconomic status.

<.001 for all comparisons.

ace/ethnicity data total 100% after addition
other/excluded race
tegory. Other variables do not total 100%
ecause of rounding.

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

Discussion

1. Pair up, pick who will be pro/con
2. 3 minutes, brainstorm
3. 5 minutes, debate

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

Open data

http://www.youtube.com/watch?v=mVZI7NBgcWM

PGP

- Goal: disentangle genotype, phenotype, environment
- Full sequence data, medical records, imaging, etc
- Ongoing
  - PGP-10: 10 volunteers
  - PGP-Harvard now over 5K, but only a few hundred complete sequences
  - PGP-Canada
  - PGP-UK: 1K so far, aim to recruit 100K (on hold?)

PGP potential risks(part)

- The risks of public disclosure of your genetic and trait information could affect your employment, insurance and financial well-being and social interactions for you and your family.
- Anyone with sufficient knowledge and resources could take your DNA sequence data and/or posted trait information and use that data, with or without modification, to: (i) infer paternity or other features of your genealogy; (ii) claim statistical evidence that could affect your employment, insurance or ability to obtain financial services; (iii) claim relatedness to criminals or incriminate relatives; (iv) make synthetic DNA and plant it at a crime scene, or otherwise use it to falsely identify you; or (v) reveal the possibility of a disease or unknown propensity for a disease.
- Whether or not it is lawful to do so, you could be subject to actual or attempted employment, insurance, financial, or other forms of discrimination or negative treatment on the basis of the public disclosure of your genetic and trait information by the PGP or by a third party.
- The distribution of your cell lines could result in the creation and further distribution by a third party of additional cell lines, organs, or tissues containing your DNA for research, commercial, clinical, or other uses, including certain forms of assisted reproduction, some of which you may find objectionable or upsetting.
- If you have previously made available or intend to make available genetic information in a confidential setting, for example in another research study or in a clinical trial, the data that you provide as part of the PGP may be used, on its own or in combination with your previously shared data, to identify you as a participant in otherwise confidential genetic research or trials.

From PGP consent form

“Anonymous” Genomes Identified

The names and addresses of people participating in the Personal Genome Project can be easily tracked down despite such data being left off their online profiles.

By Dan Cossins | May 3, 2013

2 Comments Like 37 Pin it 0 Link this Stumble Tweet this



WIKIMEDIA, GEORGE GASTIN

Data privacy researchers have been able to identify the names of hundreds of participants in the Personal Genome Project (PGP) using demographic data from their profiles, according to a paper out this week on the *arXiv* preprint server. The authors also suggest ways in which contributors can increase their privacy.

Launched in 2006, the PGP aims to collect genetic data as well as health and lifestyle information from 100,000 people to help researchers tease apart the interactions between genotype, environment, and phenotype. The project does not guarantee privacy, reported *MIT Technology Review*, and participants can choose to disclose as much personal data as they want, including ZIP code, birth date, and gender, on their online PGP profile. But these profiles are “de-identified,” meaning their names and addresses are not made public.

For next week

- For this Friday, complete CITI training
- For next Friday, reading