

Machine Learning for Healthcare

6.S897, HST.S53

Lecture 1: What makes healthcare unique?

Prof. David Sontag
MIT EECS, CSAIL, IMES



**Massachusetts
Institute of
Technology**

Outline for today's class

1. **Brief history of AI and ML in healthcare**
2. Why *now*?
3. Examples of machine learning in healthcare
4. What is *unique* about ML in healthcare?
5. Overview of class syllabus and projects

1970's: MYCIN expert system

- 1970's (Stanford): MYCIN expert system for identifying bacteria causing severe infections
- Proposed a good therapy in ~69% of cases. Better than infectious disease experts

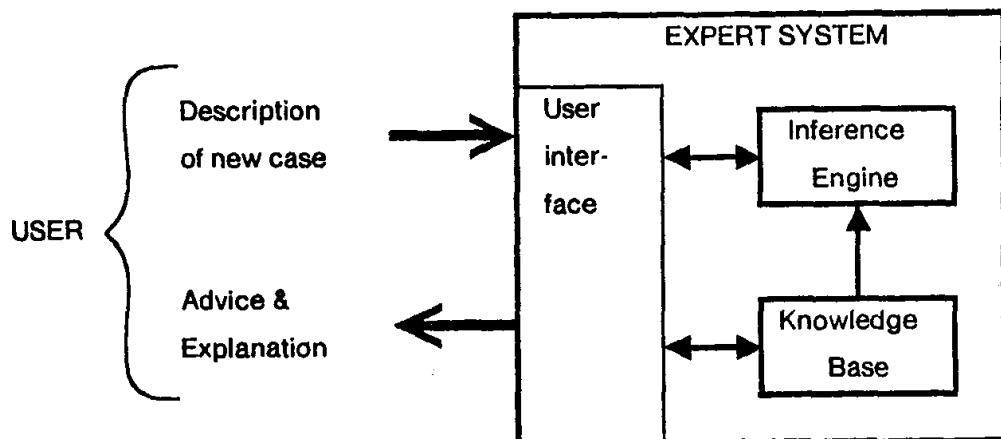


FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.

Dialogue interface

I am ready

** THIS IS A 26 YEAR OLD MALE PATIENT

My understanding is:

The age of the patient is 26

The sex of the patient is male

** FIVE DAYS AGO, HE HAD RESPIRATORY-TRACT SYMPTOMS

What is his name?

** JO

My understanding is:

The name of the patient is Jo

Respiratory-tract is one of the symptoms that the patient had

** A COUPLE OF DAYS BEFORE THE ADMISSION, HE HAD A MALAISE

Please give me the date of admission

** MARCH 12, 1979

My understanding is:

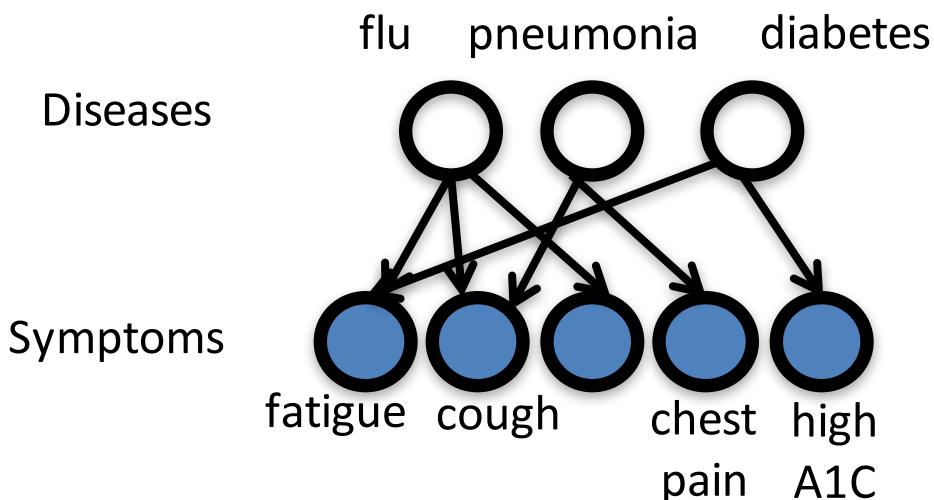
The patient was admitted at the hospital 3 days ago

Malaise is one of the symptoms that the patient had 5 days ago

FIGURE 33-1 Short sample dialogue. The physician's inputs appear in capital letters after the double asterisks.

1980's: INTERNIST-1/QMR model

- 1980's (Univ. of Pittsburgh): INTERNIST-1/Quick Medical Reference
- Diagnosis for internal medicine



Probabilistic model relating:

570 binary disease variables

4,075 binary symptom variables

45,470 directed edges

Elicited from doctors:

15 person-years of work

**Led to advances in ML & AI
(Bayesian networks, approximate inference)**

- Problems:**
1. Clinicians entered symptoms *manually*
 2. Difficult to maintain, difficult to generalize

[Miller et al., '86, Shwe et al., '91]

1980's: automating medical discovery

RX PROJECT: AUTOMATED KNOWLEDGE ACQUISITION

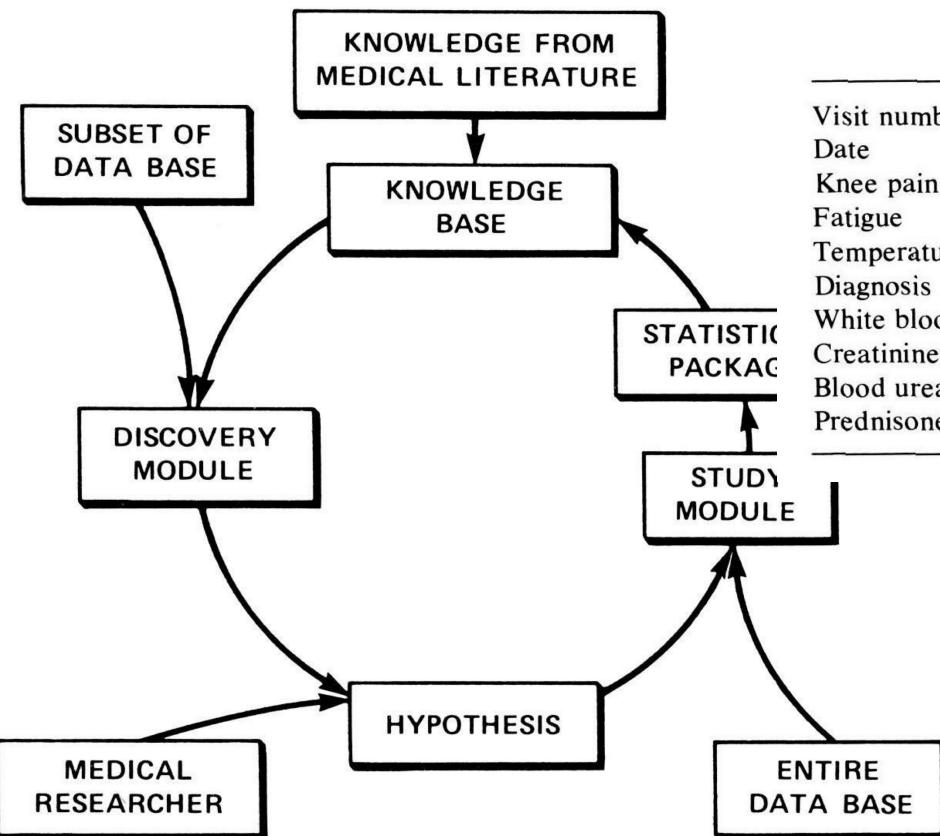


TABLE 1
HYPOTHETICAL TIME-ORIENTED RECORD FOR ONE PATIENT

Visit number	1	2	3
Date	January 17, 79	June 23, 79	July 1, 79
Knee pain	Severe	Mild	Mild
Fatigue	Moderate	—	Moderate
Temperature	38.5	37.5	36.9
Diagnosis	Systemic lupus		
White blood count	3500	4700	4300
Creatinine clearance	45	—	65
Blood urea nitrogen	36	33	—
Prednisone	30	25	20

Discover that prednisone elevates cholesterol
(Annals of Internal Medicine, '86)

[Robert Blum, "Discovery, Confirmation and Incorporation of Causal Relationships from a Large Time-Oriented Clinical Data Base: The RX Project". Dept. of Computer Science, Stanford. 1981]

1990's: neural networks in medicine

- Neural networks with clinical data took off in 1990, with 88 new studies that year
- Small number of features (inputs)
- Data often collected by chart review

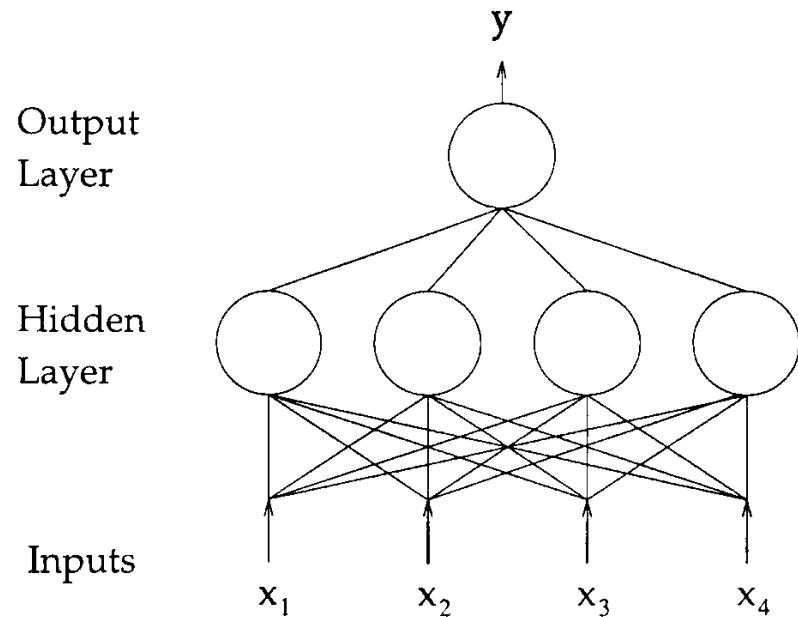


FIGURE 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.

Problems: 1. Did not fit well into clinical workflow
2. Poor generalization to new places

[Penny & Frost, Neural Networks in Clinical Medicine. Med Decis Making, 1996]

Table 1 • 25 Neural Network Studies in Medical Decision Making*

Subject	No. of Examples					Accuracy§	
	Training	Test	P†	Network	D‡	Neural	Other
Breast cancer ⁴	57	20	60	9-15-2	0.6	80	75
Vasculitis ²	404	403	73	8-5-1	8.0	94	—
Myocardial infarction ⁶	351	331	89	20-10-10-1	1.1	97	84
Myocardial infarction ⁸	356	350	87	20-10-10-1	1.1	97	84
Low back pain ¹¹	100	100	25	50-48-2	0.2	90	90
Cancer outcome ¹³	5,169	3,102	—	54-40-1	1.4	0.779	0.776
Psychiatric length of stay ¹⁷	957	106	73	48-400-4	0.2	74	76
Intensive care outcome ²³	284	138	91	27-18-1	0.5	0.82	0.82
Skin tumor ²¹	150	100	80	18	—	80	90
Evoked potentials ³⁵	100	67	52	14-4-3	3.8	77	77
Head injury ⁴⁷	500	500	50	6-3-3	20	66	77
Psychiatric outcome ⁵⁴	289	92	60	41-10-1	0.7	79	—
Tumor classification ⁵⁵	53	6	38	8-9-3	1.4	99	88
Dementia ⁵⁷	75	18	19	80-10-7-7	0.6	61	—
Pulmonary embolism ⁵⁹	607	606	69	50-4-1	2.9	0.82	0.83
Heart disease ⁶²	460	230	54	35-16-8-2	3	83	84
Thyroid function ⁶²	3,600	1,800	93	21-16-8-3	22	98	93
Breast cancer ⁶²	350	175	66	9-4-4-2	10	97	96
Diabetes ⁶²	384	192	65	8-4-4-2	12	77	75
Myocardial infarction ⁶³	2,856	1,429	56	291-1	9.8	85	—
Hepatitis ⁶⁵	39	42	38	4-4-3	3.3	74	79
Psychiatric admission ⁷⁶	319	339	85	53-1-1	6.0	91	—
Cardiac length of stay ⁸³	713	696	73	15-12-1	3.5	0.70	—
Anti-cancer agents ⁸⁹	127	141	25	60-7-6	1.5	91	86
Ovarian cancer ⁹¹	75	98	—	6-6-2	2.6	84	81
MEDIAN VALUE	350	175	71	20	2.8		

*For reference citations, see the reference list.

†P = prior probability of most prevalent category.

‡D = ratio of training examples to weights per output.

§A single integer in the accuracy column denotes percentage overall classification rate and a single real number between 0 and 1 indicates the AUROC value. Neural = accuracy of neural net, Other = accuracy of best other method.

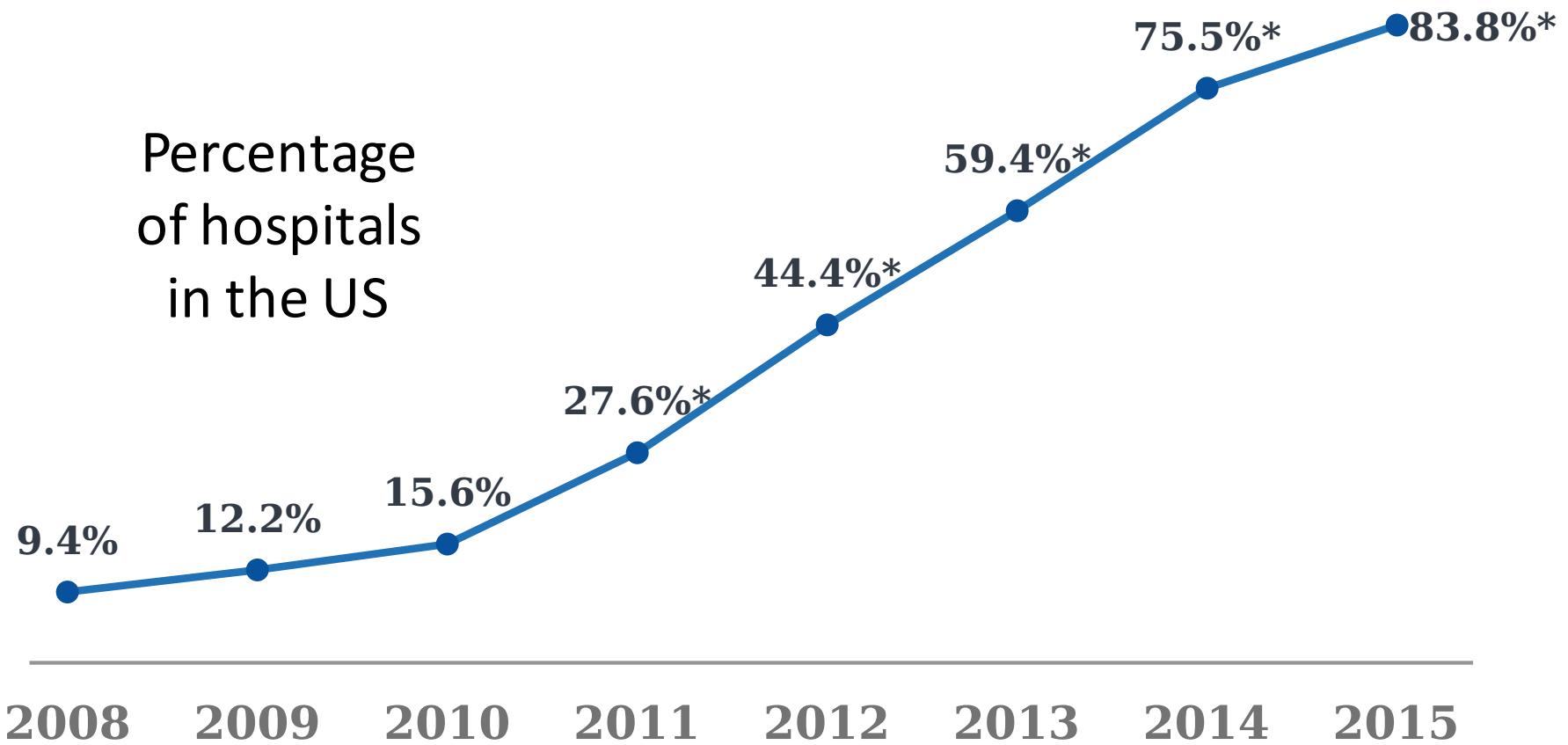
Outline for today's class

1. Brief history of AI and ML in healthcare
2. **Why now?**
3. Examples of machine learning in healthcare
4. What is *unique* about ML in healthcare?
5. Overview of class syllabus and projects

Why now?

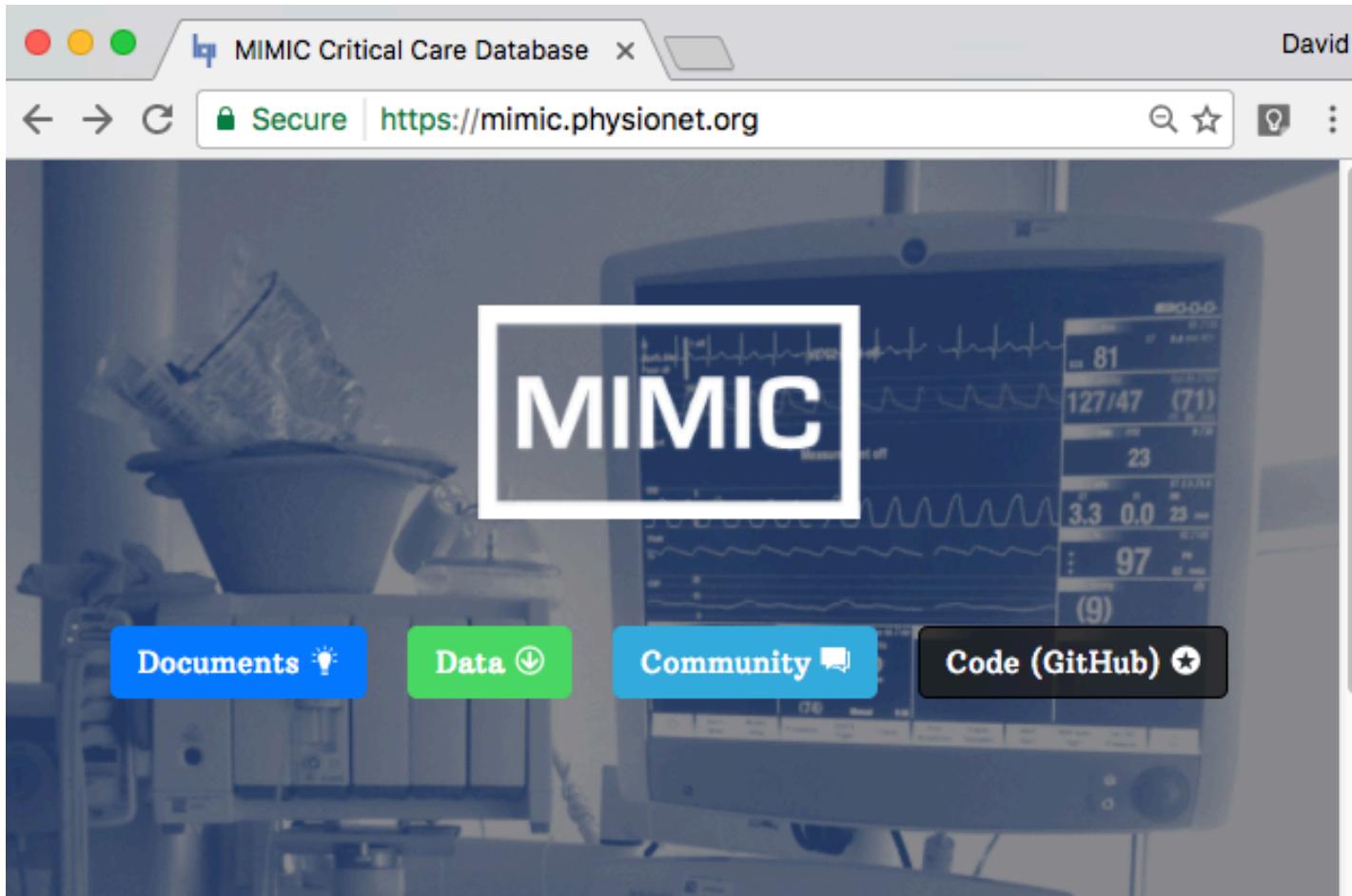
DATA

Adoption of Electronic Health Records (EHR) has increased 9x since 2008



[Henry et al., ONC Data Brief, May 2016]

Large datasets



If you use MIMIC data or code in your work, please cite the following publication:

MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. *Scientific Data* (2016). DOI: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35). Available from: <http://www.nature.com/articles/sdata201635>



De-identified
health data from
~40K critical care
patients

Demographics,
vital signs,
laboratory tests,
medications,
notes, ...

Large datasets

A screenshot of a web browser showing the URL truvnhealth.com/markets/life-sciences/products/data-tools/marketscan-databases. The page header includes links for MEDIA ROOM, SUPPORT, and CAREER. The main navigation bar features the TRUVEN logo, an IBM Company, and links for SOLUTIONS, EVENTS, KNOWLEDGE, and AB.

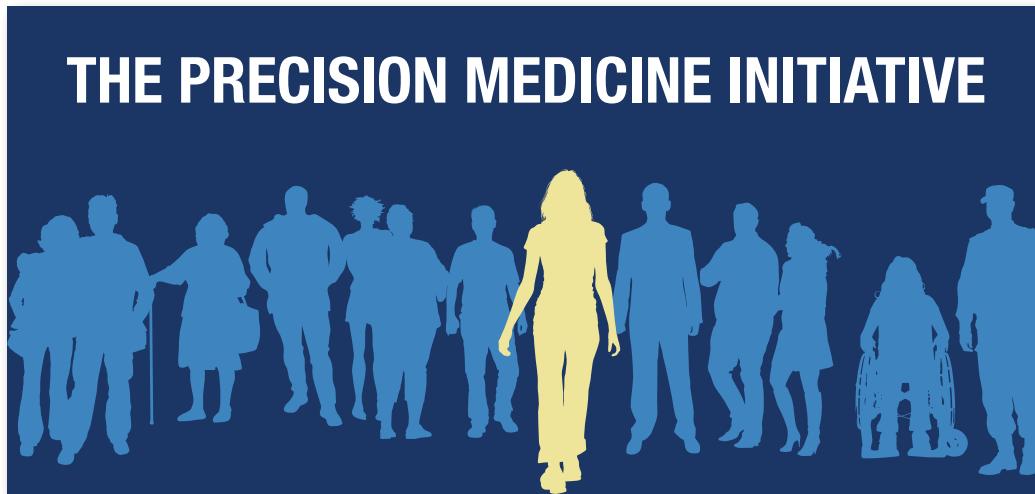
A screenshot of the Truven Health Analytics Life Sciences page. The top navigation bar shows "Life Sciences" and the full URL "Home » Life Sciences » Data & Tools » MarketScan Databases". Below the navigation is a photograph of a person's hands interacting with a tablet displaying a graph. To the left is a sidebar with links: Market Knowledge, Real World Evidence, Stakeholder Management, Data & Tools (which is highlighted in grey), MarketScan Databases, Treatment Pathways, Inpatient/Outpatient View, PULSE, and Heartbeat Profiler. The main content area features the title "Putting Research Data Into Your Hands with the MarketScan Databases" and a subtext: "The Family of MarketScan® Research Databases is the largest of its kind in the industry, with data on nearly 230 million unique patients since 1995."

“Data on nearly
230 million
unique patients
since 1995”

\$\$\$

Large datasets

President Obama's initiative to create a 1 million person research cohort

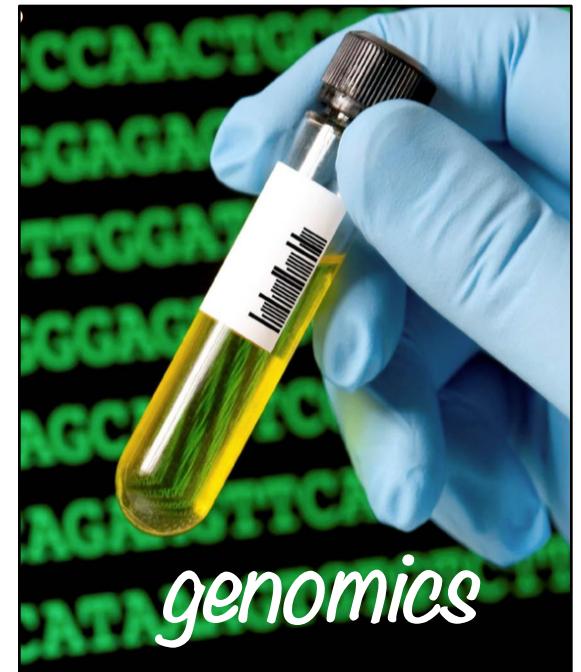
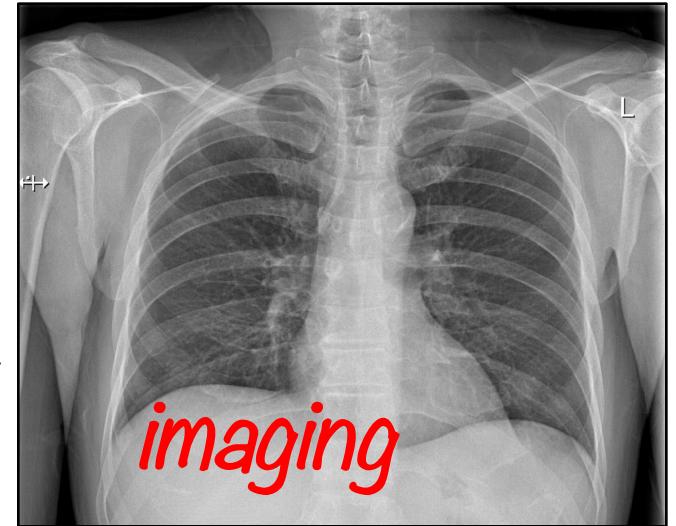
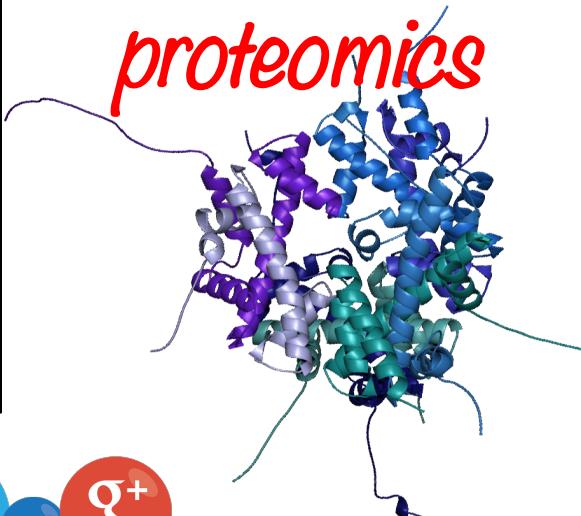


Core data set:

- Baseline health exam
- Clinical data derived from electronic health records (EHRs)
- Healthcare claims
- Laboratory data

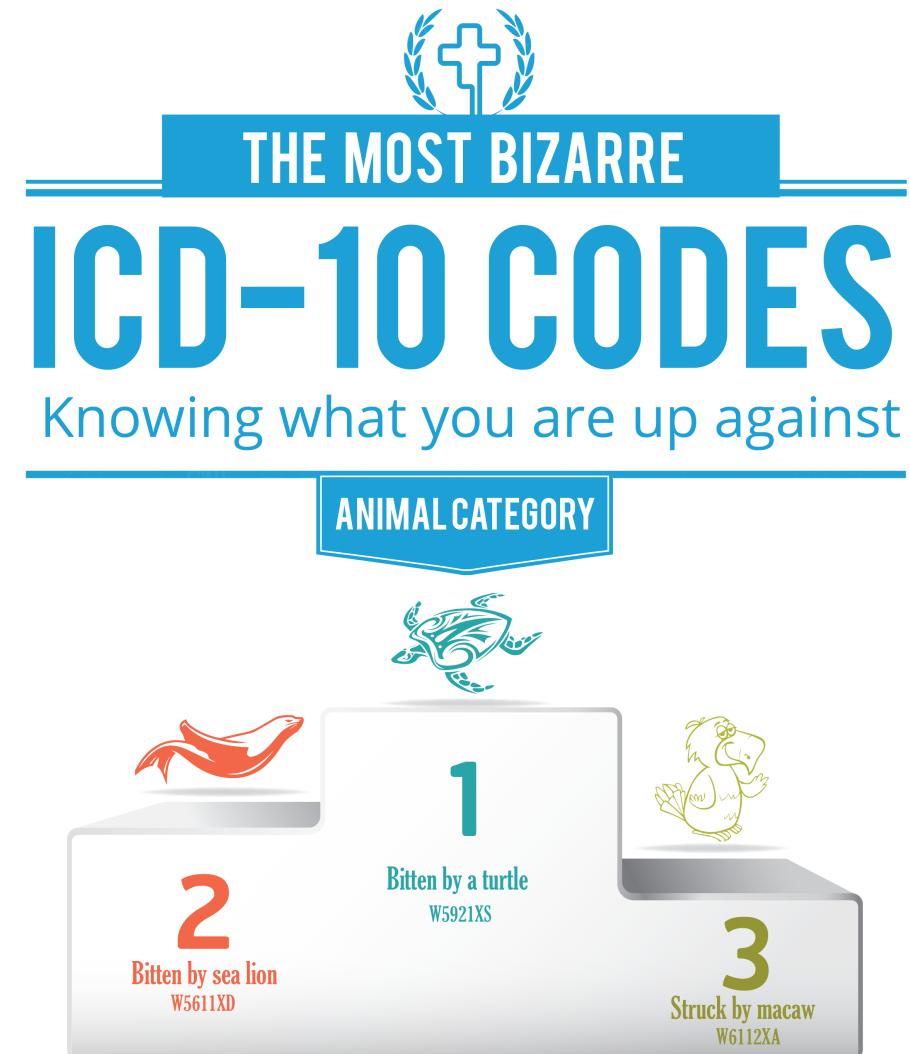
[Precision Medicine Initiative (PMI) working Group Report, Sept. 17 2015]

Diversity of digital health data



Standardization

- Diagnosis codes: ICD-9 and ICD-10 (International Classification of Diseases)



ICD-9 codes 290–319: mental disorders

ICD-9 codes 320–359: diseases of the nervous system

ICD-9 codes 360–389: diseases of the sense organs

ICD-9 codes 390–459: diseases of the circulatory system

ICD-9 codes 460–519: diseases of the respiratory system

ICD-9 codes 520–579: diseases of the digestive system

ICD-9 codes 580–629: diseases of the genitourinary system

ICD-9 codes 630–679: complications of pregnancy, childbirth,

...

[https://en.wikipedia.org/wiki/List_of_ICD-9_codes]

[<https://blog.curemd.com/the-most-bizarre-icd-10-codes-infographic/>]

Standardization

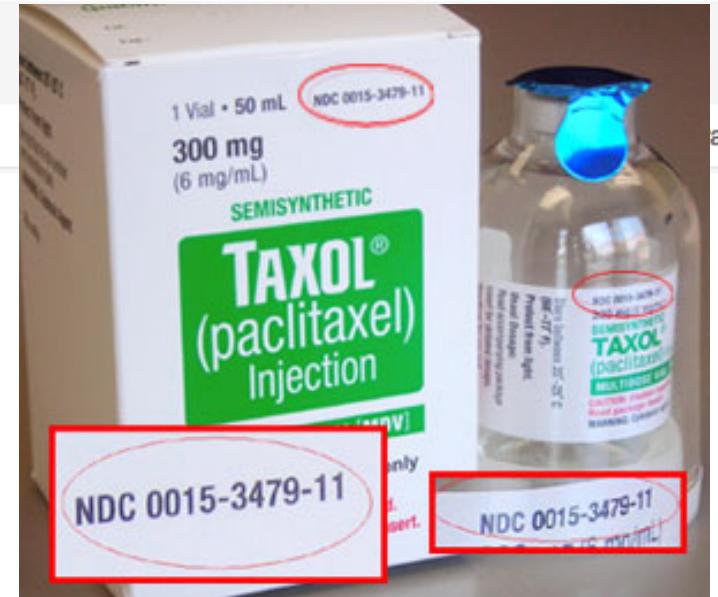
- Diagnosis codes: ICD-9 and ICD-10 (International Classification of Diseases)
- Laboratory tests: LOINC codes
- Pharmacy: National Drug Codes (NDCs)
- Unified Medical Language System (UMLS): millions of medical concepts

LOINC® from Regenstrief

glucose

1 / 5

LOINC	LongName
27353-2	Glucose mean value [Mass/volume] in Blood Estimated from glycated hemoglobin
2352-3	Glucose in CSF/Glucose plas
49689-3	Glucose tolerance [Interpretation] in Serum or Plasma Narrative—post 100 g glucose PO
49688-5	
72650-5	



NDC 0015-3479-11

TAXOL® (paclitaxel) Injection

SEMISYNTHETIC

1 Vial • 50 mL
300 mg
(6 mg/mL)

NDC 0015-3479-11

NDC 0015-3479-11

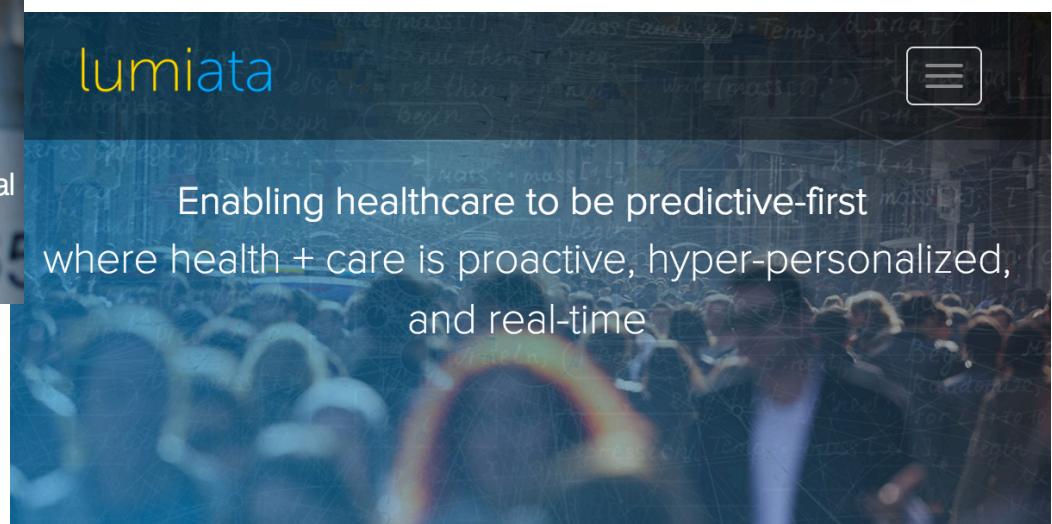
Why now?

ALGORITHMS

Advances in machine learning

- Major advances in ML & AI
 - Learning with high-dimensional features (e.g., l1-regularization)
 - Semi-supervised and unsupervised learning
 - Modern deep learning techniques (e.g. convnets, variants of SGD)
- Democratization of machine learning
 - High quality open-source software, such as Python's scikit-learn, TensorFlow, Torch, Theano

Industry interest in AI & healthcare



Outline for today's class

1. Brief history of AI and ML in healthcare
2. Why *now*?
- 3. Examples of machine learning in healthcare**
4. What is *unique* about ML in healthcare?
5. Overview of class syllabus and projects



Emergency Department:

- Limited resources
- Time sensitive
- Critical decisions

Data in Emergency Department (ED)

Electronic records for over 300,000 ED visits

Triage Information
(Free text)



30 min

MD comments
(free text)



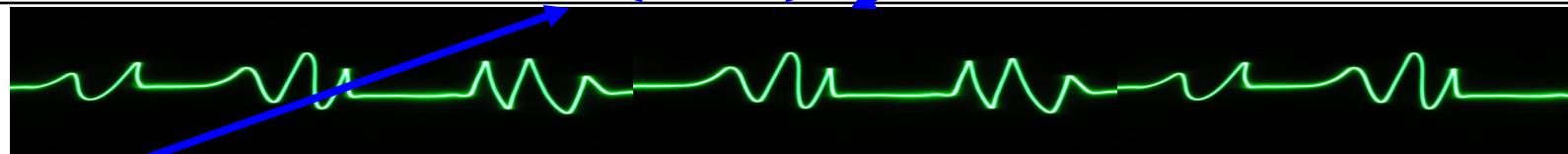
Specialist consults



Physician documentation



T=0



2 hrs



Lab results
(Continuous valued)

Repeated vital signs
(continuous values)
Measured every 30 s

Disposition

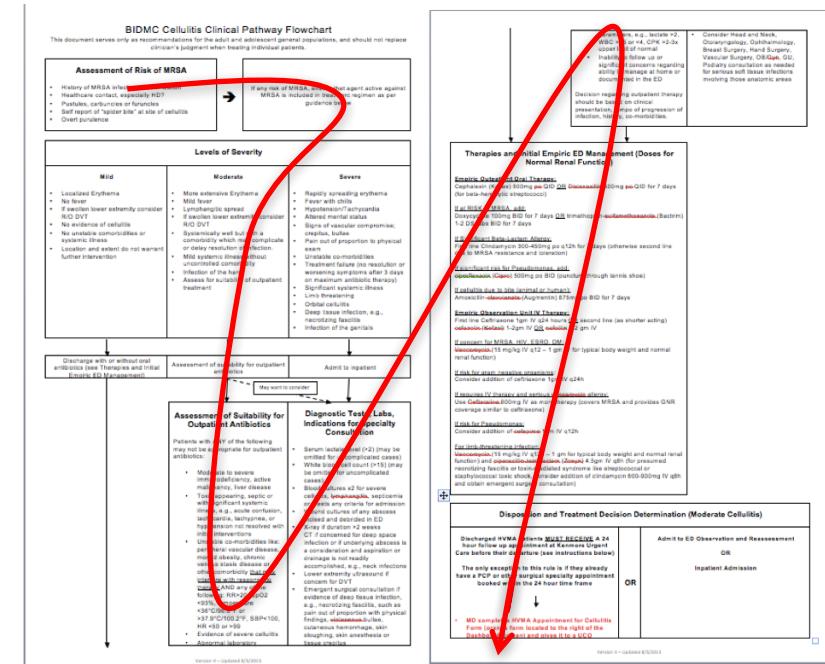
Collaboration with
Steven Horng, MD



Opportunities for machine learning

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

BIDMC Cellulitis Clinical Pathway Flowchart



Pathways have been shown to reduce in-hospital complications without increasing costs
[Rotter et al 2010]

Opportunities for machine learning

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

Automating triggers

Don't rely on the user's knowledge
that the pathway exists!

The ED Dashboard decision support algorithms have determined that this patient may be eligible for the Atrius Cellulitis pathway. Please choose from the following options:

You can include a comment for the reviewers: *Mandatory if Declining*

Below are links to the pathway and/or other supporting documents:

[Atrius Cellulitis Pathway](#)

Our task:

Determine whether a patient has or is suspected to have cellulitis

Opportunities for machine learning

- Triggering clinical pathways
- **Context-specific displays**
- Risk stratification
- Improving clinical documentation

Automatically place specialized order sets on patient displays

- Chest Pain Order Set

To be drawn immediately Add-on

Initial

Place IV (saline lock); flush per protocol
 Continuous Cardiac monitoring
 Continuous Pulse oximetry

EKG (pick 1)

Indication: Chest Pain
 Indication: Dyspnea

Laboratory

CBC + Diff
 Chem-7
 Troponin

Aspirin (pick 1)

Aspirin 324 mg PO chewed
 Aspirin 243 mg PO chewed
 Aspirin taken before arrival

Imaging

XR Chest PA & Lateral

Our task:

Determine whether patient complained of chest pain, or is a psych patient

- Psych Order Set

To be drawn immediately Add-on

Laboratory

CBC + Diff
+ Chem-7
+ Serum Tox
+ Urine Tox

Order

Opportunities for machine learning

- Triggering clinical pathways
 - Context-specific displays
 - **Risk stratification**
 - Improving clinical documentation
- Ex 1: Likelihood of mortality or admission to ICU
- Ex 2: Early detection of severe sepsis

(Topic of next week's lecture)

Real-time predictions in BIDMC emergency department

<u>History</u>	<u>Acute</u>		
Alcoholism	Abdominal pain	Deep vein thrombosis	Laceration
Anticoagulated	Allergic reaction	Employee exposure	Motor vehicle accident
Asthma/COPD	Ankle fracture	Epistaxis	Pancreatitis
Cancer	Back pain	Gastroenteritis	Pneumonia
Congestive heart failure	Bicycle accident	Gastrointestinal bleed	Psych
Diabetes	Cardiac etiology	Geriatric fall	Obstruction
HIV+	Cellulitis	Headache	Septic shock
Immunosuppressed	Chest pain	Hematuria	Severe sepsis
Liver malfunction	Cholecystitis	Intracerebral hemorrhage	Sexual assault
	Cerebrovascular accident	Infection	Suicidal ideation
		Kidney stone	Syncope
			Urinary tract infection

[Halpern, Horng, Choi, Sontag, JAMIA '16]



Opportunities for machine learning

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- **Improving clinical documentation**

Improving documentation: Chief complaints

Changed workflow to have chief complaints assigned *last*. Predict them.

KERMIT,F [69 / M]

Temp 99 HR 102 BP 150/70 RR 24 O2sat 99%

69 y/o M Patient with severe intermittent RUQ pain. Began soon after eating.
Also is a heavy drinker.

Chief Complaints:

- RUQ abdominal pain
- Allergic reaction
- L Knee pain
- Rectal pain
- Right sided abdominal pain

Transfer
MCI

Enter Cancel

Triage note

Predicted
chief
complaints

KERMIT,F [69 / M]

Temp 99 HR 102 BP 150/70 RR 24 O2sat 99%

69 y/o M Patient with severe intermittent RUQ pain. Began soon after eating.
Also is a heavy drinker.

Chief Complaints:

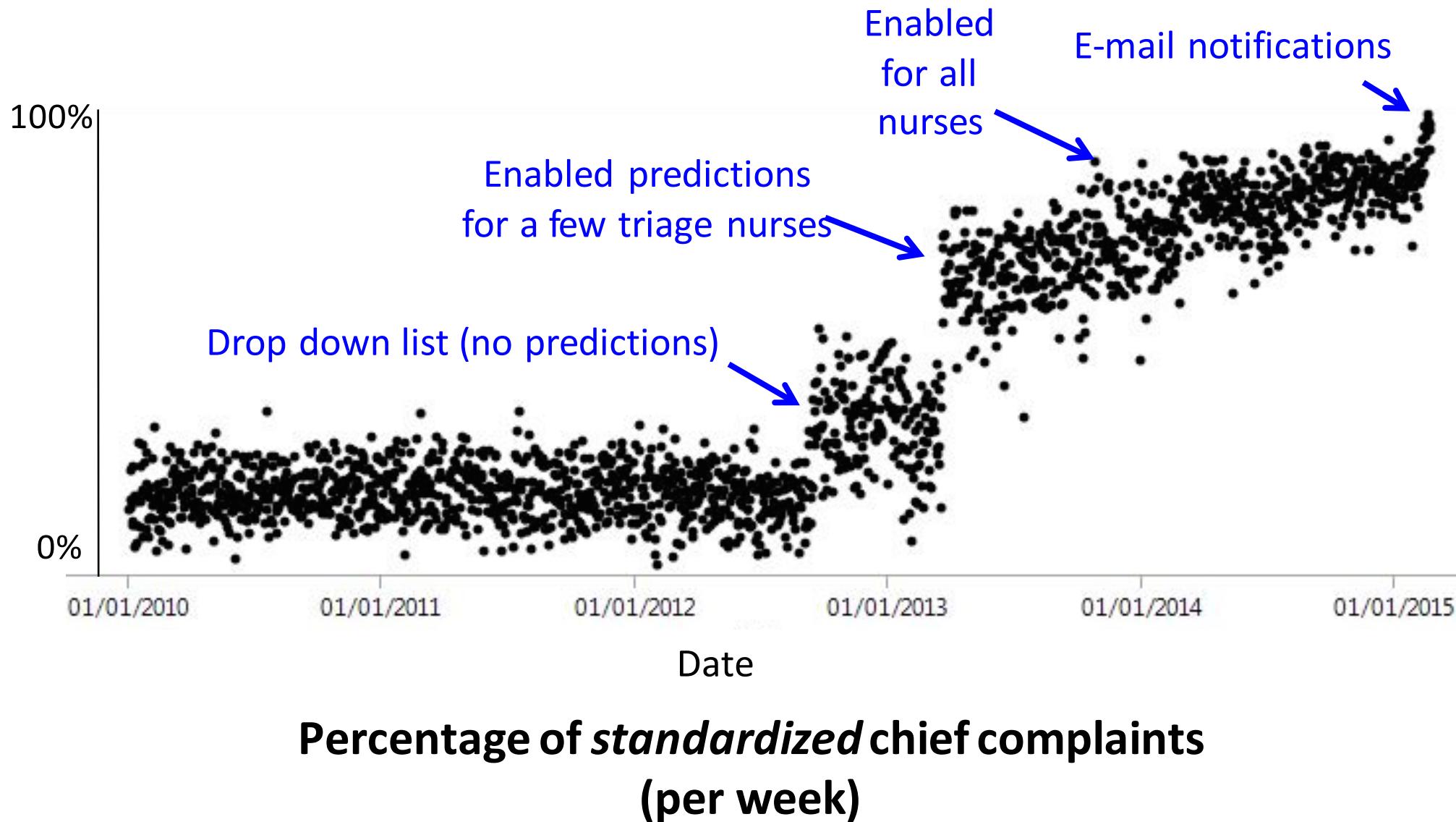
- RIGHT UPPER QUADRANT PAIN
- RUQ ABDOMINAL PAIN
- RUQ PAIN
- ALLERGIC REACTION
- L KNEE PAIN
- RECTAL PAIN
- RIGHT SIDED ABD PAIN
- RIGHT SIDED ABDOMINAL PAIN
- L WRIST PAIN
- RIGHT SIDED CHEST PAIN
- TESTICULAR PAIN
- KNEE PAIN
- ELBOW PAIN
- RIB PAIN
- L ELBOW PAIN
- HAND PAIN
- VAGINAL PAIN

Enter Cancel

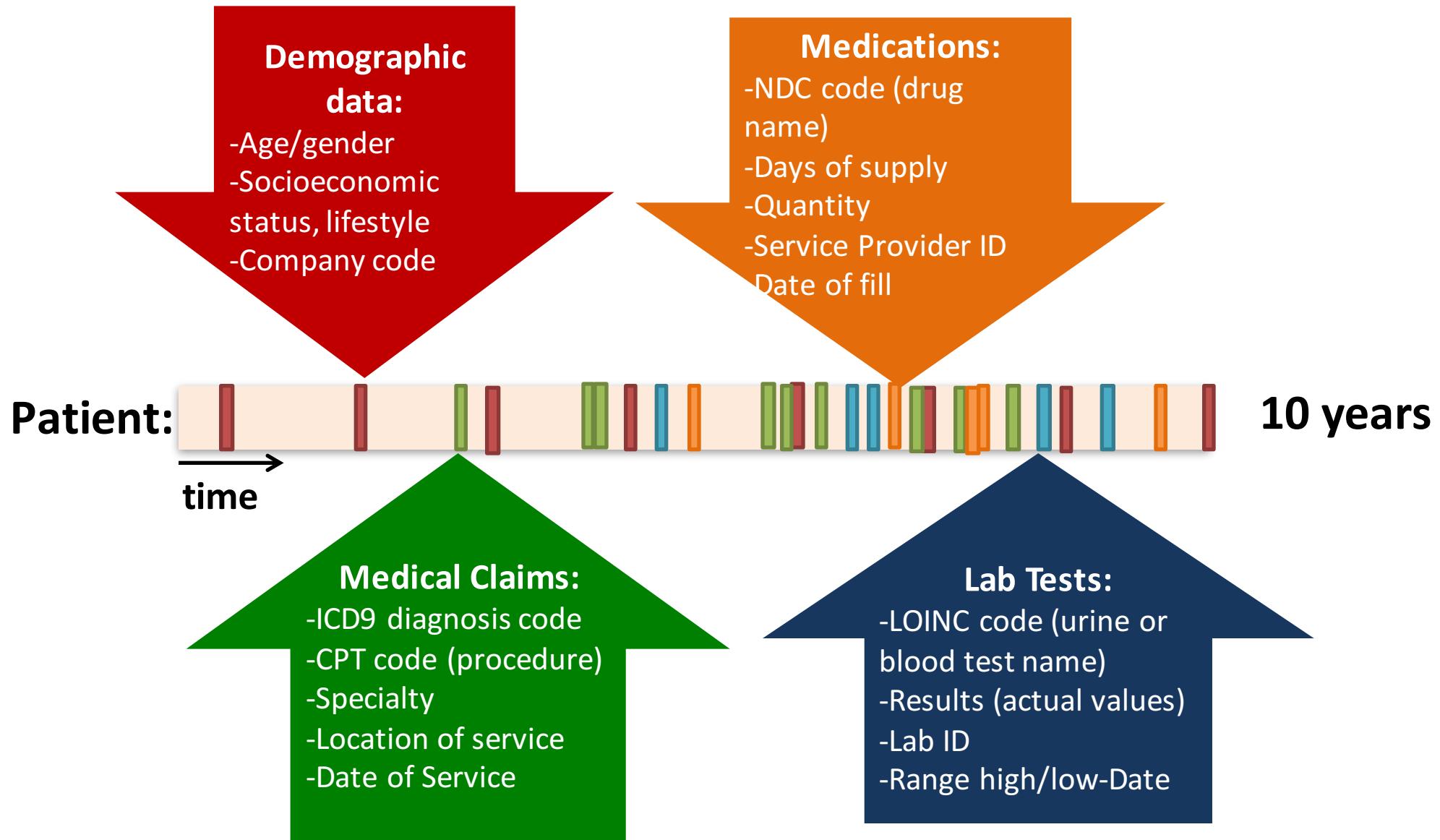
Contextual
auto-
complete

Using for all 55,000 patients/year that present at BIDMC ED

Improving documentation: Chief complaints



Zooming out...



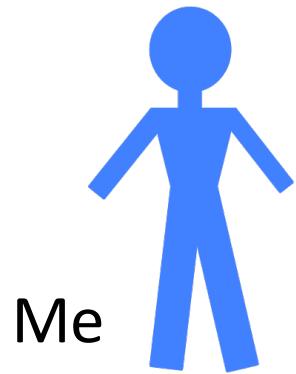
Collaboration with:



Independence
Blue Cross

Temporal modeling of disease progression

- Find markers of disease stage and progression, statistics of what to expect when
 - *What is the “typical trajectory” of a female diagnosed with Sjögren’s syndrome at the age of 19?*
- Estimate a patient’s future disease progression
 - *When will a specific individual with smoldering multiple myeloma (a rare blood cancer) transition to full-blown multiple myeloma?*
 - *Which second-line diabetes treatment should we give to a patient?*

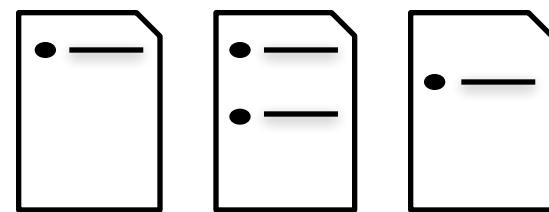


?????

Me



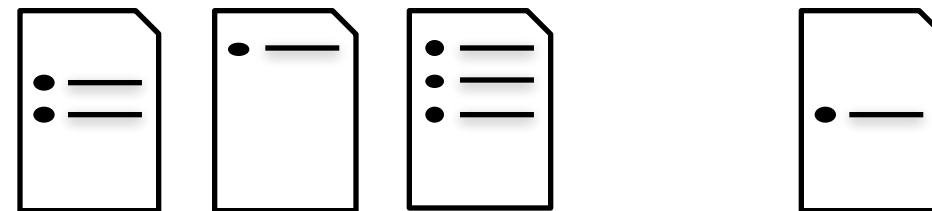
Patient 1



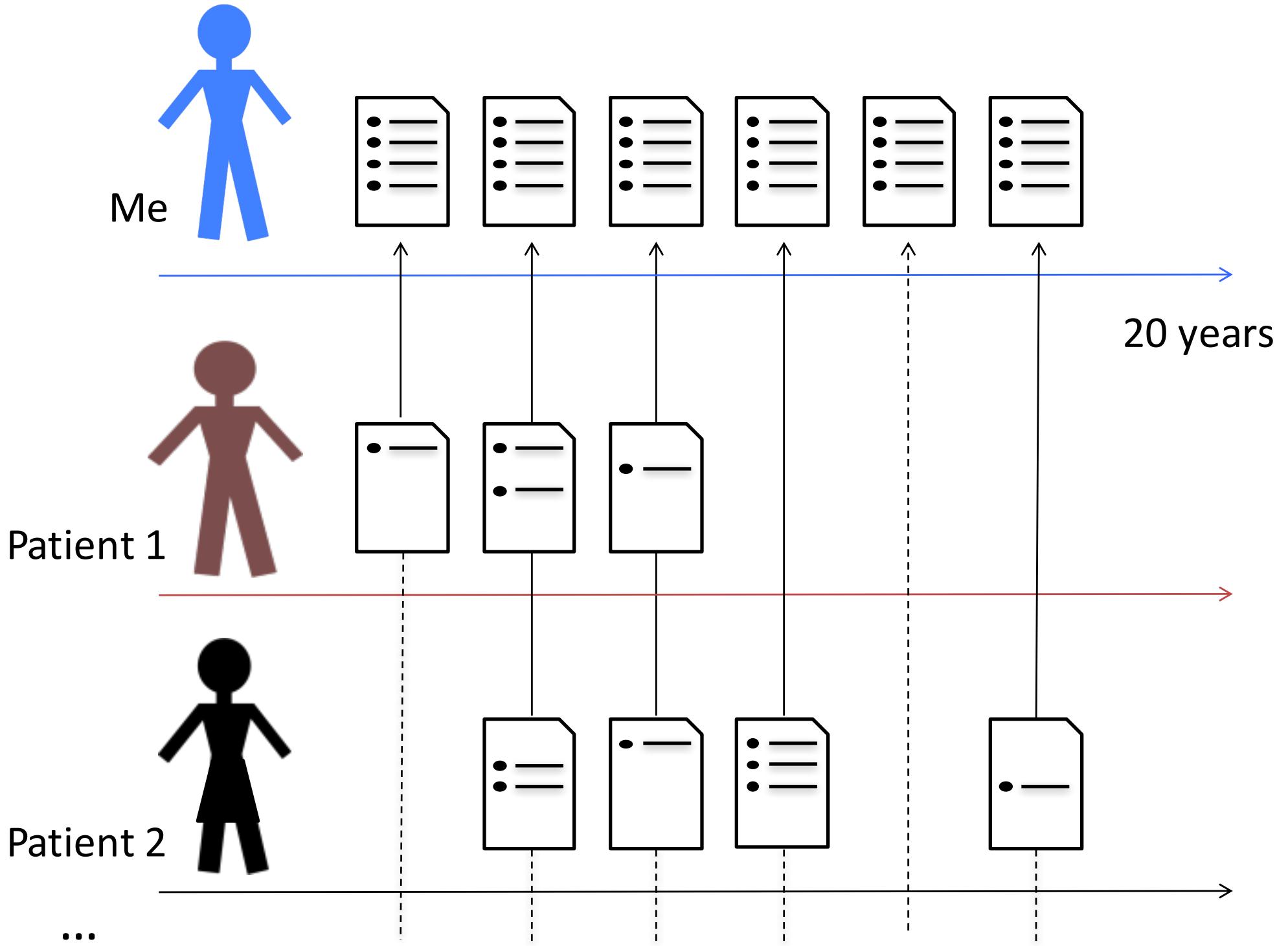
20 years

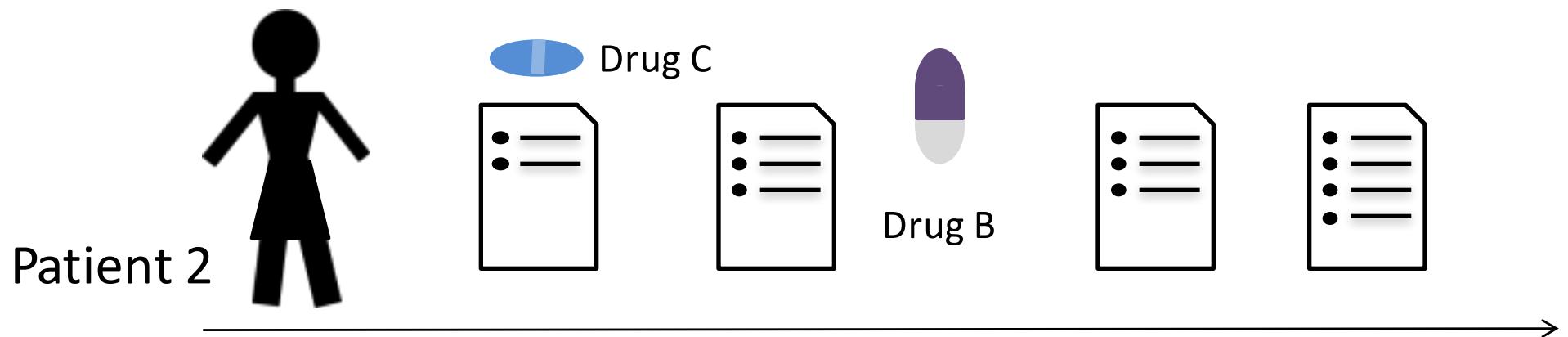
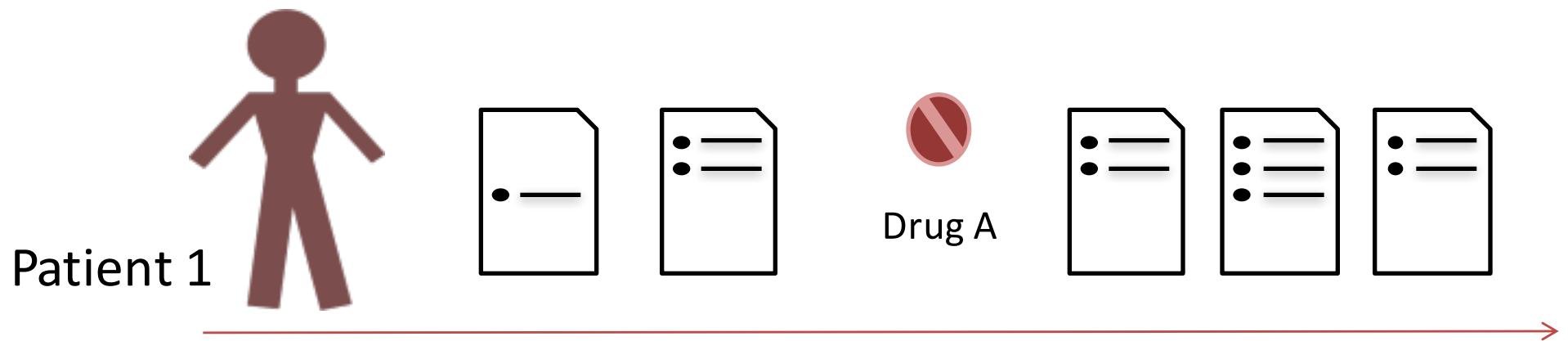
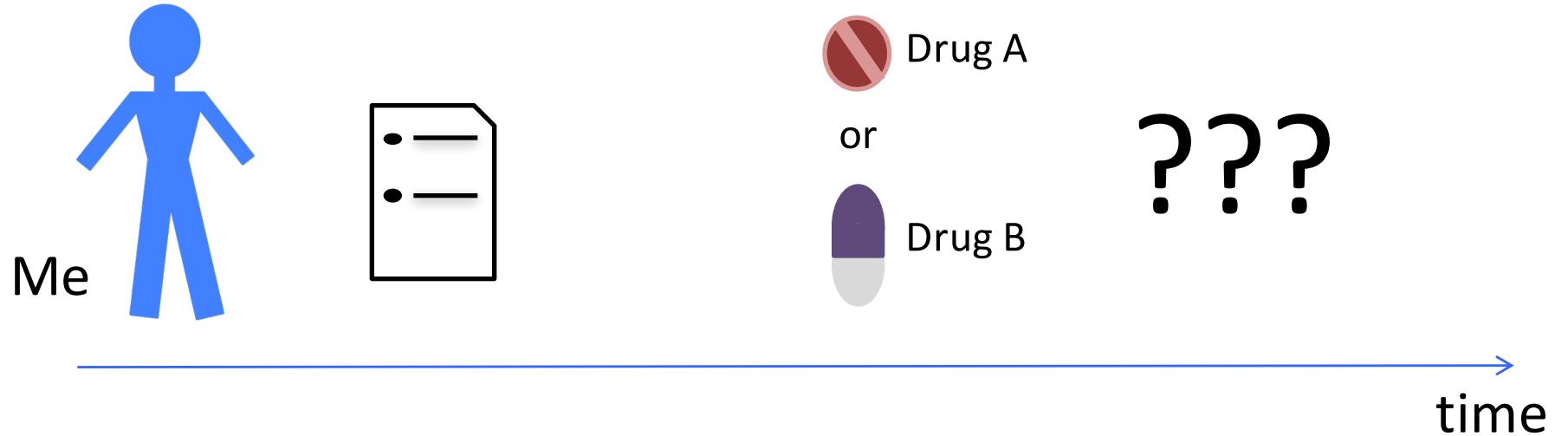


Patient 2



...





Outline for today's class

1. Brief history of AI and ML in healthcare
2. Why *now*?
3. Examples of machine learning in healthcare
4. **What is *unique* about ML in healthcare?**
5. Overview of class syllabus and projects

What makes healthcare different?

- Life or death decisions
 - Need **robust** algorithms
 - Checks and balances built into ML deployment
 - (Also arises in other applications of AI such as autonomous driving)
 - Need **fair** and **accountable** algorithms
- Many questions are about unsupervised learning
 - Discovering disease subtypes, or answering question such as “characterize the types of people that are highly likely to be readmitted to the hospital”?
- Many of the questions we want to answer are *causal*
 - Naïve use of supervised machine learning is insufficient

What makes healthcare different?

- Often very little labeled data (e.g., for clinical NLP)
 - Motivates semi-supervised learning algorithms
- Sometimes small numbers of samples (e.g., a rare disease)
 - Learn as much as possible from other data (e.g. healthy patients)
 - Model the problem carefully
- Lots of missing data, varying time intervals, censored labels

What makes healthcare different?

- Difficulty of de-identifying data
 - Need for data sharing agreements and sensitivity
- Difficulty of deploying ML
 - Commercial electronic health record software is difficult to modify
 - Data is often in silos; everyone recognizes need for interoperability, but slow progress
 - Careful testing and iteration is needed

Outline for today's class

1. Brief history of AI and ML in healthcare
2. Why *now*?
3. Examples of machine learning in healthcare
4. What is *unique* about ML in healthcare?
5. **Overview of class syllabus and projects**

Course staff

- David Sontag (instructor)
 - Assistant professor in EECS, joint IMES & CSAIL
 - PhD MIT, then 5 years as professor at NYU
 - Leads clinical machine learning research group
- Maggie Makar (teaching assistant)
 - PhD student with John Guttag, studying ML for healthcare
 - Before PhD, worked for 2.5 yrs as researcher at Brigham and Women's hospital
- We prefer Piazza to e-mail. If e-mail necessary, please send to 6.s897hst.s53@gmail.com



Prerequisites

- Must submit pre-req quiz (on course website) by 11:59PM EST today
- We assume previous undergraduate-level ML class, and comfort with:
 - Machine learning methodology (e.g. generalization, cross-validation)
 - Supervised machine learning techniques (e.g. L1-regularized logistic regression, SVMs, decision trees)
 - Optimization for ML (e.g. stochastic gradient descent)
 - Clustering (e.g. k-means)
 - Statistical modeling (e.g. Gaussian mixture models)

Logistics

- Course website:
<http://people.csail.mit.edu/dsontag/courses/mlhc17/>
- All announcements made via Piazza – make sure you are signed up for it!
- Office hours will be announced next week
- Grading:
 - 25% homework (2-3 problem sets)
 - 25% participation
 - 50% course project
- **Because of space limitations, auditors must obtain permission of course staff (e-mail 6.s897hst.s53@gmail.com)**

Homework (tentative)

- PS0 (this week): CITI “Data or Specimens Only Research” training
[https://mimic.physionet.org/gettingstarted/ac
cess/](https://mimic.physionet.org/gettingstarted/access/)
- PS1: Supervised ML on real-world clinical data, survival analysis, causal inference
- PS2: Neural nets for diagnosis from medical images and/or time series
- PS3: Disease progression modeling

Readings

- 2-4 required readings most weeks
 - Research articles, ranging from applied to theoretical
 - Required response to readings (short questions; fast) that you submit prior to next class
- Background videos (optional)
 - Neural networks (convnets, recurrent neural nets)
 - Bayesian networks
 - We will assume that you have watched these before the relevant lecture

Projects

- This will be the most interesting part of class, and where you will learn the most
- Teams of 4-5 students
- Use real-world clinical data!
- Two types of projects:
 - 6-8 projects proposed by clinical mentors, working closely with them on **their** data
 - Your own design, using publicly available data

#1: When does deployed ML break?

Clinical
mentor:



Adam Wright, PhD

Brigham and Women's Hospital

Associate Professor of Medicine, Harvard Medical School

Goal: anomaly detection system to identify clinical decision support malfunctions

The screenshot shows a clinical software interface with a header bar containing patient information (BWHlmmrmapitest, Four, 24252934 (BWH), 07/15/1939 (75 yrs.) F) and navigation links (Home, Select, Desktop, Pt Chart: Summary, Oncology, Custom, Reports, Admin, Sign, Other EMRs, Results). Below the header is a 'Reminders' section with a list of clinical alerts. A red box highlights the bottom two items in this list:

- ⓘ Patient 65 yrs or older, may be due for Pneumococcal. Please verify historical entries.
- ⓘ Patient due for seasonal influenza vaccination
- ⓘ Recommend bone densitometry every 2 years and appropriate treatment for patients at high risk for osteoporosis.
- ⓘ Pt on Thiazide for > 365 consecutive days. Checking K+ is recommended.
- ⓘ Pt on Amiodarone for > 365 consecutive days. Checking TSH level is recommended.
- ⓘ Pt on Amiodarone for > 365 consecutive days. Checking ALT is recommended.
- ⓘ No documented height in last year. Please enter height in flowsheet.

[Wright A, et al. "Analysis of clinical decision support system malfunctions: a case series and survey." J Am Med Inform Assoc (2016) 23 (6): 1068-1076]

#1: When does deployed ML break?

Clinical
mentor:

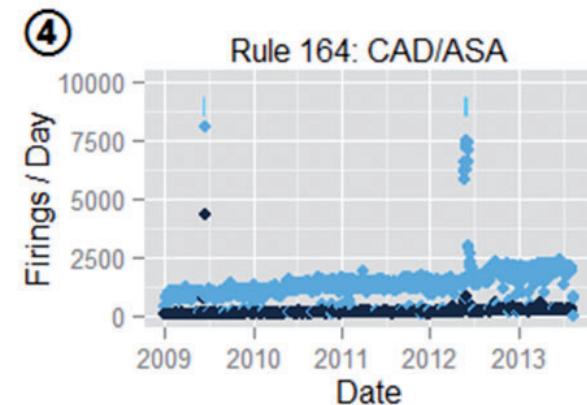
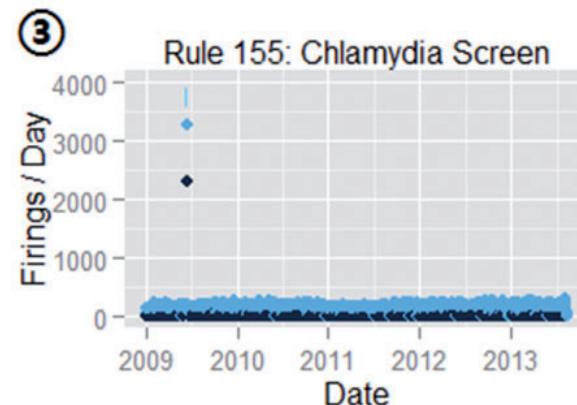
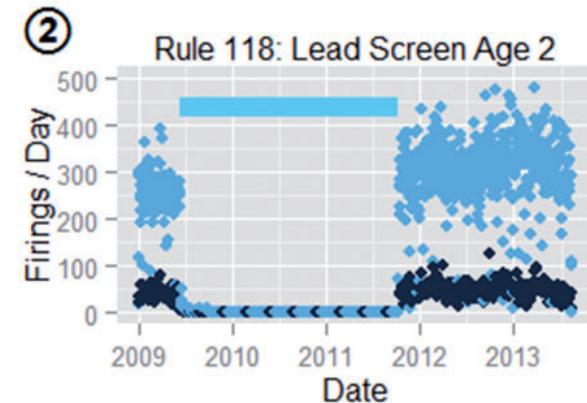
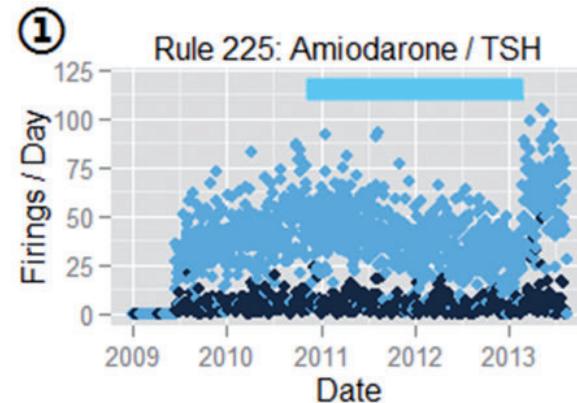


Goal: anomaly
detection system
to identify clinical
decision support
malfunctions

Adam Wright, PhD

Brigham and Women's Hospital

Associate Professor of Medicine, Harvard Medical School



[Wright A, et al. "Analysis of clinical decision support system malfunctions: a case series and survey." J Am Med Inform Assoc (2016) 23 (6): 1068-1076]

#2: Improving accuracy of CDS alerts

Clinical
mentor:



Adam Wright, PhD

Brigham and Women's Hospital

Associate Professor of Medicine, Harvard Medical School

- Most clinical decision support (CDS) systems are simple & rule-based (“If the patient is over 65 and has not received a vaccination, suggest one”)
- Once deployed, we gather data on when CDS alerts are ignored or overridden by users
- **Goal: use machine learning to improve accuracy of alerts.**

Other angles we might consider:

- Clustering to understand *why* alerts were overridden
- Tackling the false negatives, i.e. broadening the alerts
- Deep learning on clinical text
- Learning interpretable models

#3 Predicting antibiotic resistance

Clinical
mentors:



Steven Horng, MD MMSc

Eugene Kim, MD

Beth Israel Deaconess Medical Center
Dept. of Emergency Medicine



Sanjat Kanjilal, MD MPH

Massachusetts General Hospital
Div. of Infectious Diseases

- Culture results can take up to 6 days
- Patients are started on empiric antibiotics based on population-level resistance patterns
- Critical patients, if started on wrong antibiotics, may not survive that long
- Can we predict a patient's personalized antibiotic resistance profile even before their culture is available?

#4 Progression of Congestive Heart Failure

Clinical
mentors:



Steven Horng, MD MMSc

Beth Israel Deaconess Medical Center
Dept. of Emergency Medicine



Sandeep Gangireddy, MD

Beth Israel Deaconess Medical Center
Cardiologist, Informatics Research Fellow

- Heart unable to pump enough blood to meet body's demands
- Heart failure hospitalizations cost the US over \$17 billion/year
 - Physicians struggle to diagnose & treat heart failure exacerbations before patients require hospitalization
- Patients with heart failure progress at different rates. It is unclear when patients will worsen, and the gold standard test is infrequently performed
- Goal: predict heart failure progression using frequently collected data in the electronic medical record
 - Vitals, medications, orders, laboratory tests, echocardiography & chest x-ray reports

Projects

PUBLICLY AVAILABLE DATASETS

Critical care (~40K patients)



If you use MIMIC data or code in your work, please cite the following publication:

MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. *Scientific Data* (2016).
DOI: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35). Available from: <http://www.nature.com/articles/sdata201635>

Multiple Myeloma (975 patients)

Secure | <https://research.themmr.org>

MM RF Researcher Gateway

Forum Explore IA8 MyCoMMpass Help Download Log Out

Latest News

MM Literature RG MMRF

TOPCODER CHALLENGE
INFORMATION Hello TopCoder
Contestants, and welcome to the
MMRF CoMMpass Researcher
Gateway! Your access will go live at
9PM EST December 19. Once you
enter the Gateway, the 4 files that
you need can be found under the
"Download" tab. Good Luck!

**Eleventh CoMMpass Data Trenche
(IA11) To Be Released in February**
2017 Data from the latest CoMMpass
interim analysis will be posted on the
Researcher Gateway by mid-
February 2017. Public users will
have access to clinical data from 975

DEMOGRAPHICS

COPY NUMBER VARIATION #

A stacked bar chart titled 'DEMOGRAPHICS' showing the 'Number of Patients' on the y-axis (0 to 50) against 'Age' groups on the x-axis. The bars are stacked by race: White (red), Black / African American (orange), and Other race (green). The chart shows a peak in the 65-69 age group for White patients.

Age Group	White	Black / African American	Other race	Total
35 - 39	3	0	0	3
40 - 44	5	0	0	5
45 - 49	9	1	0	10
50 - 54	12	2	1	15
55 - 59	17	5	2	24
60 - 64	25	5	3	33
65 - 69	42	7	3	52
70 - 74	32	5	3	40
75 - 79	18	3	2	23
80 - 84	14	3	1	18
85 - 89	2	0	0	2
90 - 94	1	0	0	1

Parkinson's disease (400+ subjects)

← → C ① www.ppmi-info.org/access-data-specimens/download-data/

Parkinson's Progression Markers Initiative



PARKINSON'S
PROGRESSION
MARKERS
INITIATIVE

Play a Part in Parkinson's Research

[About PPMI](#) [Study Design](#) [Access Data & Specimens](#) [Publications & Presentations](#) [PPMI News](#)

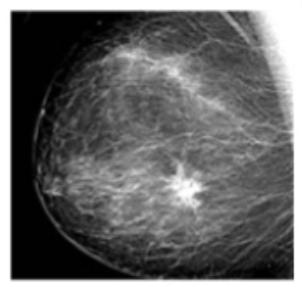
[Request Specimens](#) [Request Cell Lines](#) [Download Data](#) [Ongoing Analysis](#) [Data & Specimens FAQ](#) [Com](#)

DOWNLOAD DATA

Through this Web site, qualified researchers may obtain access to all clinical, imaging and biomarker data collected in PPMI. This includes raw and processed MRI and SPECT images. All data are de-identified to protect patient privacy.

Mammography (86K subjects)

The Digital Mammography DREAM Challenge



Build a model to help reduce the recall rate for breast cancer screening

Learn more & register to participate here: www.synapse.org/Digital_Mammography_DREAM_Challenge



Coding4Cancer
Challenge for improving cancer screening

GroupHealth.
RESEARCH INSTITUTE



Icahn School of Medicine
at Mount Sinai



Competitive Period Launch: Nov 18, 2016

Competitive Period Close: May 9, 2017

Out of 1000 women screened, only 5 will have breast cancer

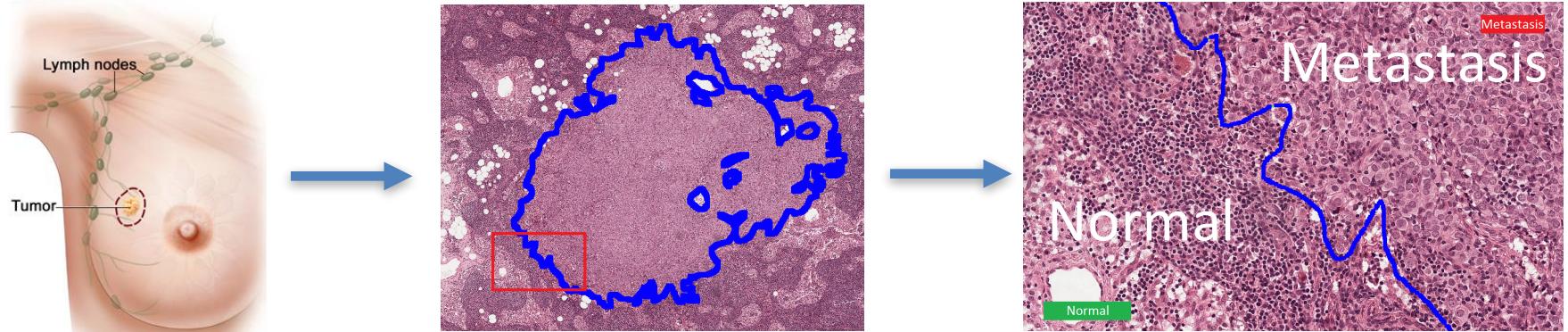
Goal: develop algorithms for risk stratification of screening mammograms that can be used to improve breast cancer detection

Pathology (200 patients)



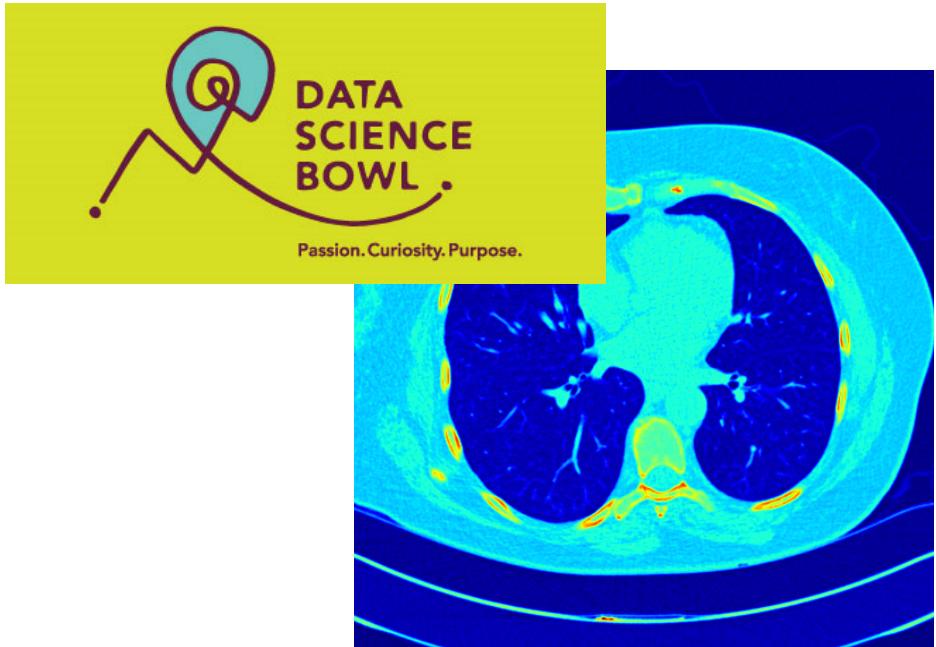
Competitive Period Launch: Nov 20, 2016

Competitive Period Close: April 1, 2017



Whole slide images with lesion-level annotations of metastases

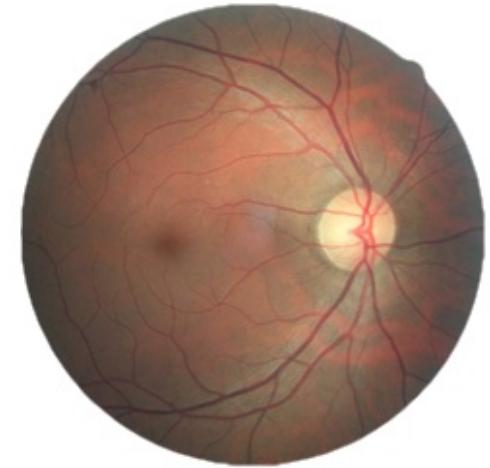
Lung cancer



Enter Competition By: Mar 31, 2017
Competitive Period Close: April 12, 2017

(Last year's challenge was on diagnosing heart disease – data also available, via Kaggle)

Diabetic retinopathy



<https://www.kaggle.com/c/diabetic-retinopathy-detection>