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

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.

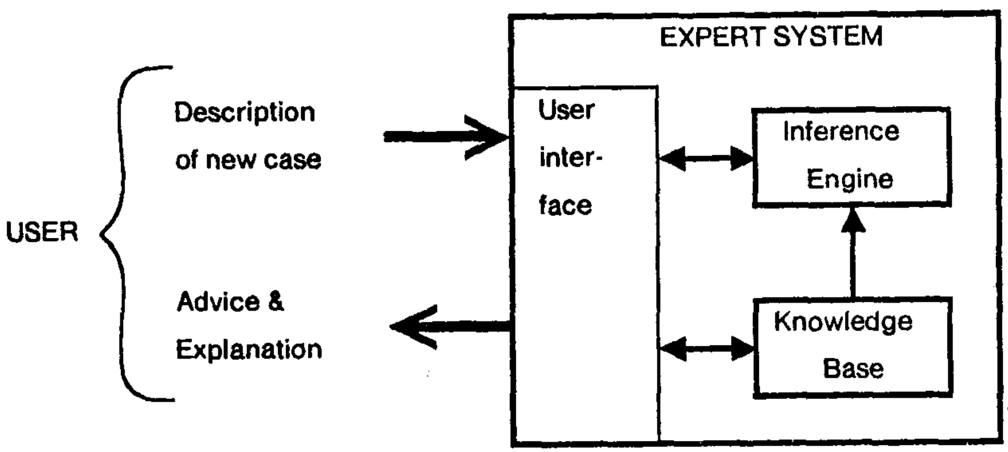
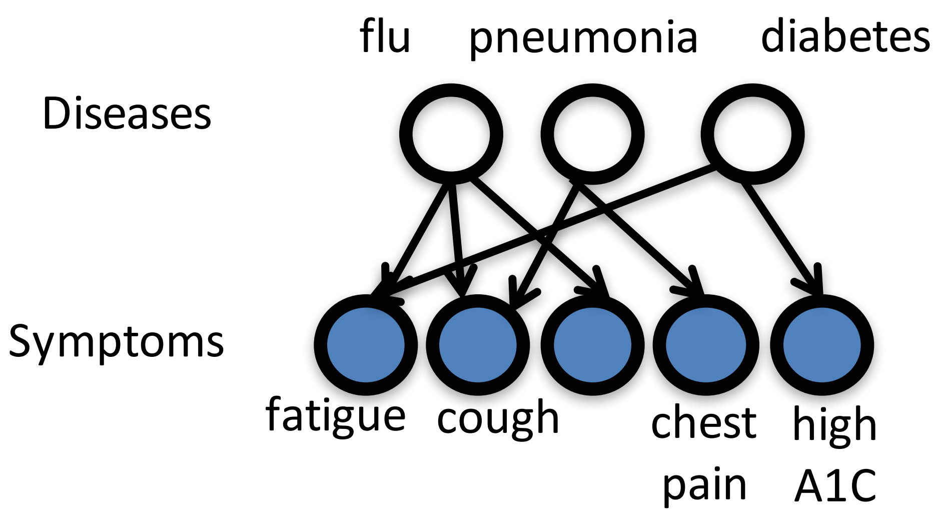


FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.

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

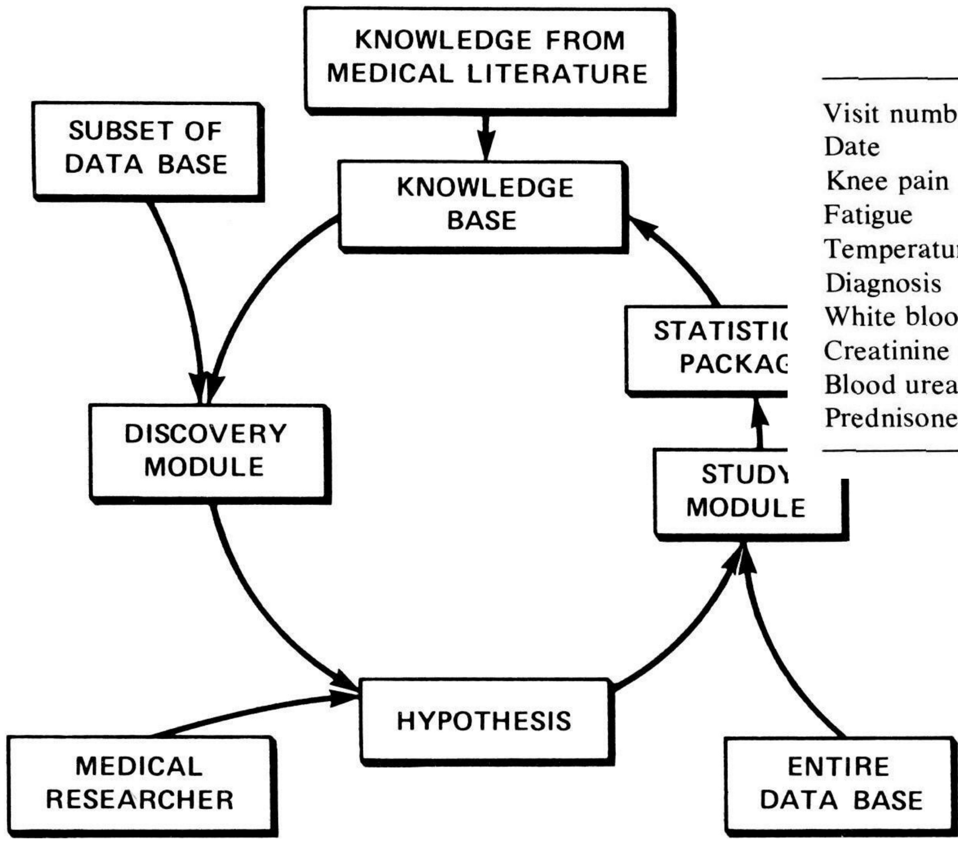


TABLE 1

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

Discovers 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

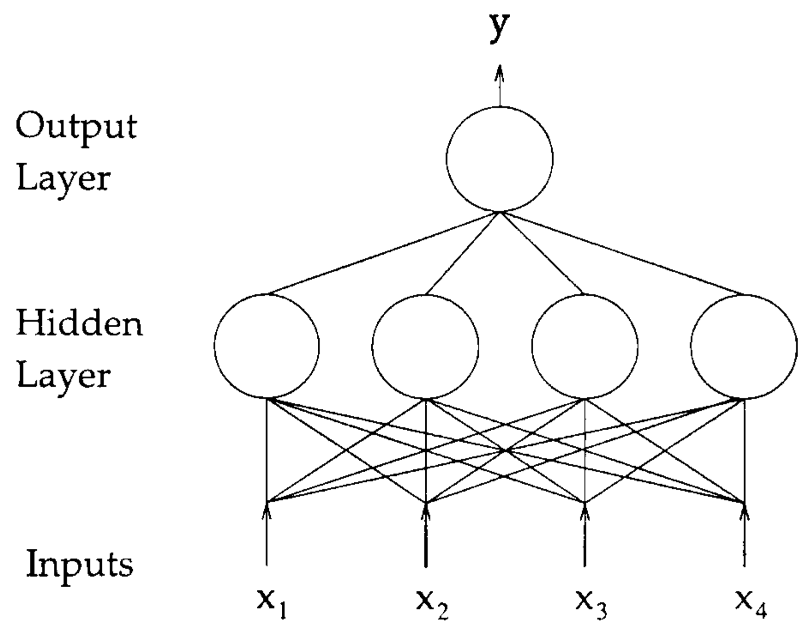


FIGURE 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.

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: Training	Test	P†	Network	D‡	Accuracy§: 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	94
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
Mycardial 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 AUROCC value. Neural = accuracy of neural net, Other = accuracy of best other method.

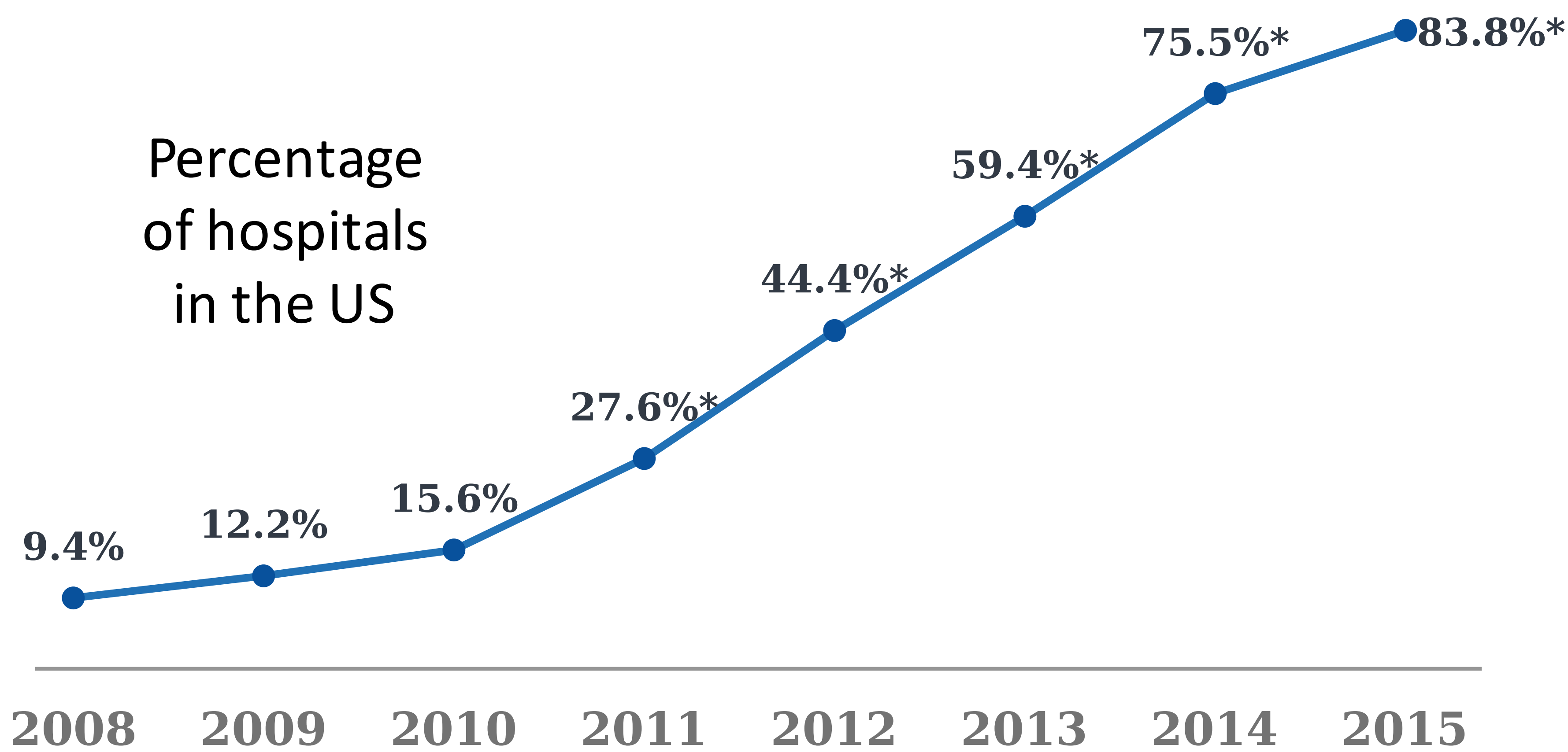
Outline for today's class

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

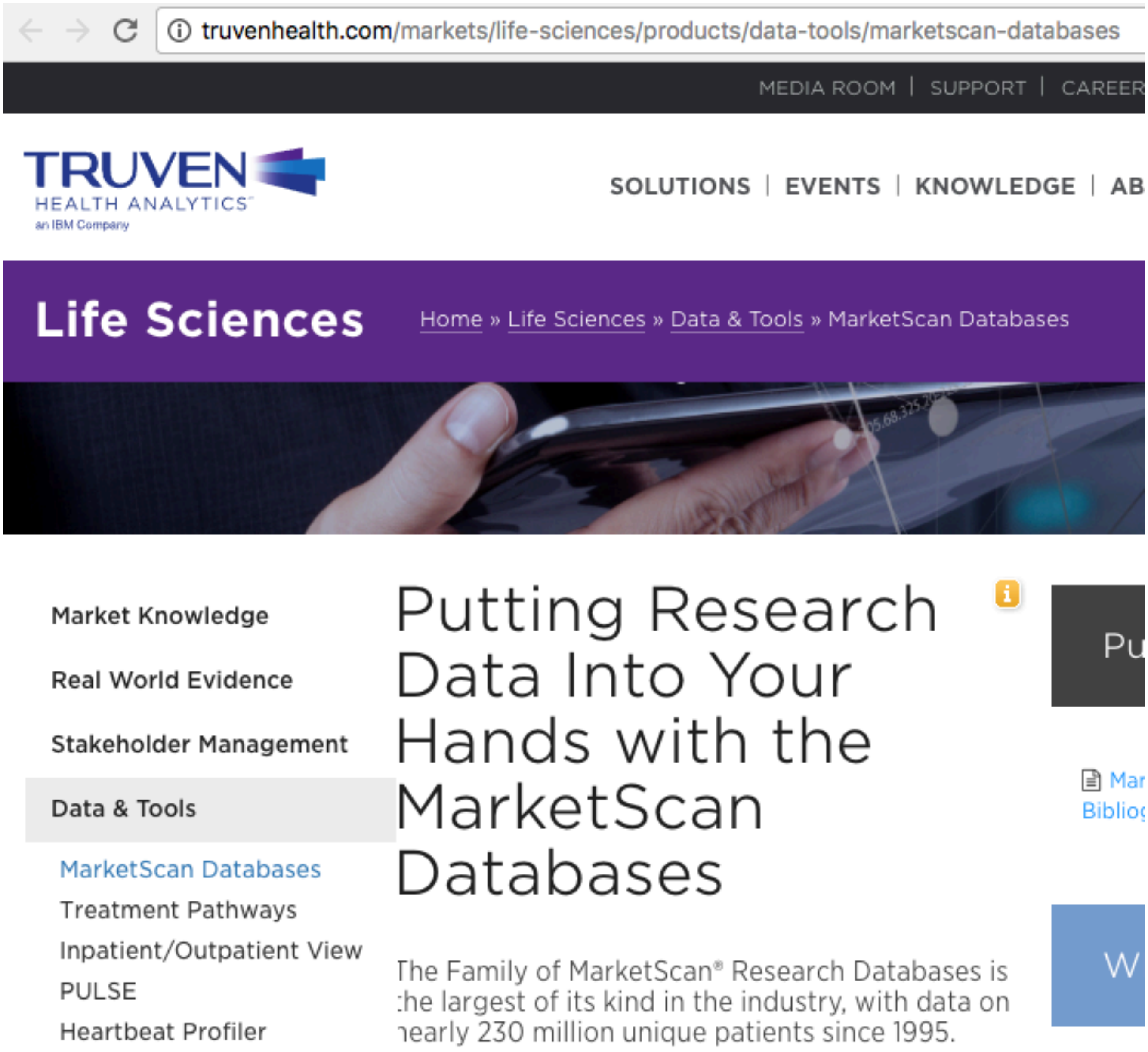
Massachusetts Institute of Technology

Laboratory for Computational Physiology

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

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

Large datasets



"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,

…

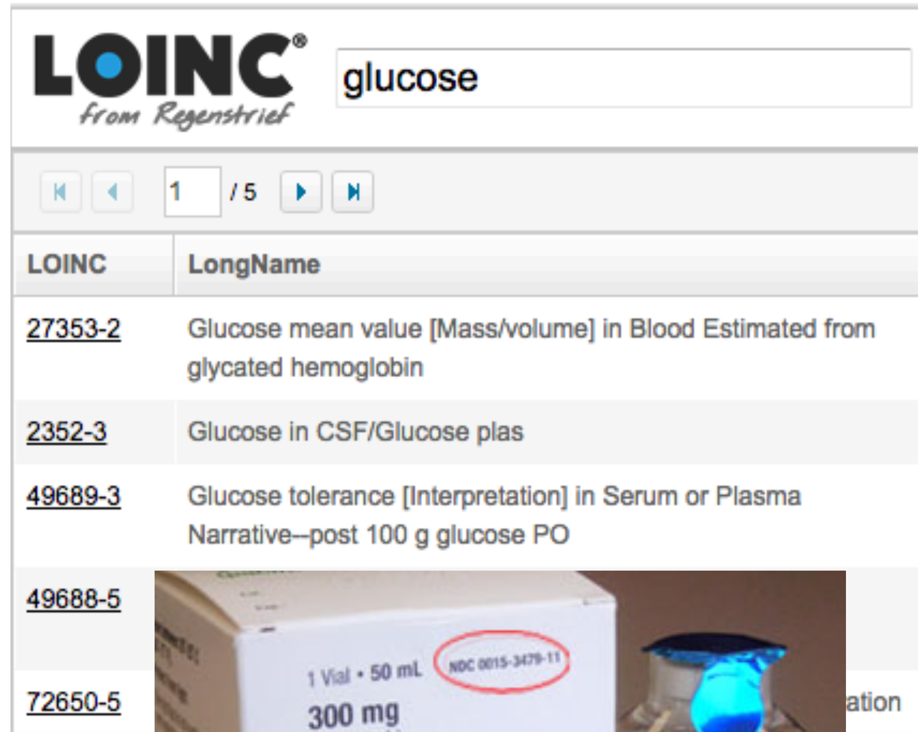
…

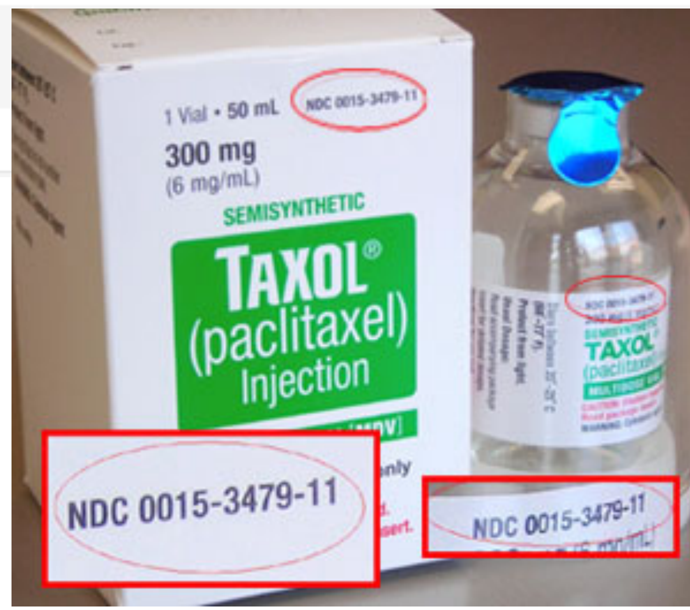
[https://en.wikipedia.org/wiki/List_of_ICD-9_codes]

[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





[http://oplinc.com/newsletter/index_May08.htm]

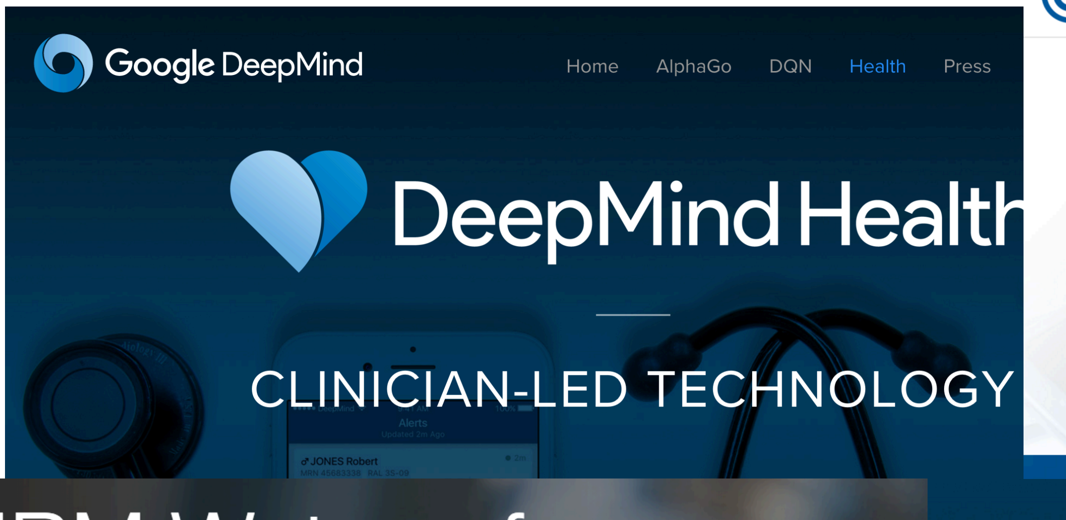
Why now?

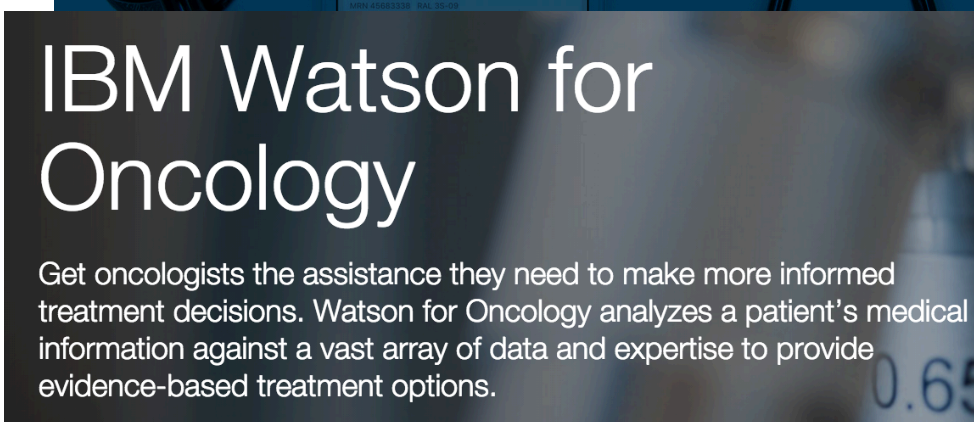
ALGORITHMS

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**

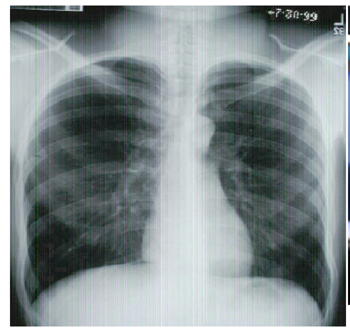
Data in Emergency Department (ED)

Electronic records for over 300,000 ED visits

Triage Information
(Free text)

MD comments
(free text)

Specialist consults



Physician
documentation

30 min

2 hrs

T=0

Disposition

Repeated vital signs
(continuous values)
Measured every 30 s

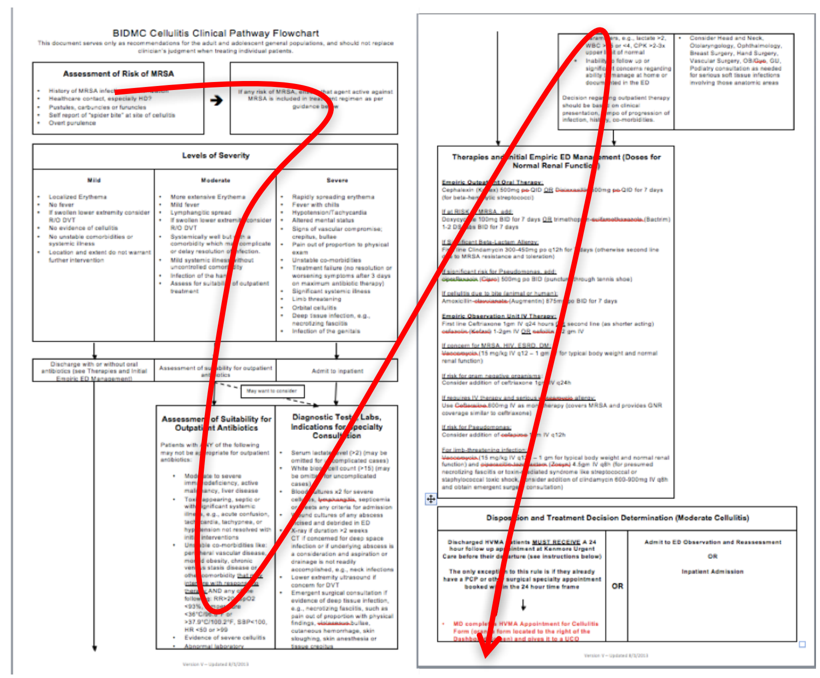
Lab results
(Continuous valued)

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



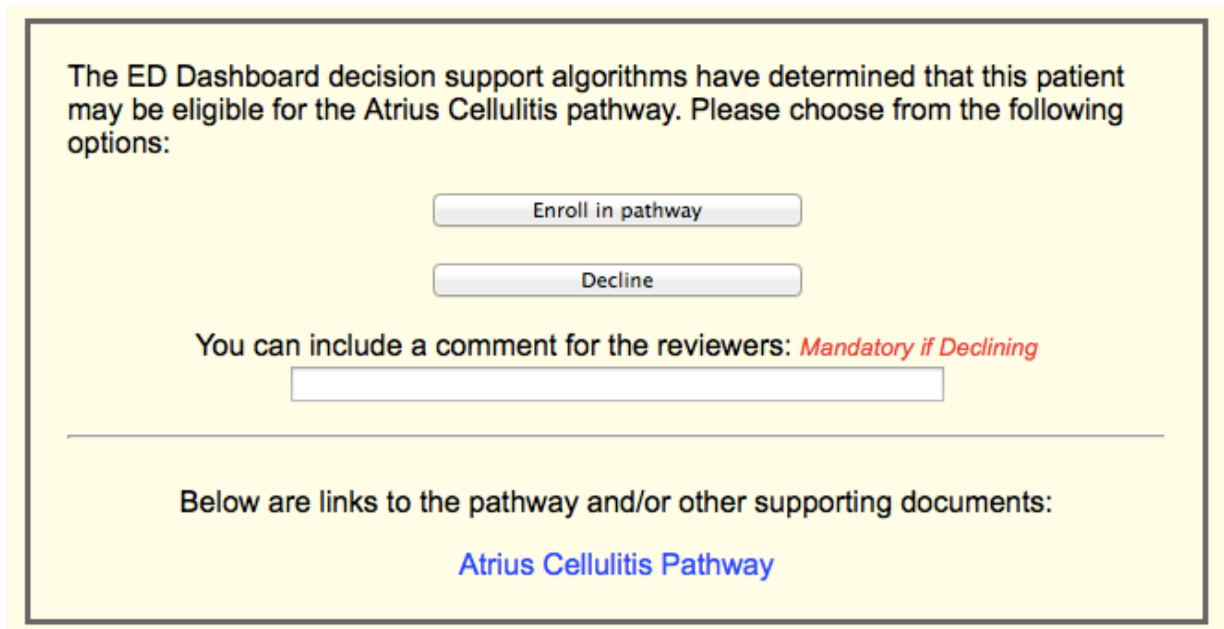
Pathways have been shown to reduce in-hospital complications without increasing costs [Rotter et al 2010]

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:

Enroll in pathway

Decline

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

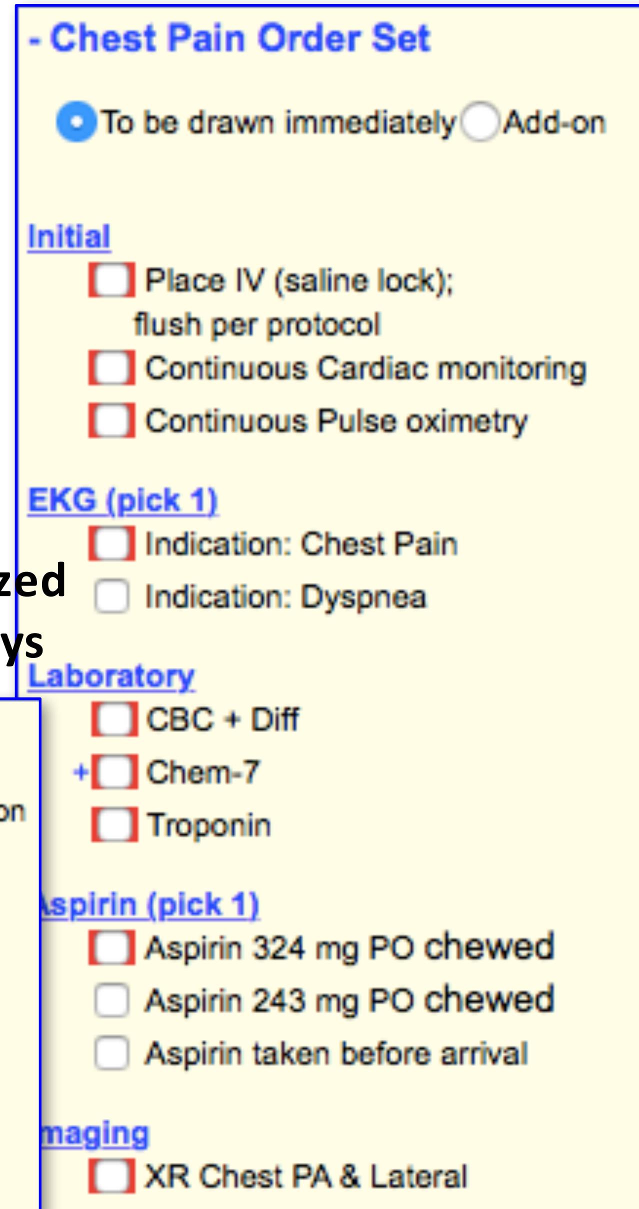
Our task:

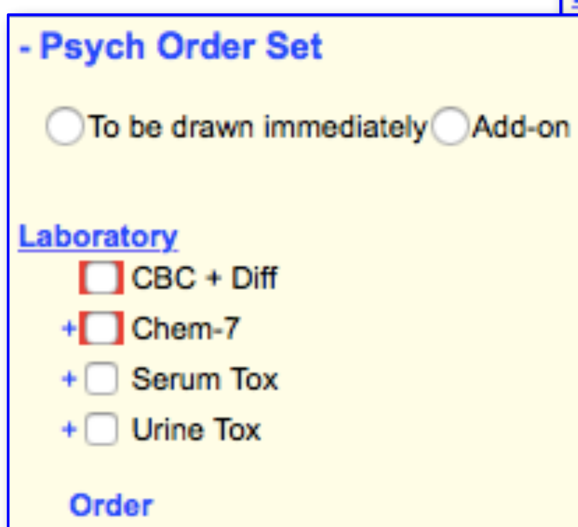
Determine whether a patient has or is suspected to have cellulitis

Opportunities for machine learning

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

Automatically place **specialized order sets on patient displays**





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

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

History

- Alcoholism
- Anticoagulated
- Asthma/COPD
- Cancer
- Congestive heart failure
- Diabetes
- HIV+
- Immunosuppressed
- Liver malfunction

Acute

- Abdominal pain
- Allergic reaction
- Ankle fracture
- Back pain
- Bicycle accident
- Cardiac etiology
- Cellulitis
- Chest pain
- Cholecystitis
- Cerebrovascular accident
- Deep vein thrombosis
- Employee exposure
- Epistaxis
- Gastroenteritis
- Gastrointestinal bleed
- Geriatric fall
- Headache
- Hematuria
- Intracerebral hemorrhage
- Infection
- Kidney stone
- Laceration
- Motor vehicle accident
- Pancreatitis
- Pneumonia
- Psych
- Obstruction
- Septic shock
- Severe sepsis
- Sexual assault
- Suicidal ideation
- Syncope
- Urinary tract infection

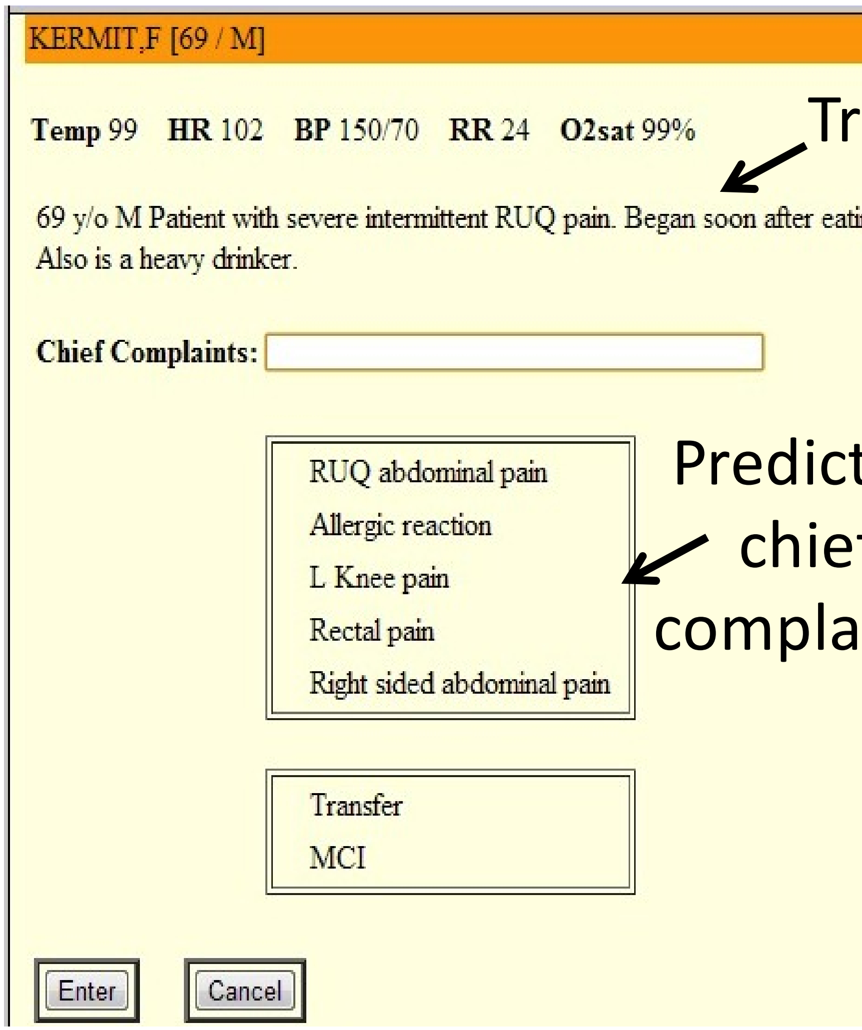
[Halpern, Horng, Choi, Sontag, JAMIA ‘16]

Opportunities for machine learning

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

Improving documentation: Chief complaints

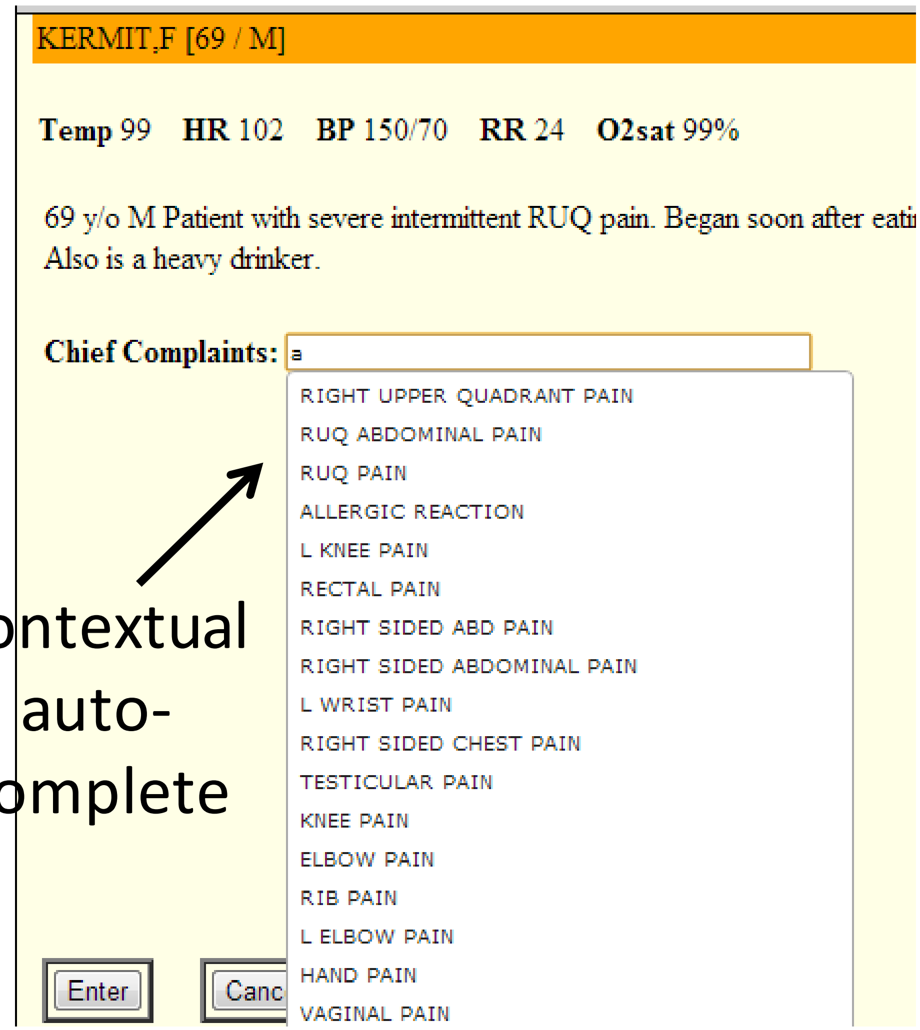
Changed workflow to have chief complaints assigned *last*. Predict them.



Triage note

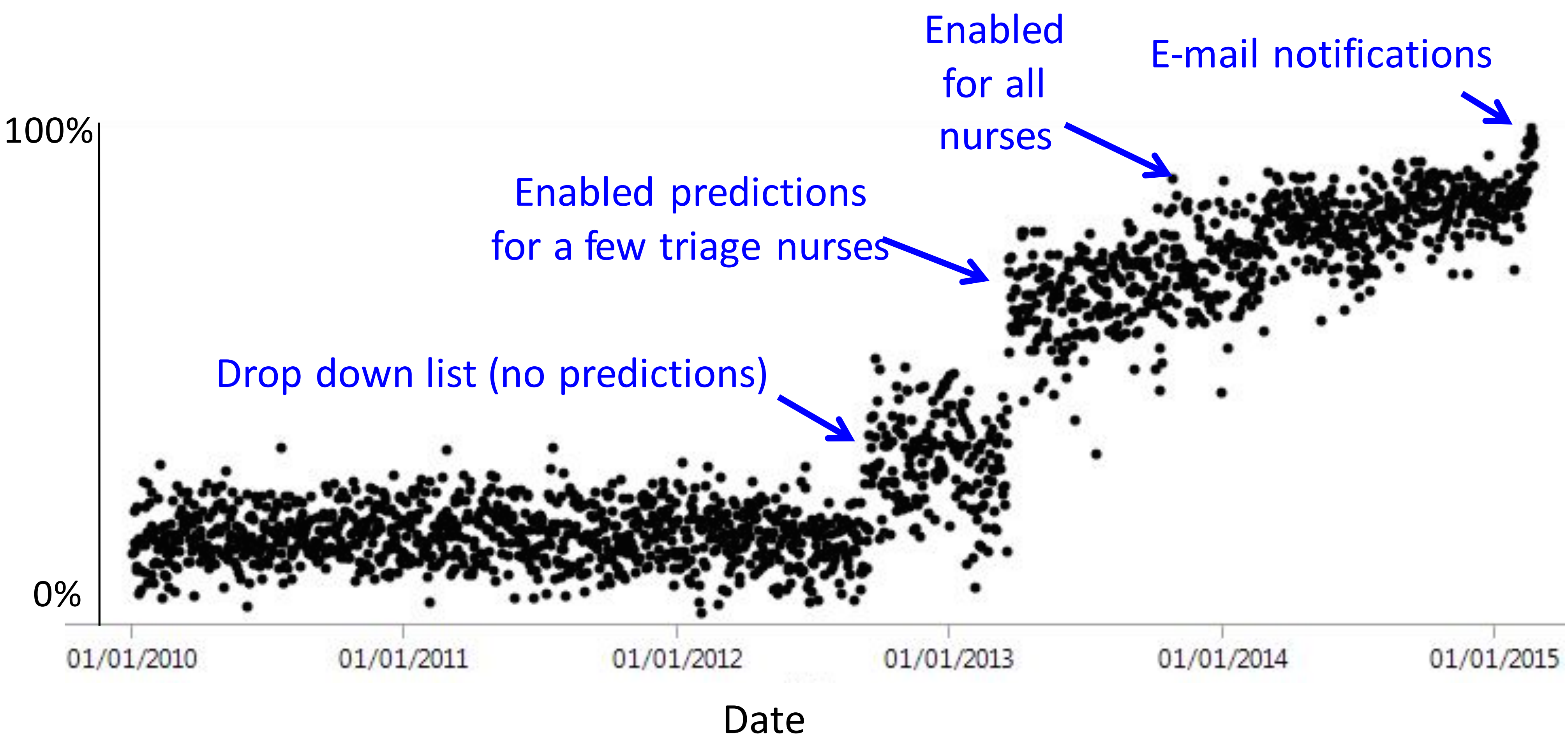
Predicted chief complaints

Contextual auto-complete



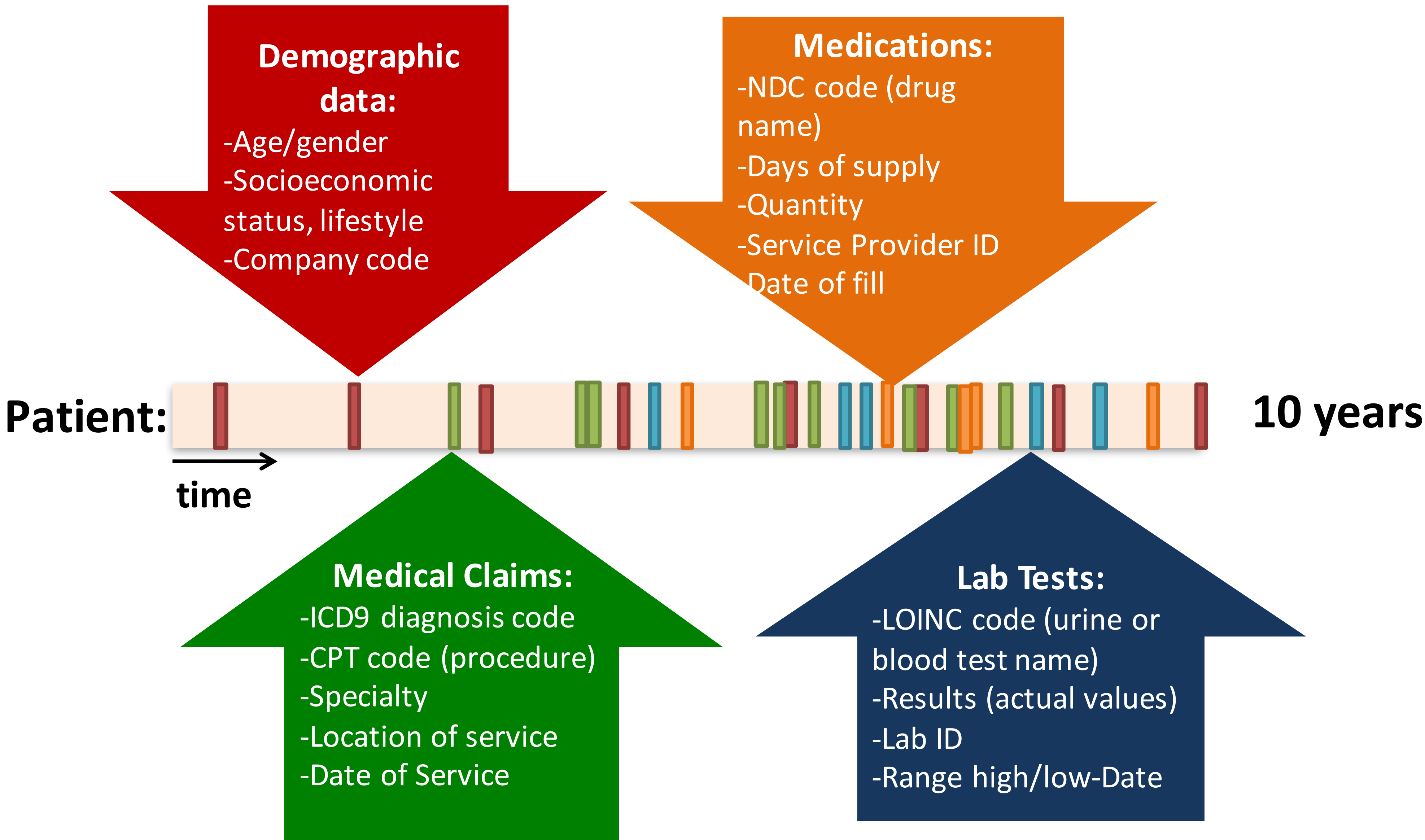
Using for all 55,000 patients/year that present at BIDMC ED

Improving documentation: Chief complaints



Percentage of *standardized* chief complaints (per week)

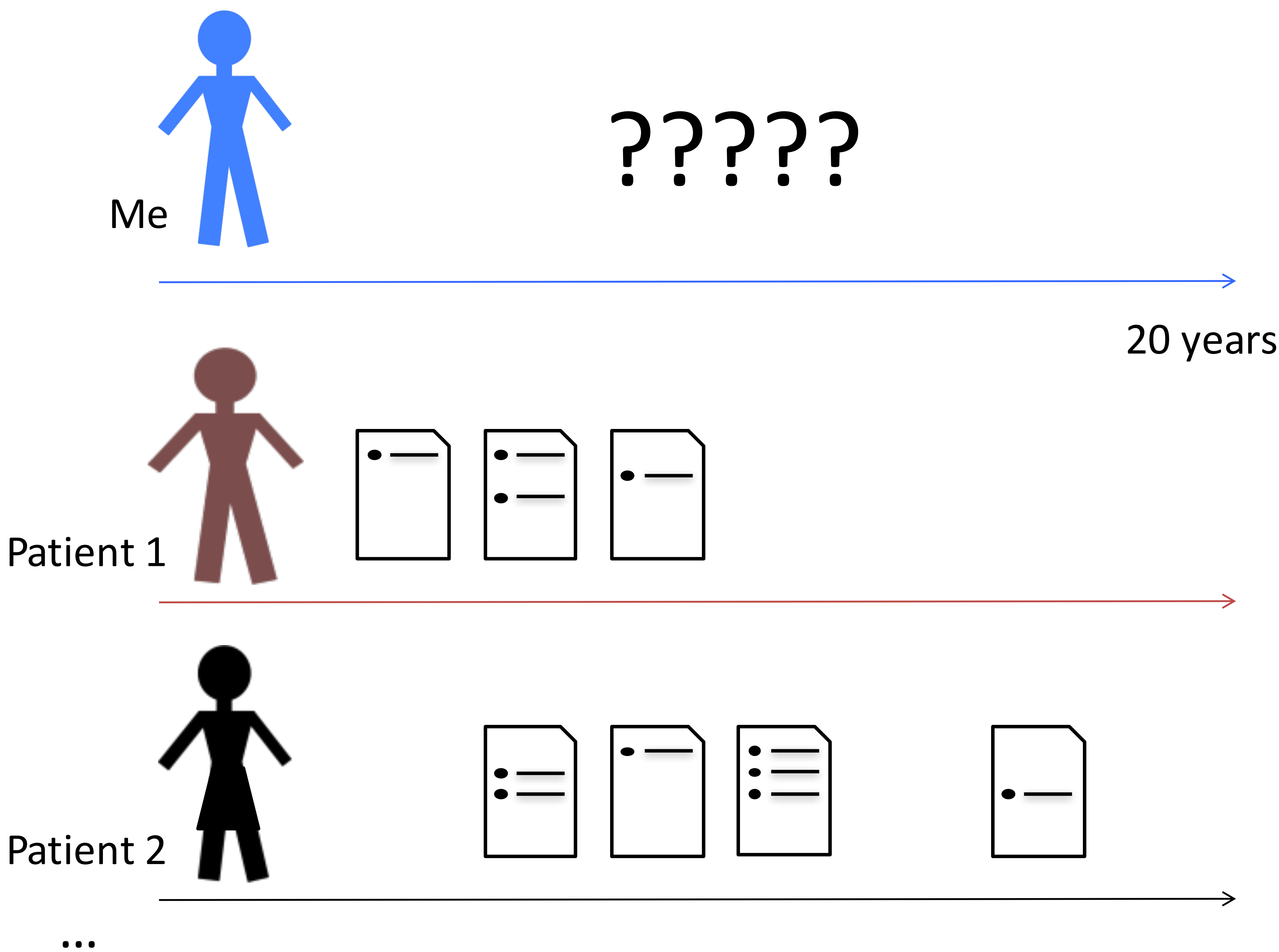
Zooming out…

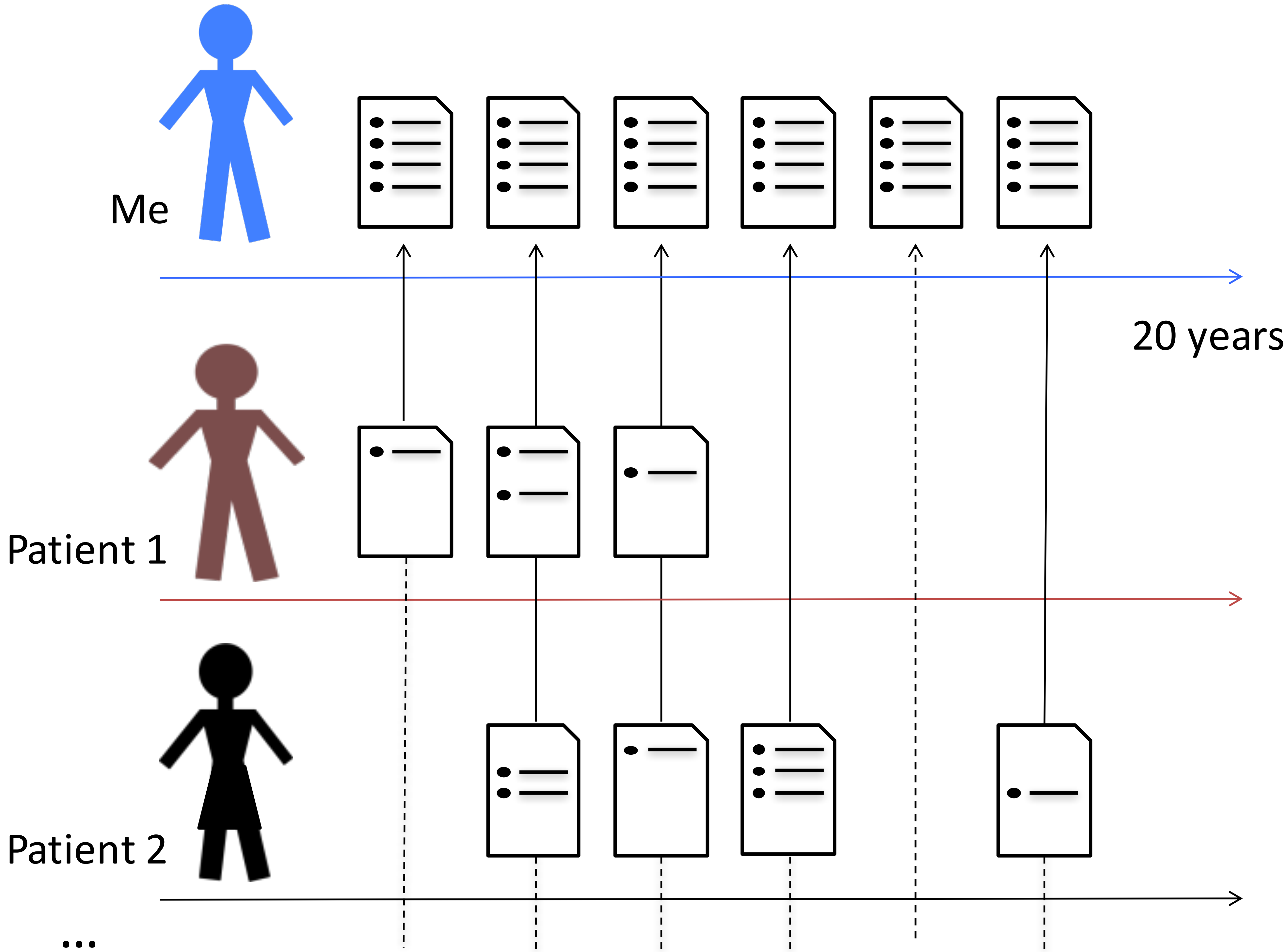


Collaboration with: Independence Blue Cross

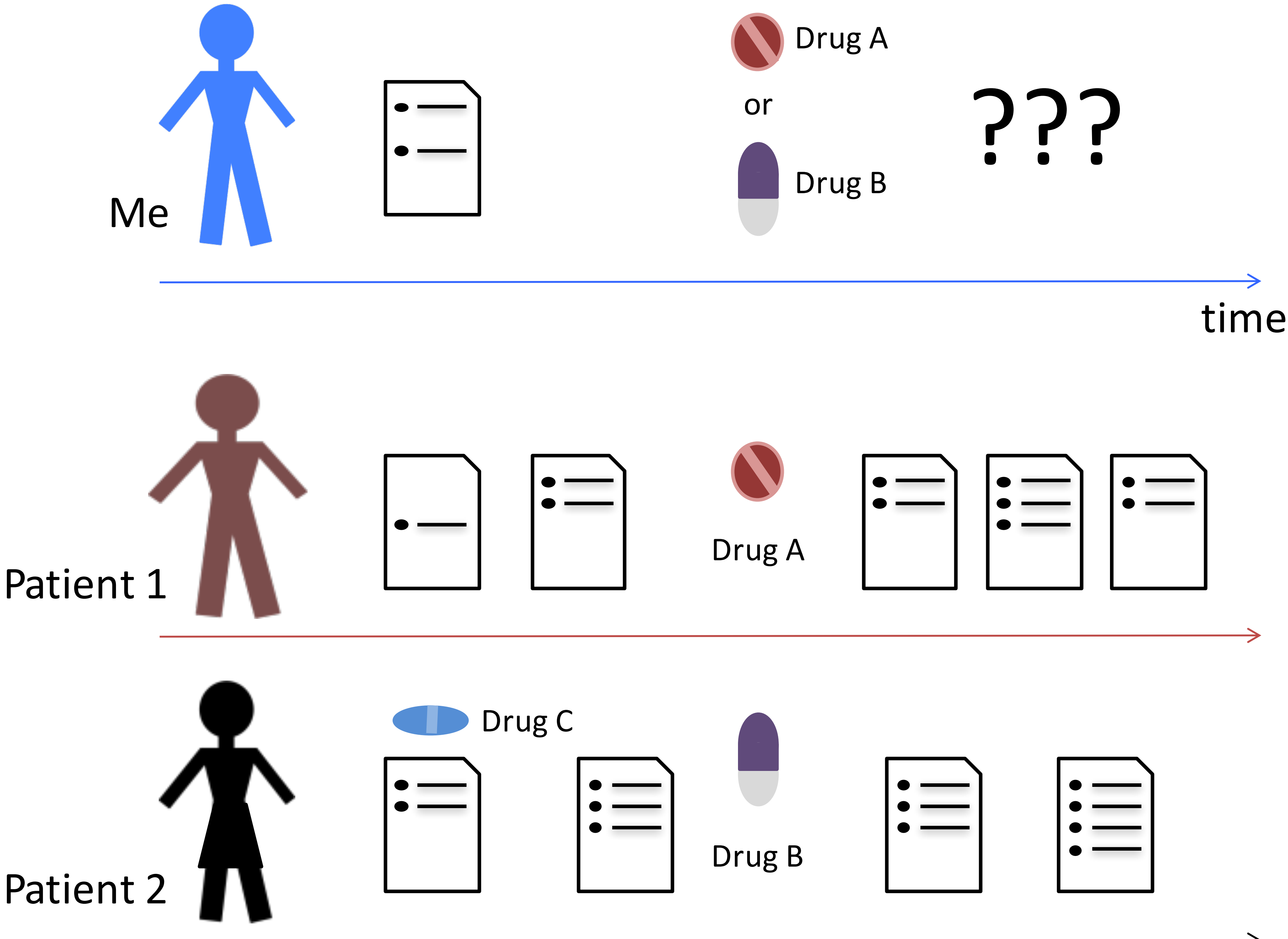
Temporal modeling of disease progression

- 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
20 years
Patient 1
Patient 2
...



Me
20 years
Patient 1
Patient 2
...


Drug A
or
Drug B
???
Me
time
Patient 1
Drug A
Drug C
Drug B
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?

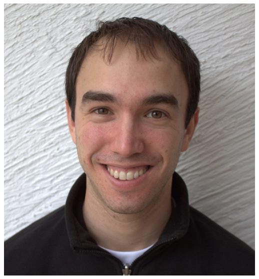
- Difficulty of de-identifying data
  - Need for data sharing agreements and sensitivity
- Difficulty of deploying ML
  - Commercial electronic health record software is difficult to modify
  - Data is often in silos; everyone recognizes need for interoperability, but slow progress
  - Careful testing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

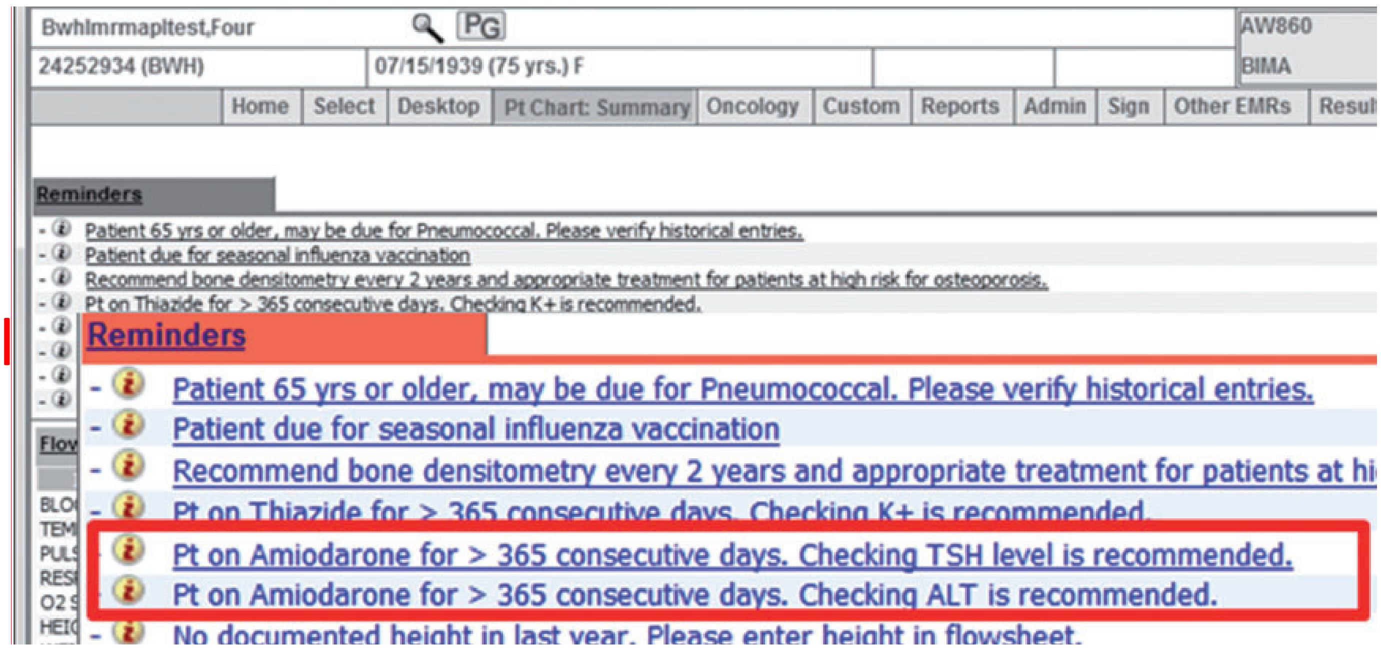
- This will be the most interesting part of class, and where you will learn the most
- Teams of 4-5 students
- Use real-world clinical data!
- Two types of projects:
  - 6-8 projects proposed by clinical mentors, working closely with them on **their** data
  - Your own design, using publicly available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

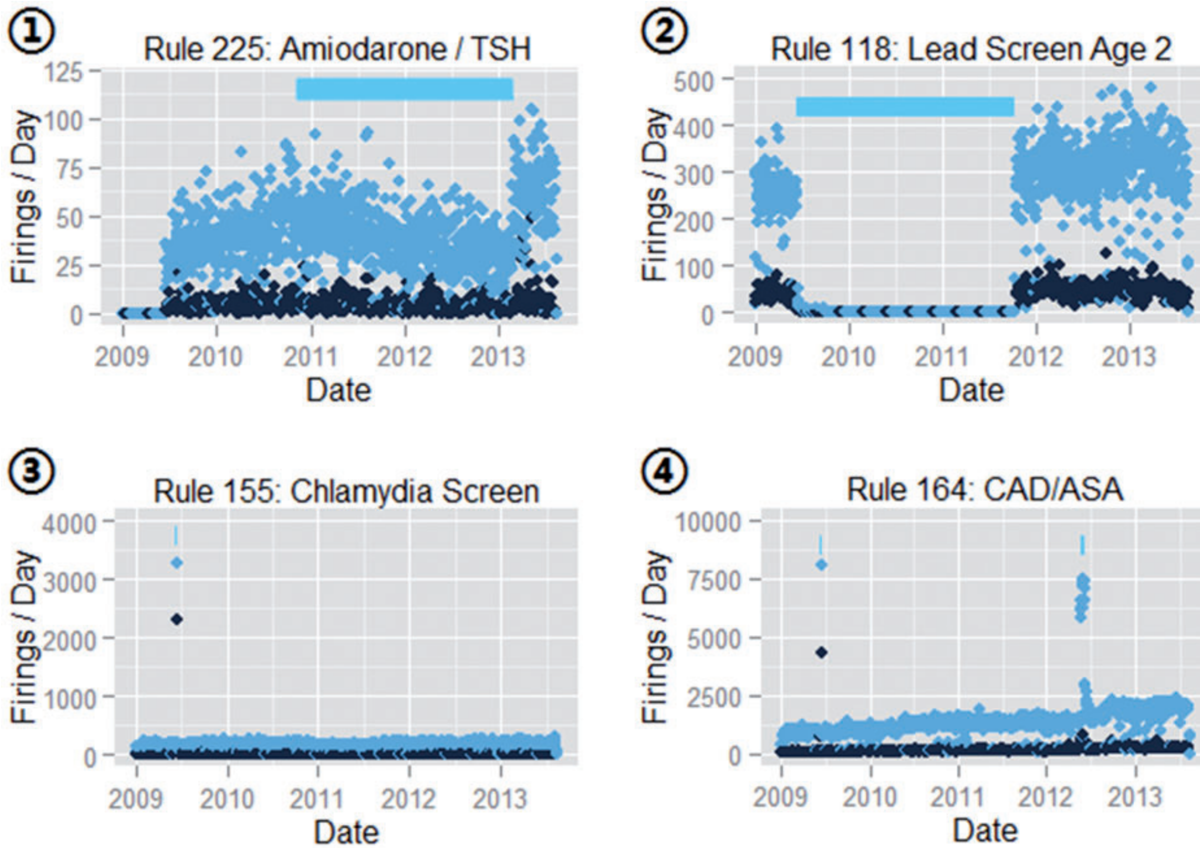


[Wright A, et al. "Analysis of clinical decision support system malfunctions: a case series and survey." J Am Med Inform Assoc (2016) 23 (6): 1068-1076]

#1: When does deployed ML break?

Clinical mentor:

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

[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

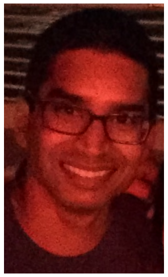
Sanjat Kanjilal, MD MPH
Massachusetts General Hospital
Div. of Infectious Diseases

- 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

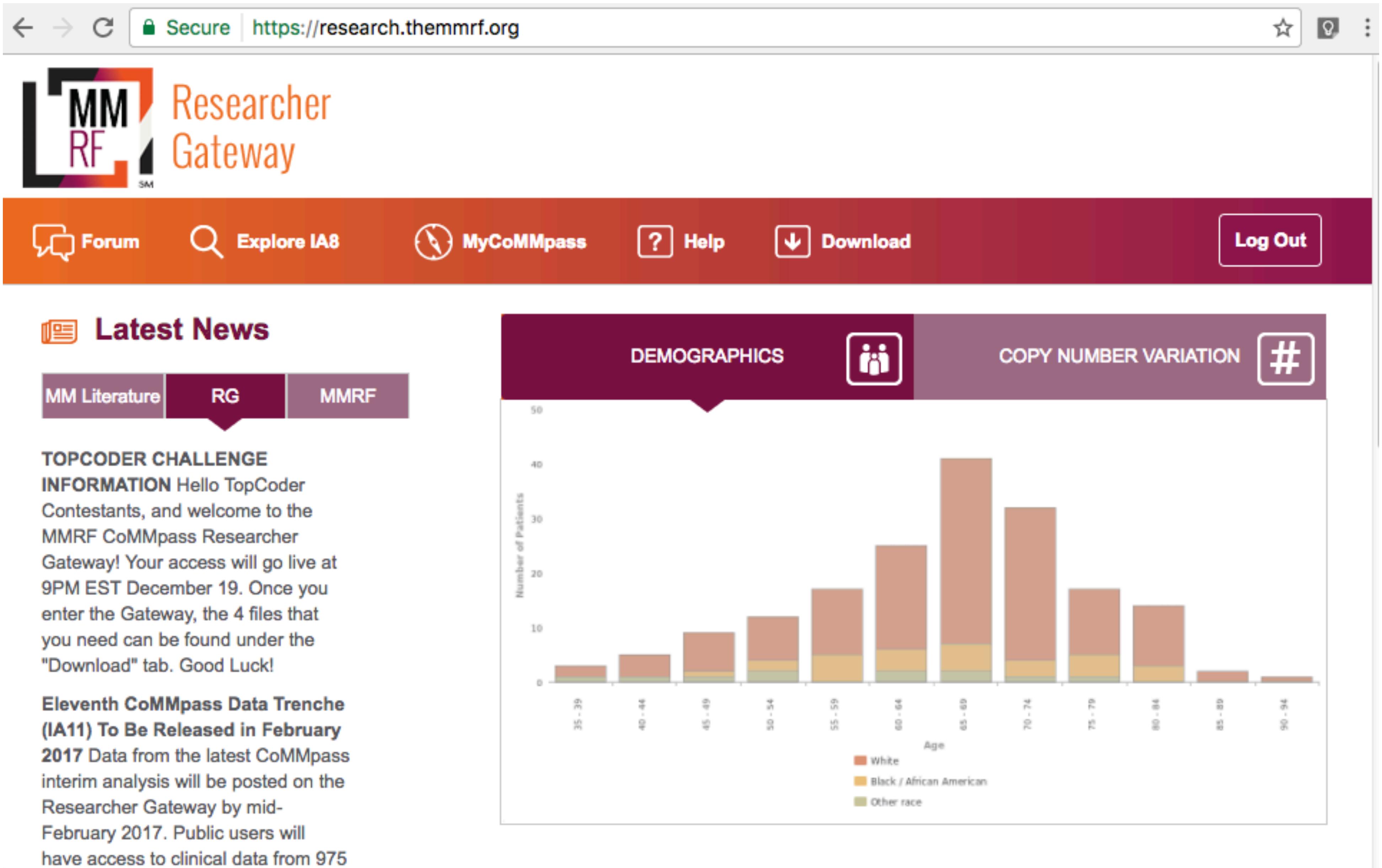
- Heart unable to pump enough blood to meet body's demands
- Heart failure hospitalizations cost the US over $17 billion/year
  - Physicians struggle to diagnose & treat heart failure exacerbations before patients require hospitalization
- Patients with heart failure progress at different rates. It is 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

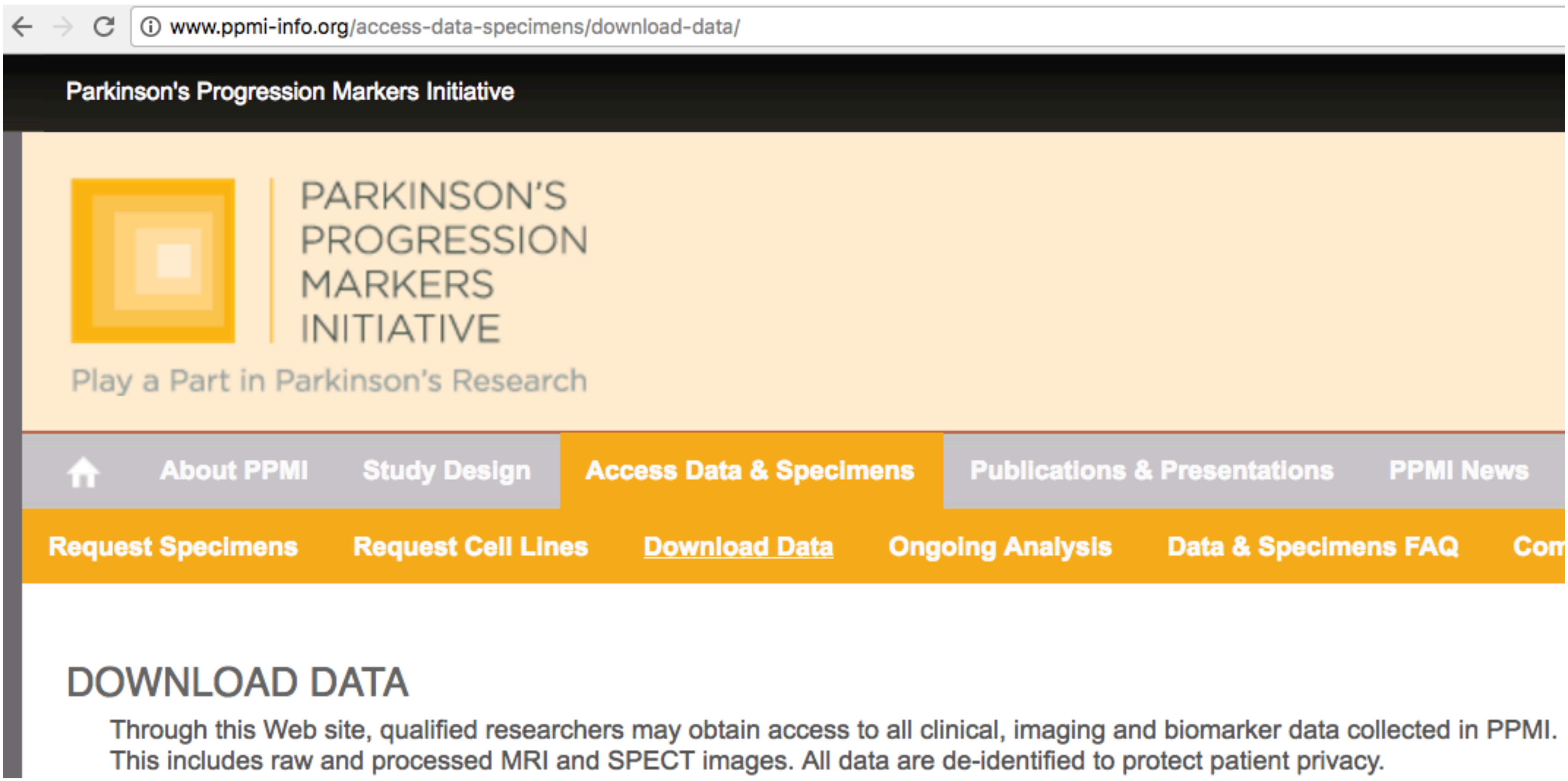
PUBLICLY AVAILABLE DATASETS

Critical care (~40K patients)

Multiple Myeloma (975 patients)



Parkinson’s disease (400+ subjects)



Mammography (86K subjects)

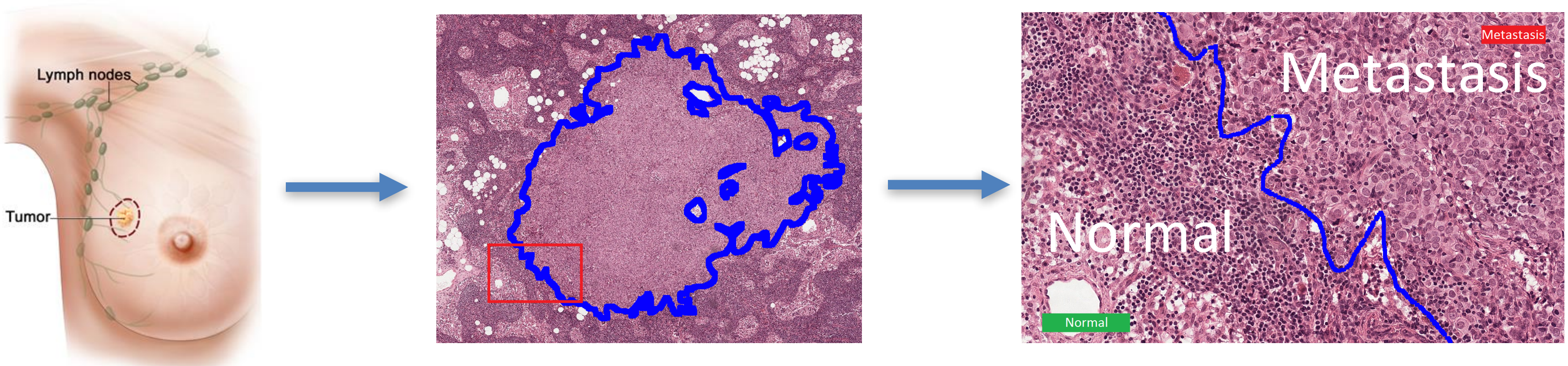
Competitive Period Launch: Nov 18, 2016
Competitive Period Close: May 9, 2017

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