Machine Learning for Healthcare

HST.956, 6.S897

Lecture 1: What makes healthcare unique?

Prof. David Sontag & Pete Szolovits

CSAIL

imes
INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

HST
HEALTH SCIENCES & TECHNOLOGY

The Problem

- Cost of health care expenditures in the US are over $3 trillion, and rising
- Despite having some of the best clinicians in the world, chronic conditions are
  - Often diagnosed late
  - Often inappropriately managed
- Medical errors are pervasive

Outline for today's class

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

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.

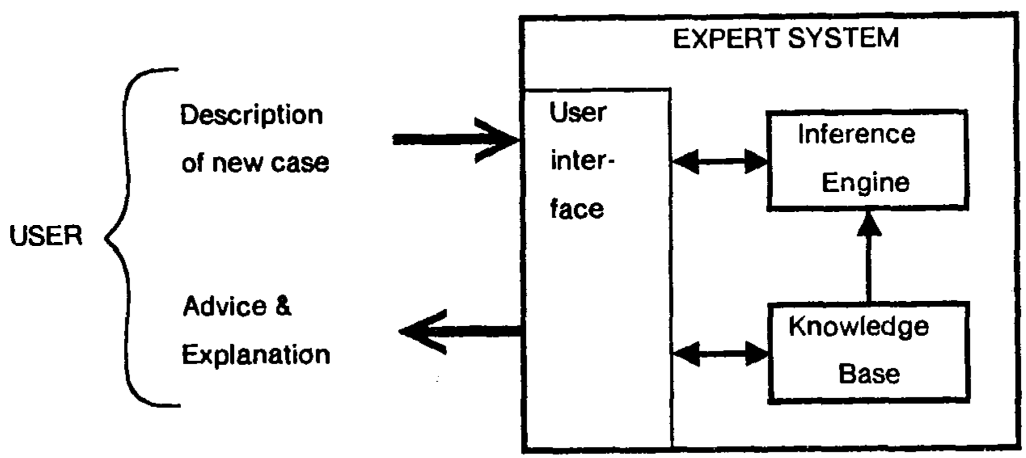
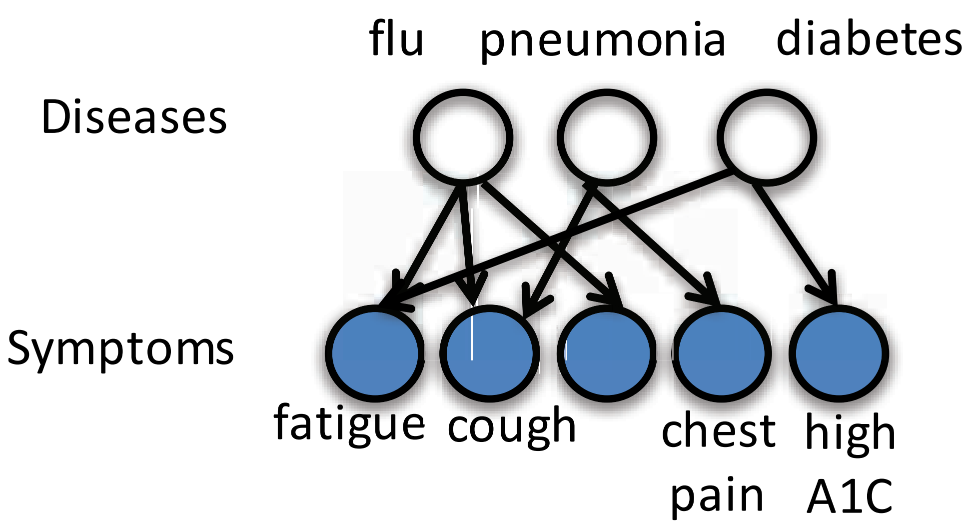


FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.

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

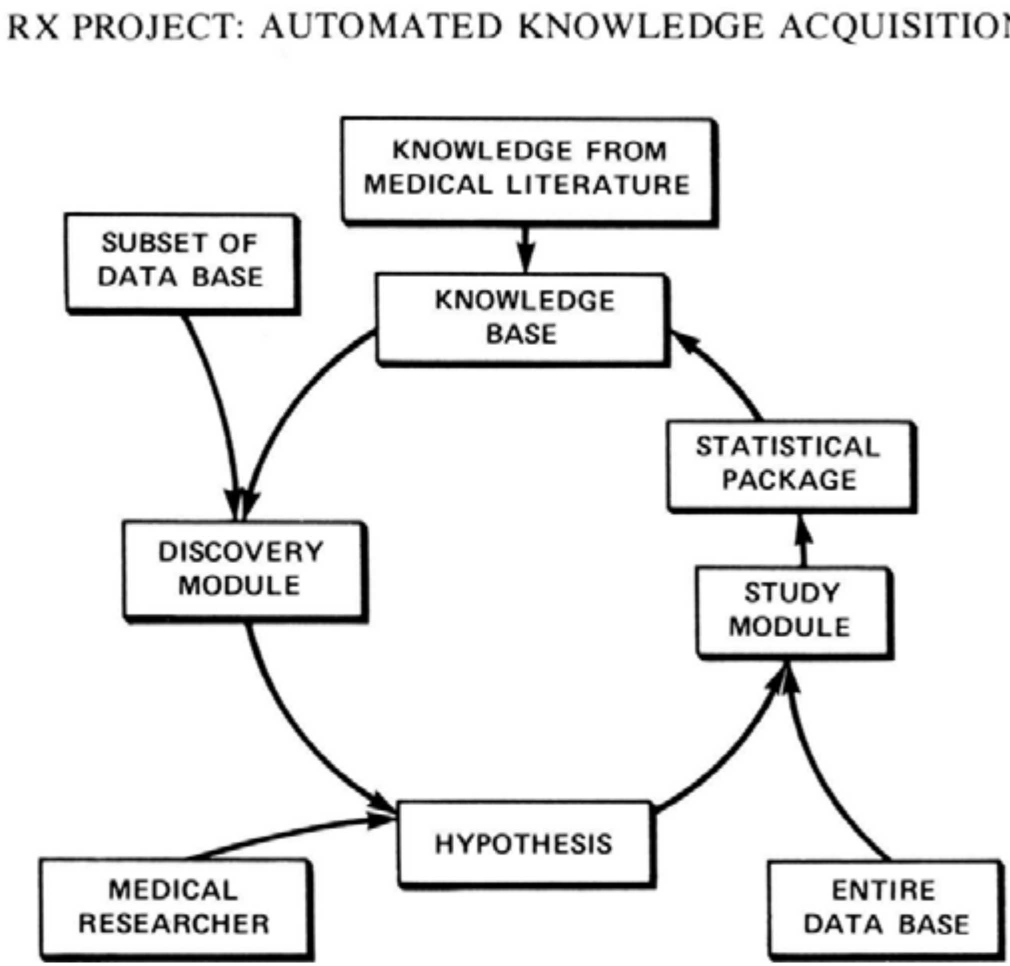
Elicited from doctors:
15 person-years of work

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



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

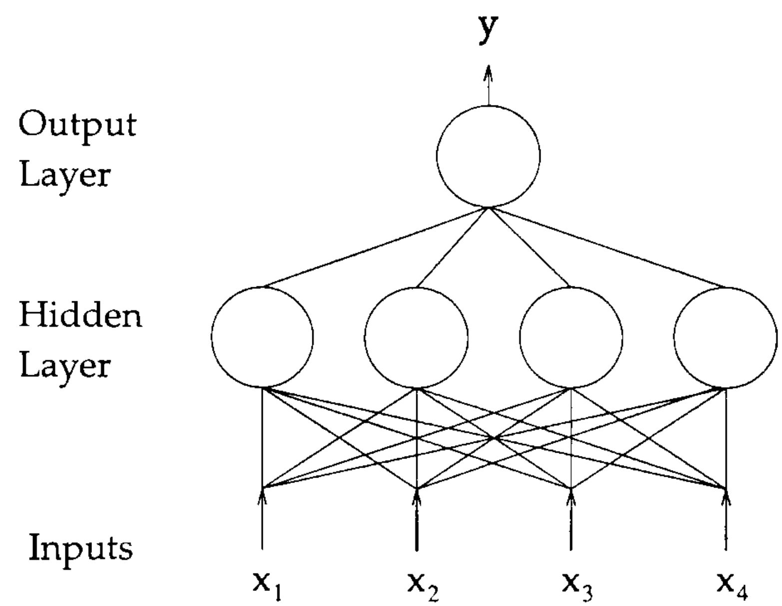


FIGURE 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.

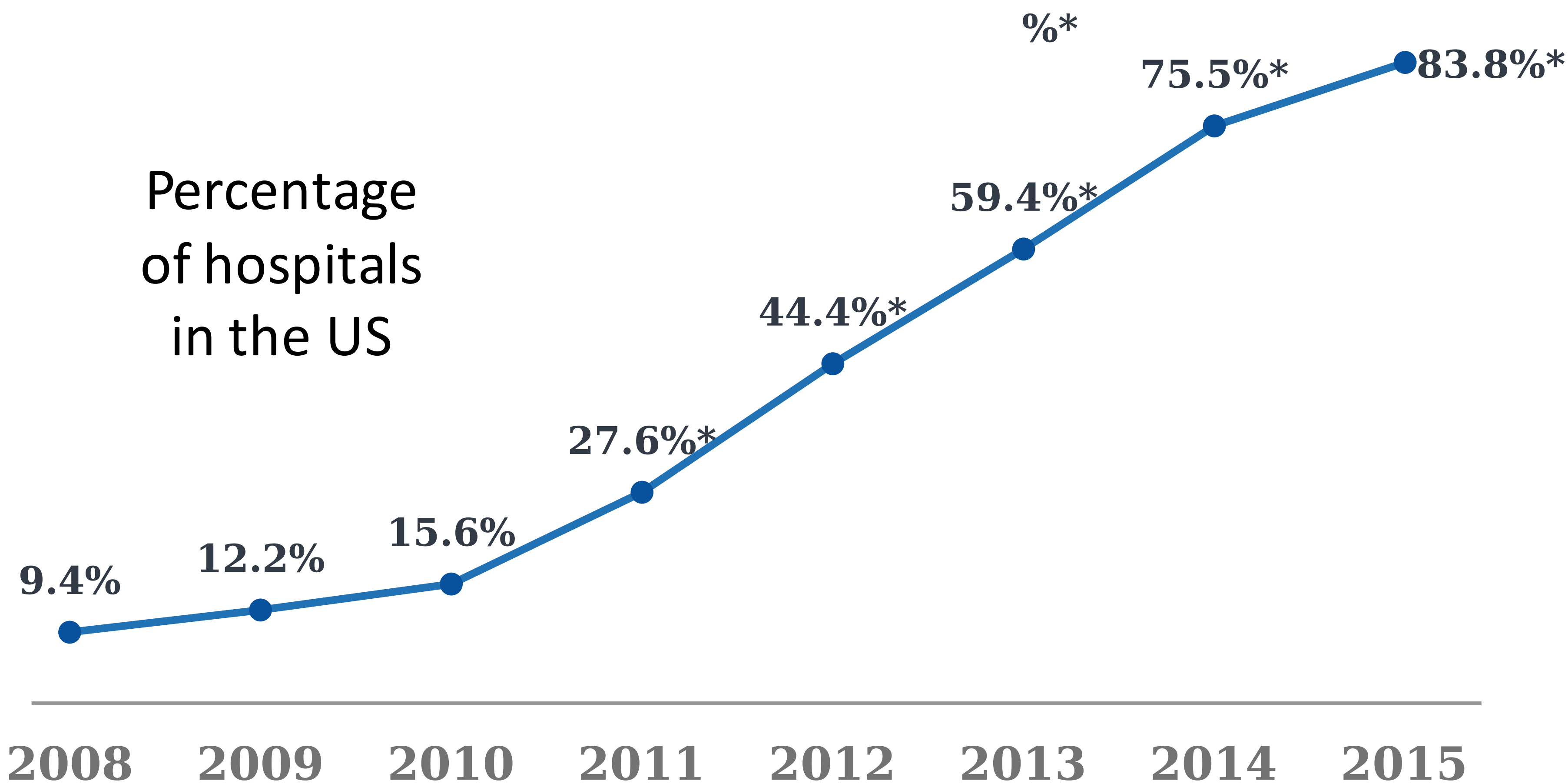
Problems:
1. Did not fit well into clinical workflow
2. Hard to get enough training data
3. Poor generalization to new places

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

Outline for today’s class

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

The Opportunity: Adoption of Electronic Health Records (EHR) has increased 9x in US 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

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,

…

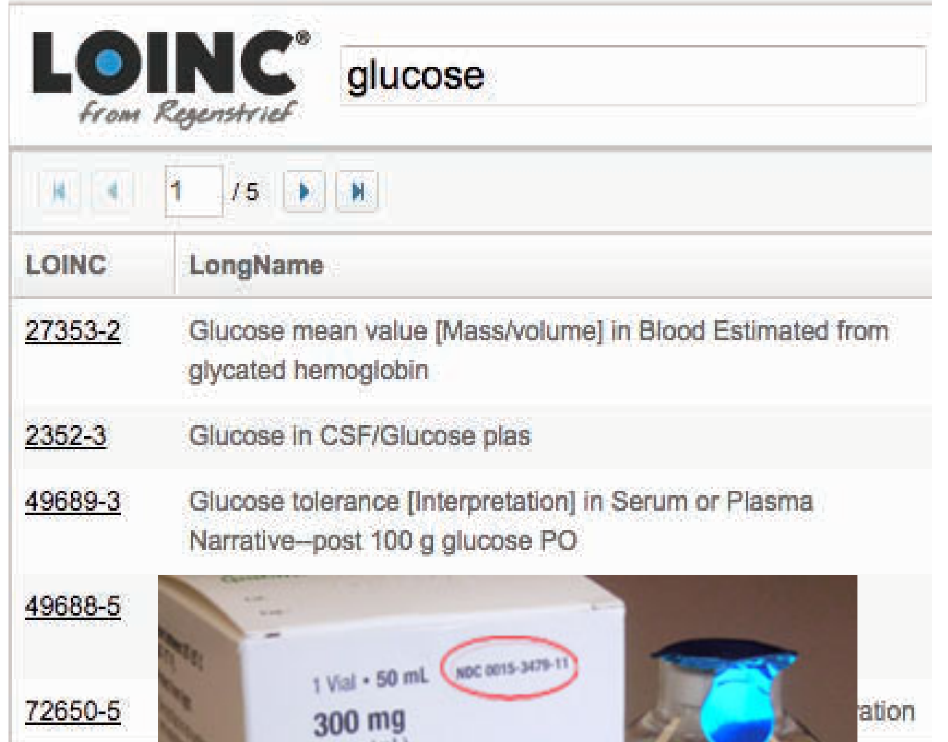
…

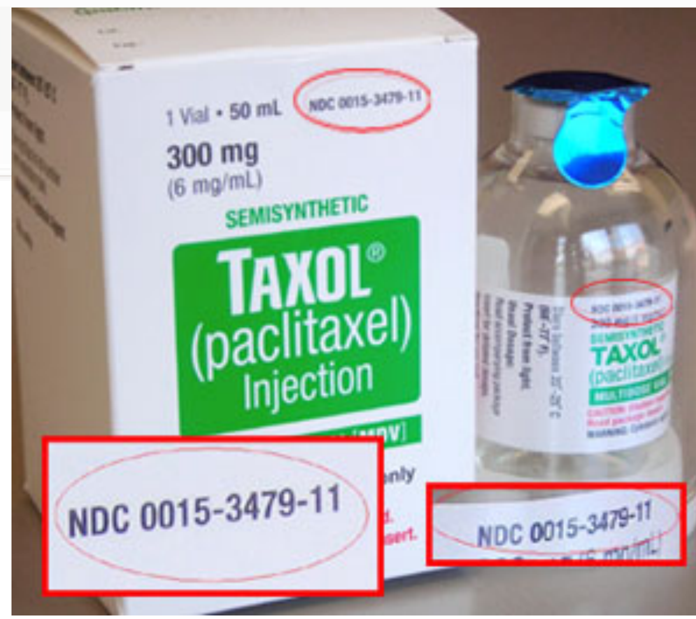
https://en.wikipedia.org/wiki/List_of_ICD-9_codes

https://blog.curemd.com/the-most-bizarre-icd-10-codes-infographic/

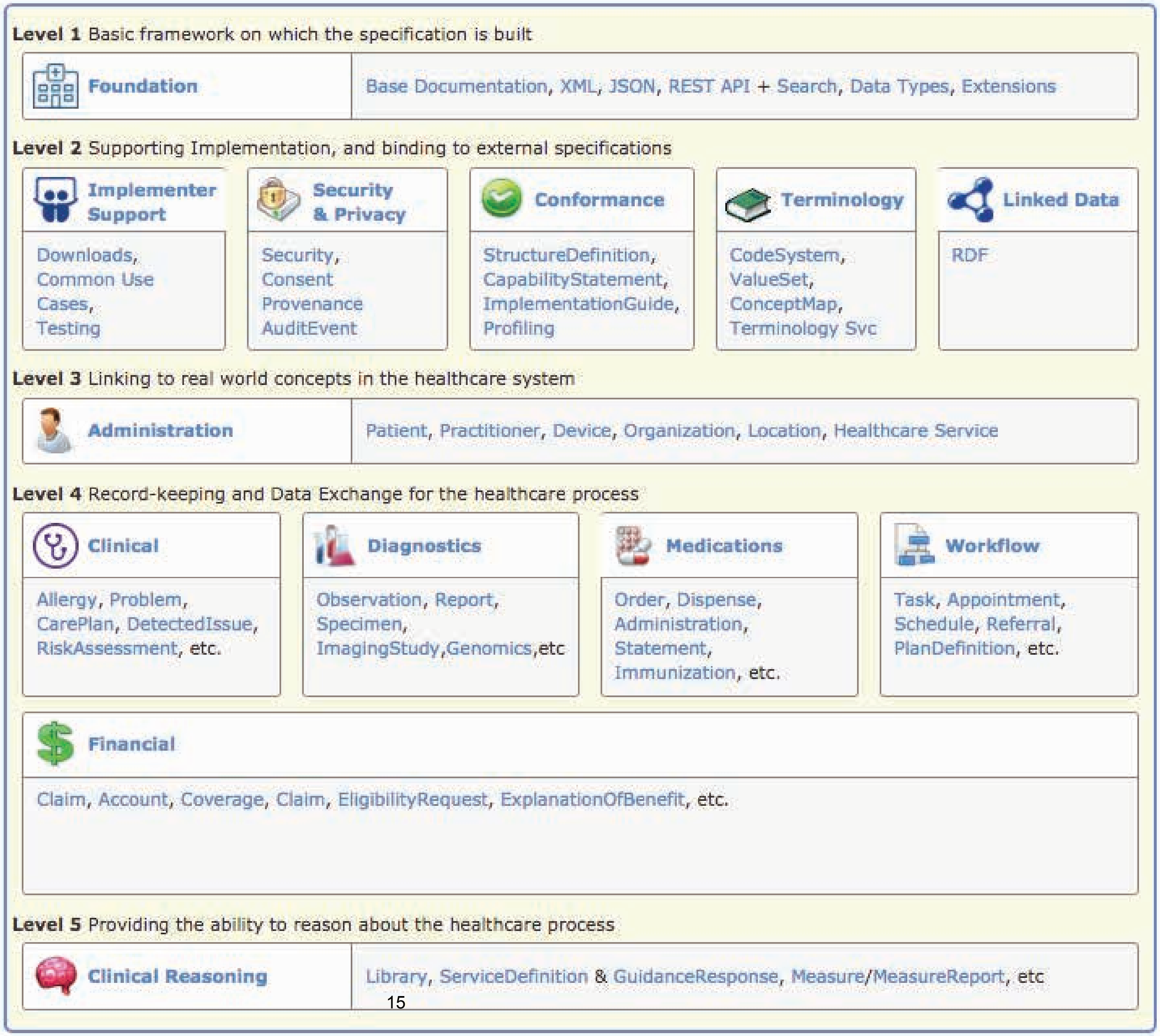
Standardization

- 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



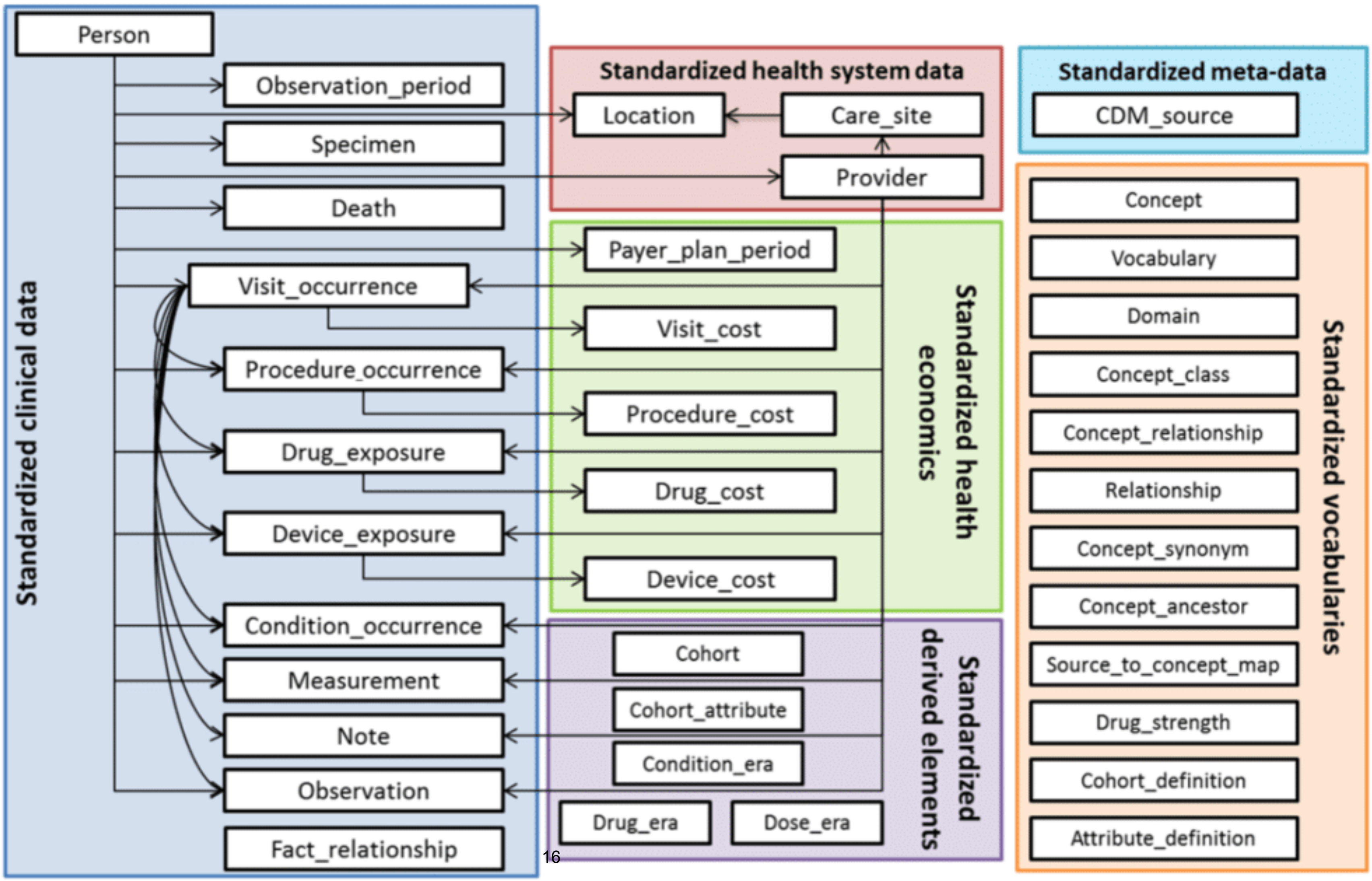


Standardization

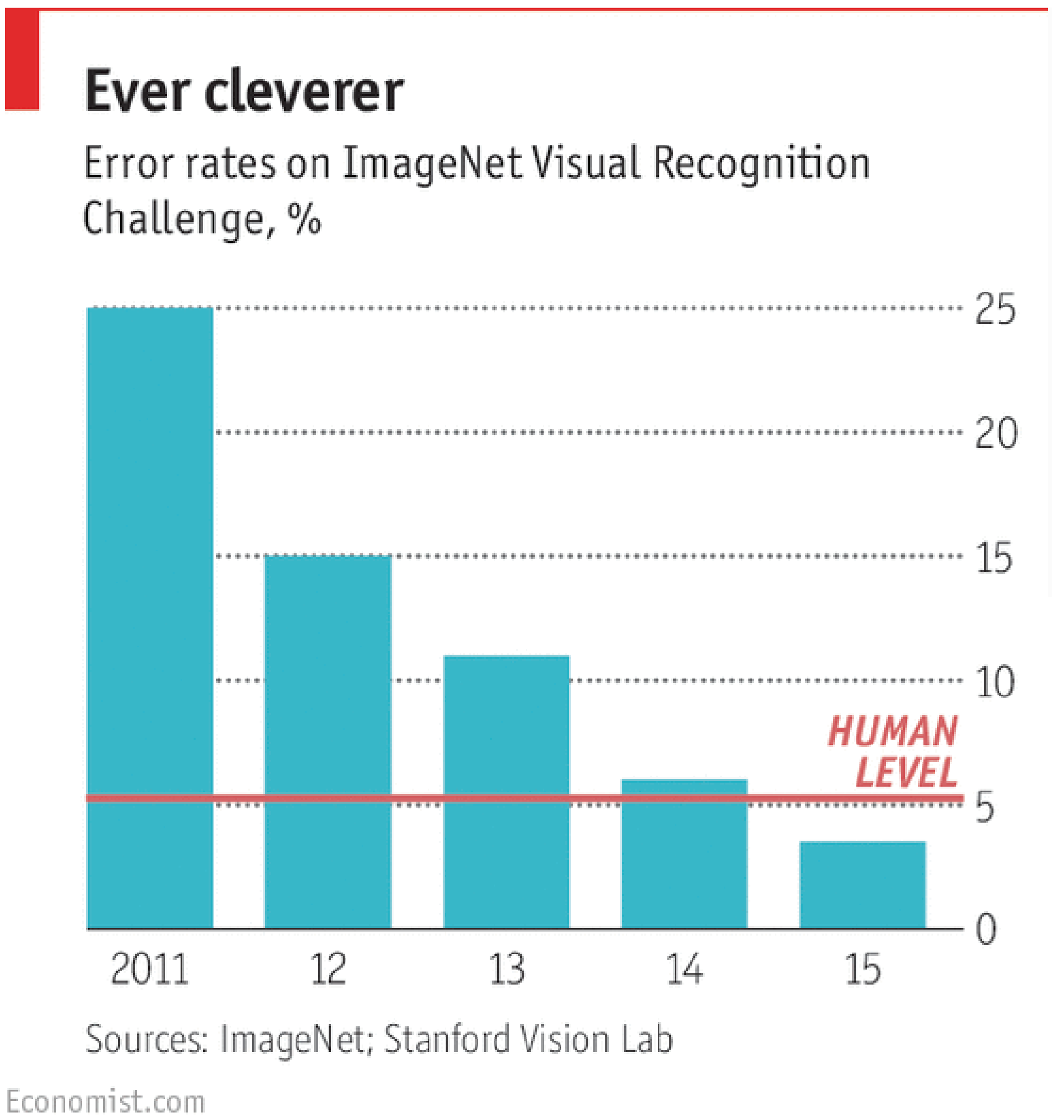


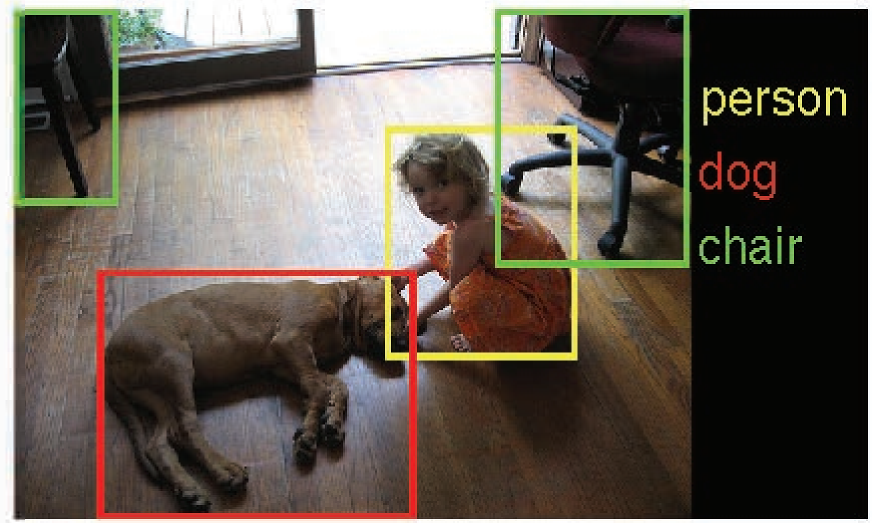
Standardization

OMOP Common Data Model v5.0



Breakthroughs in machine learning





Why now?

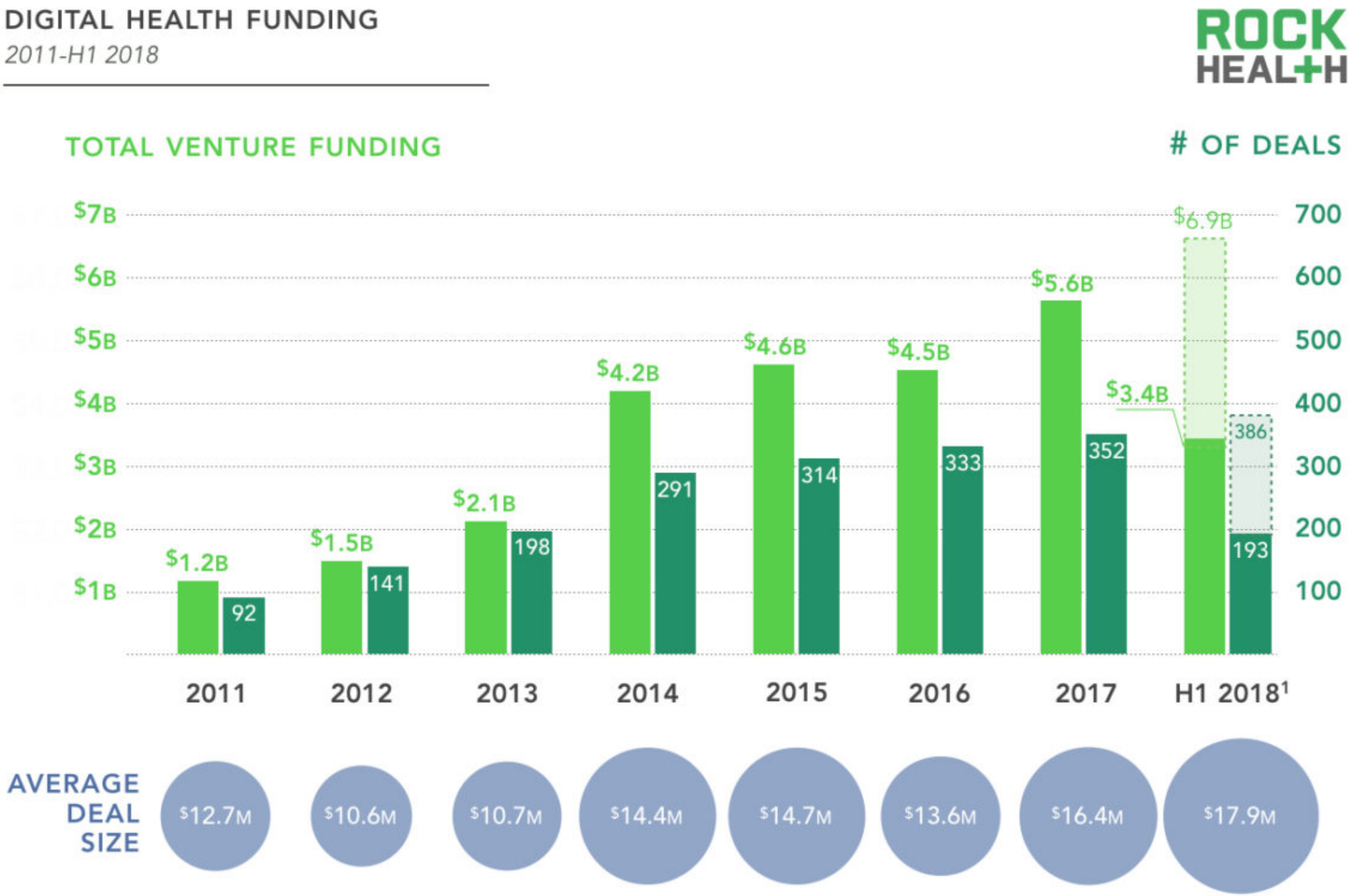
- Big data
- Algorithmic advances
- Open-source software

Breakthroughs 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

DIGITAL HEALTH FUNDING

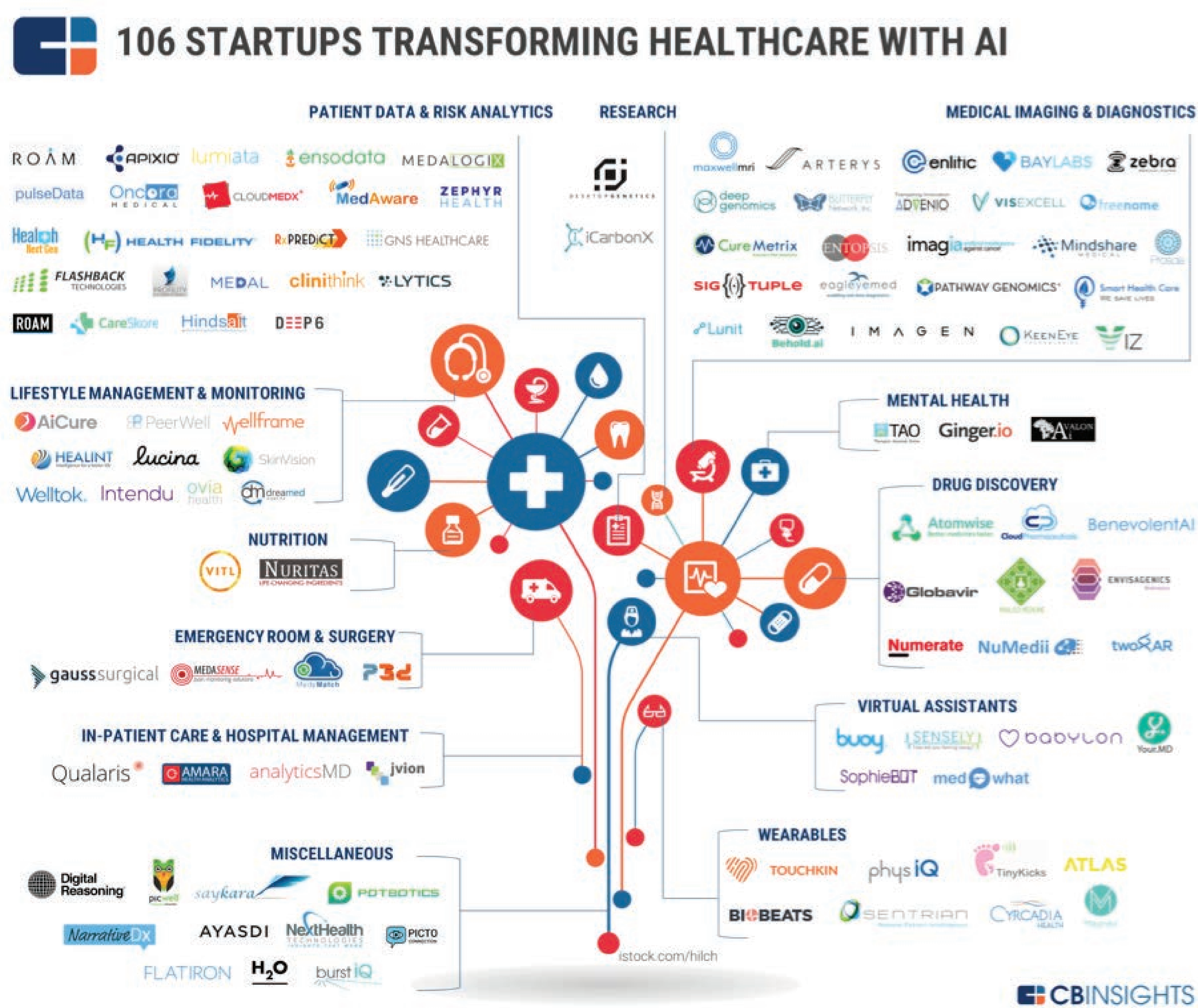
2011-H1 2018



Source: Rock Health Funding Database

1: Shadowed portion shows projections for entire year of 2018, assuming current funding pace continues.

Note: Only includes U.S. deals >$2M; data through June 30, 2018


106 STARTUPS TRANSFORMING HEALTHCARE WITH AI
PATIENT DATA & RISK ANALYTICS
ROAM
APIXIO
lumiata
ensodata
MEDALOGIX
pulseData
Oncora MEDICAL
CLOUDMEDX
MedAware
ZEPHYR HEALTH
Health Next Gen
HEALTH FIDELITY
RxPREDiCT
GNS HEALTHCARE
FLASHBACK TECHNOLOGIES
MEDAL
clinithink
LYTICS
ROAM
CareSkore
Hindsait
DEEP6
RESEARCH
iCarbonX
MEDICAL IMAGING & DIAGNOSTICS
maxwellmri
ARTERYS
enlitic
BAYLABS
zebra
deep genomics
ADVENIO
VISEXCELL
freenome
CureMetrix
imagia
Mindshare MEDICAL
SIGTUPLE
eagleyemed
PATHWAY GENOMICS
Smart Health Care
Lunit
Behold.ai
IMAGEN
KEENEYE
IZ
LIFESTYLE MANAGEMENT & MONITORING
AiCure
PeerWell
wellframe
HEALINT
Lucina
SkinVision
Welltok
Intendu
ovia health
dreamed
NUTRITION
VITL
NURITAS
EMERGENCY ROOM & SURGERY
gausssurgical
MEDASENSE
MedyMatch
P3d
IN-PATIENT CARE & HOSPITAL MANAGEMENT
Qualaris
AMARA
analyticsMD
jvion
MISCELLANEOUS
Digital Reasoning
pic well
saykara
POTBOTICS
NarrativeDx
AYASDI
NextHealth TECHNOLOGIES
PICTO
FLATIRON
H2O
burst iQ
MENTAL HEALTH
TAO
Ginger.io
AVALON AI
DRUG DISCOVERY
Atomwise
Cloud Pharmaceuticals
BenevolentAI
Globavir
ENVISAGENICS
Numerate
NuMedii
twoXAR
VIRTUAL ASSISTANTS
buoy
SENSELY
babylon
Your.MD
SophieBOT
med what
WEARABLES
TOUCHKIN
physIQ
TinyKicks
ATLAS
BIOBEATS
SENTRIAN
CYRCADIA HEALTH
istock.com/hilch
CBINSIGHTS

Industry interest in ML & healthcare

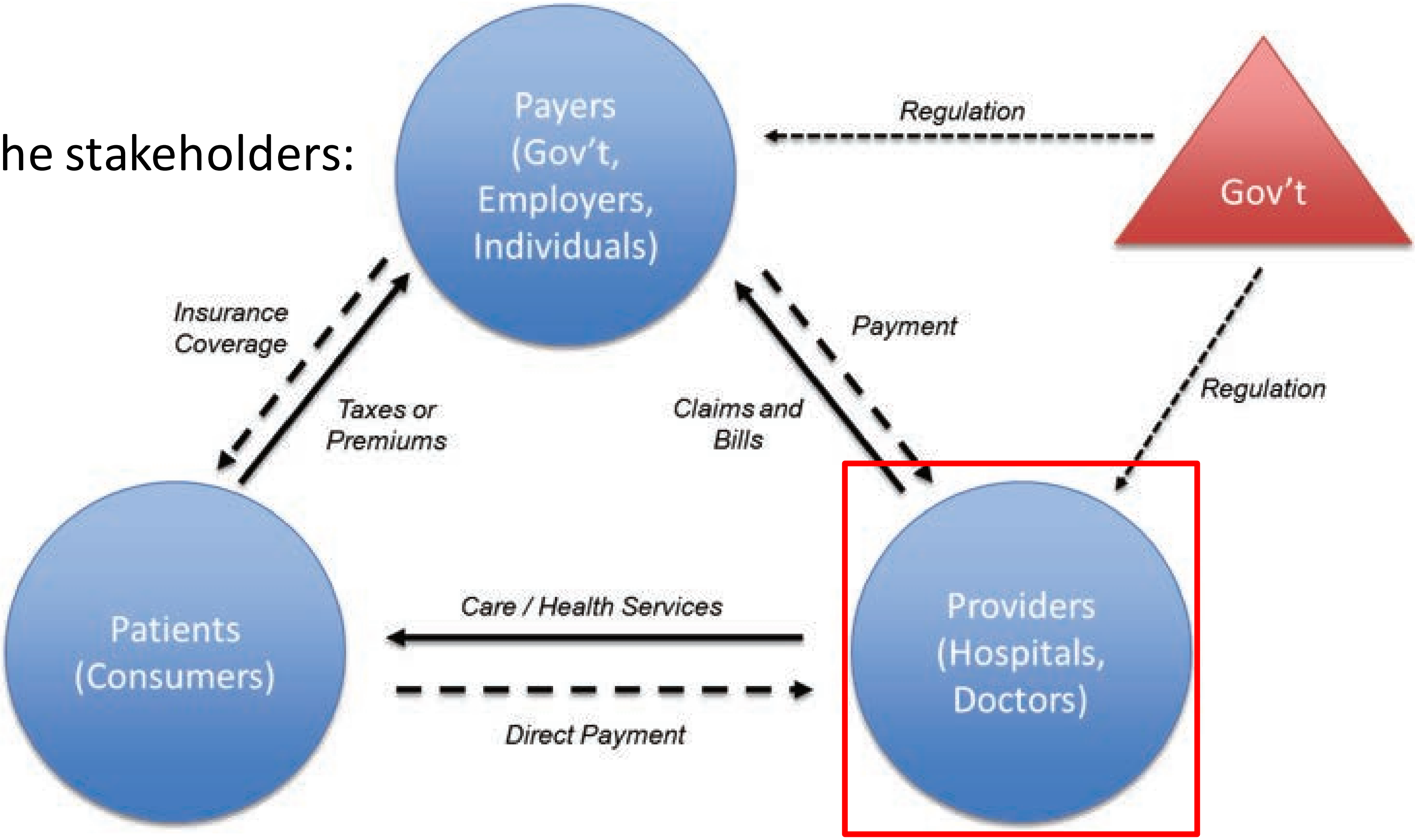
- Major acquisitions to get big data for ML:
  - Merge ($1 billion purchase by IBM, 2015)
    medical imaging
  - Truven Health Analytics ($2.6 billion purchase by IBM, 2016)
    health insurance claims
  - Flatiron Health ($1.9 billion purchase by Roche, 2018)
    electronic health records (oncology)

Outline for today’s class

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

ML will transform every aspect of healthcare

The stakeholders:

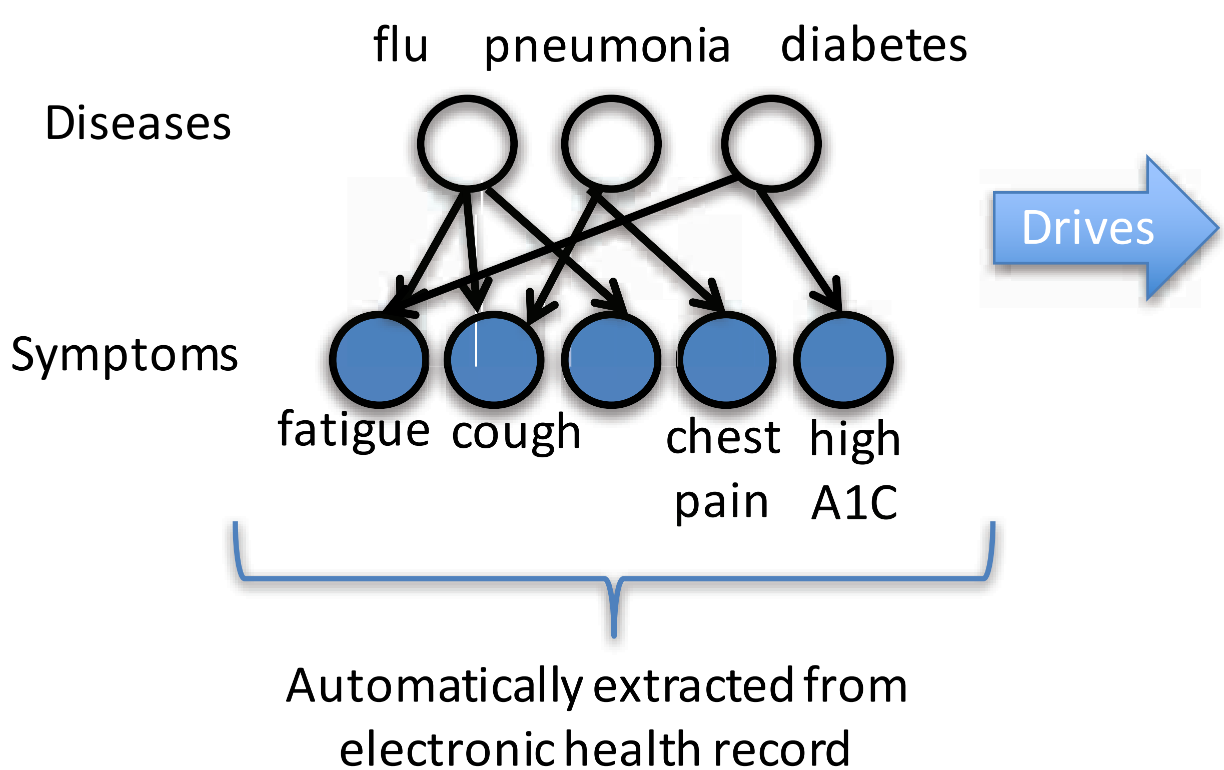


Emergency Department:

- **Limited resources**
- **Time sensitive**
- **Critical decisions**

What will the ER of the future be like?

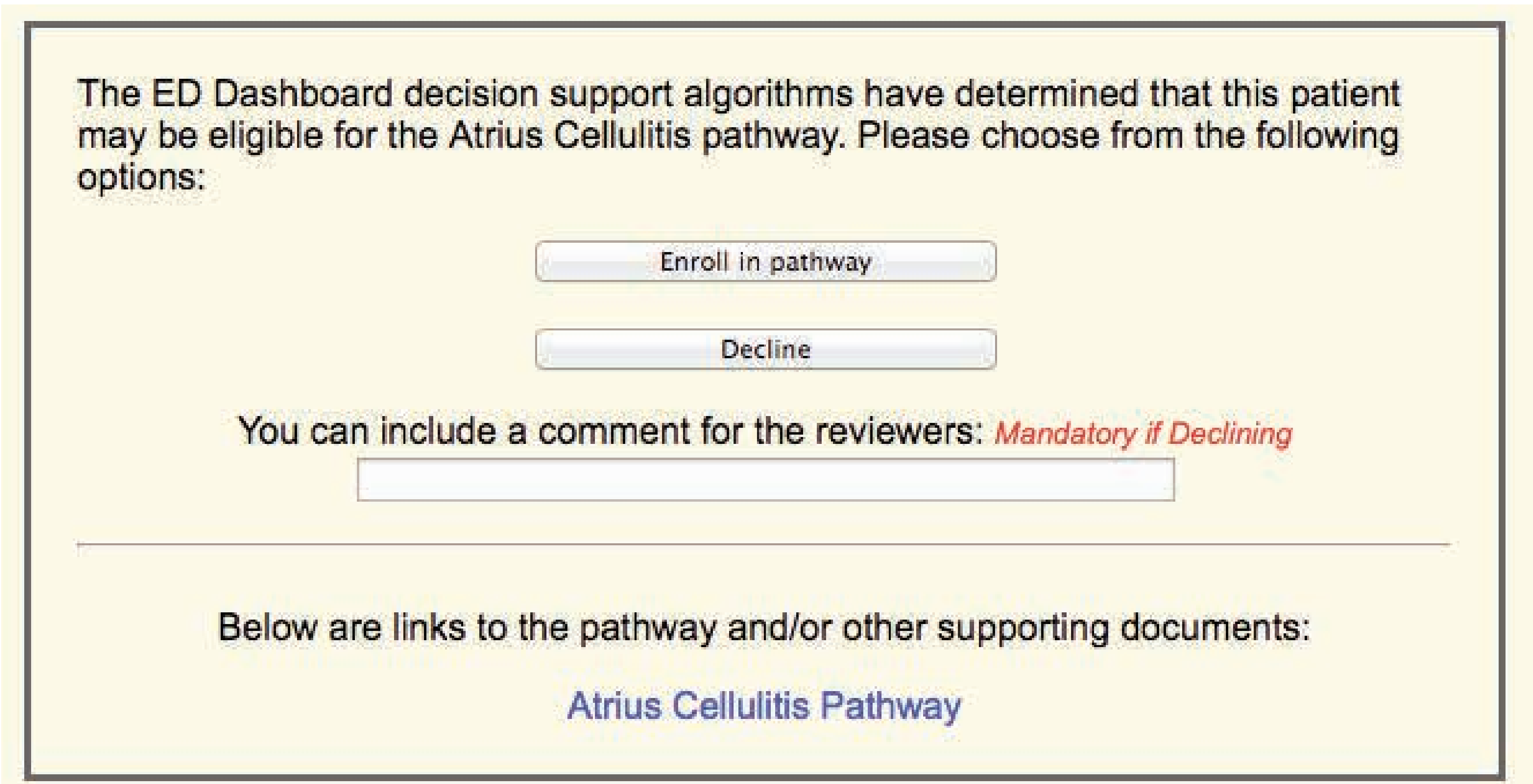
Behind-the-scenes reasoning about the patient's conditions (current and future)



- Better triage
- Faster diagnosis
- Early detection of adverse events
- Prevent medical errors

What will the ER of the future be like?

Propagating best practices


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

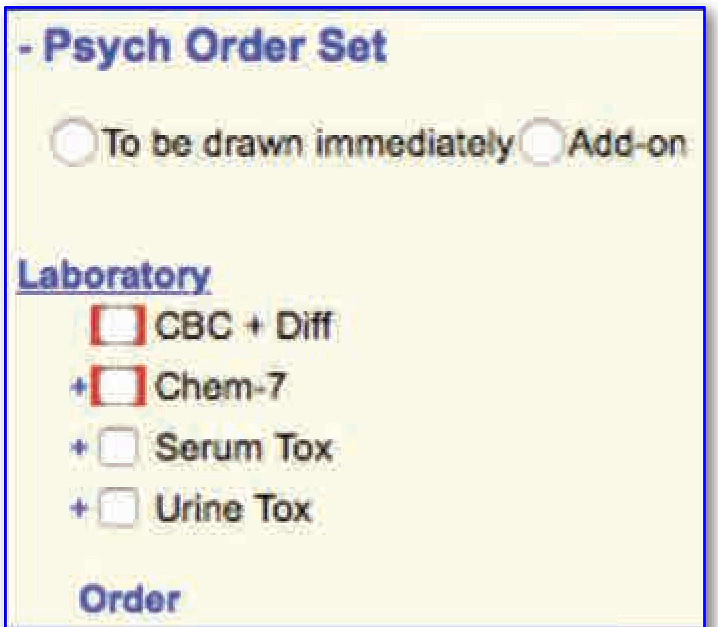
You can include a comment for the reviewers: *Mandatory if Declining*

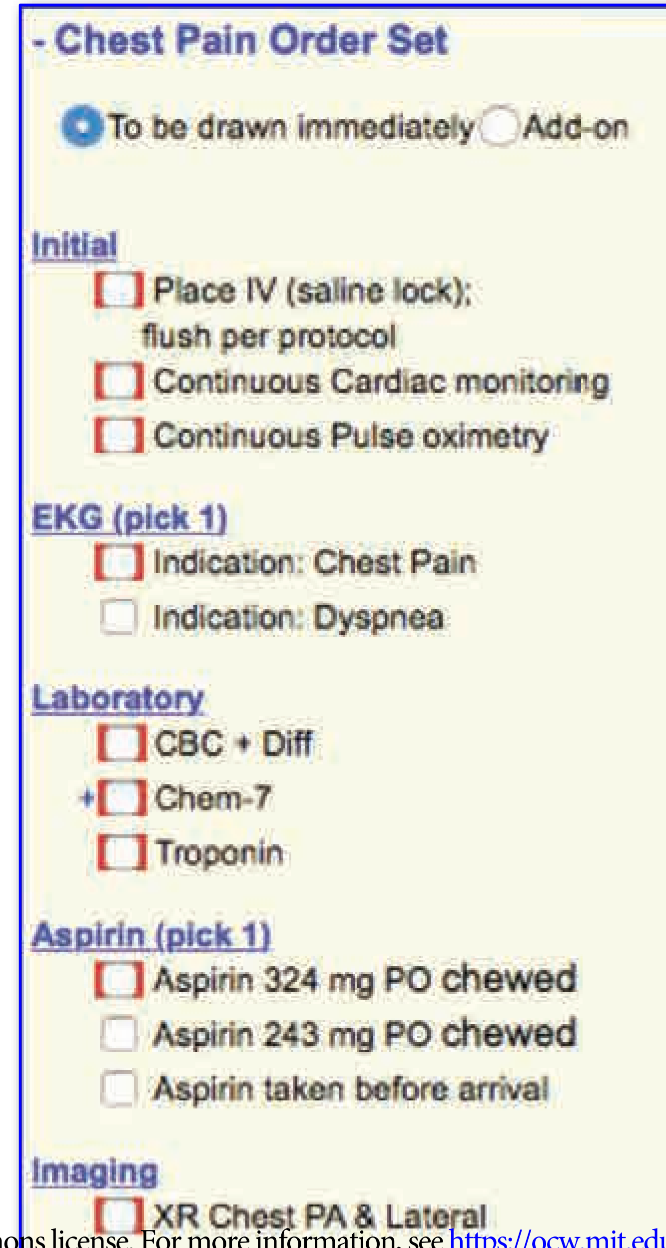
Below are links to the pathway and/or other supporting documents:

Atrius Cellulitis Pathway

What will the ER of the future be like?

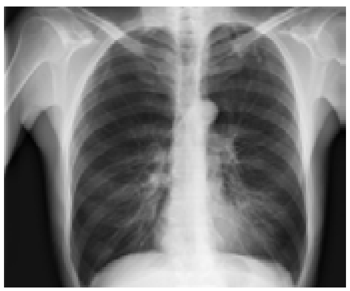
Anticipating the clinicians' needs





What will the ER of the future be like?

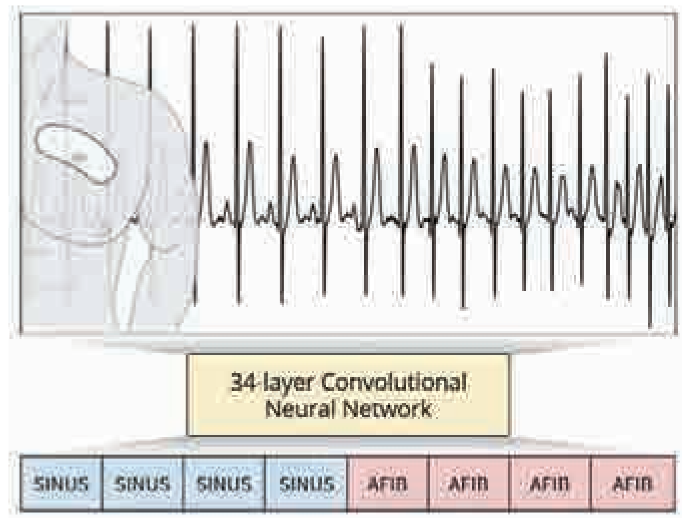
Reducing the need for specialist consults



Input
Chest X-Ray Image

CheXNet
121-layer CNN

Output
Pneumonia Positive (85%)

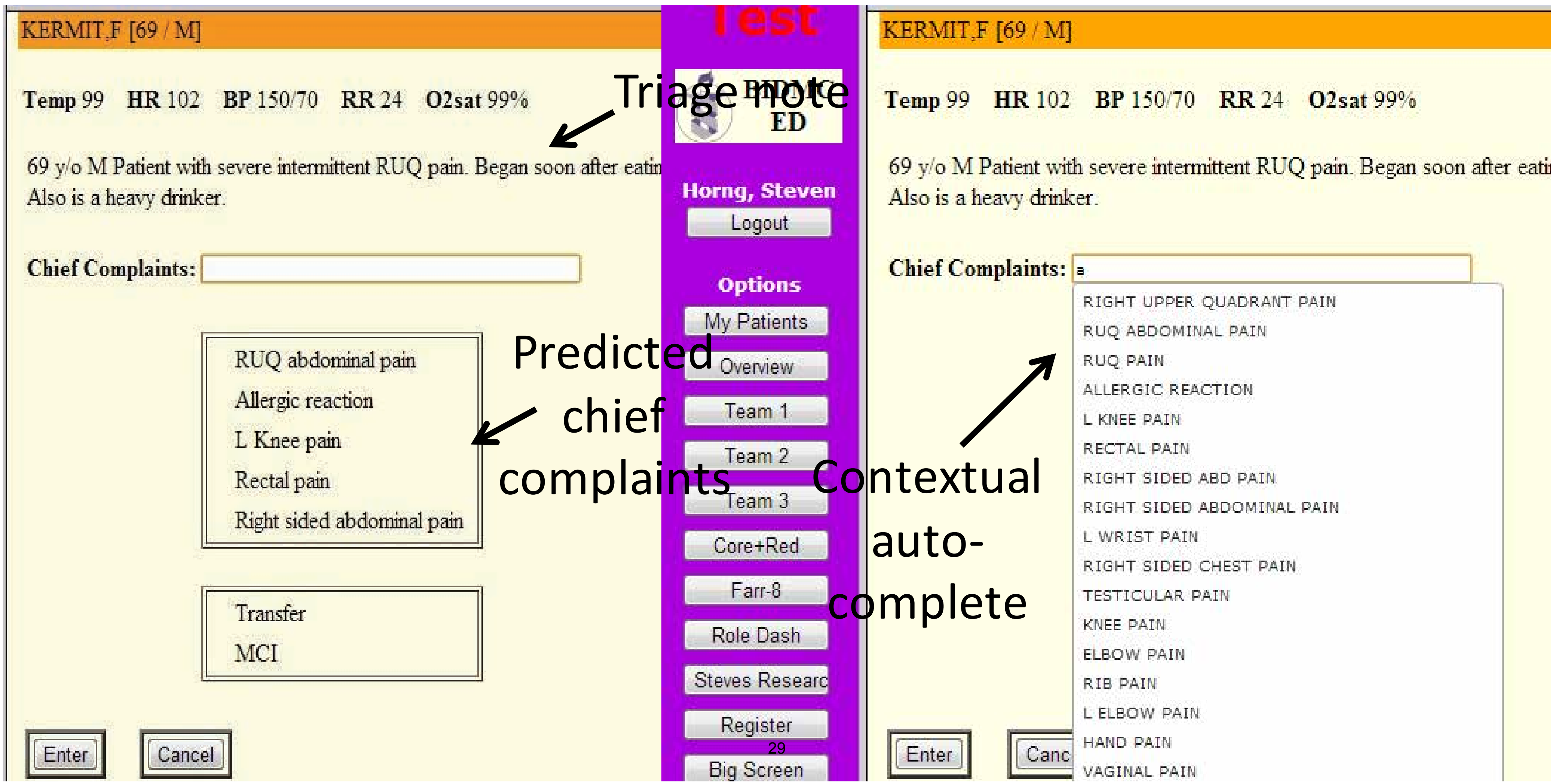


Arrhythmia?

Figure sources: Rajpurkar et al., arXiv:1711.05225 '17
Rajpurkar et al., arXiv:1707.01836, '17

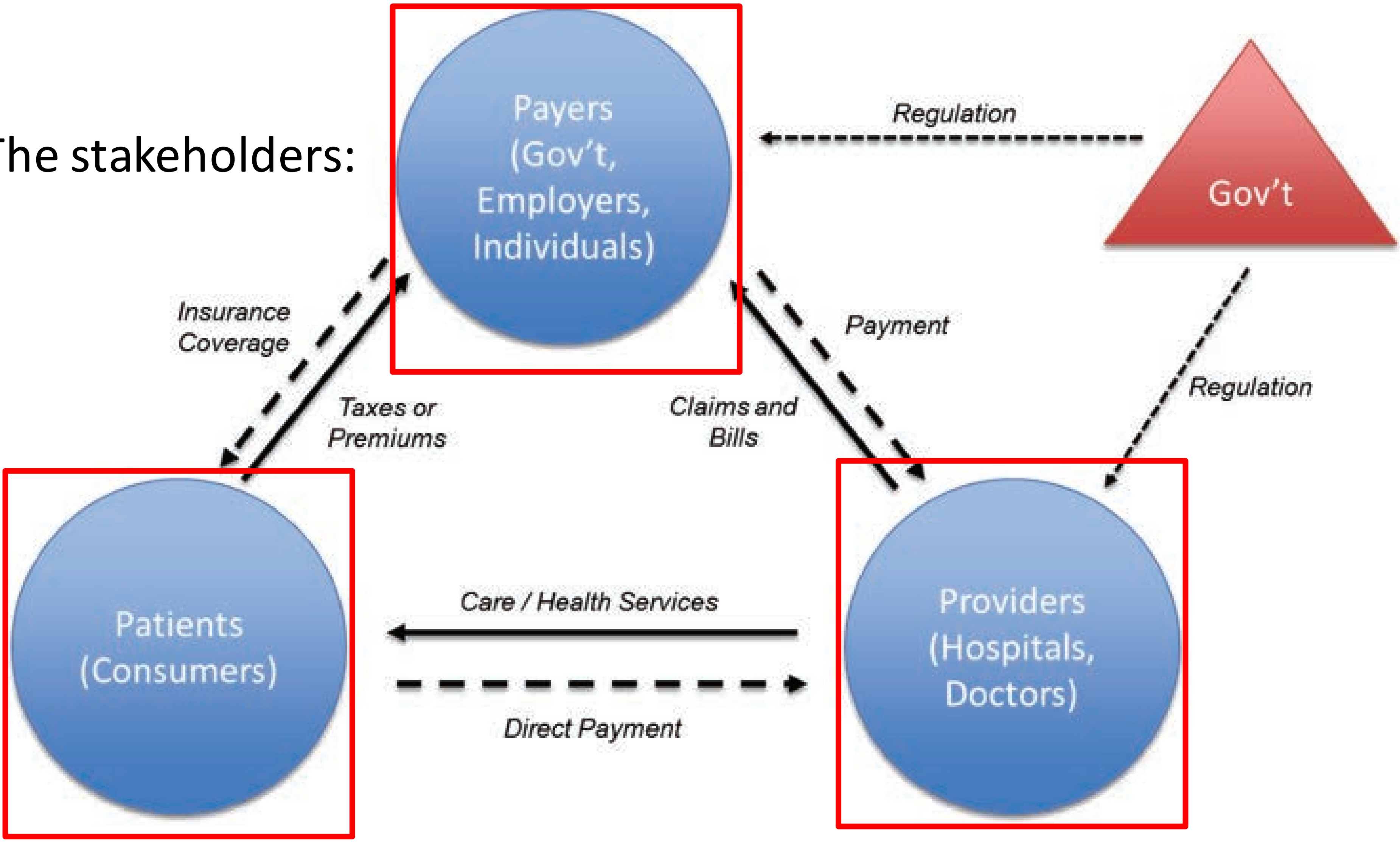
What will the ER of the future be like?

Automated documentation and billing



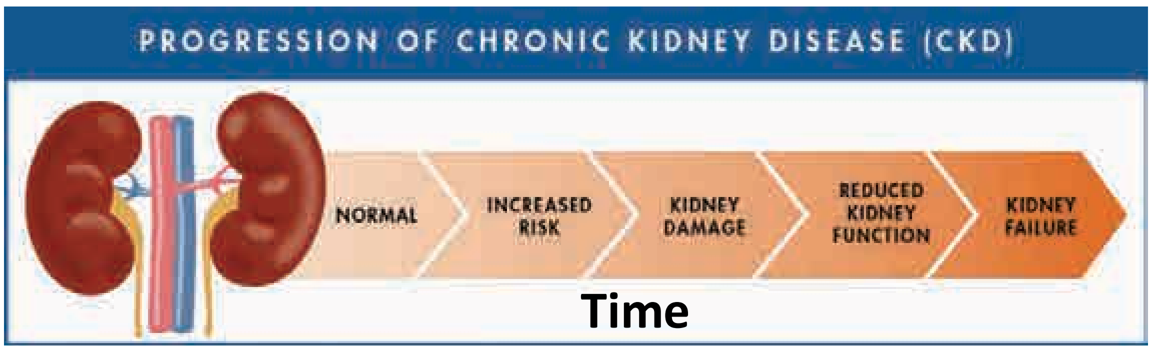
ML will transform every aspect of healthcare

The stakeholders:



What is the future of how we treat chronic disease?

- Predicting a patient's future disease progression



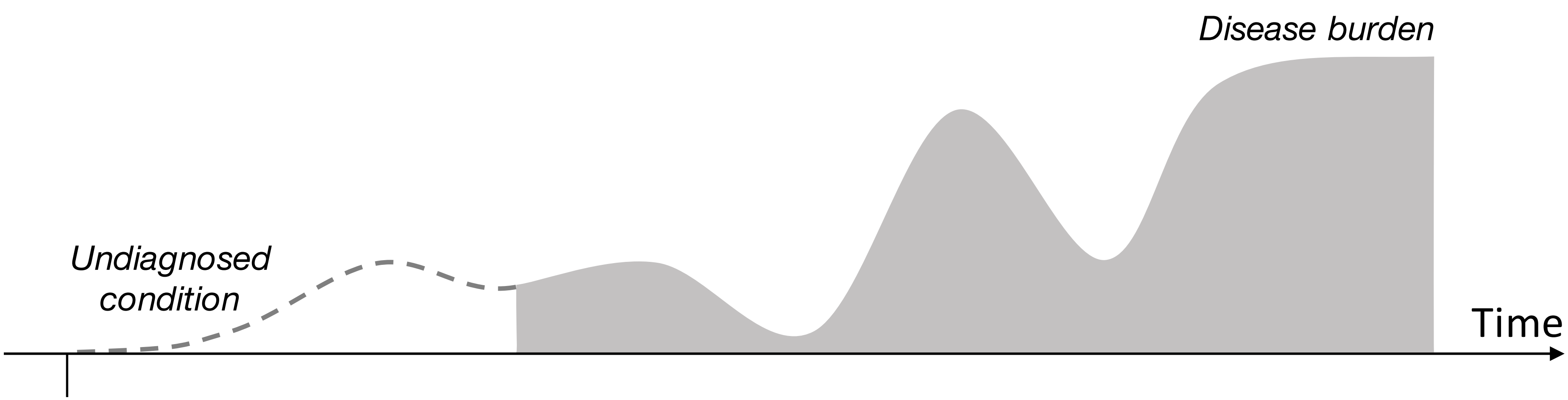


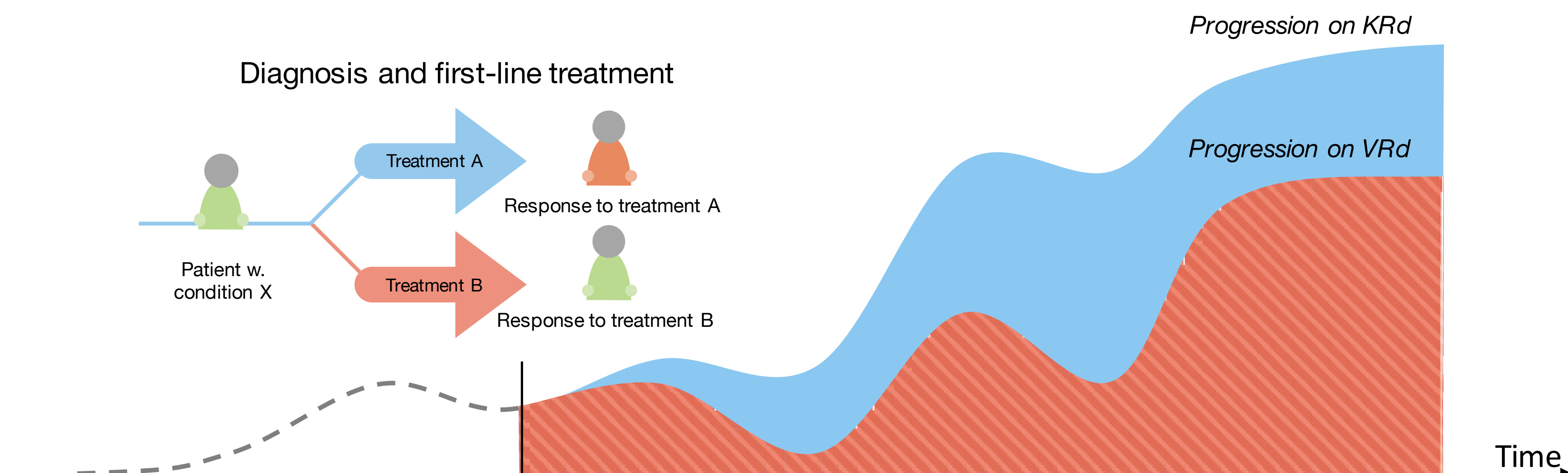
Figure credit: https://www.cdc.gov/kidneydisease/prevention-risk.html

What is the future of how we treat chronic disease?

- Predicting a patient's future disease progression
- Precision medicine

Choosing first line therapy in multiple myeloma

A) KRd: carfilzomib-lenalidomide-dexamethasone, **B) VRd:** bortezomib-lenalidomide-dexamethasone



What is the future of how we treat chronic disease?

- Early diagnosis, e.g. of diabetes, Alzheimer's, cancer
- Continuous monitoring and coaching, e.g. for the elderly, diabetes, psychiatric disease
- Discovery of new disease subtypes; design of new drugs; better targeted clinical trials

Outline for today's class

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

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?

- Very little labeled data
  - 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

Goals for the semester

- Intuition for working with healthcare data
- How to set up as machine learning problems
- Understand which learning algorithms are likely to be useful and when
- Appreciate subtleties in safely & robustly applying ML in healthcare
- Set the research agenda for the next decade

6.S897/HST.956 vs 6.874

- Our class will focus on **clinical data** and its use to improve health care
- For reasons of time & scope, we will not go deep into applications in the life sciences
  - For this, we recommend taking **6.874 Computational Systems Biology: Deep Learning in the Life Sciences**