# BMI 702: Biomedical Artificial Intelligence

Foundations of Biomedical Informatics II, Spring 2024

Lecture 2: Introduction to AI on clinical datasets

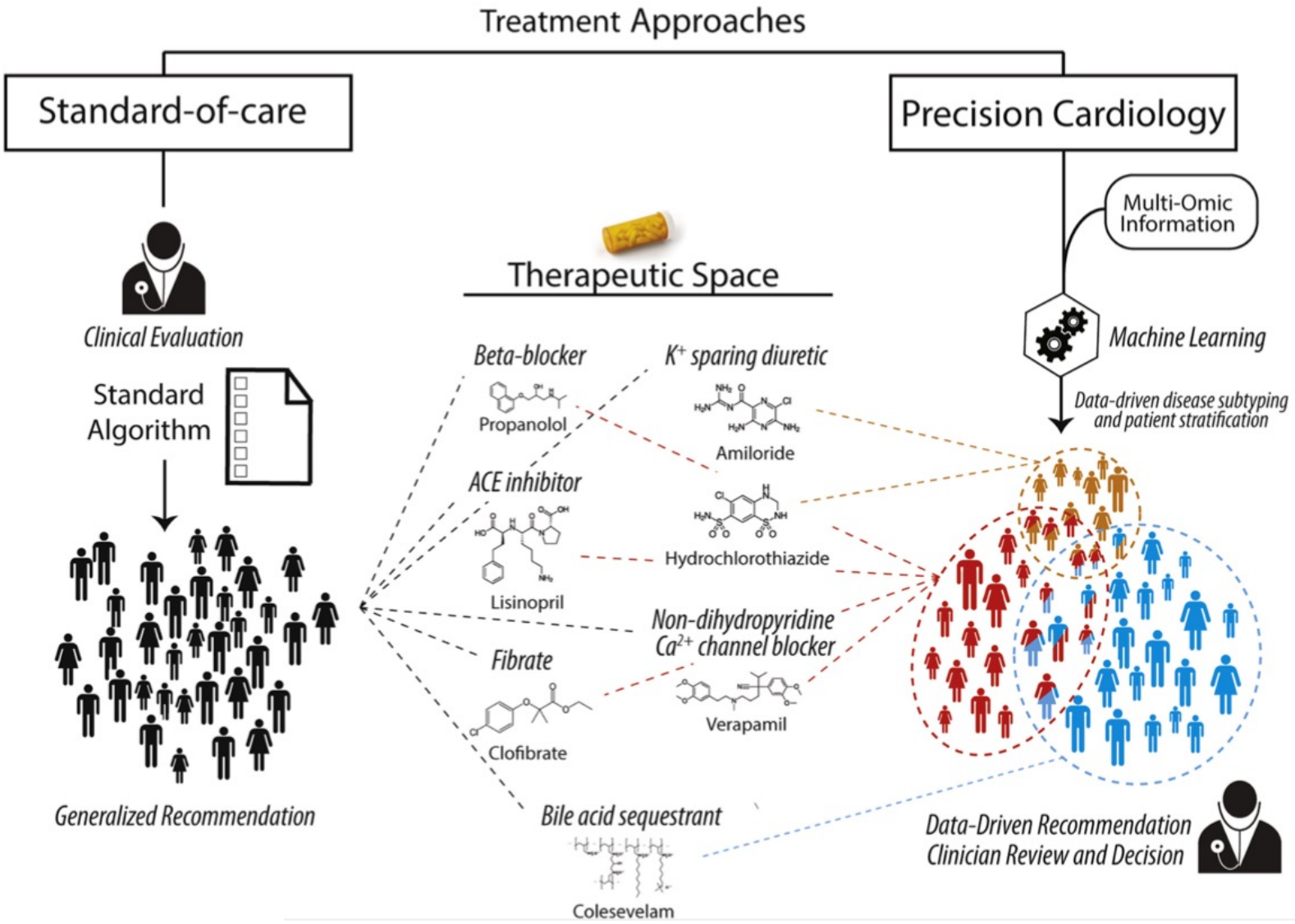

HARVARD
MEDICAL SCHOOL

Marinka Zitnik
marinka@hms.harvard.edu

# Outline for today's class

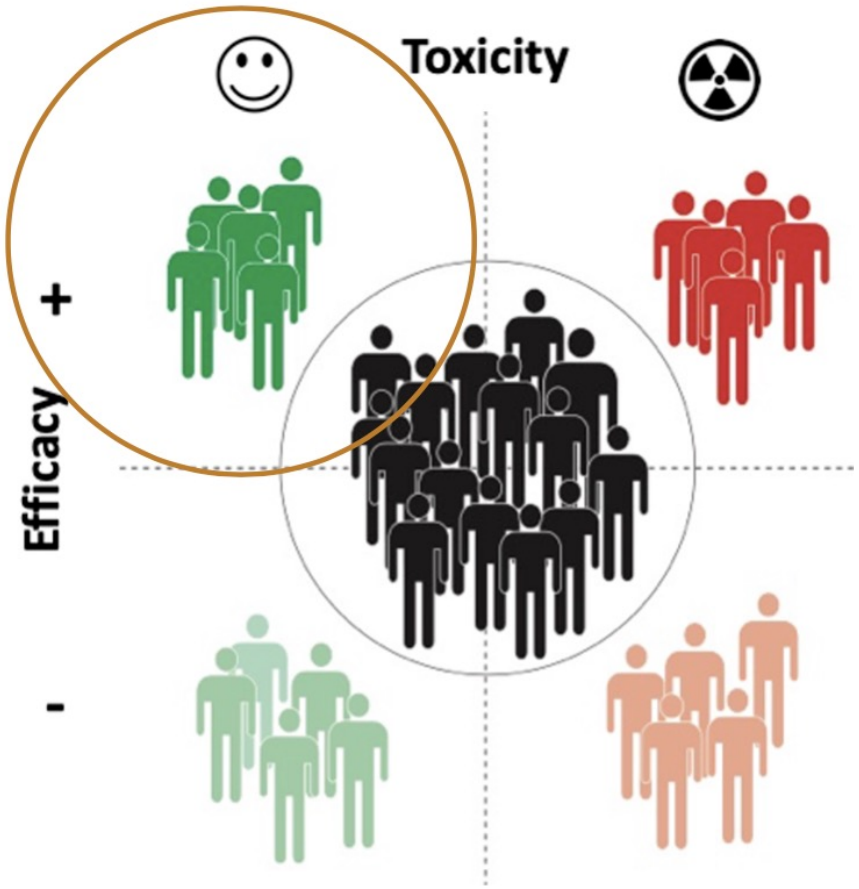
1. AI/ML for precision medicine
2. What are EHR data useful for?
3. Limitations & biases of EHR data
4. Highlights of ML on EHR data:
   - Polypharmacy and adverse drug events
   - Modeling disease progression

# General vs. personalized medicine




Johnson et al., JACC Basic Transl Sci, 2017

# Precision medicine goals



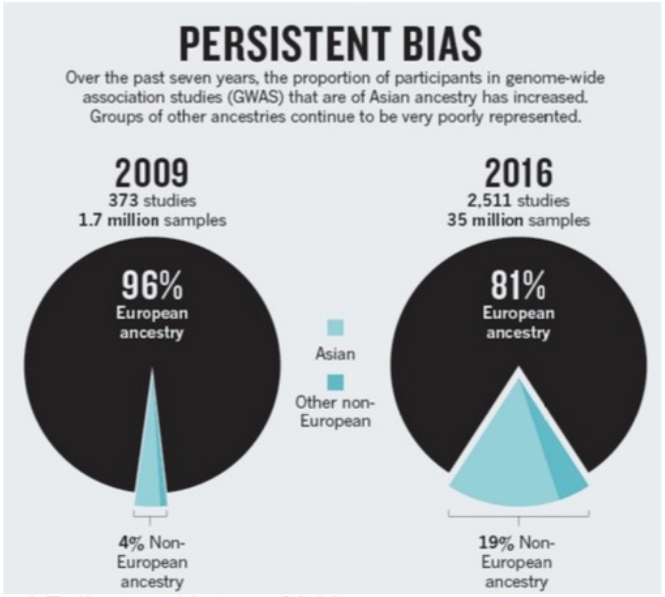


http://hitconsultant.net/2014/04/03/infographic-the-rise-of-personalized-medicine/

# Why are these goals relevant?

## Problem: Underrepresentation in clinical research

### Genomics




Popejoy and Fullerton, *Nature*, 2016

### Clinical Trials

**Participation in Cancer Clinical Trials**
Race-, Sex-, and Age-Based Disparities

**Table 1.** Participants in National Cancer Institute Cooperative Group Breast, Colorectal, Lung, or Prostate Cancer Therapeutic Trials, 1996-2002 (N = 75 215)*

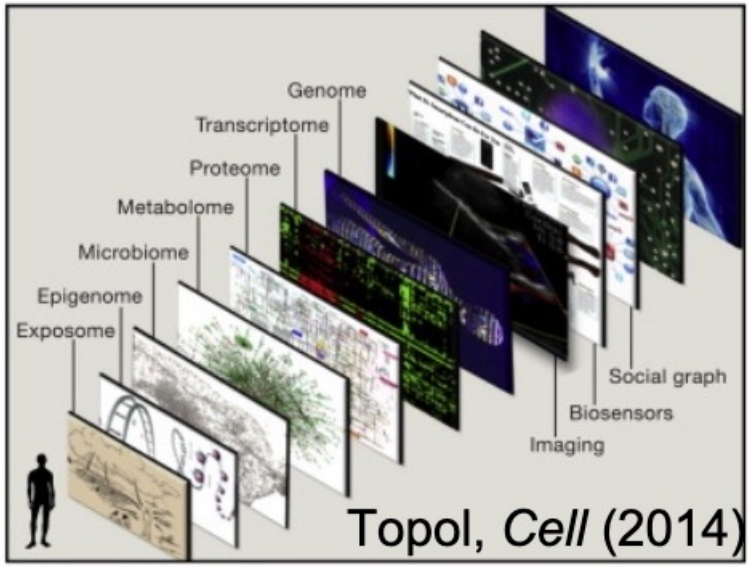
| Characteristic | Trial Participants, No. (%) | Proportion of Incident Cancer Patients, %† | Proportion of US Population, %† |
|---|---|---|---|
| Race/ethnicity | | | |
| White non-Hispanic | 64 355 (85.6) | 83.1 | 75.7 |
| Hispanic | 2292 (3.1) | 3.8 | 9.1 |
| Black | 6882 (9.2) | 10.9 | 10.8 |
| Asian/Pacific Islander | 1446 (1.9) | 2.0 | 3.8 |
| American Indian/Alaskan Native | 240 (0.3) | 0.2 | 0.7 |

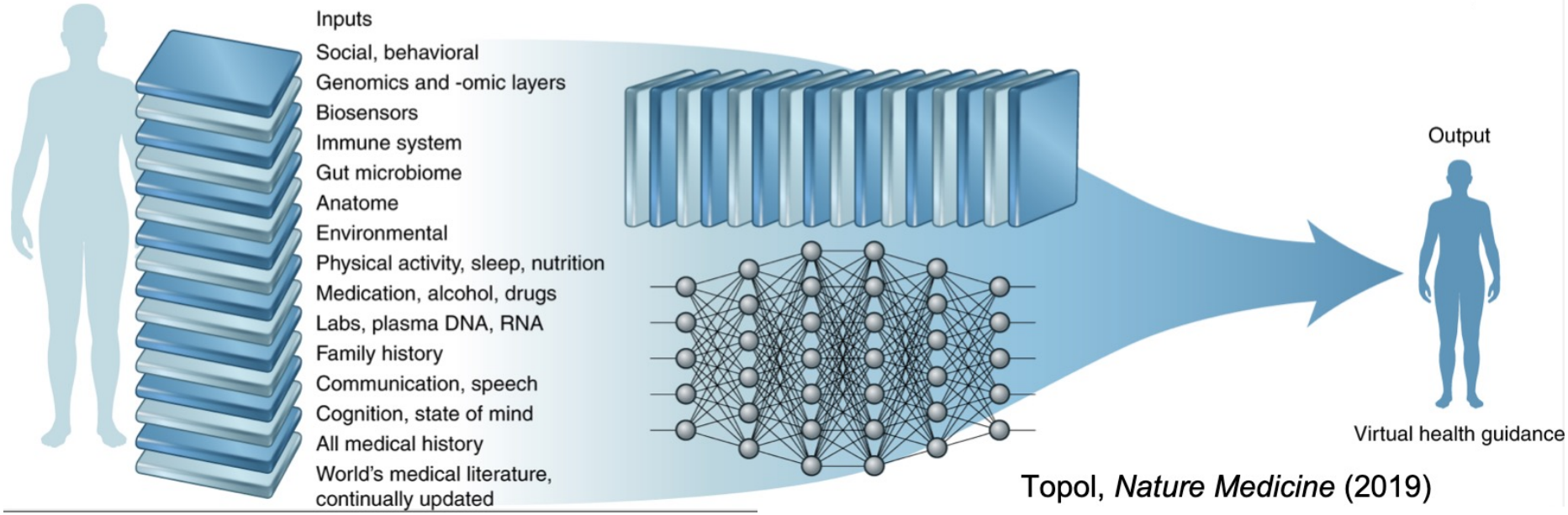
Murthy et al., *JAMA*, 2004.


Most clinical decisions involve bridging the **inferential gap**: Clinicians are required to "fill in" where they lack knowledge or where no knowledge yet exists:

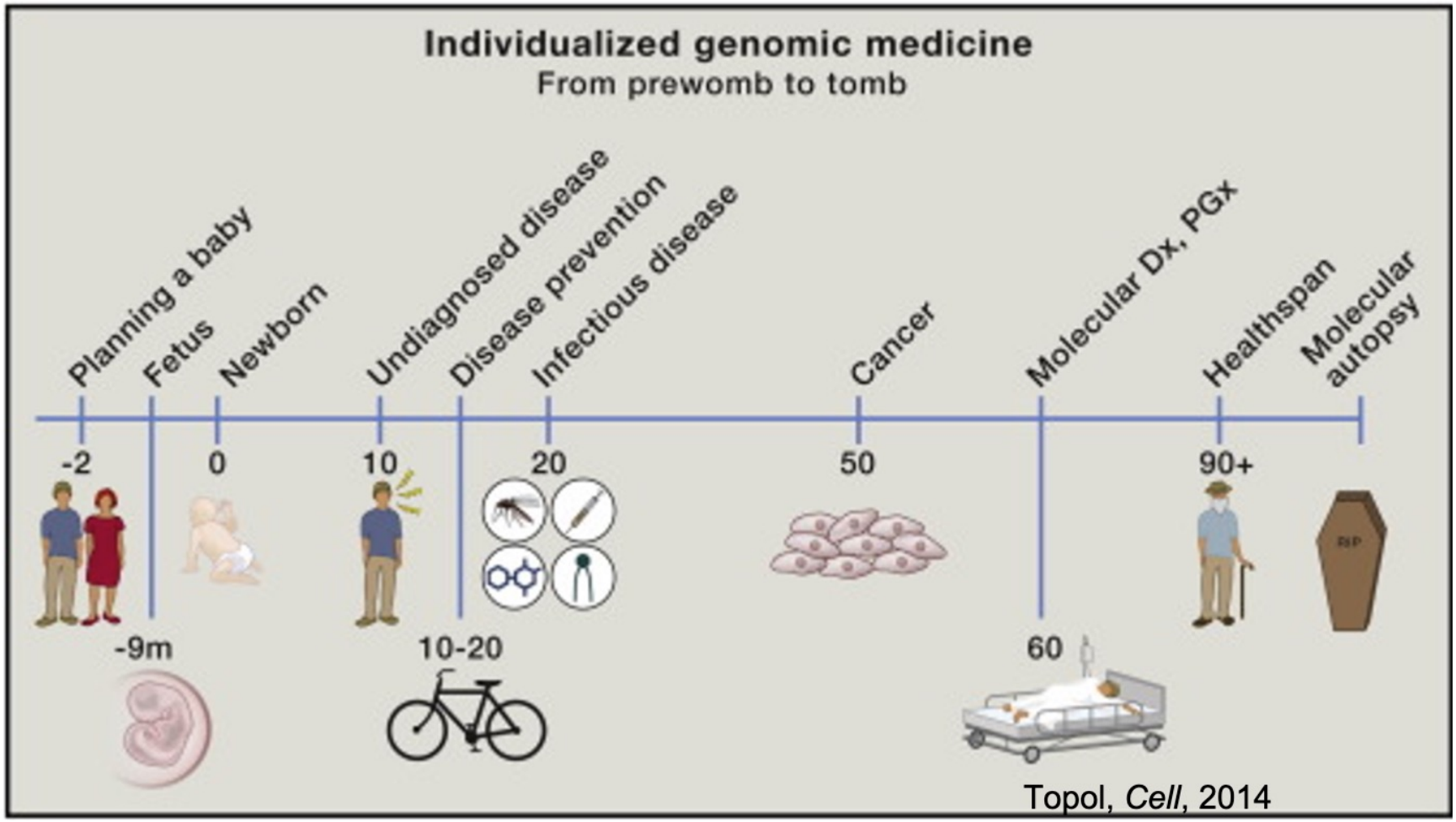
- Misdiagnoses, medical errors, prescription errors, surgical errors, under-treatments, over-treatments, unnecessary lab tests can be due to inferential gaps
- Late diagnosis of cancer can be due to the inferential gaps at the primary care
- Crisis caused by misuse, underuse, or overuse of antibiotics is in part due to serious inferential gaps

# Precision medicine requires a multi-level understanding of health and disease…

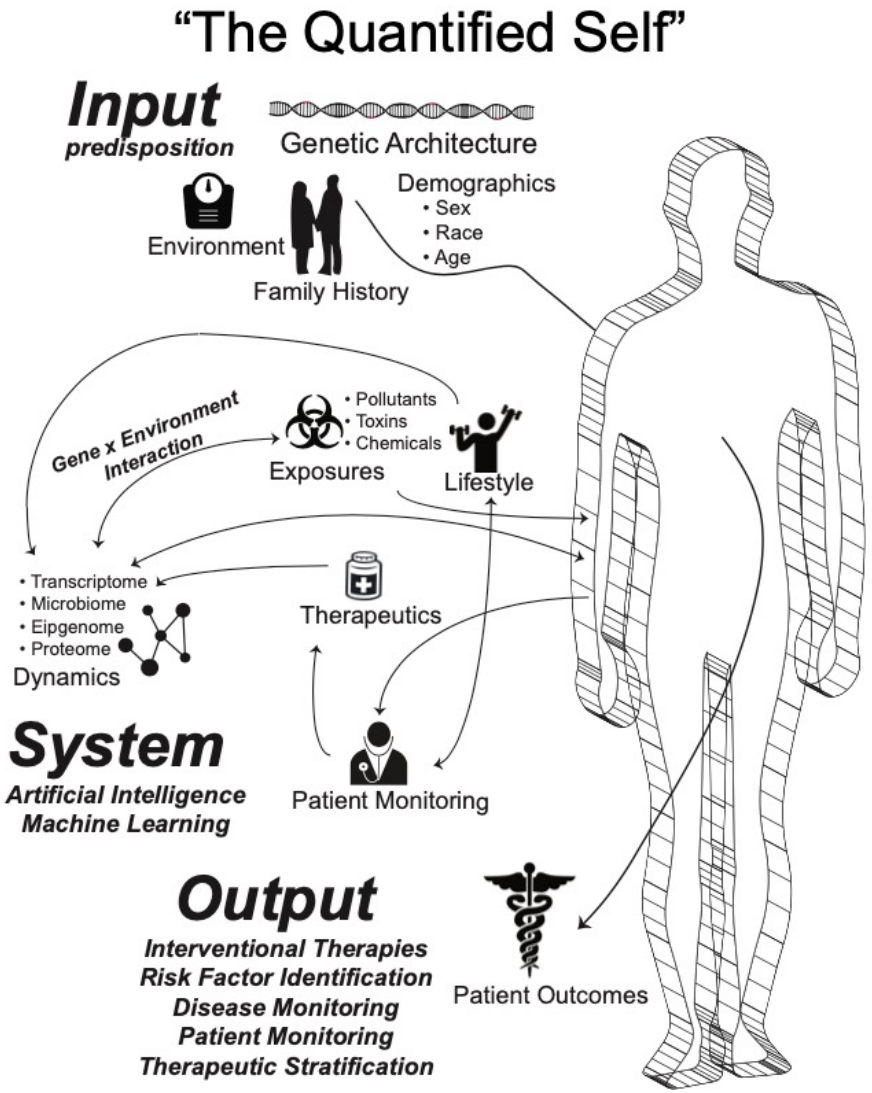


# …und understanding how health and disease states evolve

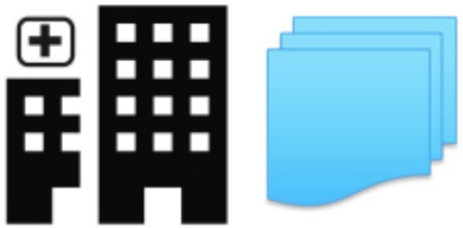


# This all-encompassing dataset does no exist…



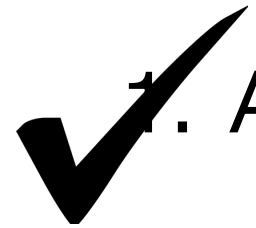

… but real-world data can serve as proxy

## Electronic Health Records



**Next:** How are electronic health records used for research?

# Outline for today's class

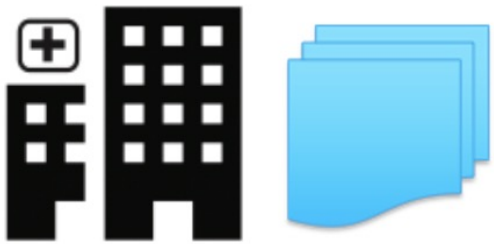


1. AI/ML for precision medicine
2. What are EHR data useful for?

3. Limitations & biases of EHR data
4. Highlights of ML on EHR data:
   - Polypharmacy and adverse drug events
   - Modeling disease progression

# Electronic health records

- The digitized paper charts
- The underlying goal/purpose of EHRs is **billing/infrastructure**
- Contains any data collected during an individual's interaction with a medical system
- Different software vendors (e.g., EPIC, Cerner)

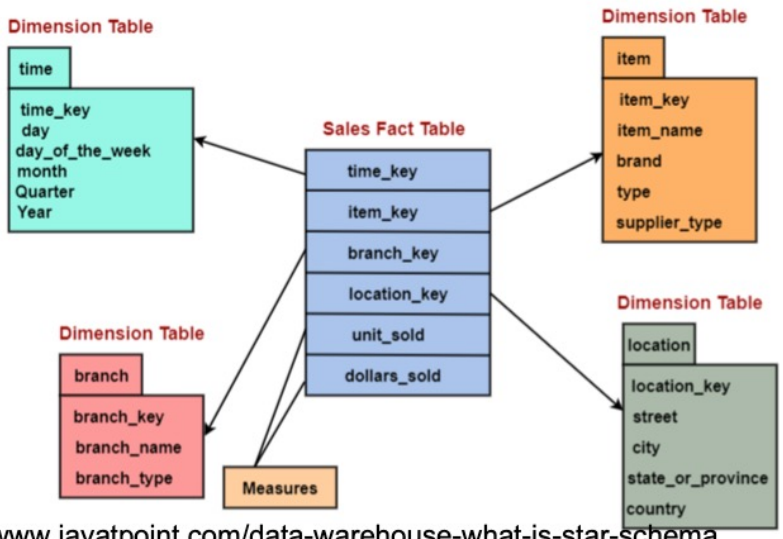


**Data type examples:**

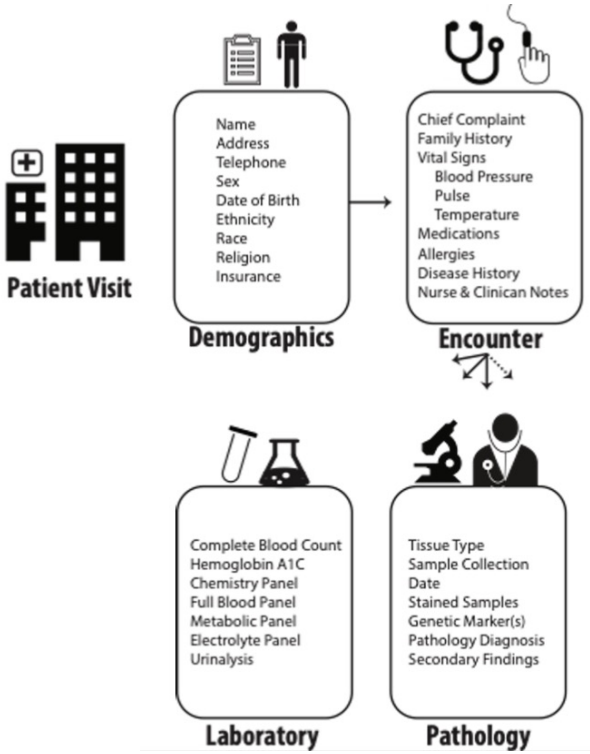
| *Clinical* | *Non-clinical* |
|---|---|
| ○ Diagnoses | ○ Demographics |
| ○ Procedures | ○ Insurance |
| ○ Lab test results | ○ Location |
| ○ Imaging | ○ Lifestyle |
| ○ Medications | |
| ○ Notes | |

# EHR data types and formats

- Made available by data warehouses
- Are often *encounter-based*
- Typically separated by modality (e.g., demographics table, lab table)
- Often in star-schema format

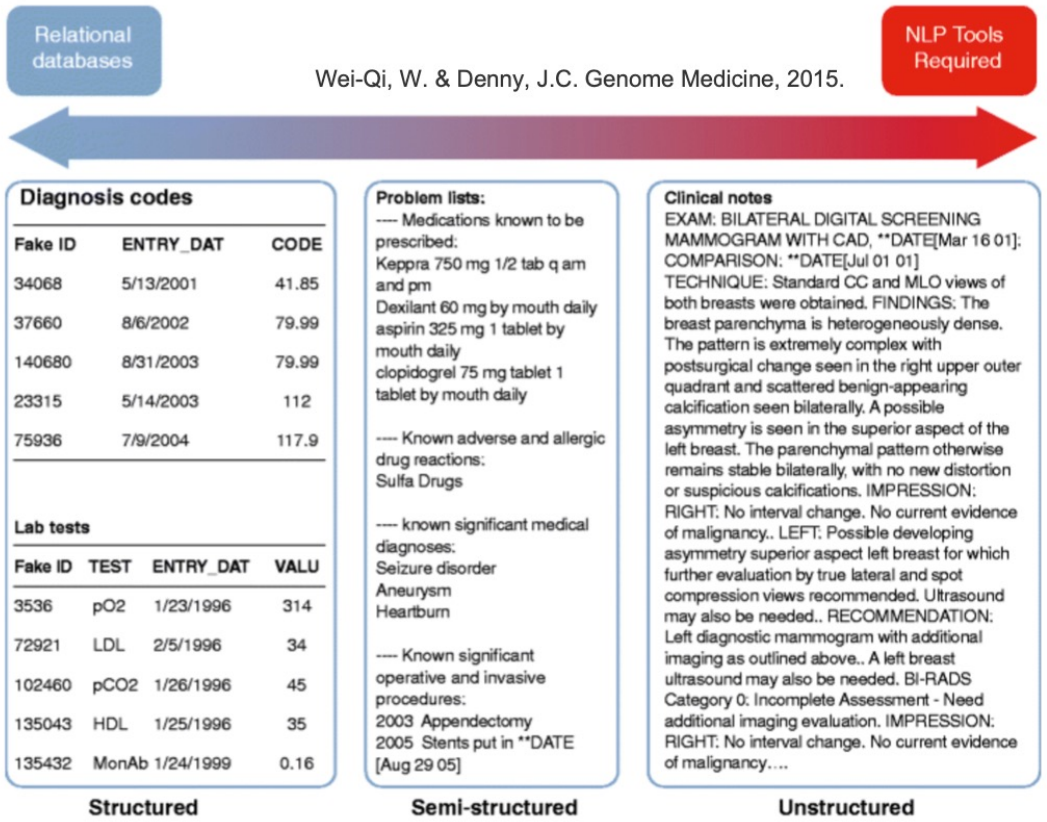



https://www.javatpoint.com/data-warehouse-what-is-star-schema



# EHR data structure

- **Structured:** labs, medications, etc.
- **Semi-structured:** smartforms, radiology impressions, echo reports
- **Unstructured:** clinical notes
- **Note:** It does not have all data!



# Types of research using EHRs?

- Characterize co-morbidities & epidemiological trends
- Identify disease sub-phenotypes
- Identify unknown drug adverse events
- Find symptom clusters
- Predict medication response
- Anticipate disease flare-ups
- Guide triage decisions
- Track treatment progression and sequelae
- Couple with other patient data modalities: genetics, images, notes, biosignals, etc.

+ countless more...

# Example: Identifying temporal disease trajectories (1/4)

**Data:** The entire spectrum of diseases covering 14.9 years of EHR data on 6.2 million patients

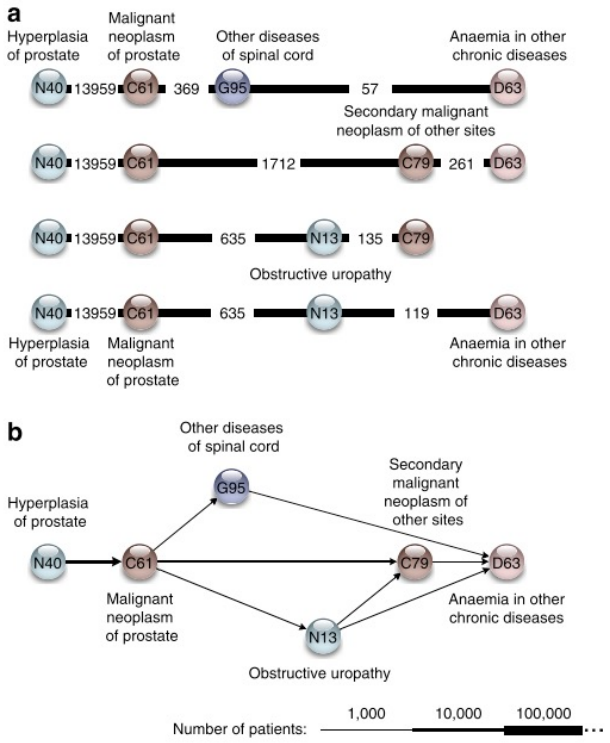



**Figure 2 | Disease trajectories and trajectory-cluster for prostate cancer.** The figure illustrates the transition from trajectories to a trajectory cluster. Each circle represents a diagnosis and is labelled with the corresponding ICD-10 code. The colours represent different ICD-10 chapters. The temporal diagnosis progression goes from left to right. (**a**) All trajectories that contribute to the prostate-cancer cluster. The number of patients, who follow the trajectory until a given diagnosis, is given in the edges. (**b**) The prostate cancer trajectory cluster that represents all the trajectories. The width of the edges corresponds to the number of patients with the directed diagnosis pair from the full population. The cluster describes a normal progression from having hyperplesia of prostate diagnosed to having prostate cancer, cancer metastasis and anaemia.

Jensen et al., Nature Communications, 2014

# Example: Identifying temporal disease trajectories (2/4)

1. Analyze temporal co-morbidity:
   - From the full data set, identify pairs (D1→D2) of diagnoses where D2 occurs within a 5-year time frame of D1
   - Test pairs for significant directionality: Identify those where a significantly higher number of patients had D1 occurring before D2 compared with the opposite direction or in the same admission
   - This analysis yielded 1,171 four-long diagnosis trajectories
2. Cluster trajectories:
   - Objective: Cluster trajectories that have large diagnosis overlap and represent variants of general patterns of disease progression.
   - Cluster trajectories based on which diagnoses they share
     - Use **Markov clustering** to assign each diagnostic code to a cluster
     - Use the **Jaccard index** as a similarity measure: **Count how many trajectories both diagnoses are part of and normalize by the total number of trajectories either is part of**
     - Combine trajectories with all diagnoses within the same cluster into directed trajectory clusters in which the patterns can be examined

Jensen et al., Nature Communications, 2014

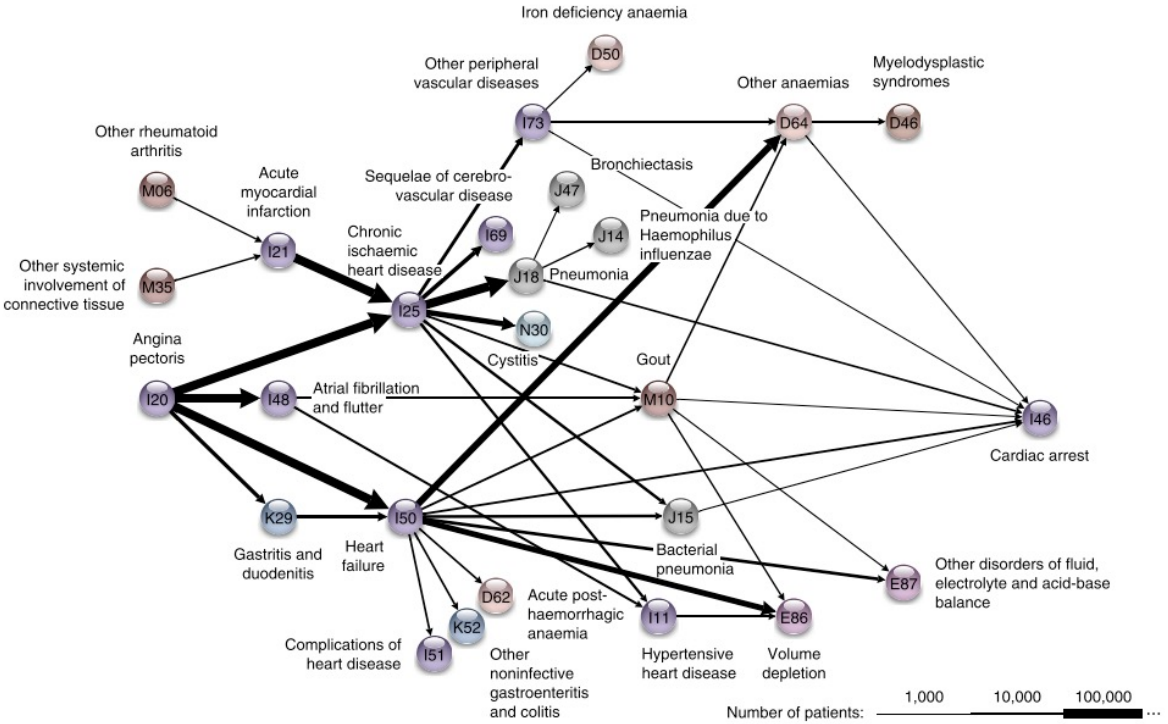
# Example: Identifying temporal disease trajectories (3/4)

- Clustering identified 15 clusters:
  - The five largest clusters covered 46, 25, 12, 9, and 8 diagnoses each
  - Each is a group of patterns centered on a small number of key diagnoses, which are central to disease progression and important to diagnose early to mitigate the risk of adverse outcomes
  - The five largest clusters were enriched for:
    - Diseases of the prostate
    - Chronic obstructive pulmonary disease
    - Cerebrovascular disease
    - Cardiovascular disease
    - Diabetes mellitus

Jensen et al., Nature Communications, 2014

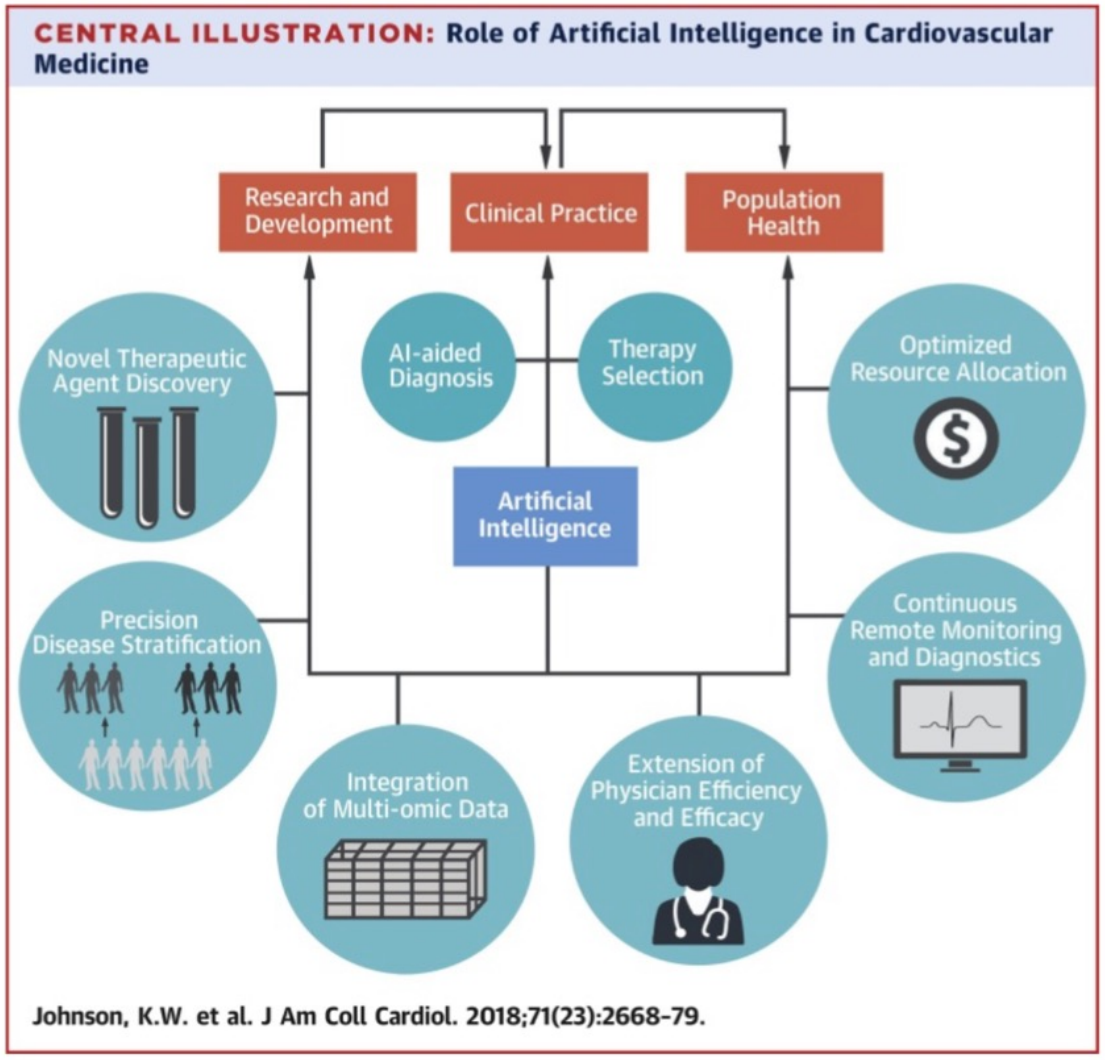
# Example: Identifying temporal disease trajectories (4/4)

**Cardiovascular cluster:** Gout is a central diagnosis in the cardiovascular cluster, supporting evidence that gout is important to progression of cardiovascular diseases




Jensen et al., Nature Communications, 2014

# Goals of ML for healthcare using EHR

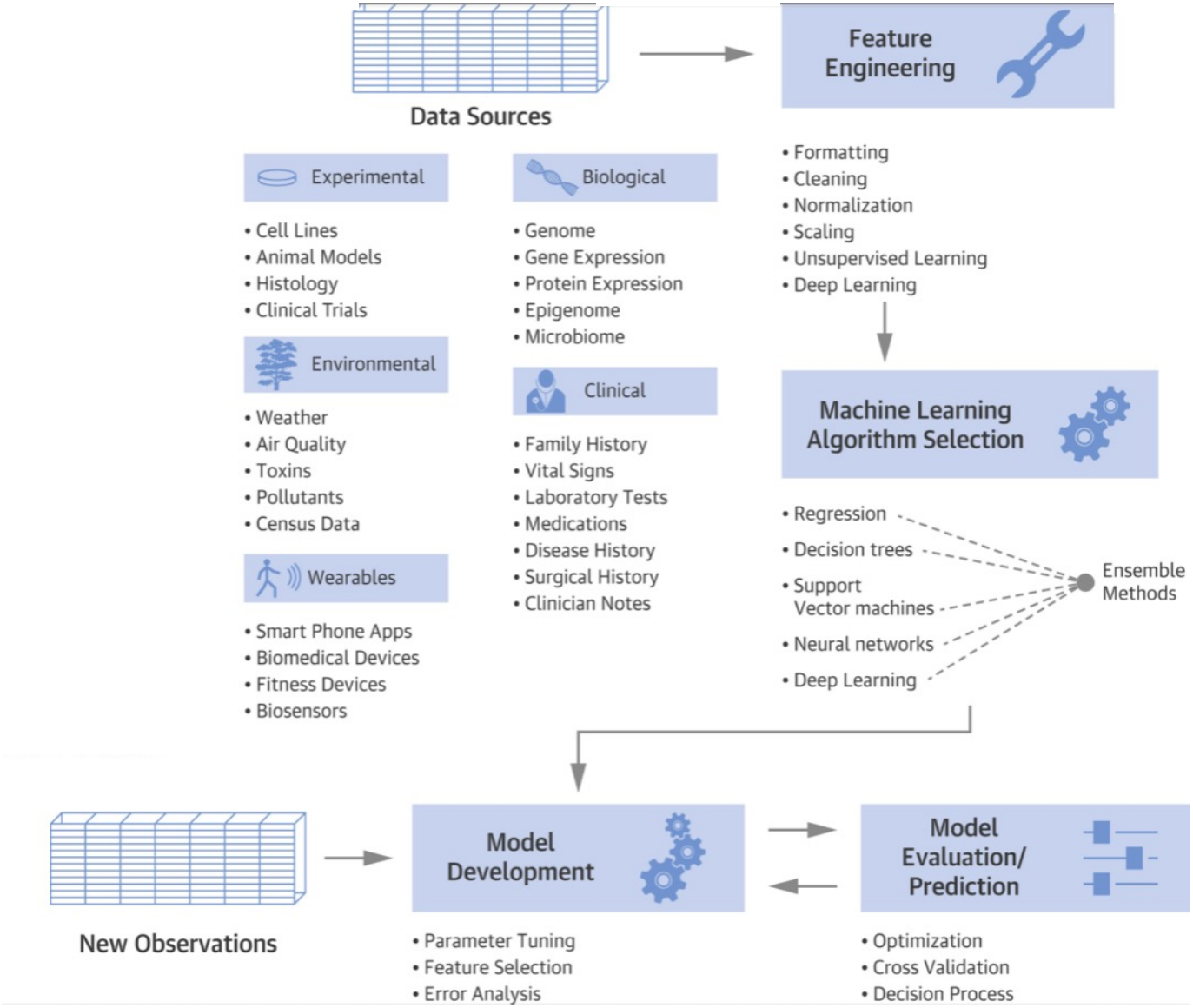


# Typical ML workflow for EHR data

- Gather (identify relevant feature)
- QC values (wrong unit?)
- Check for/address missingness
- Phenotype and design cohort
- Define outcome (label) and study period
- Use relevant ML techniques
- Pre-process data to fit the ML technique
- Refine and repeat

Johnson et al., JACC, 2018

# Typical ML workflow for EHR data



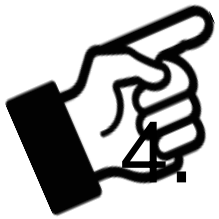

Johnson et al., JACC, 2018

# Outline for today's class

1. AI/ML for precision medicine

2. What are EHR data useful for?

3. Limitations & biases of EHR data



4. Highlights of ML on EHR data:
   - Polypharmacy and adverse drug events
   - Modeling disease progression

# What is a disease?

- A disease is not easily defined in EHRs!
- Many ways in which a disease can be represented (and often wrong)
- Phenotyping algorithms and standardized concepts to the rescue: accurately identify patients with a specific observable trait from imperfect EHR data

# How well do various data types define a disease? (1/3)

- **Goal:** Evaluate phenotyping performance of major EHRs
  - Diagnosis codes
  - Primary notes
  - Medication lists
- **Approach:**
  - Select ten diseases: atrial fibrillation, Alzheimer's disease, breast cancer, gout, human immunodeficiency virus infection, multiple sclerosis, Parkinson's disease, rheumatoid arthritis, and T1D/T2D
  - For each disease, classify patients into seven categories based on the presence of evidence for disease in a) diagnosis codes, b) primary notes, and c) specific medications
  - For each disease, select 175 patients for **manual chart review**
  - Use review results to estimate **positive predictive value (PPV)** for each EHR data type alone and in combination

Wei et al., JAMIA, 2016

# How well do various data types define a disease? (2/3)

- PPV is the ratio of patients that truly have the disease according to manual chart review to all patients who had been identified as having the disease
- PPVs on single data types were inadequate for accurate phenotyping (0.06–0.71)
- Using two or more ICD codes improved the average PPV to 0.84

Positive prediction values of various categories based on chart review results

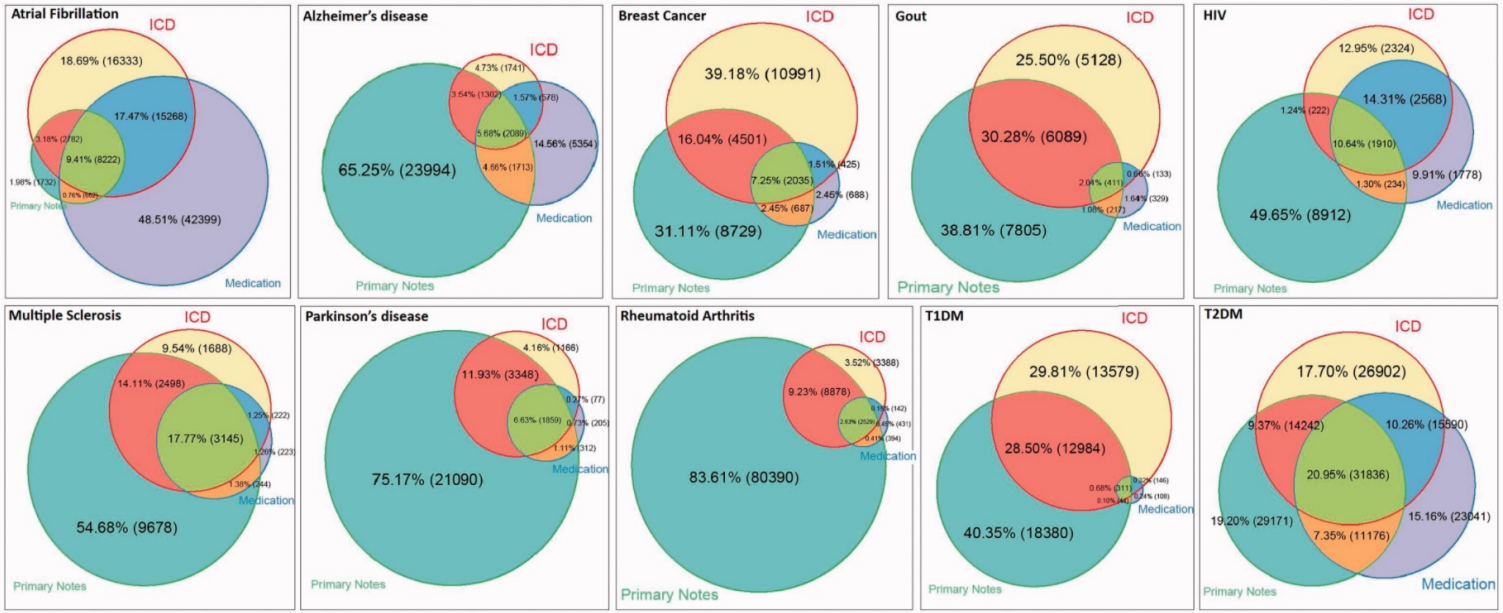
| Disease | ICD-9 Only | PN Only | Meds Only | ICD-9+Meds | ICD-9+PN | Meds+PN | ICD-9+both | ICD-9 | Meds | PN | ≥2 ICD-9 s | ≥2 Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AFIB | 0.52 | 0.72 | 0.08 | 0.72 | 1.00 | 1.00 | 1.00 | 0.72 | 0.35 | 0.96 | 0.88 | 0.84 |
| Alzheimer's | 0.28 | 0.20 | 0.00 | 0.80 | 0.88 | 0.92 | 0.88 | 0.69 | 0.40 | 0.32 | 0.74 | 0.88 |
| Breast CA | 0.12 | 0.72 | 0.04 | 0.88 | 0.96 | 1.00 | 1.00 | 0.45 | 0.81 | 0.84 | 1.00 | 0.97 |
| Gout | 0.56 | 0.84 | 0.00 | 0.92 | 1.00 | 1.00 | 1.00 | 0.81 | 0.69 | 0.91 | 0.93 | 1.00 |
| HIV | 0.52 | 0.00 | 0.00 | 0.92 | 0.84 | 0.88 | 1.00 | 0.81 | 0.69 | 0.20 | 0.89 | 0.95 |
| MS | 0.20 | 0.08 | 0.12 | 0.88 | 0.88 | 0.88 | 1.00 | 0.78 | 0.93 | 0.41 | 0.86 | 0.94 |
| Parkinson | 0.48 | 0.16 | 0.04 | 0.84 | 1.00 | 0.88 | 0.96 | 0.89 | 0.87 | 0.33 | 0.94 | 0.98 |
| RA | 0.36 | 0.20 | 0.00 | 0.64 | 0.76 | 0.88 | 0.84 | 0.68 | 0.73 | 0.27 | 0.77 | 0.78 |
| T1DM | 0.28 | 0.12 | 0.04 | 0.16 | 0.92 | 0.84 | 0.76 | 0.59 | 0.49 | 0.45 | 0.62 | 0.91 |
| T2DM | 0.36 | 0.68 | 0.24 | 0.60 | 0.80 | 1.00 | 0.84 | 0.65 | 0.65 | 0.80 | 0.73 | 0.81 |
| Average | 0.37 | 0.37 | 0.06 | 0.74 | 0.90 | 0.93 | 0.93 | 0.71 | 0.66 | 0.55 | 0.84 | 0.91 |
| Standard Deviation | 0.15 | 0.32 | 0.08 | 0.23 | 0.09 | 0.06 | 0.09 | 0.13 | 0.20 | 0.29 | 0.12 | 0.08 |

PN, primary notes

Wei et al., JAMIA, 2016

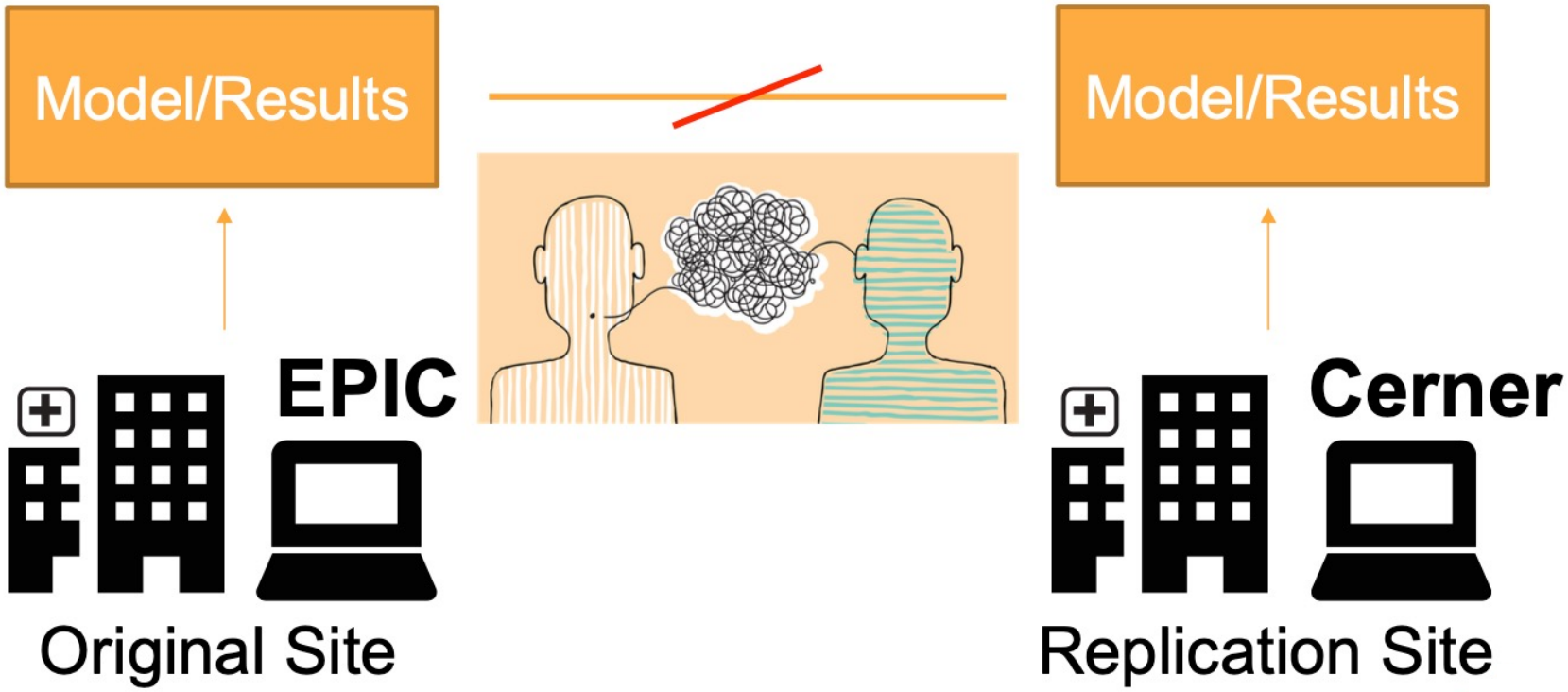
# How well do various data types define a disease? (3/3)

- Multiple data types provide a more consistent and higher performance than a single one
- Use multiple EHR data types for disease phenotyping

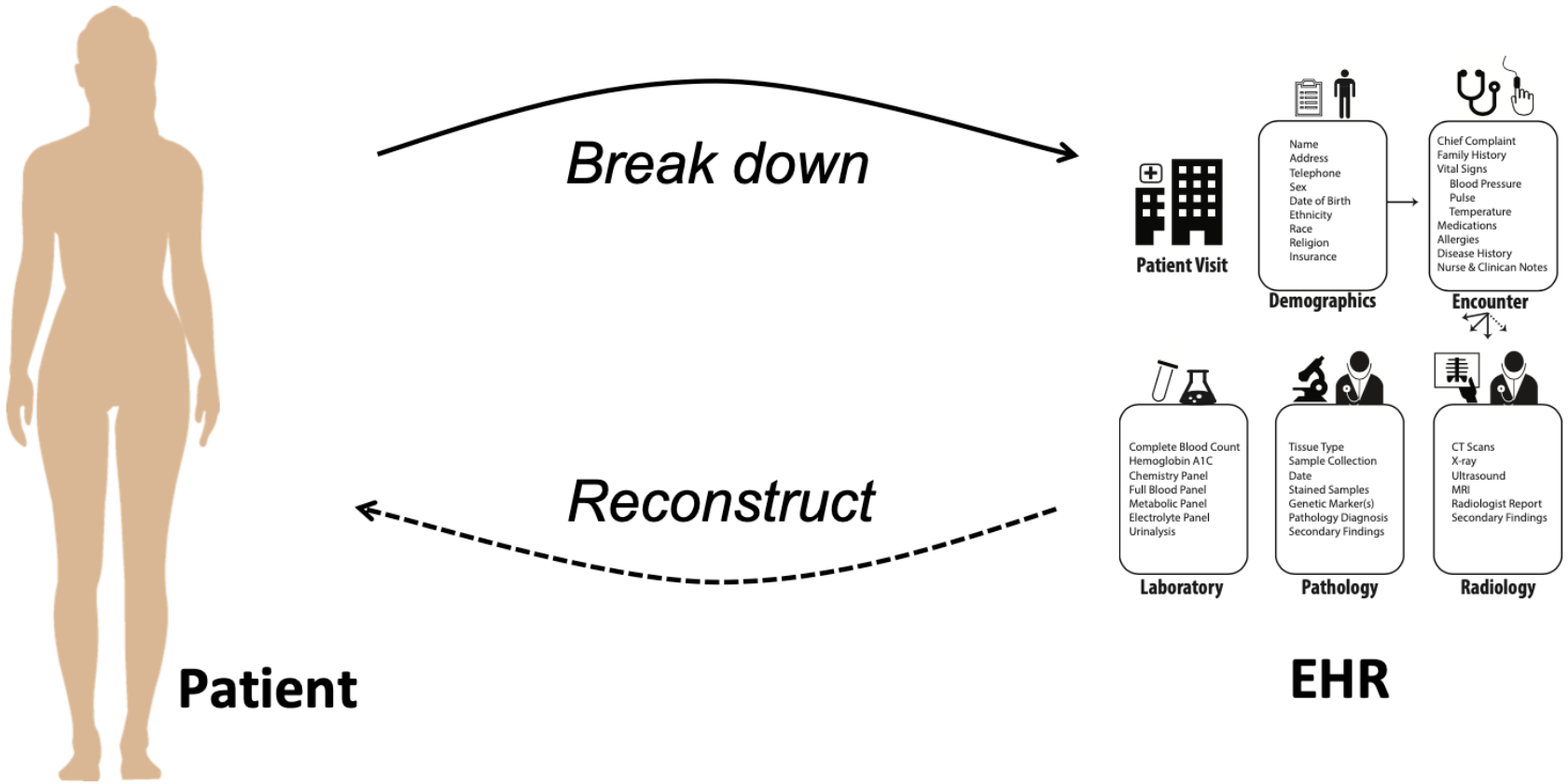



Wei et al., JAMIA, 2016

# External replication is necessary but not easy to facilitate



# It is challenging to capture health state from EHR



# ML models can learn the wrong information

RESEARCH

OPEN ACCESS

## Biases in electronic health record data due to processes within the healthcare system: retrospective observational study

Denis Agniel,[1] Isaac S Kohane,[1,2] Griffin M Weber[1,3]

**RESULTS**

The presence of a laboratory test order, regardless of any other information about the test result, has a significant association (P<0.001) with the odds of survival in 233 of 272 (86%) tests. Data about the timing of when laboratory tests were ordered were more accurate than the test results in predicting survival in 118 of 174 tests (68%).

**CONCLUSIONS**

Healthcare processes must be addressed and accounted for in analysis of observational health data. Without careful consideration to context, EHR data are unsuitable for many research questions. However, if explicitly modeled, the same processes that make EHR data complex can be leveraged to gain insight into patients' state of health.

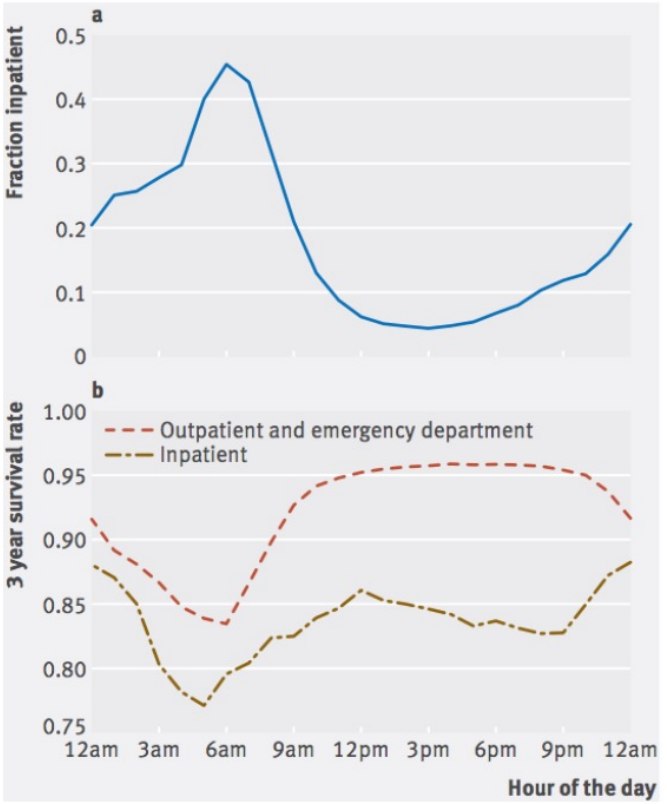



**Fig 4 | White blood cell count by hour of the day. Note that (b) was smoothed using a three point running average**

# ML models can "cheat" (1/3)

- **Objective:** Hip fractures are a leading cause of death and disability among older adults
  - Most commonly missed diagnosis on pelvic radiographs
  - Delayed diagnosis leads to higher cost & worse outcomes

**Deep learning predicts hip fracture using confounding patient and healthcare variables**

Marcus A. Badgeley, John R. Zech, Luke Oakden-Rayner, Benjamin S. Glicksberg, Manway Liu, William Gale, Michael V. McConnell, Bethany Percha, Thomas M. Snyder & Joel T. Dudley

- **Data:** Collect 23,602 hip radiographs from 9,024 patients, patient and hospital process EHR data:
  - Prevalence of fracture is 3% (779/23,602)
  - Patients with fractures were more likely to report a recent fall and less likely to report pain
  - Features: image (IMG), disease (fracture) class, 5 patient (PT) features, 14 hospital process (HP) features

Badgeley et al., NPJ Digital Medicine, 2019

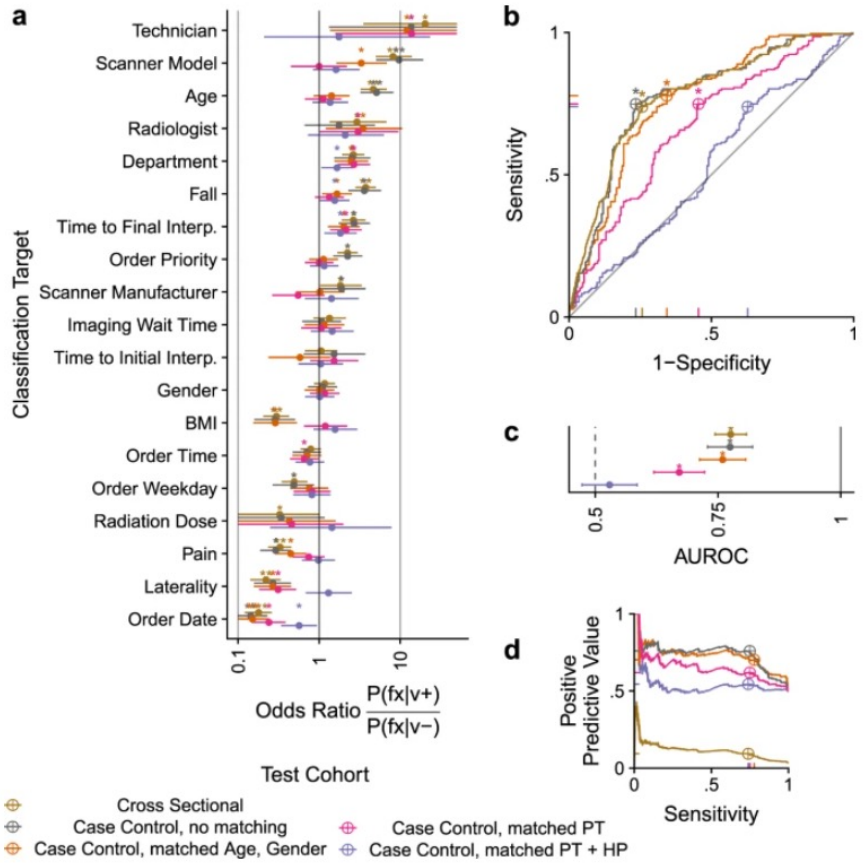
# ML models can “cheat” (2/3)

- **ML model:** Train a neural network on radiographs to classify fracture
- **Results:** Fracture is predicted:
  - Moderately well from the **IMG** data alone (AUC=0.78)
  - Better when combining **IMG + PT** (AUC=0.86)
  - Better when combining **IMG + PT + HP** (AUC=0.91)
- **Follow-up analysis:**
  - Test ML model whether it can directly detect fracture versus indirectly predict fracture by detecting confounding variables associated with fracture
  - On a test set with fracture risk balanced across PT and HP variables, fracture detector is no better than random (AUC=0.52)

Badgeley et al., NPJ Digital Medicine, 2019

# ML models can “cheat” (3/3)

- On test set with fracture risk balanced across PT and HP features, fracture detector is no better than random (AUC=0.52)
- Confounding variable (e.g., time since prior lab order, or which scanner in a hospital is used to acquire a radiograph) is associated with both:
  - **Explanatory variable** (acuity of a patient’s illness, or a patient’s clinically predicted risk of fracture)
  - **Outcome** (mortality, or the likelihood of a radiograph’s pixels containing patterns suggestive of fracture)




Badgeley et al., NPJ Digital Medicine, 2019

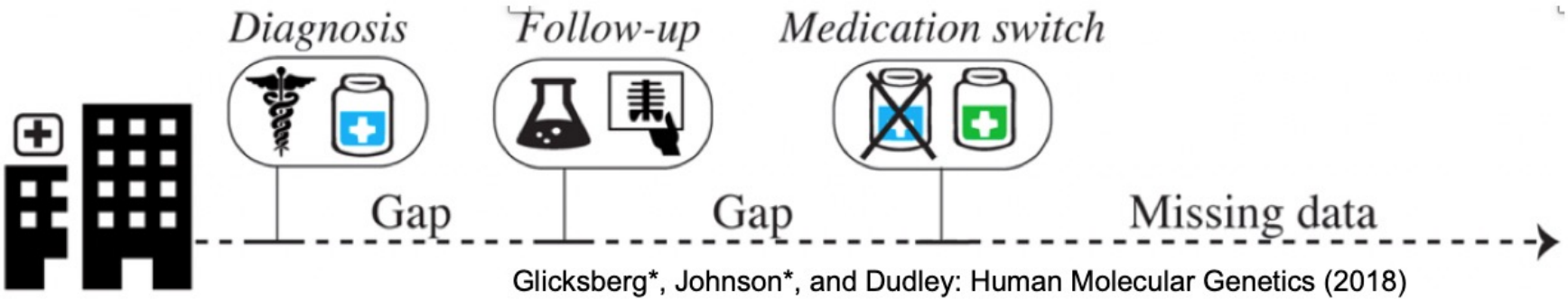
# Limitations & biases of EHR

- Diseases are not easily defined in EHRs!
- External replication is not easy to facilitate
- It is challenging to capture health state from EHR
- ML algorithms can learn the wrong information
- ML algorithms can “cheat”
- ML algorithms can fail on other patient populations
- Biased real-world data can lead to real-world consequences

Ben Glicksberg, Mount Sinai

# Fine print of using EHRs

- In USA (and elsewhere), the healthcare is fragmented and EHRs do not extend beyond specific health system
- EHRs capture only data that is entered and how it is entered: “Garbage in, garbage out”
- EHR systems are messy, redundant, incomplete, heterogenous, erroneous, etc.
- Interfacing with EHR data is challenging and requires domain expertise
- Biases are propagated through!
- Poorly encoded key information: i.e., social determinants of health
- The “missing phenome”




Glicksberg*, Johnson*, and Dudley: Human Molecular Genetics (2018)

# Quick Check

https://forms.gle/N85jAoUVPuBFyG3U8

## BMI 702: Biomedical Artificial Intelligence

*Foundations of Biomedical Informatics II, Spring 2024*

Quick check quiz for lecture 2: Introduction to AI on clinical datasets

Course website: https://zitniklab.hms.harvard.edu/BMI702

* Indicates required question

First and last name *

Your answer

Harvard email address *

Your answer

Give an example of inferential gap in clinical decision making *

Your answer

Select one EHR-based research project from slide 13 and briefly describe how a typical ML workflow (slide 19) for the project would look like *

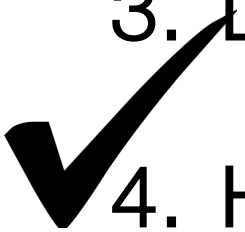
Your answer

# Outline for today's class

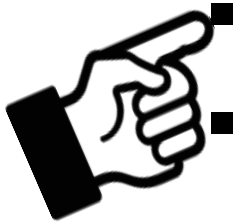
1. AI/ML for precision medicine

2. What are EHR data useful for?

3. Limitations & biases of EHR data



4. Highlights of ML on EHR data:
   - Polypharmacy and adverse drug events
   - Modeling disease progression



# Polypharmacy

Patients take multiple drugs to treat complex or co-existing diseases

46% of people over 65 years take more than 5 drugs

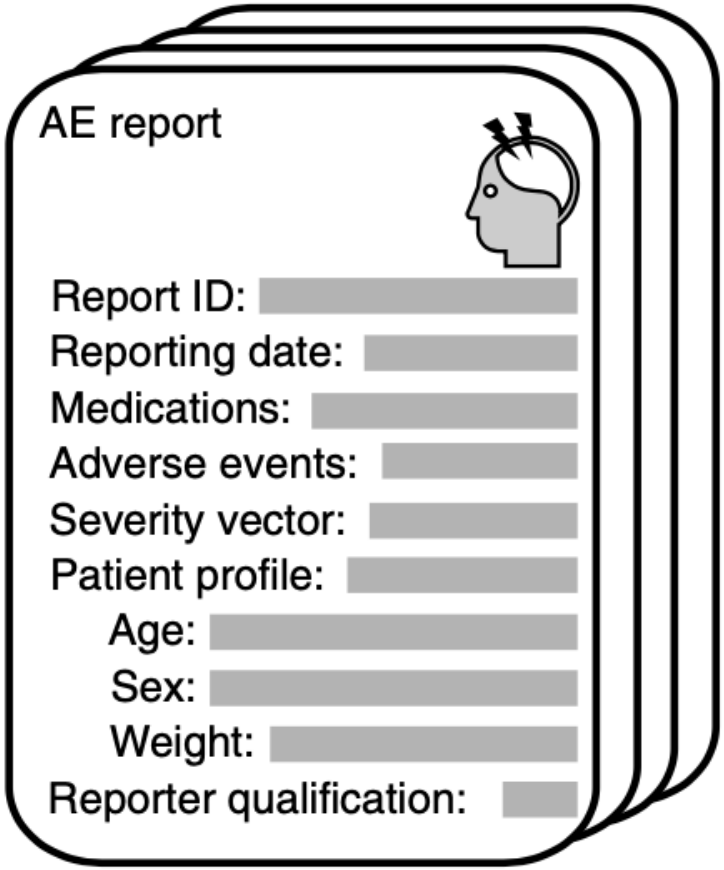
Many take more than 20 drugs to treat heart diseases, depression or cancer

15% of the U.S. population affected by unwanted side effects

Annual costs in treating side effects exceed $177 billion in the U.S. alone

Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# FDA adverse event reporting




FDA U.S. FOOD & DRUG ADMINISTRATION

(hypothetical scenario)

# Unwanted side effects

The FDA Adverse Event Reporting System (FAERS)

| **Drugs taken** | **Unwanted side effects** |
|---|---|
|  | Peliosis hepatis (0.2%), Heart rate increased (0.5%), Aortic aneurysm (0.1%) |
|  | Joint stiffness (3%), Joint swelling (1%), Bone marrow fibrosis (0.01%) |
|  | Anaemia (1%), Bone marrow fibrosis (0.5%), Intestinal ulcer (0.001%) |
|  | Anaemia (1%), Bone marrow fibrosis (0.1%), Intestinal ulcer (0.01%), Joint stiffness (3%), Joint swelling (1%) |
|  | Peliosis hepatis (0.2%), Heart rate increased (0.5%), Aortic aneurysm (0.1%), Joint stiffness (3%), Joint swelling (1%), Bone marrow fibrosis (0.01%) |
| ... | ... |

(hypothetical scenario)

# Unwanted side effects

The FDA Adverse Event Reporting System (FAERS)

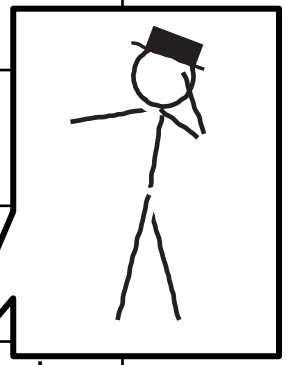
| Drugs taken | Unwanted side effects |
|---|---|
|  | Peliosis hepatis (0.2%), Heart rate increased (0.5%), Aortic aneurysm (0.1%) |
|  | Joint stiffness (3%), Joint swelling (1%), Bone marrow fibrosis (0.01%) |
|  | Anaemia (1%), Bone marrow fibrosis (0.5%), Intestinal ulcer (0.001%) |
|  | Anaemia (1%), Bone marrow fibrosis (0.1%), Intestinal ulcer (0.01%), Joint stiffness (3%), Joint swelling (1%) |
|  | Peliosis hepatis (0.2%), Heart rate increased (0.5%), Aortic aneurysm (0.1%), Joint stiffness (3%), Joint swelling (1%), Bone marrow fibrosis (0.01%) |
| ... | ... |

(hypothetical scenario)

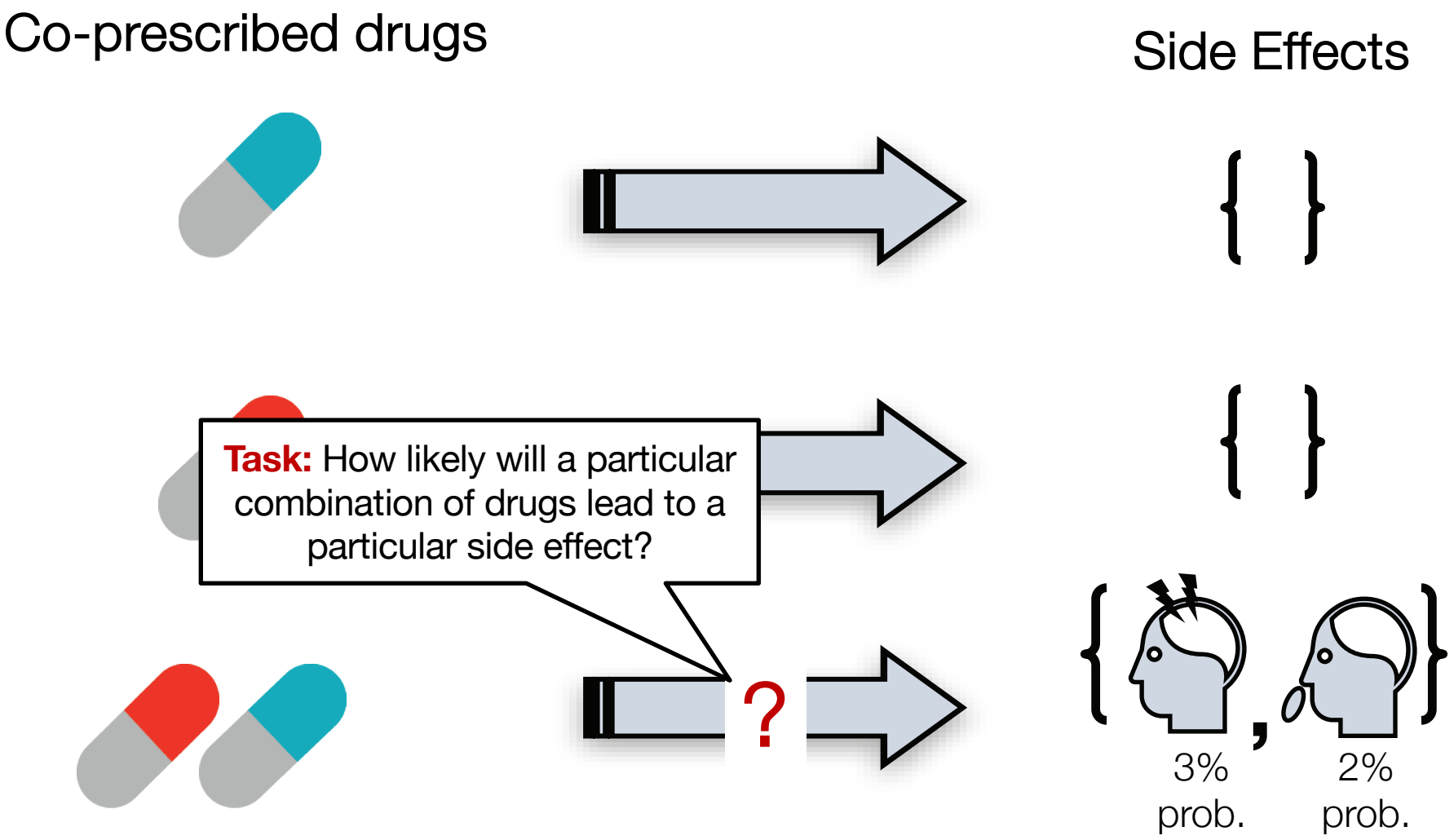
# Unwanted side effects

The FDA Adverse Event Reporting System (FAERS)

| **Drugs taken** | **Unwanted side effects** |
|---|---|
|  | Peliosis hepatis (0.2%), Heart rate increased (0.5%), Aortic aneurysm (0.1%) |
|  | Joint stiffness (3%), Joint swelling (1%), Bone marrow fibrosis (0.01%) |
|  | Anaemia (1%), Bone marrow fibrosis (0.5%), Intestinal (0.001%) |
|  | Anaemia (1%), Bone marrow fibrosis (0.1%), Intestinal ulcer (0.01%), Joint stiffness (3%), Joint swelling (1%), **Colon cancer (0.1%), Fatigue (2%)** |
|  | Peliosis hepatis (0.2%), Heart rate increased (0.5%), Aortic aneurysm (0.1%), Joint stiffness (3%), Joint swelling (1%), Bone marrow fibrosis (0.01%) |
| ... | ... |



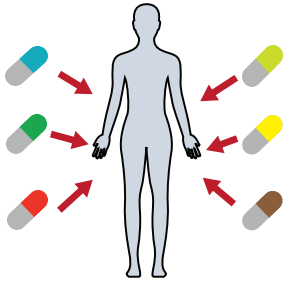
# Unexpected drug Interactions



# Why is modeling polypharmacy a hard problem?

## Combinatorial explosion

- >13 million possible combinations of 2 drugs
- >20 billion possible combinations of 3 drugs

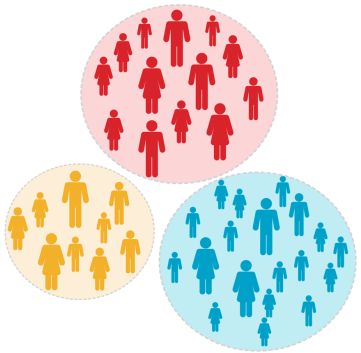


## Non-linear & non-additive interactions

- Different effect than the additive effect of individual drugs

## Small subsets of patients

- Side effects are interdependent
- No info on drug combinations not yet used in patients



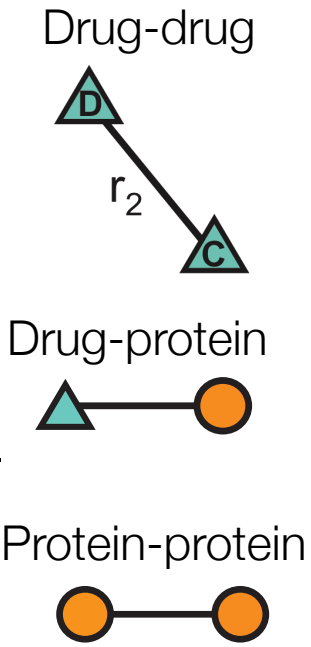
Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Polypharmacy dataset

Objective: Capture molecular, drug, and patient data for all drugs prescribed in the U.S.

Dataset:

- 4,651,131 drug-drug edges: Patient data from adverse event system, tested for confounders [FDA]
- 18,596 drug-protein edges
- 719,402 protein-protein edges: Physical, metabolic enzyme-coupled, and signaling interactions
- Drug and protein features: drugs' chemical structure, proteins' membership in pathways




Gives polypharmacy network with over 5 million edges separated into 1,000 different edge types

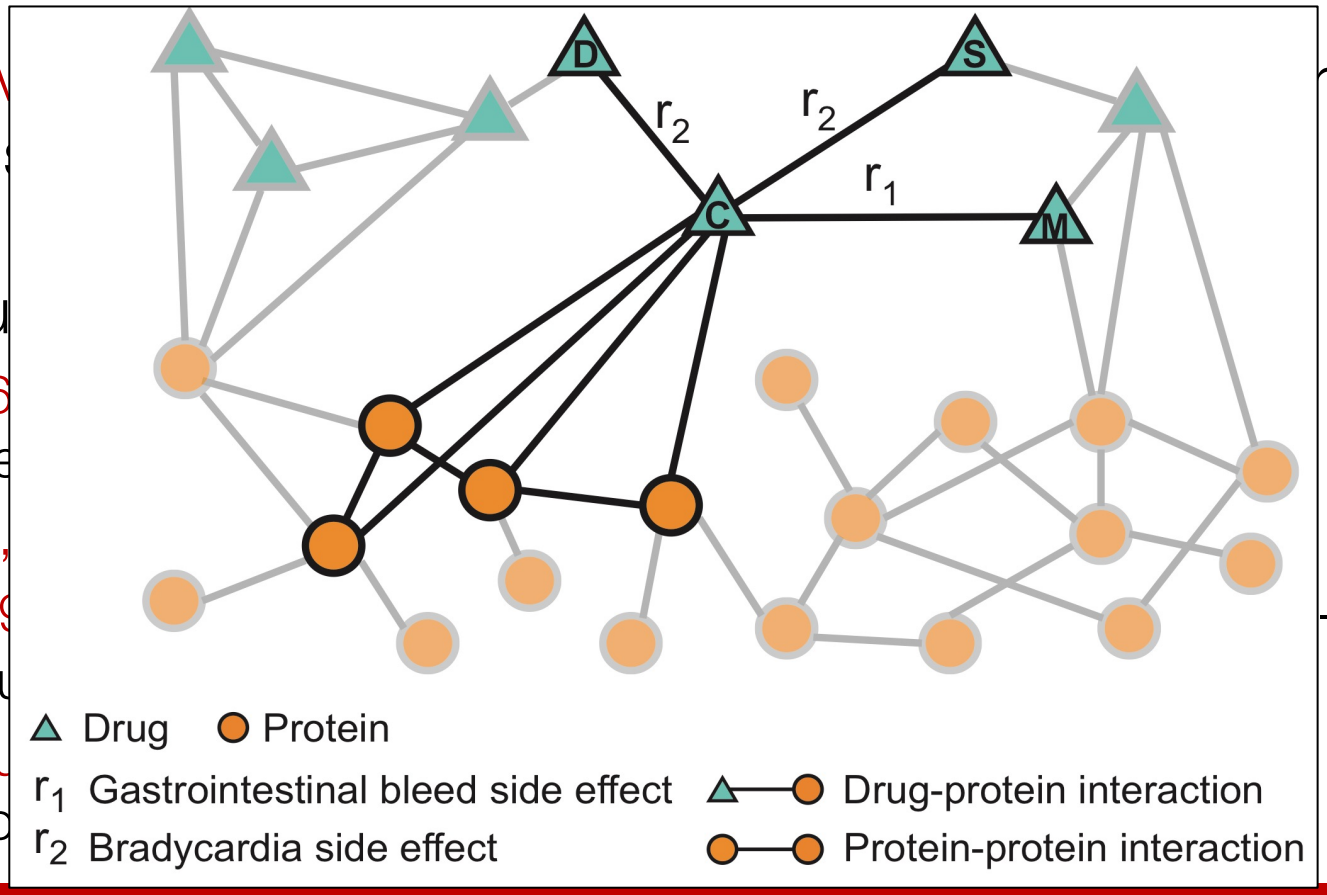
Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Polypharmacy dataset

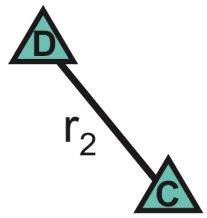
Objectiv
all drug

A uniqu

- 4,6
  eve
- 18,
- 719
  cou
- Dru
  pro




Drug-drug



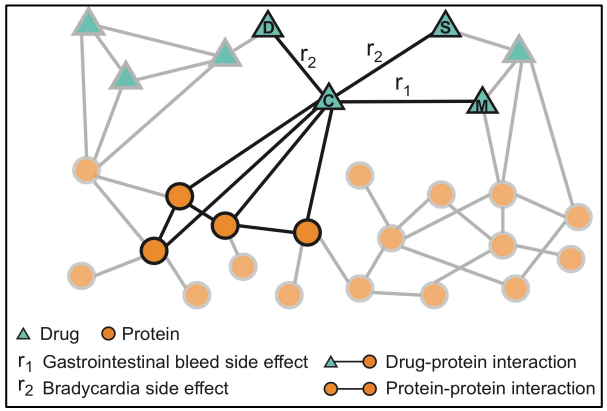

Drug-protein

Protein-protein

Gives polypharmacy network with over 5 million edges separated into 1,000 different edge types

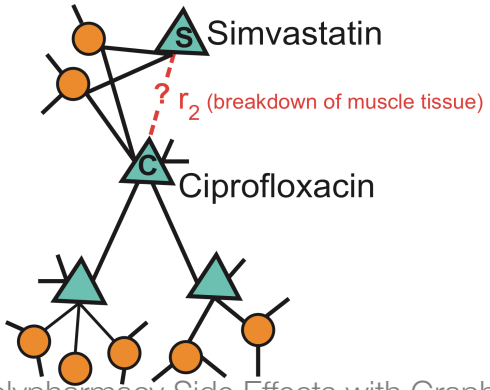
Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Overall ML approach




Two main stages:

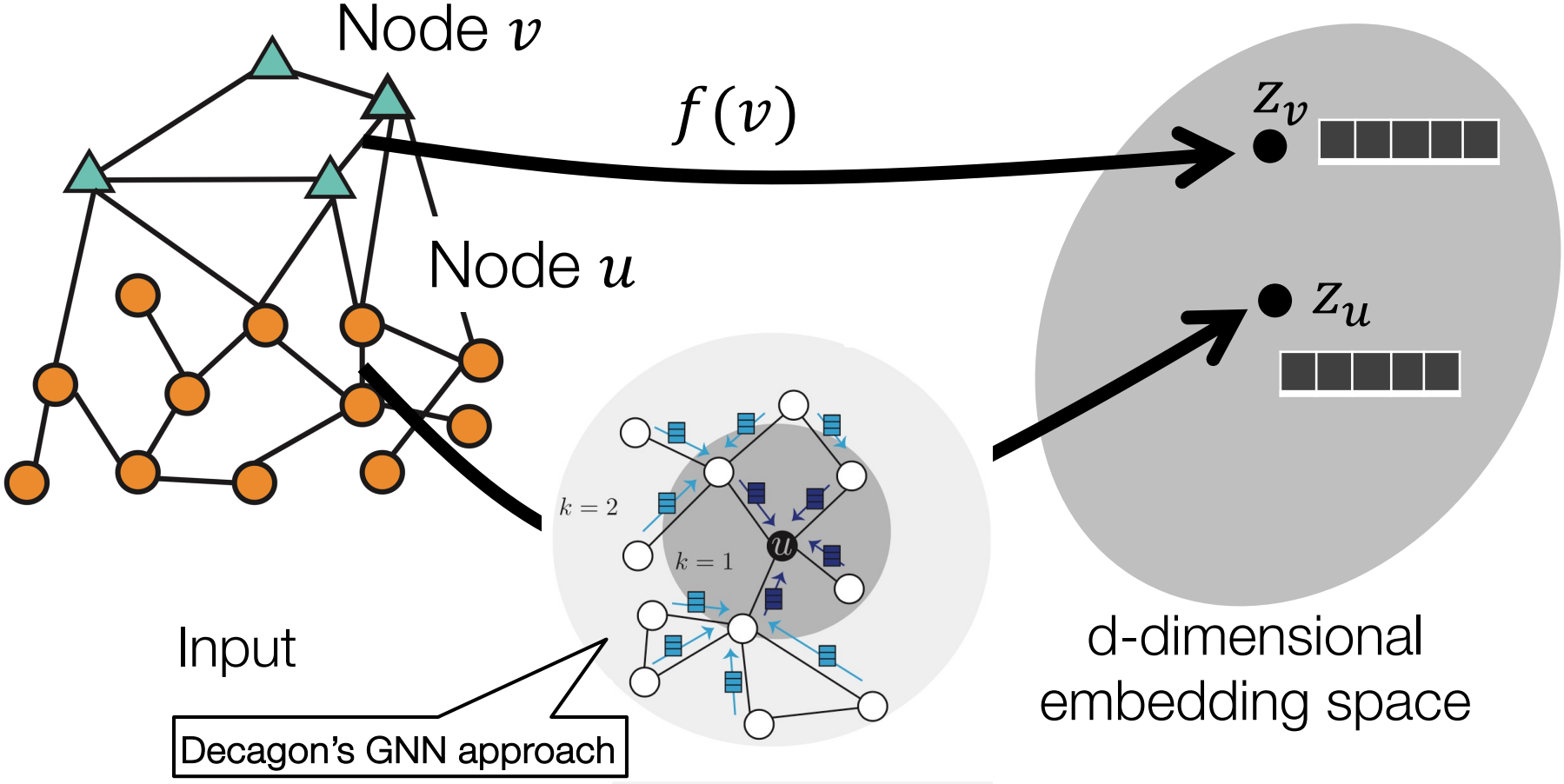
1. Learn an **embedding** for every node in polypharmacy dataset
2. Predict a score for every drug-drug, drug-protein, protein-protein pair in the test set based on the embeddings




Example: How likely will Simvastatin and Ciprofloxacin, when taken together, break down muscle tissue?

Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Approach: Graph neural network

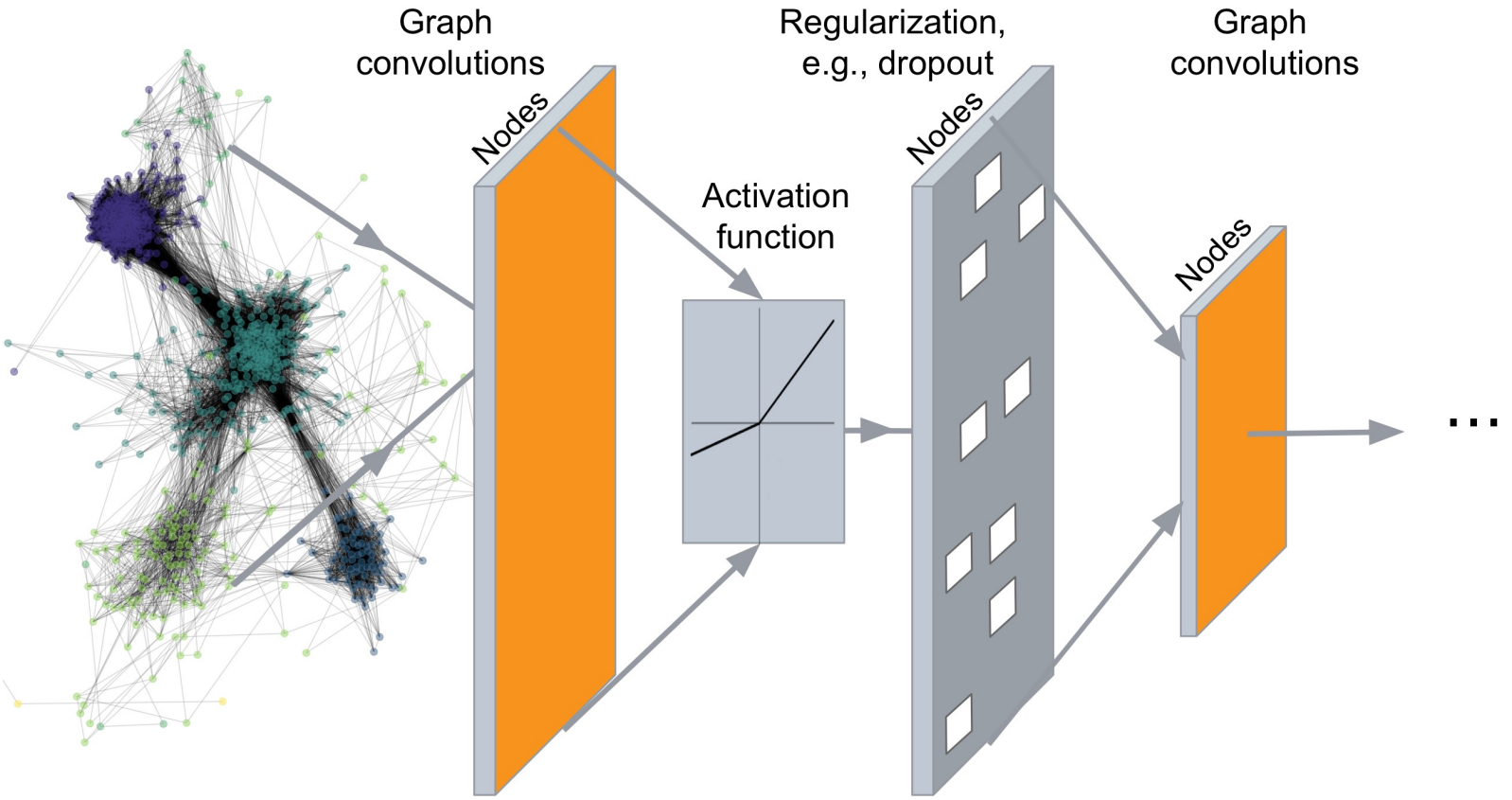



Map nodes to d-dimensional embeddings such that nodes with similar network neighborhoods are embedded close together

Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Graph neural networks

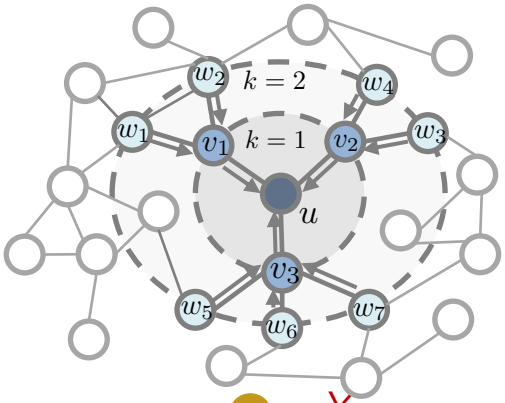
- Encoder: Multiple layers of nonlinear transformation of graph structure



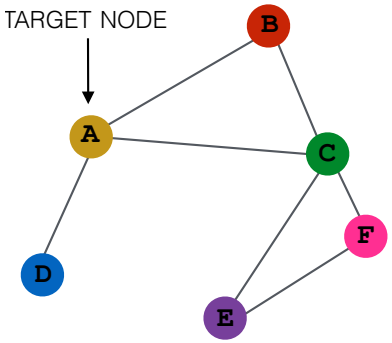

Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Graph neural networks

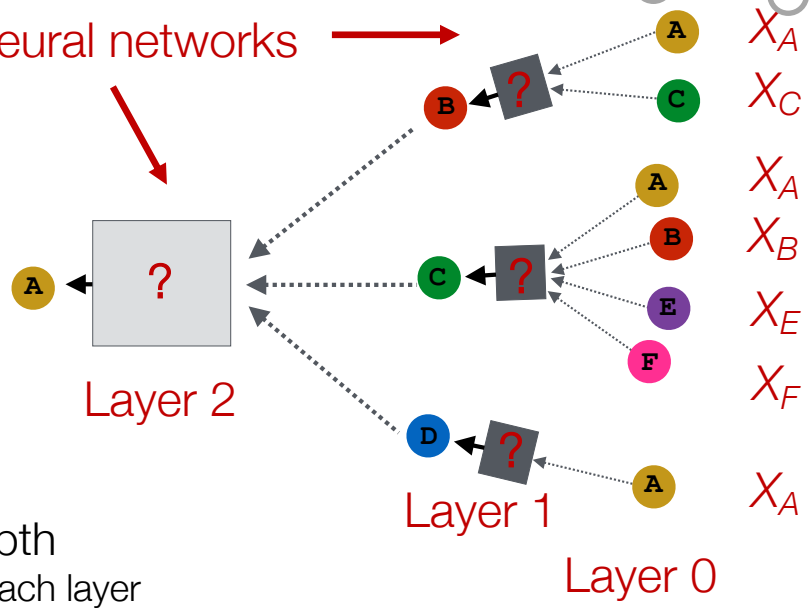
- Intuition:
  - Each node's neighborhood defines a computational graph
  - Generate node embeddings based on local network neighborhoods




- Neighborhood aggregation:




INPUT GRAPH

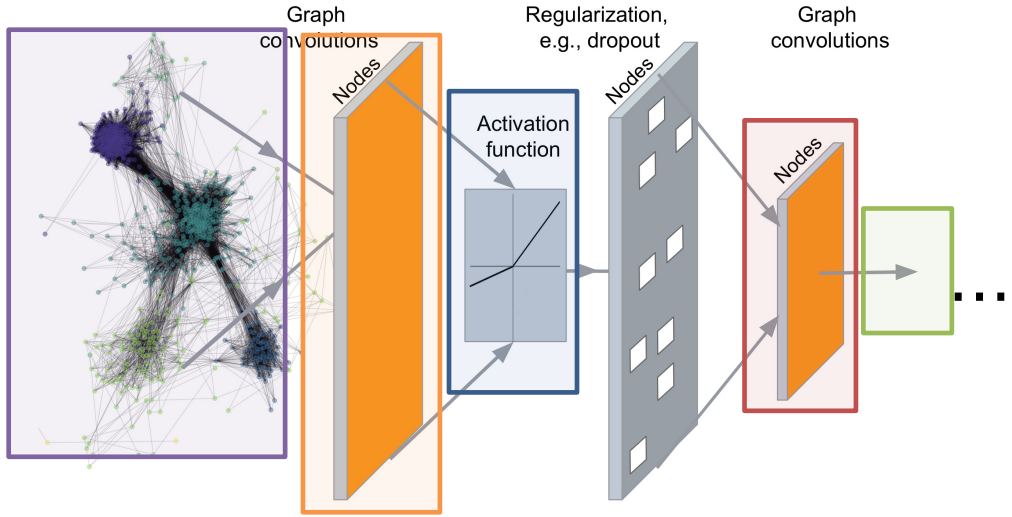



  - Model can be of arbitrary depth
    - Nodes have embeddings at each layer
    - Layer 0 embedding of node *u* is its input features $X_u$

- Basic neighborhood aggregation approach: Average information from neighbors and apply a neural network

Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Basic GNN approach




Initial 0-th layer embeddings are equal to node features

$$\mathbf{h}_v^0 = \mathbf{x}_v$$

Previous layer embedding of $v$

$$\mathbf{h}_v^k = \sigma\left(\mathbf{W}_k \sum_{u \in N(v)} \frac{\mathbf{h}_u^{k-1}}{|N(v)|} + \mathbf{B}_k \mathbf{h}_v^{k-1}\right), \quad \forall k \in \{1, ..., K\}$$

Average of neighbor's previous layer embeddings

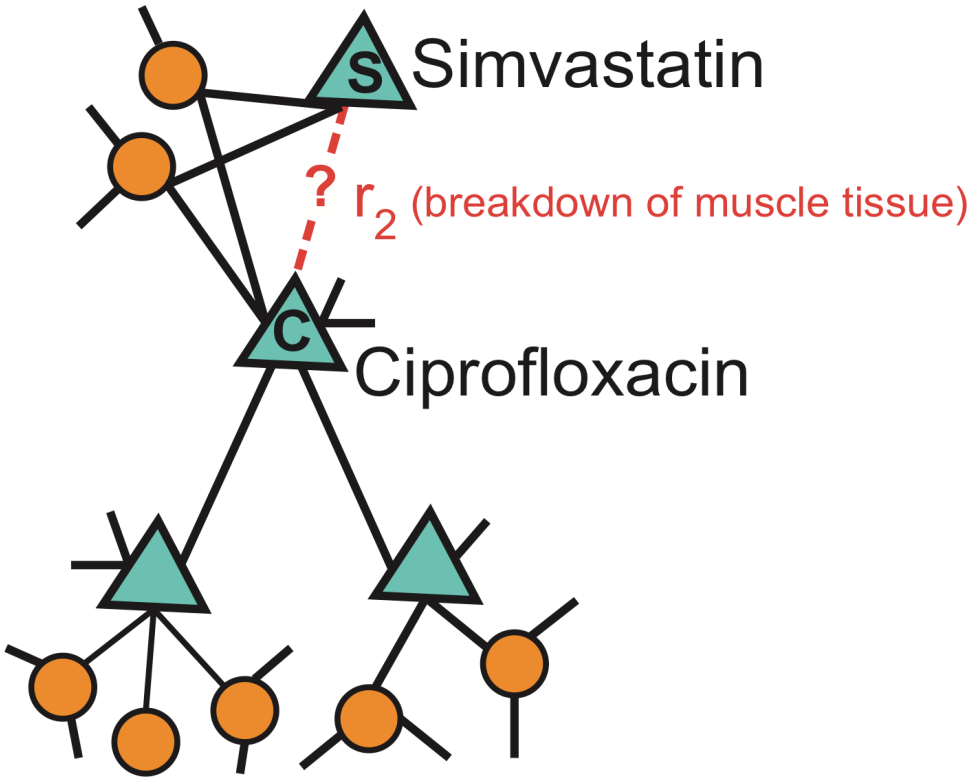
$$\mathbf{z}_v = \mathbf{h}_v^K$$

Embedding after K layers of neighborhood aggregation

Non-linearity (e.g., ReLU)

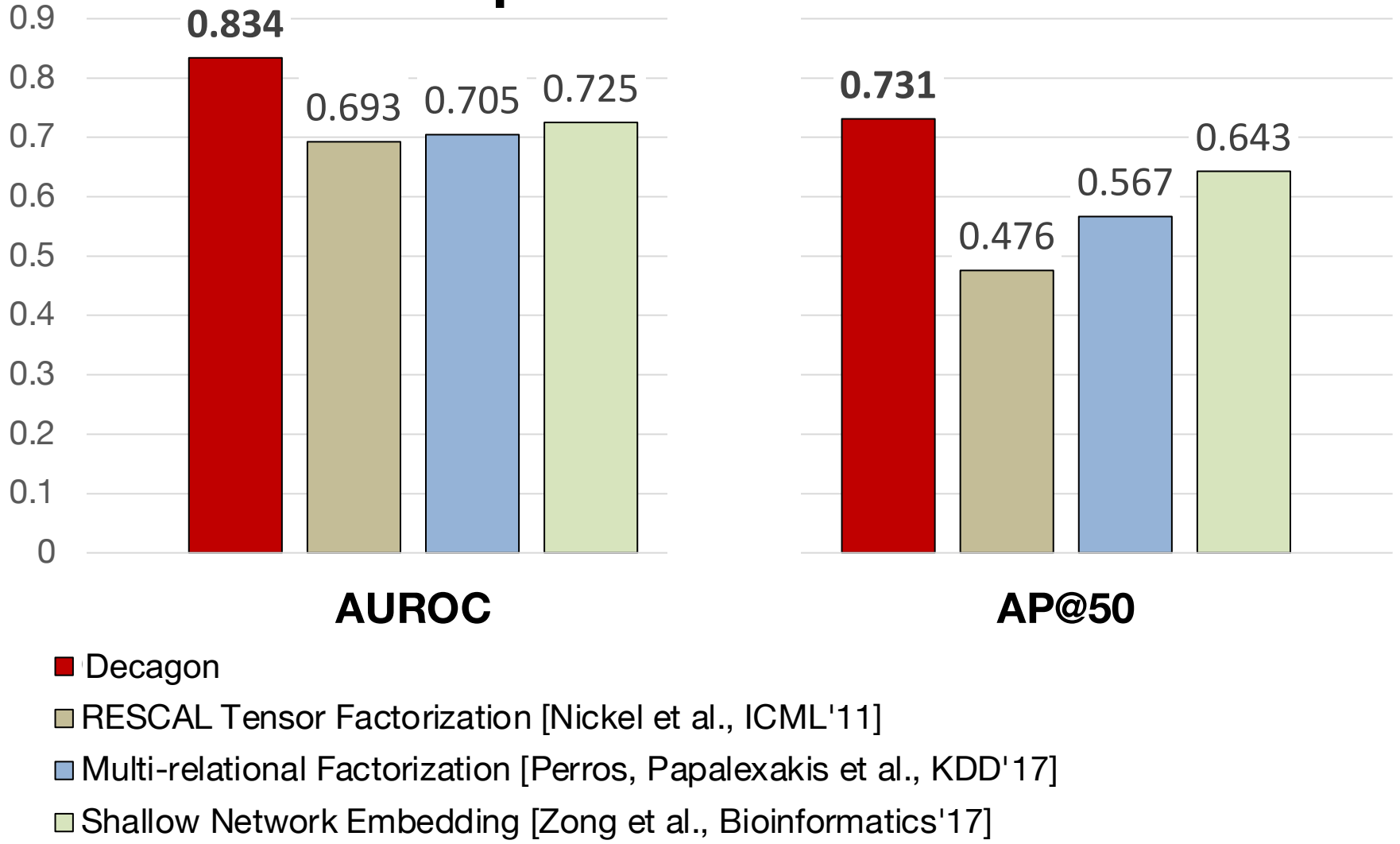
# Apply Decagon's GNN to the polypharmacy dataset

E.g.: How likely will Simvastatin and Ciprofloxacin, when taken together, break down muscle tissue?




Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Results: Polypharmacy side effect prediction

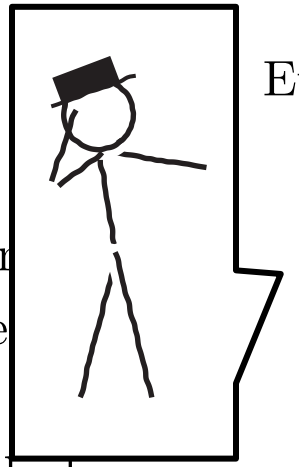



Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics,* 2018

# Polypharmacy side effect prediction

Approach:

1) Train deep model on data generated **prior to 2012**
2) How many **predictions** have been **confirmed after 2012**?



| Rank | Drug | Drug | Side effect | Evidence found |
|---|---|---|---|---|
| **1** | Pyrimethamine | Aliskiren | Sarcoma | |
| **2** | Tigecycline | Bimatoprost | Autonomic r | |
| **3** | Telangiectases | Omeprazole | Dacarbazine | |
| **4** | Tolcapone | Pyrimethamine | Blood brain | |
| | | | eadache | |
| | | | ular acidosis | |
| **7** | Anagr | Azelaic acid | Cerebral thrombosis | |
| **8** | Atorvastatin | Amlodipine | Muscle inflammation | |
| **9** | Aliskiren | Tioconazole | Breast inflammation | |
| **10** | Estradiol | Nadolol | Endometriosis | |

*Case Report*

**Severe Rhabdomyolysis due to Presumed Drug Interactions between Atorvastatin with Amlodipine and Ticagrelor**

Modeling Polypharmacy Side Effects with Graph Convolutional Networks, *Bioinformatics*, 2018

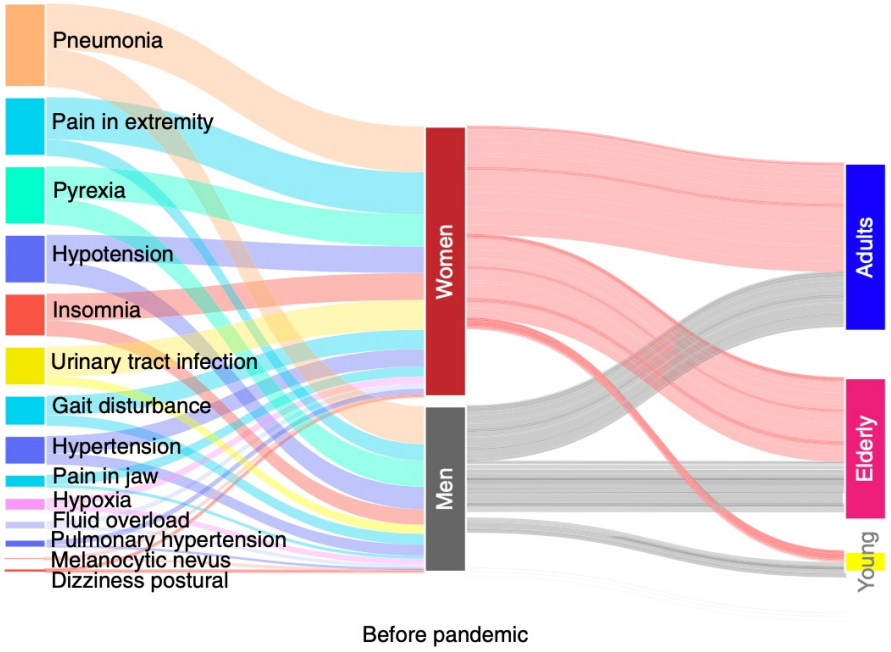
# Where do we go from here?

- Adverse events from medications accounted for over 110,000 deaths in the US alone in 2019
- It remains largely unknown:
  - How a nationwide pandemic (such as COVID-19) can influence patient safety
  - What inequalities in patients are exacerbated more than expected had the pandemic not occurred
- Dependencies between aspects of the pandemic, **drug effects,** and **patient characteristics** create additional challenges for understanding patient safety during a public health emergency

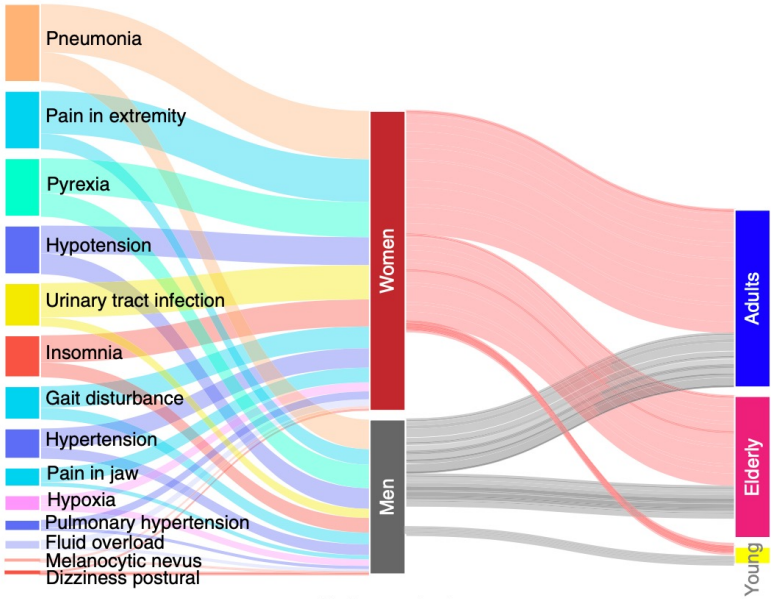
Population-scale identification of differential adverse events before and during a pandemic, *Nature Computational Science,* 2021

# Variation of adverse events across patient groups

- Substantial variation in adverse events before and during the pandemic:
  - Among 64 adverse events identified by our analyses, 54 have increased incidence rates during the pandemic, even though adverse event reporting decreased by 4.4% overall relative to 2019




Before pandemic



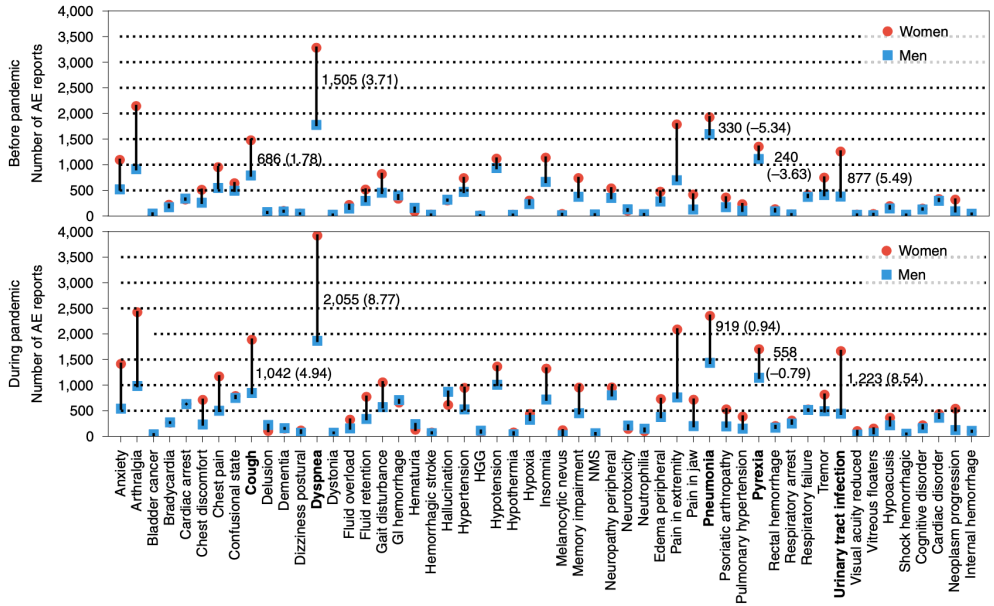

During pandemic

Population-scale identification of differential adverse events before and during a pandemic, *Nature Computational Science,* 2021

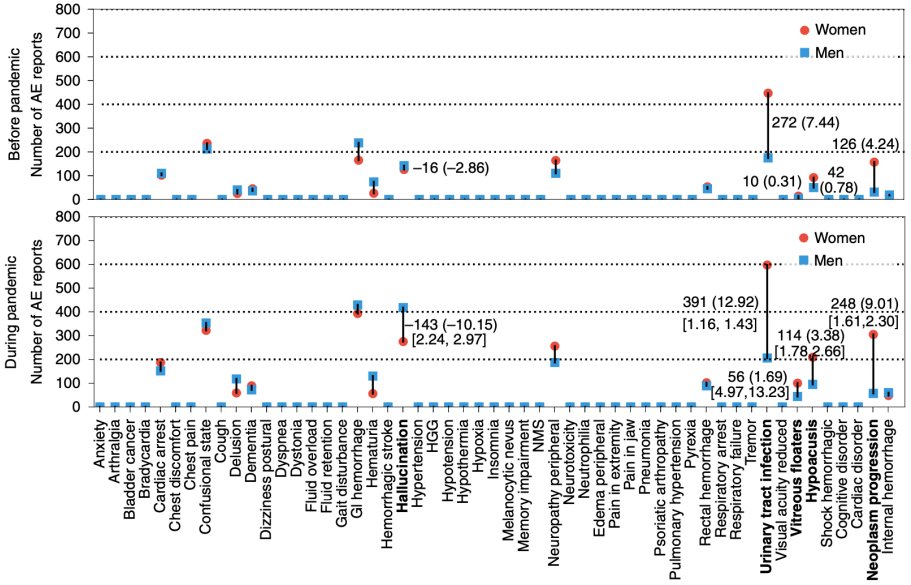
# Variation of adverse events across patient groups

- Adverse events whose reporting frequency has changed relative to pre-pandemic levels tend to be reported considerably more often than expected:
  - Pre-pandemic gender differences are exaggerated during the pandemic
  - Women suffer from more adverse events than men relative to pre-pandemic, across all ages
  - Anxiety and insomnia were disproportionately increased in women and elderly

All patients




Elderly



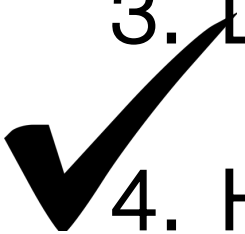

Population-scale identification of differential adverse events before and during a pandemic, *Nature Computational Science,* 2021

# Outline for today's class

1. AI/ML for precision medicine

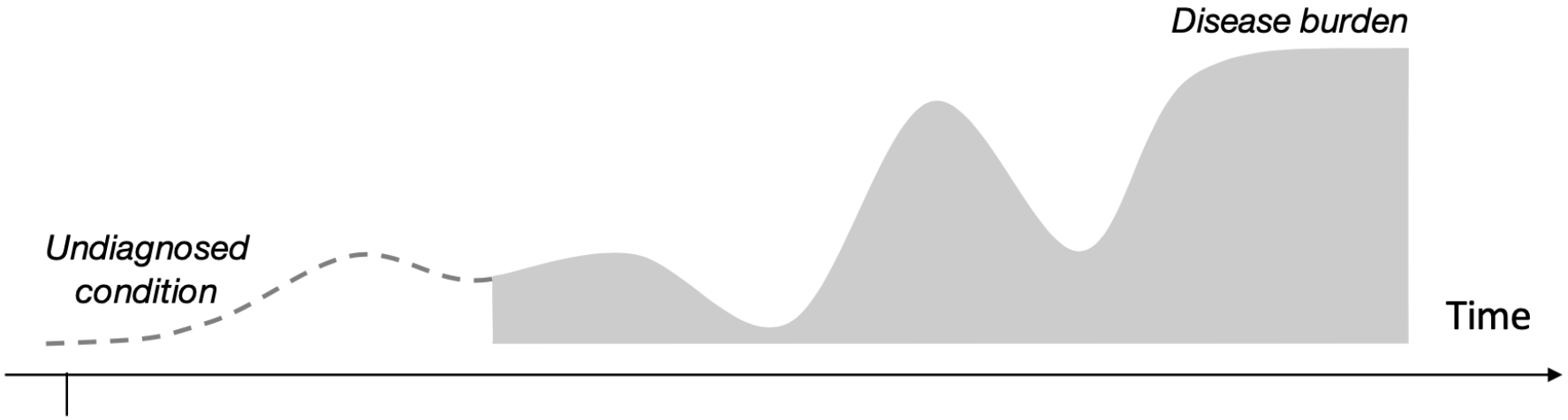
2. What are EHR data useful for?

3. Limitations & biases of EHR data

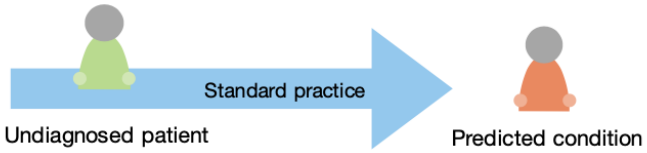


4. Highlights of ML on EHR data:
   - Polypharmacy and adverse drug events
   - Modeling disease progression

# **Prognosis:** Where is a patient in their disease trajectory? When will the disease progress? How will treatment affect disease progression?



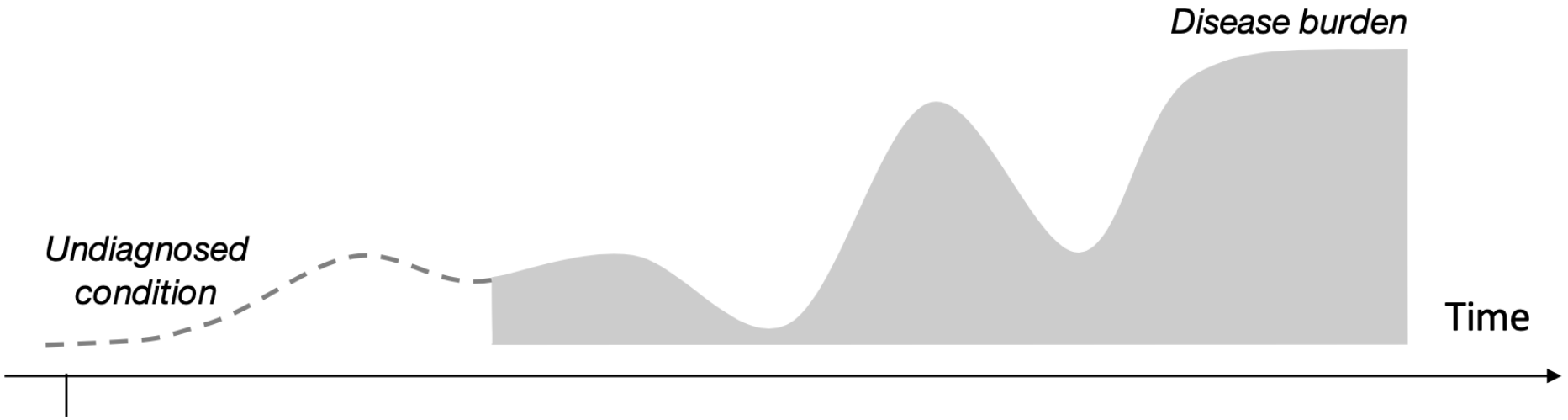

Predicted risk of developing disease or predicting outcome




**Example:** Multiple myeloma

- Rare blood cancer
- MMRF CoMMpass Study has ~1000 patients

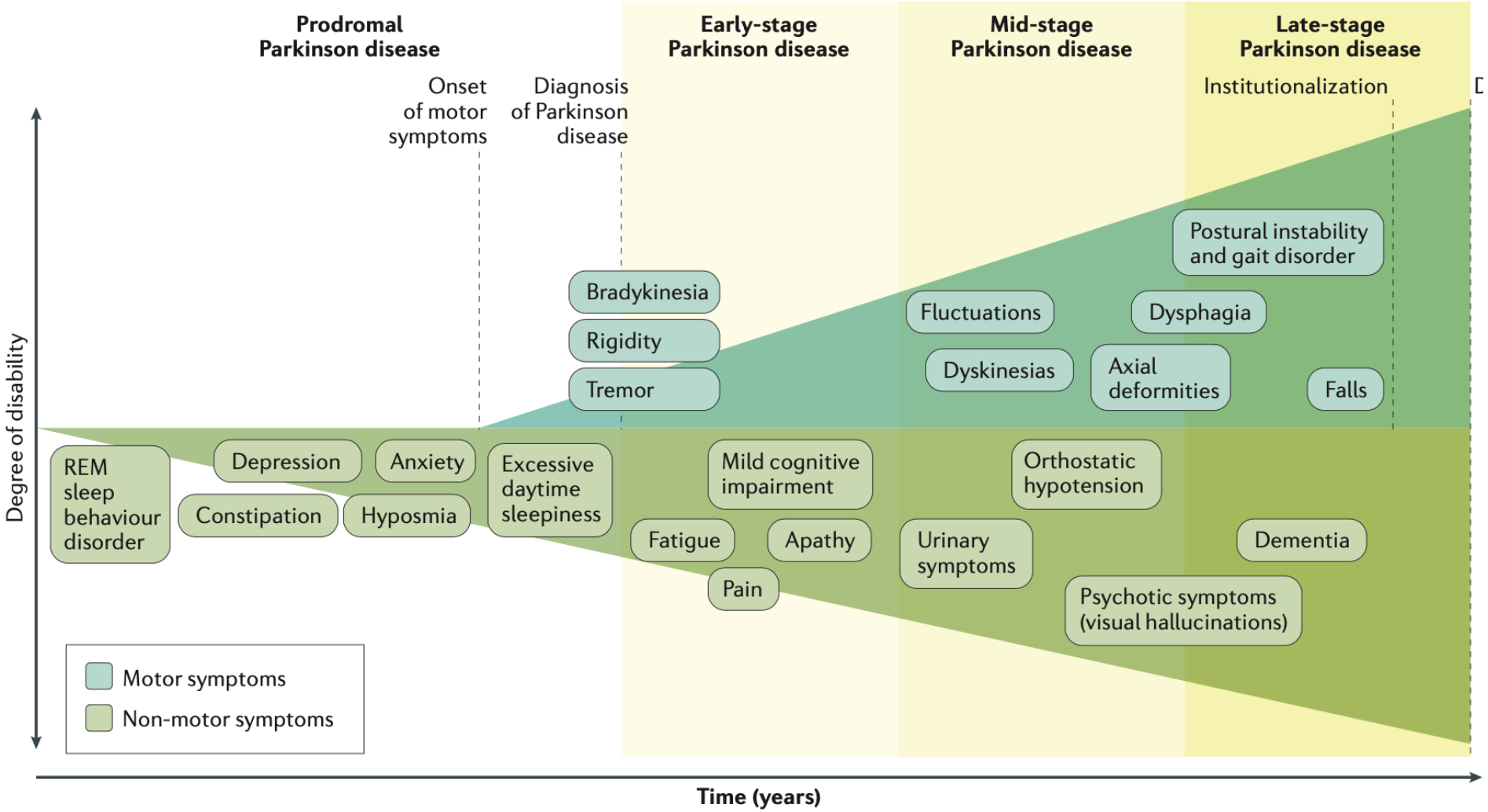
# **Descriptive:** What does a typical trajectory look like?




**Example:** Parkinson's

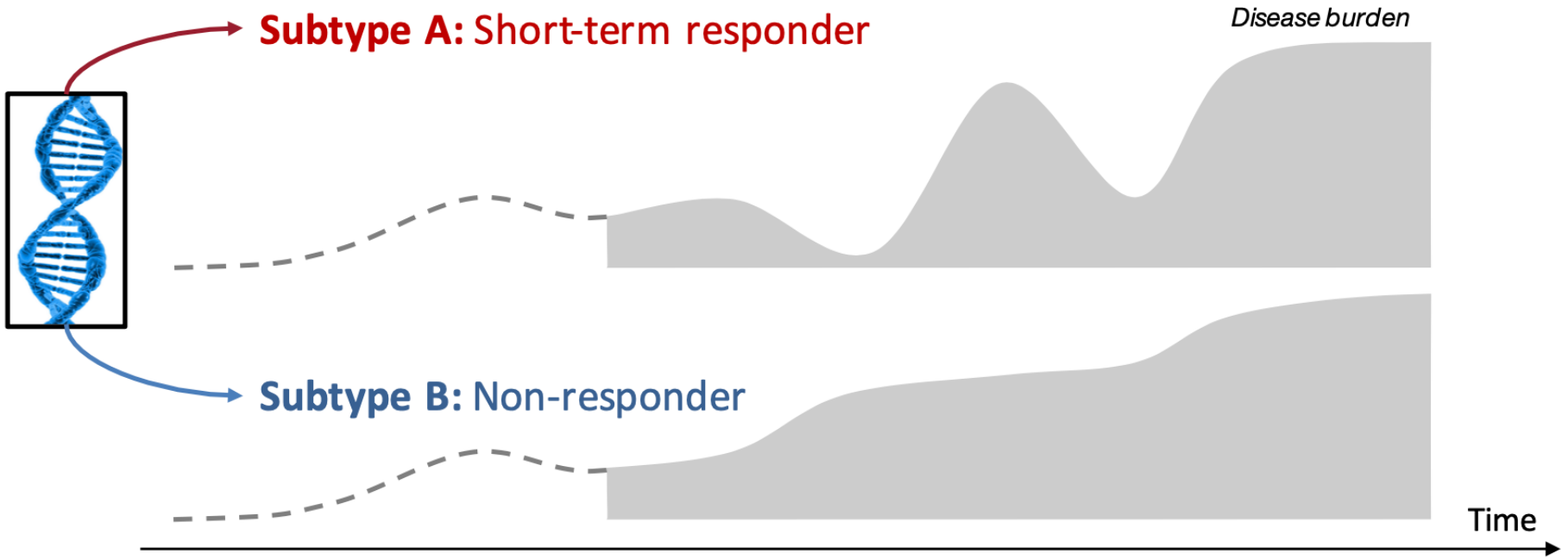
- Progressive nervous system disorder
- Affects 1 in 100 people over age 60
- PPMI dataset follows patients across time

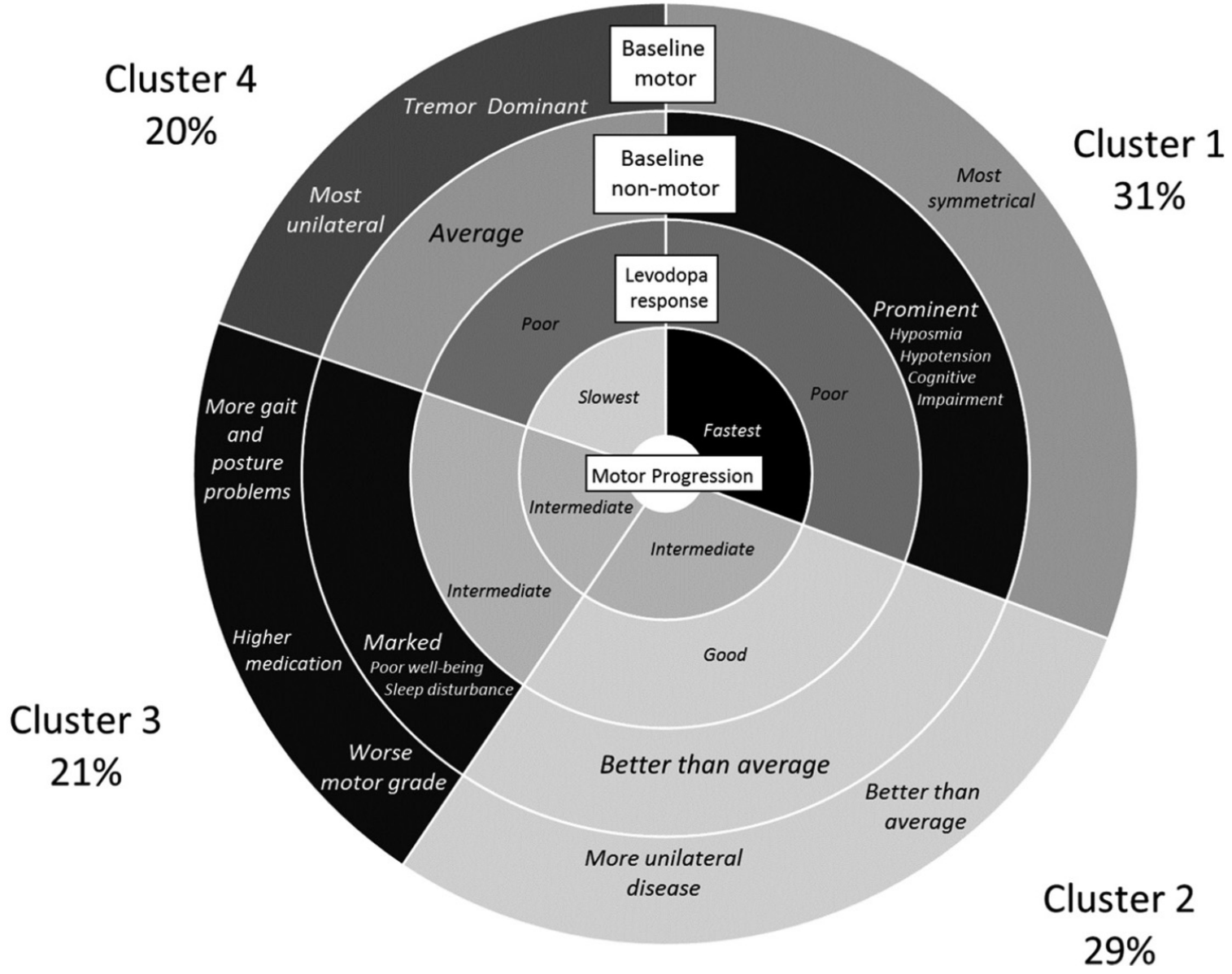
# Clinical symptoms associated with Parkinson's disease progression




[Poewe et al., Parkinson's disease. Nature Reviews Disease Primers, 2017]

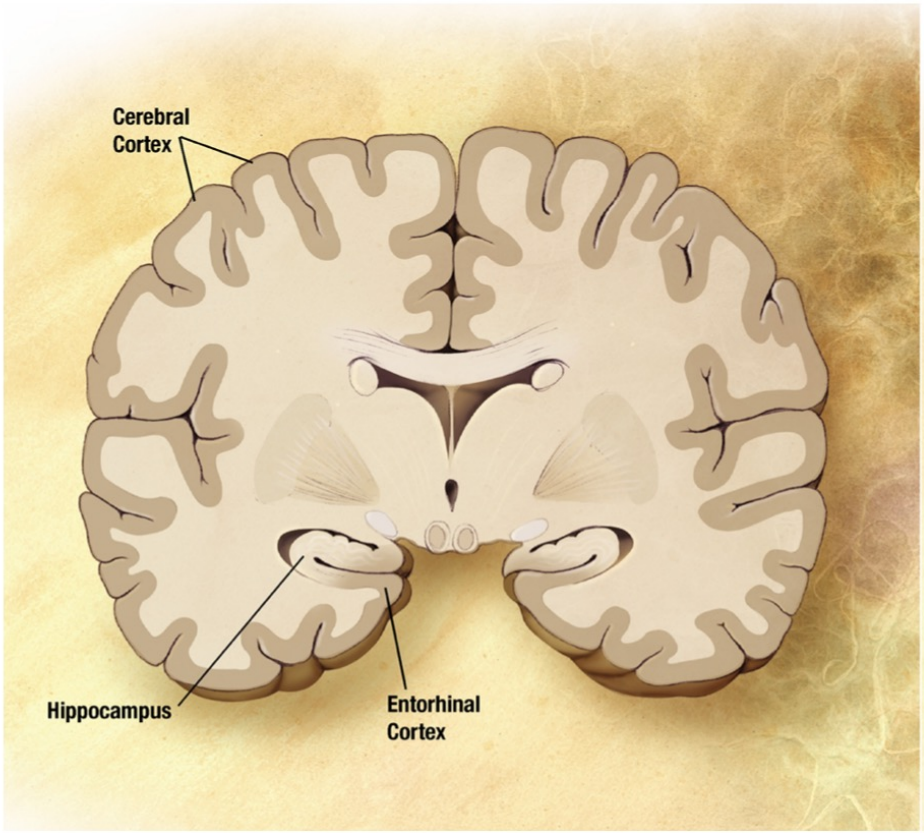
# **Subtyping:** Can we re-define the disease altogether?

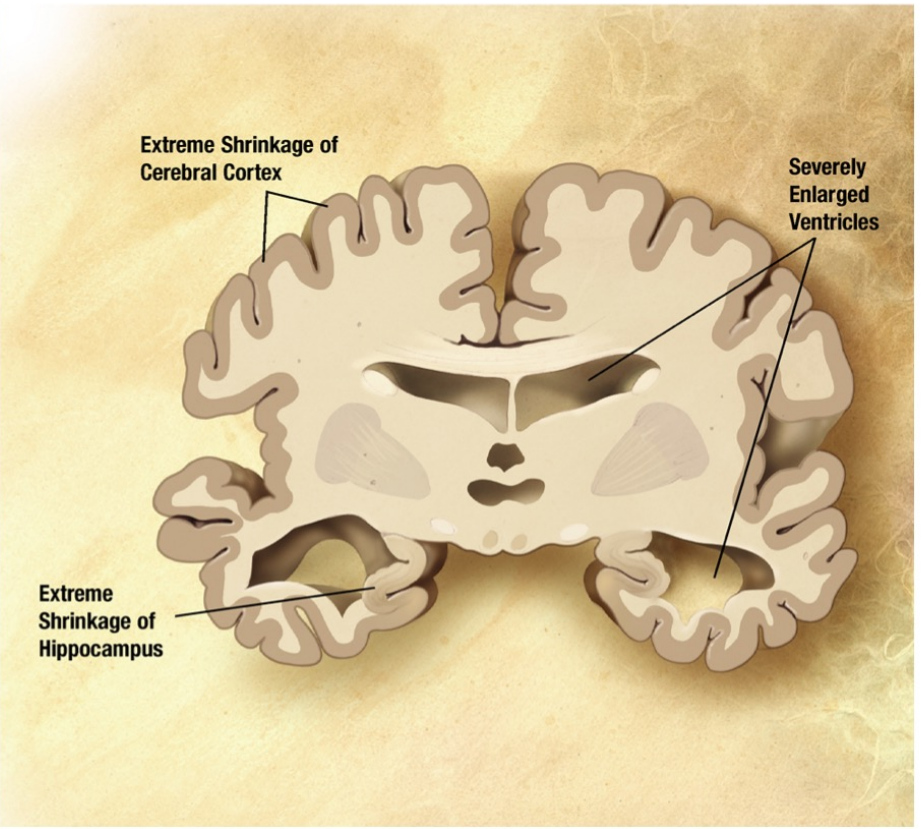





[Lawton et al., Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J Neurol Neurosurg Psychiatry*, 2018]

# Predicting disease progression in Alzheimer's disease




[Image credit: Wikipedia; "Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging."]

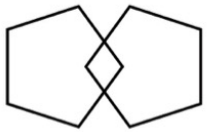
Disease status quantified by cognitive score (continuous valued)

# MINI MENTAL STATE EXAMINATION (MMSE)

Name:

DOB:

Hospital Number:

| One point for each answer | DATE: | | |
|---|---|---|---|
| **ORIENTATION** Year Season Month Date Time | ...../ 5 | ...../ 5 | ...../ 5 |
| Country Town District Hospital Ward/Floor | ...../ 5 | ...../ 5 | ...../ 5 |
| **REGISTRATION** Examiner names three objects (e.g. apple, table, penny) and asks the patient to repeat (1 point for each correct. THEN the patient learns the 3 names repeating until correct). | ...../ 3 | ...../ 3 | ...../ 3 |
| **ATTENTION AND CALCULATION** Subtract 7 from 100, then repeat from result. Continue five times: 100, 93, 86, 79, 65. (Alternative: spell "WORLD" backwards: DLROW). | ...../ 5 | ...../ 5 | ...../ 5 |
| **RECALL** Ask for the names of the three objects learned earlier. | ...../ 3 | ...../ 3 | ...../ 3 |
| **LANGUAGE** Name two objects (e.g. pen, watch). | ...../ 2 | ...../ 2 | ...../ 2 |
| Repeat "No ifs, ands, or buts". | ...../ 1 | ...../ 1 | ...../ 1 |
| Give a three-stage command. Score 1 for each stage. (e.g. "Place index finger of right hand on your nose and then on your left ear"). | ...../ 3 | ...../ 3 | ...../ 3 |
| Ask the patient to read and obey a written command on a piece of paper. The written instruction is: "Close your eyes". | ...../ 1 | ...../ 1 | ...../ 1 |
| Ask the patient to write a sentence. Score 1 if it is sensible and has a subject and a verb. | ...../ 1 | ...../ 1 | ...../ 1 |
| **COPYING:** Ask the patient to copy a pair of intersecting pentagons  | ...../ 1 | ...../ 1 | ...../ 1 |
| **TOTAL:** | ...../ 30 | ...../ 30 | ...../ 30 |

**MMSE scoring**
24-30: no cognitive impairment
18-23: mild cognitive impairment
0-17: severe cognitive impairment

OME Oxford Medical Education

# Patient dataset: 371 features

MRI scans (white matter parcellation volume, etc.) +

| | |
|---|---|
| Demographic | age, years of education, gender |
| Genetic | ApoE-$\varepsilon$4 information |
| Baseline cognitive scores | MMSE, ADAS-Cog, ADAS-MOD, ADAS sub-scores, CDR, FAQ, GDS, Hachinski, Neuropsychological Battery, WMS-R Logical Memory |
| Lab tests | RCT1, RCT11, RCT12, RCT13, RCT14, RCT1407, RCT1408, RCT183, RCT19, RCT20, RCT29, RCT3, RCT392, RCT4, RCT5, RCT6, RCT8 |

Zhou et al., KDD'12

# Progression of Alzheimer's

- **Goal:** Predict disease status in 6, 12, 24, 36, and 48 months
- Five different regression tasks?
- Challenge: **data sparsity**
  - Total number of patients is small
  - Labels are noisy
  - Due to censoring, fewer patients at later time points

Zhou et al., KDD'12

# Predicting disease progression in Alzheimer's disease

- **Goal:** Predict disease status in 6, 12, 24, 36, and 48 months
- **Approach:**
  - Five regression tasks: M06, M12, M24, M36, M48?
- Challenge: **Small sample size**

**Number of patients M months after baseline (Alzheimer's Disease Neuroimaging Initiative)**

| M06 | M12 | M24 | M36 | M48 |
|---|---|---|---|---|
| 648 | 642 | 569 | 389 | 87 |

**M06 = 6 months after baseline**

Zhou et al., KDD'12

# Approach: Multi-task learning

- **Goal:** Predict disease status in 6, 12, 24, 36, and 48 months
- Rather than learning 5 independent models, we can formulate the problem as multi-task learning:
  - Select a common set of biomarkers for all time points
  - Allow for specific set of biomarkers at different time points → candidate disease state biomarkers
  - Encourage temporal smoothness in models when making predictions for neighboring time points

Zhou et al., KDD'12

# Approach: Fused sparse group lasso

- Simultaneously learn all 5 models by solving the following optimization problem:

Feature importance values: Weight matrix that we want to learn

$$\min_W L(W) + \lambda_1 \|W\|_1 + \lambda_2 \left\|RW^T\right\|_1 + \lambda_3 \|W\|_{2,1}$$

- Squared loss: $L(W) = \|S \odot (XW - Y)\|_F^2$
  (*S* is a mask to account for labels missing in some)

Matrix of patient features, demographics, genetics, cognitive scores, lab tests

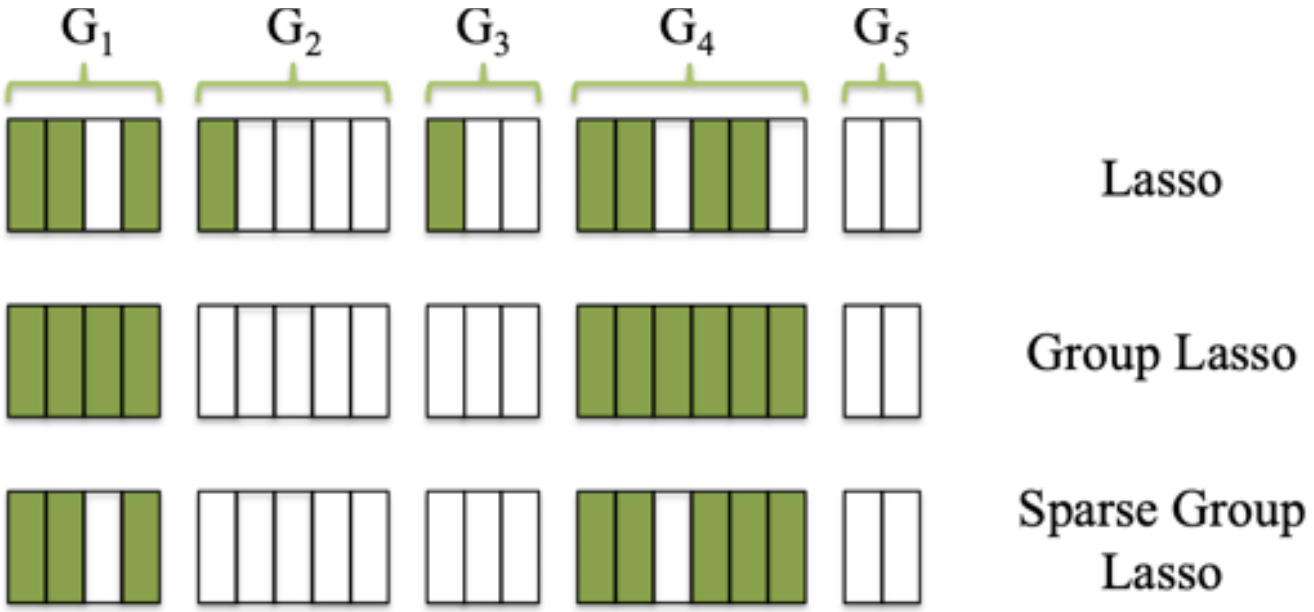
Ground-truth outcomes

- Group Lasso penalty $\|W\|_{2,1}$ given by $\sum_{i=1}^{d} \sqrt{\sum_{j=1}^{t} W_{ij}^2}$
- R =

| | 5 |
|---|---|
| 4 | 1 -1<br>1 -1<br>1 -1 |

Zhou et al., KDD'12

# Approach: Sparse group lasso




- "Fused" version of lasso **penalizes the norm of both the coefficients and their successive differences**
  - It encourages sparsity of the coefficients and sparsity of their differences—local constancy of the coefficient profile

Zhou et al., KDD'12

# Averaged results across five time points

Baseline – independent regressors

Temporal smoothing helps!

| | Ridge | cFSGL1 ($\lambda_2 = 20$) | cFSGL2 ($\lambda_2 = 50$) | cFSGL3 ($\lambda_2 = 100$) |
|---|---|---|---|---|
| | | Target: MMSE | | |
| nMSE | $0.548 \pm 0.057$ | $0.428 \pm 0.052$ | $0.400 \pm 0.053$ | $\mathbf{0.395 \pm 0.052}$ |
| R | $0.689 \pm 0.030$ | $0.772 \pm 0.030$ | $0.790 \pm 0.032$ | $\mathbf{0.796 \pm 0.031}$ |

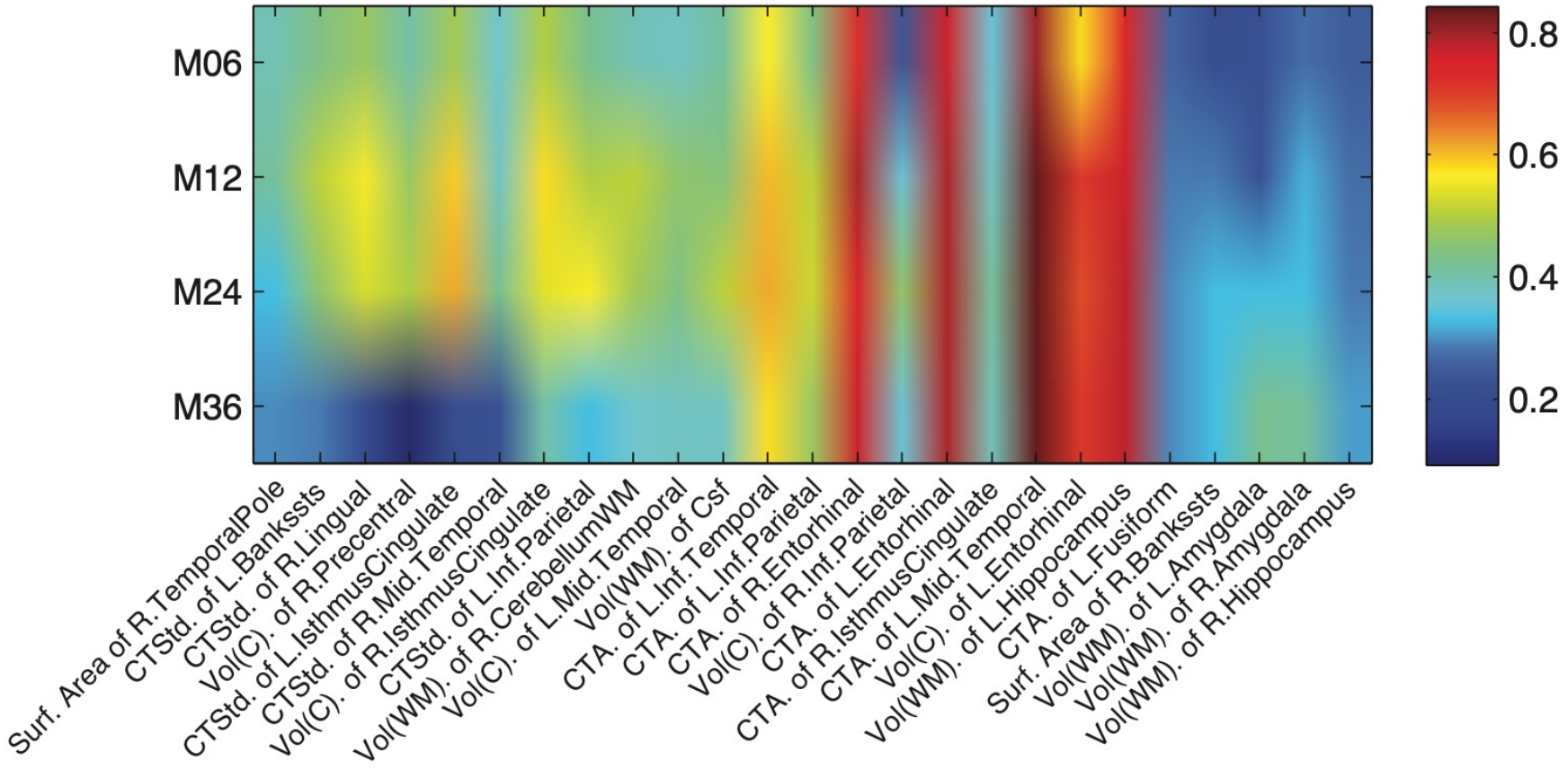
nMSE – normalized mean squared error. Smaller is better

R – average $R^2$ (correlation coefficient). Larger is better

$$\min_W L(W) + \lambda_1 \|W\|_1 + \lambda_2 \left\|RW^T\right\|_1 + \lambda_3 \|W\|_{2,1}$$

Zhou et al., KDD'12

# Predictive importance of features vary across time




(a) Target: ADAS-Cog (25 stable features)

Zhou et al., KDD'12

# Outline for today's class

1. AI/ML for precision medicine
2. What are EHR data useful for?
3. Limitations & biases of EHR data
4. Highlights of ML on EHR data:
   - Polypharmacy and adverse drug events
   - Modeling disease progression