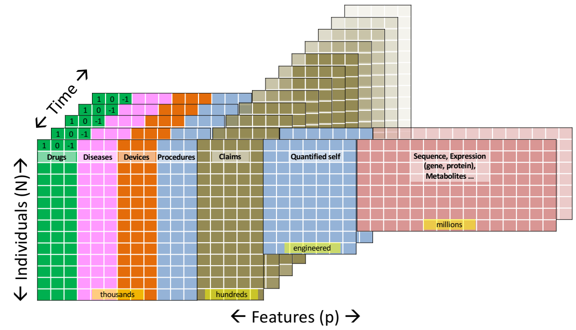
EMR Data mining

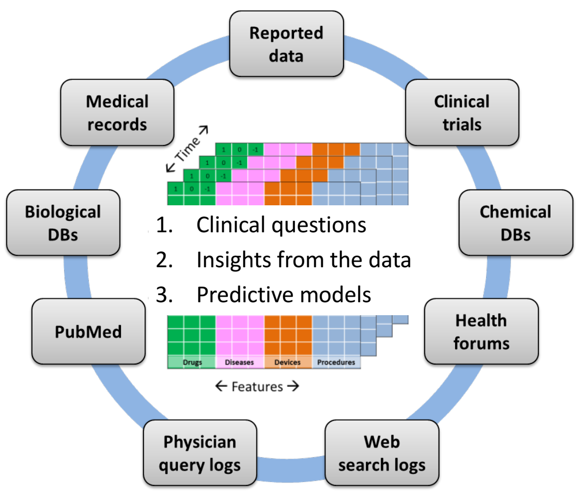
Nigam Shah
nigam@stanford.edu

STANFORD
SCHOOL OF MEDICINE

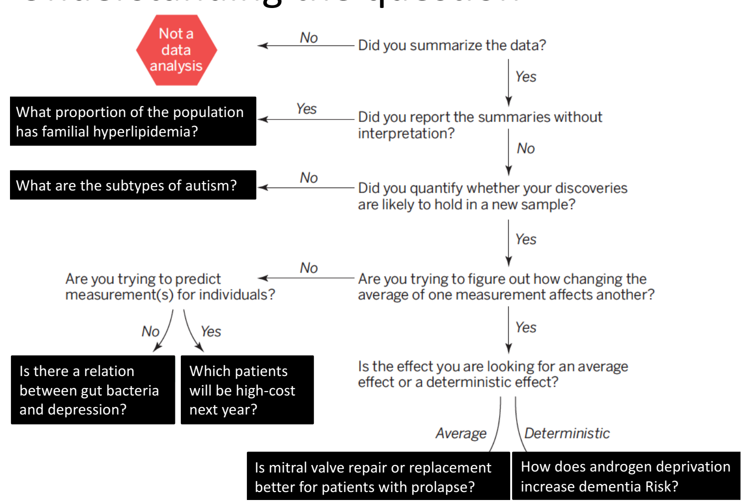
The Data and the problem

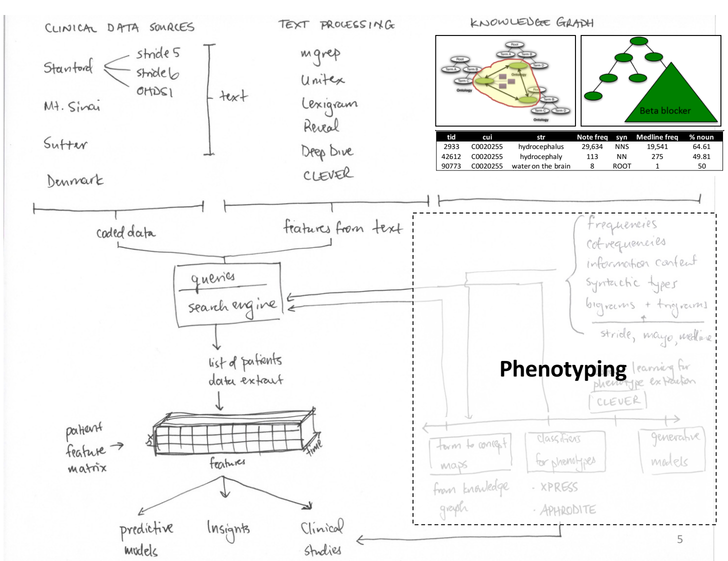


We do a lot, without knowing what works.



Understanding the question

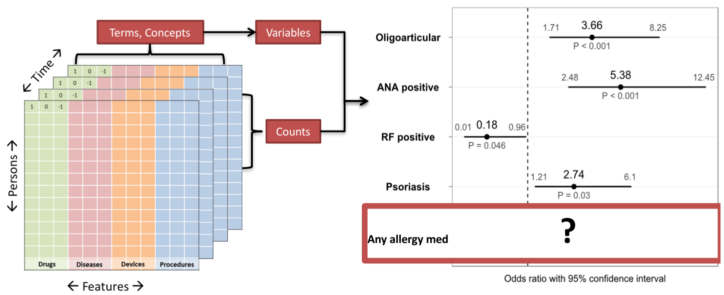


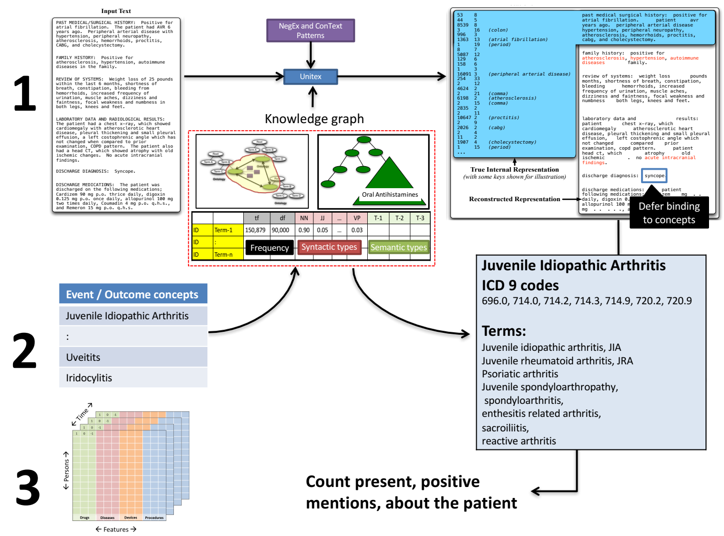


tid	cui	str	Note freq	syn	Medline freq	% noun
2933	C0020255	hydrocephalus	29,634	NNS	19,541	64.61
42612	C0020255	hydrocephaly	113	NN	275	49.81
90773	C0020255	water on the brain	8	ROOT	1	50

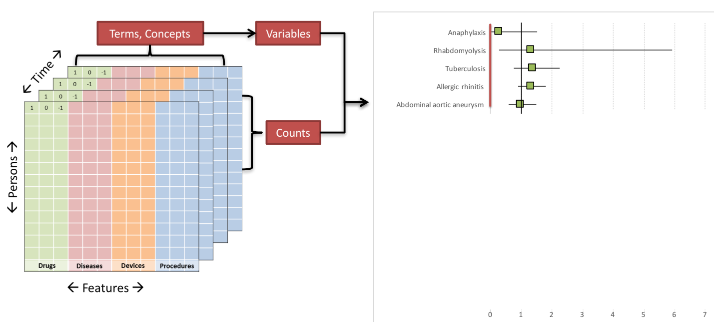
Clinical studies

Profiling risk factors for chronic uveitis



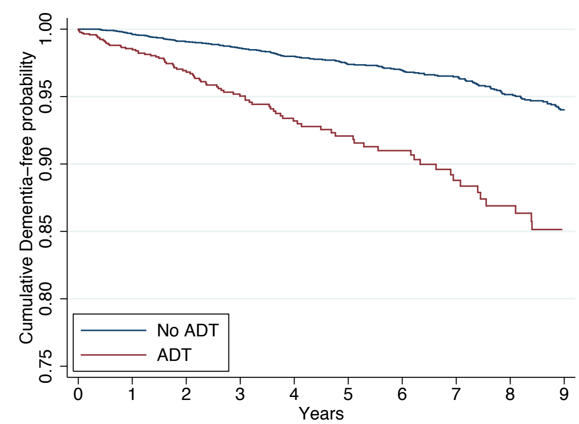


Androgen deprivation & Alzheimer's risk



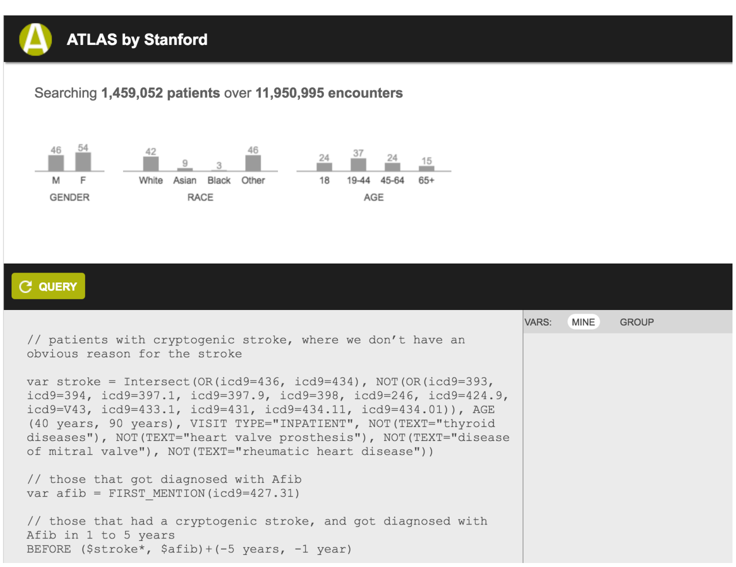
www.tinyurl.com/JCO-ADT

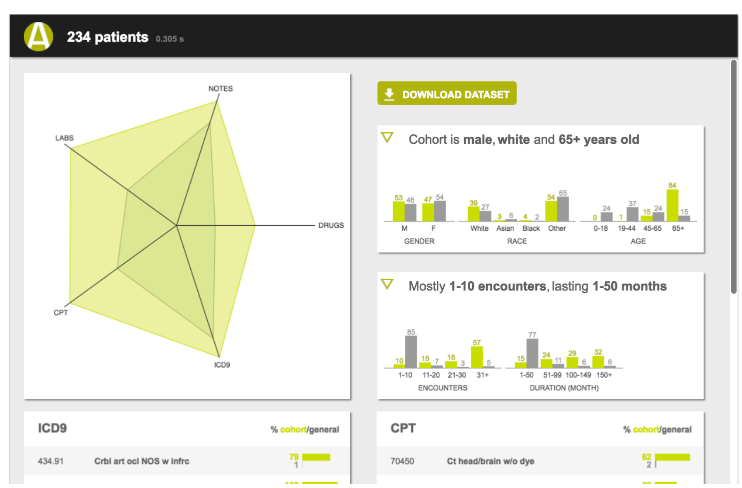
Androgen deprivation & Dementia risk



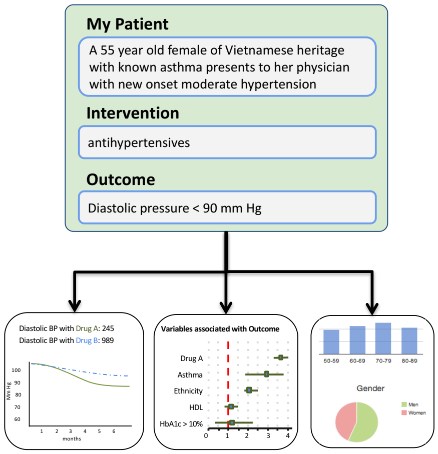
Number at risk

No ADT	7446	5448	4587	3889	3231	2708	2233	1833	1518	1214
ADT	1826	1177	872	645	470	367	282	211	160	124



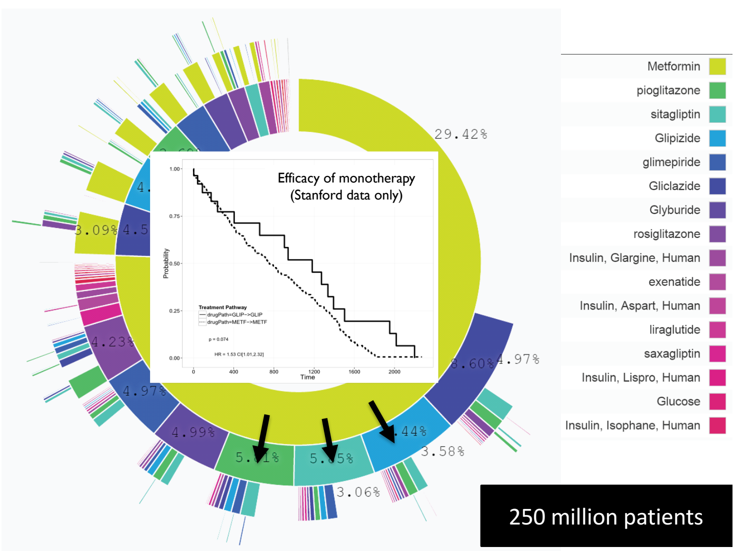


http://tinyurl.com/inf-consult

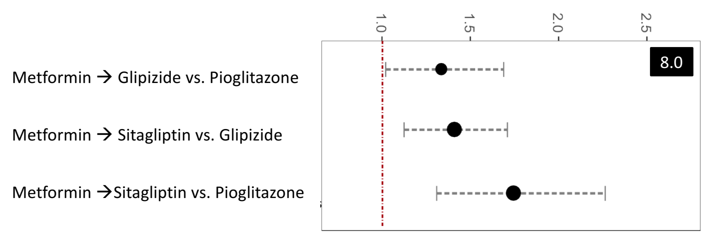


Example – 1: Choosing diabetes drugs

Scenario: Which second line drug to use for treating diabetics who have high HbA1c one to two months after first line treatment?



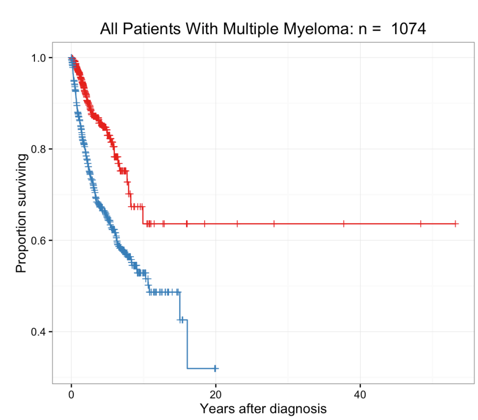
Effective treatment pathways



Example – 2: Choosing chemotherapy

Scenario: For 55-60 year old white male patient with newly diagnosed plasma cell leukemia (PCL), what is the difference in overall survival between patients treated with intensive versus less intensive chemotherapy?

Example – 2: Personalized estimate



Outline of an informatics consult

Descriptive summary

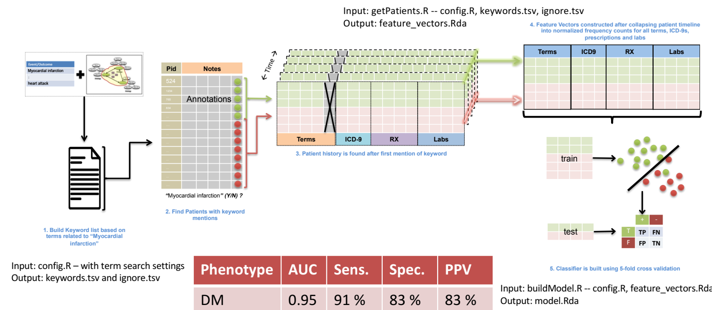
- What happened after treatment?

Making recommendations

- What treatment choices are typically made, given prior medical history? What are typical outcomes?

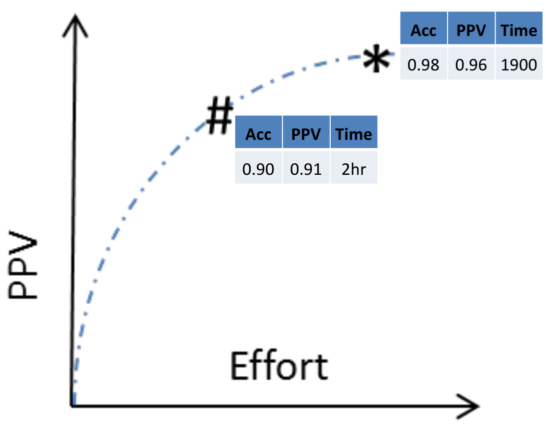
- Estimation: What is the effect of treatment choice X on outcome Y?

XPRESS- EXtraction of Phenotypes from clinical Records using Silver Standards



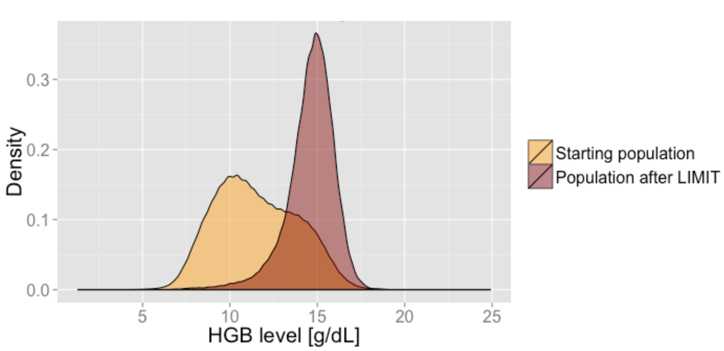
Phenotype	AUC	Sens.	Spec.	PPV
DM	0.95	91 %	83 %	83 %
MI	0.91	89 %	91 %	91 %
FH	0.90	76.5%	93.6%	~20%
Celiac	0.75	40 %	90 %	~4 %

Phenotyping - effort precision trade off

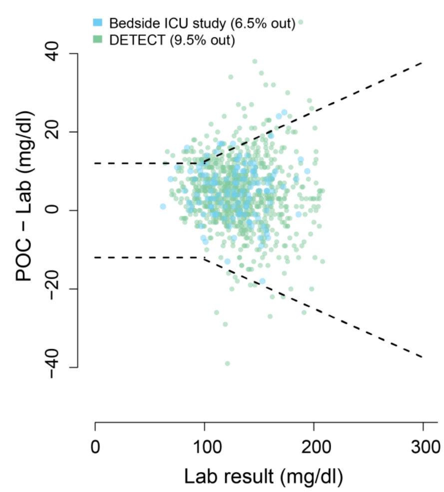


Insights

Learning true ranges of lab tests

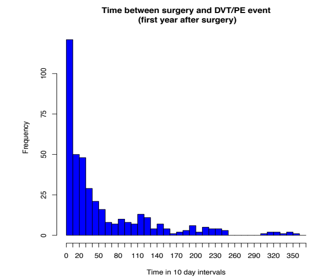


DETECT: Data mining EMRs To Evaluate Coincident Testing

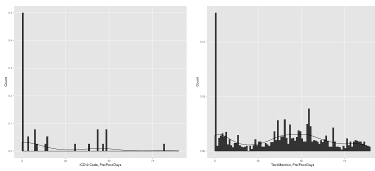


Quality metrics

- Post surgery DVT rates

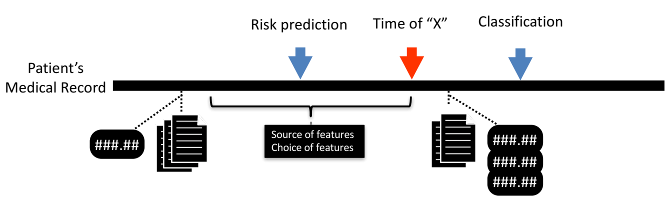


- Urinary Incontinence after prostate cancer surgery

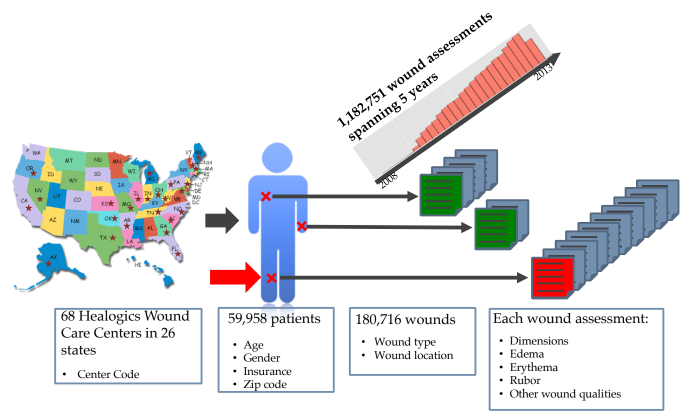


Predictive Modeling

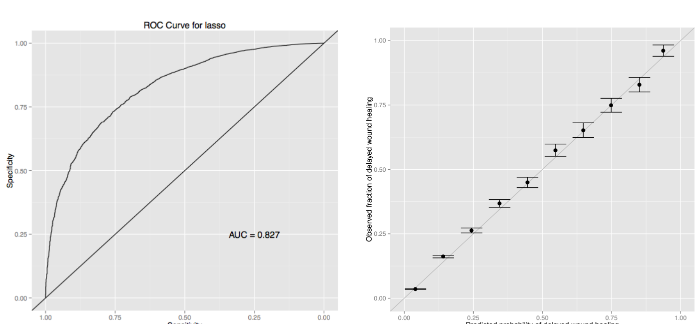
Classification or Prediction?



Predicting delayed healing wounds



Clinically usable performance



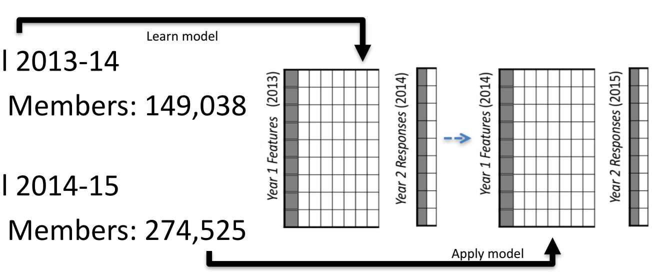
Predicting Cost Blooms

Panel 2013-14

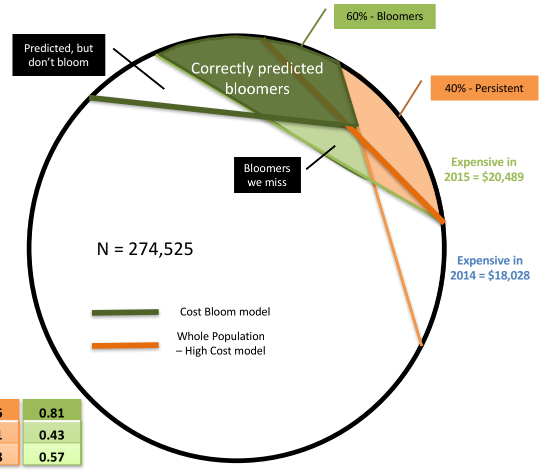
Total Members: 149,038

Panel 2014-15

Total Members: 274,525



Predicting Cost Blooms



AUC	0.86	0.81
PPV	0.51	0.43
Cost Cap.	0.68	0.57

Fruitful areas of activity

- Risk stratification: cost, latent disease, decompensation
- Personalizing evidence: risk for me, what treatment will work
- Insights into disease progression
- Using passively collected data: sensors, feature engineering
- Practice management: predicting missed appointments, medication adherence ...

Open research problems

- Data nonstationarity
- Local vs. Global models
- Handling unstructured data (text, images, time traces)
- Outcome ascertainment (and censoring)
- Evaluation: Beyond discrimination
  - calibration, net-reclassification
- Bridging the "last mile"

Acknowledgements

Group Members:

- **Scientists:** Suzanne Tamang, Ken Jung, Alison Callahan, Juan Banda
- **Fellows**: Rainer Winnenburg, Rohit Vashisht
- **Engineer:** Vladimir Polony
- **BMI Students:** Sarah Poole, Alejandro Schuler, Albee Ling, Vibhu Agarwal, David Stark
- **Med Students:** Greg Gaskin, Jassi Pannu

Alums: Anna Bauer-Mehren (Roche), Srini Iyer (Facebook), Amogh Vasekar (Citrix), Sandy Huang (Berkeley), Paea LePendu (Lexigram), Rave Harpaz (Oracle), Tyler Cole (Barrow Inst.), Sam Finlayson (Harvard), Will Chen (Yale), Yen Low (Netflix), Elsie Gyang (Fellowship in Surgery)

Funding:

- NIH – NLM, NIGMS, NHGRI, NINDS, NCI, FDA
- Stanford Internal – Dept. of Medicine, Population Health Sciences, Clinical Excellence Research Center, Dean's office
- Fellowships – Med Scholars, Siebel Scholars Foundation, Stanford Graduate Fellowship
- Industry – Apixio, CollabRx, Healogics, Janssen R&D, Oracle, Baidu USA, Amgen

IT: Alex Skrenchuk, SCCI team