EMR Data mining

Nigam Shah
nigam@stanford.edu

STANFORD
SCHOOL OF MEDICINE

The Data and the problem

Understanding the question

Clinical questions

Insights from the data

Predictive models

Not a data analysis

What proportion of the population has familial hyperlipidemia?

What are the subtypes of autism?

Are you trying to predict measurement(s) for individuals?

Is there a relation between gut bacteria and depression?

Which patients will be high-cost next year?

Is mitral valve repair or replacement better for patients with prolapse?

How does androgen deprivation increase dementia risk?

We do a lot, without knowing what works.

The Data and the problem

Clinical studies

Phenotyping
Profiling risk factors for chronic uveitis

Androgen deprivation & Alzheimer’s risk

Androgen deprivation & Dementia risk

ATLAS by Stanford

Searching 1,459,052 patients over 11,950,995 encounters

// patients with cryptogenic stroke, where we don’t have an obvious reason for the stroke
var stroke = Intersect(OR(icd9=438, icd9=434), NOT(OR(icd9=393, icd9=394, icd9=397, icd9=397.1, icd9=397.9, icd9=398, icd9=465, icd9=466, icd9=468)), NOT(OR(icd9=432, icd9=431.1, icd9=431.9, icd9=439, icd9=439.1, icd9=439.9, icd9=440, icd9=440.1, icd9=440.9)));

// those that got diagnosed with AFib
var afib = FIRST_MENTION(icd9=427.31);

// those that had a cryptogenic stroke, and got diagnosed with AFib in 1 to 5 years
BEFORE (stroke, afib)+(-5 years, -1 year)
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?

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?
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- EXtration of Phenotypes from clinical Records using Silver Standards

Phenotyping - effort precision trade off

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

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 Winneburg, Rohit Vashishth
- Engineer: Vladimir Polony
- BMI Students: Sarah Poole, Aleksandra Schuler, Albee Ling, Vibhu Agarwal, David Stark
- Med Students: Greg Gaskin, Jassi Pannu

Alums: Anna Bauer-Mehren (Roche), Srinivas Iyer (Facebook), Anoop Vasu (Citrix), Sandy Huang (Berkeley), Pawa Lepre (maxigram), Rave Haroz (Oracle), Tyler Cole (Barrow Inst.), Sam Finlaysen (Harvard), Will Chen (Yale), Yen Low (Netflix), Eilese 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, SCO team