Does Machine Learning Have a Place In a Learning Health System?

Grand Rounds: Rethinking Clinical Research
Friday, December 15, 2017

Michael J. Pencina, PhD
Professor of Biostatistics and Bioinformatics, Duke University
Director of Biostatistics, Duke Clinical Research Institute

Duke Clinical Research Institute | FROM THOUGHT LEADERSHIP TO CLINICAL PRACTICE

How evidence spreads today

Learning Health System

NIH COLLABORATORY

LIVING TEXTBOOK

of Pragmatic Clinical Trials

DESIGN | CONDUCT | DISSEMINATION

Learning healthcare systems, as defined by the Institute of Medicine (IOM), are characterized by a number of core attributes [1]. Particularly important is a consistent emphasis on a collaborative approach that shares data and insights across boundaries to drive better, more efficient medical practice and patient care. Key to this vision is the creation of system linked by a common EHR and shared databases. This interconnected system in turn can be supported by new methods of clinical research and data analysis and would rely on modern information technology and informatics to manage and communicate data that would help guide the decisions made by health systems, care providers, and patients and their families.

In this Topic:

- The Institute of Medicine and the Learning Healthcare System Concept
- Attributes of a Learning Healthcare System
- Using Clinical Data to Drive Learning Healthcare
- Ethical and Regulatory Oversight of Learning Healthcare Systems
- Patient and Public Engagement
- EHRs, Information Technology & Information Infrastructure
- Challenges to Building the Learning Healthcare System
- Early Examples
- Bibliography

http://www.rethinkingclinicaltrials.org/resources/learning-healthcare-systems/

Everyone is doing it

“The average initial increase in profits from big data investments was 6 percent for the companies we studied. That increased to 9 percent for investments spanning five years, since the companies that made them presumably benefited from the greater diffusion of data analytics over that period.”

McKinsey Quarterly, 2016

Duke | FORGE
HEALTH DATA IN ACTION

Health Data Science at Duke: Why "Forge"?

Where art, craft, and science meet to ask:
- How does this really work?
- What tools can make it better?

A practical laboratory where prototypes are made, tested, improved, and re-worked.

Data streams combine with expertise to forge new approaches in health, systems, behavior, medicine, policy and technology to improve health.

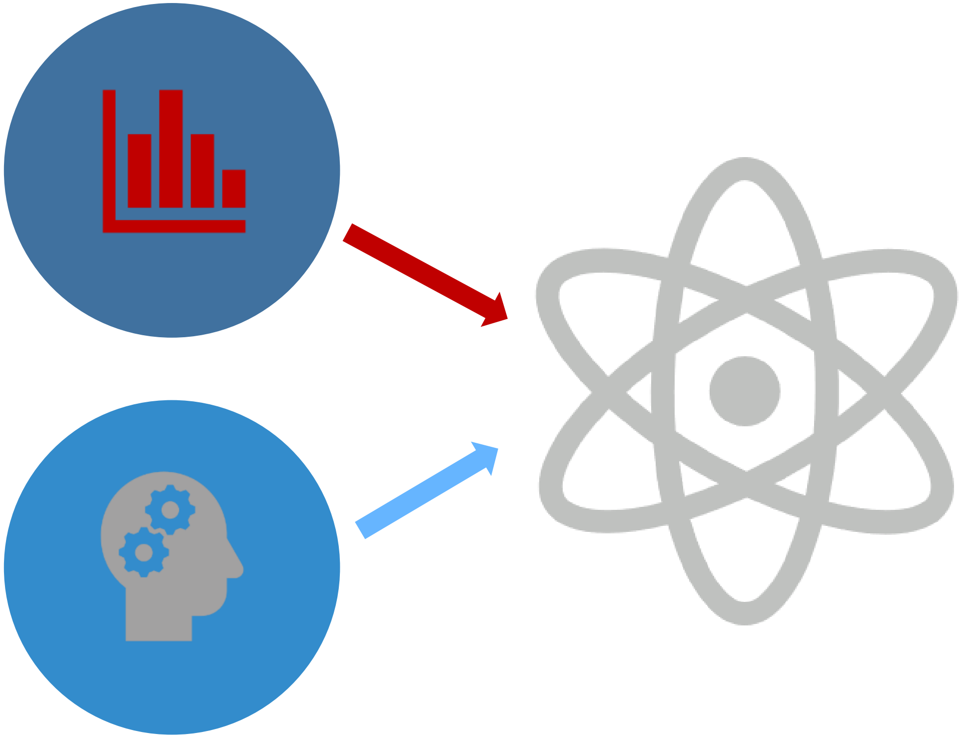
Led by **Robert M. Califf, MD,** Vice Chancellor for Health Data Science at Duke Health and part of the senior management team for Verily Health Sciences, an Alphabet company.

Grand Fusion:

Melding strengths across disciplines and between professionals

Fostering the **comprehensive toolbox across the spectrum** including frequentist statistics, Bayesian statistics, machine learning, and deep learning

Developing **the right framework for teams** including clinicians and quantitative expertise



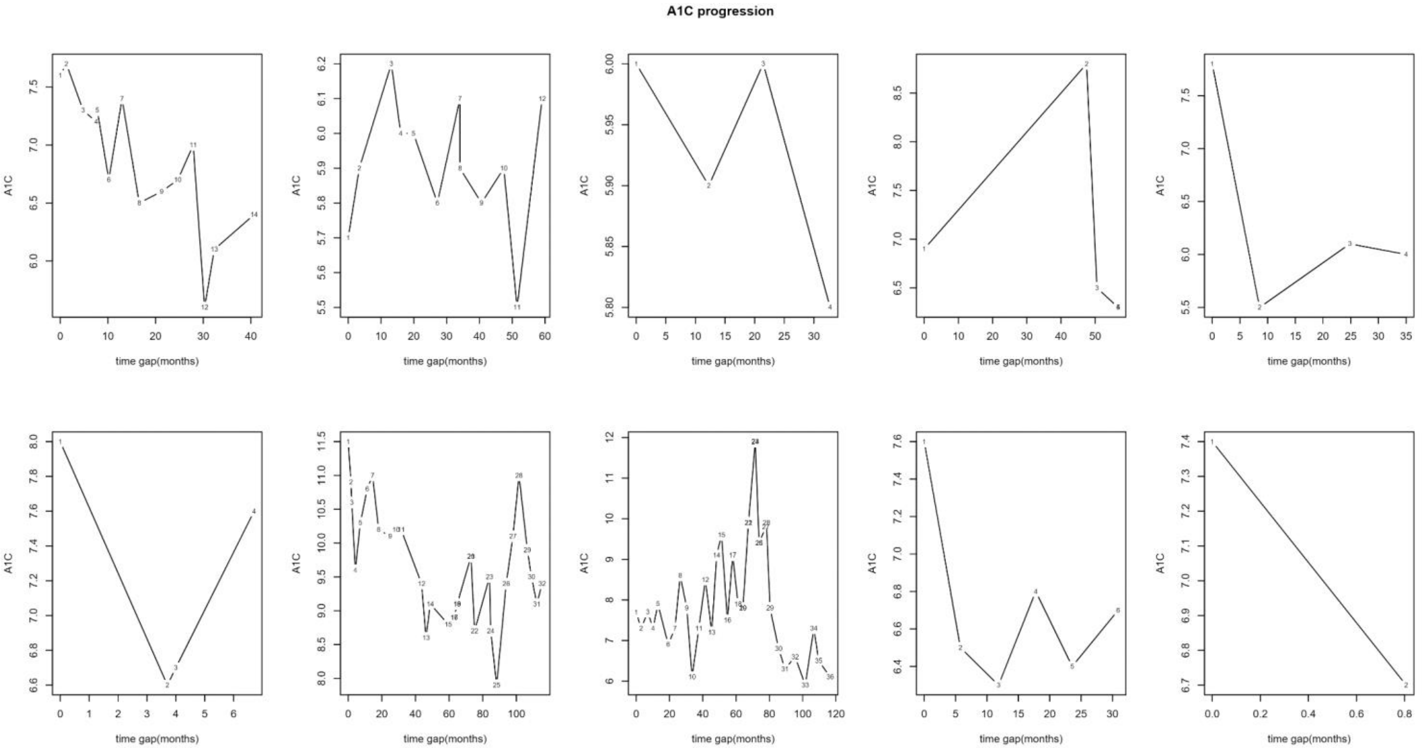
Learning Health System process

- Identify the problem
- Formulate steps to solve it
- Find the right data and perform analysis
- Test the proposed solution
- Implement or modify

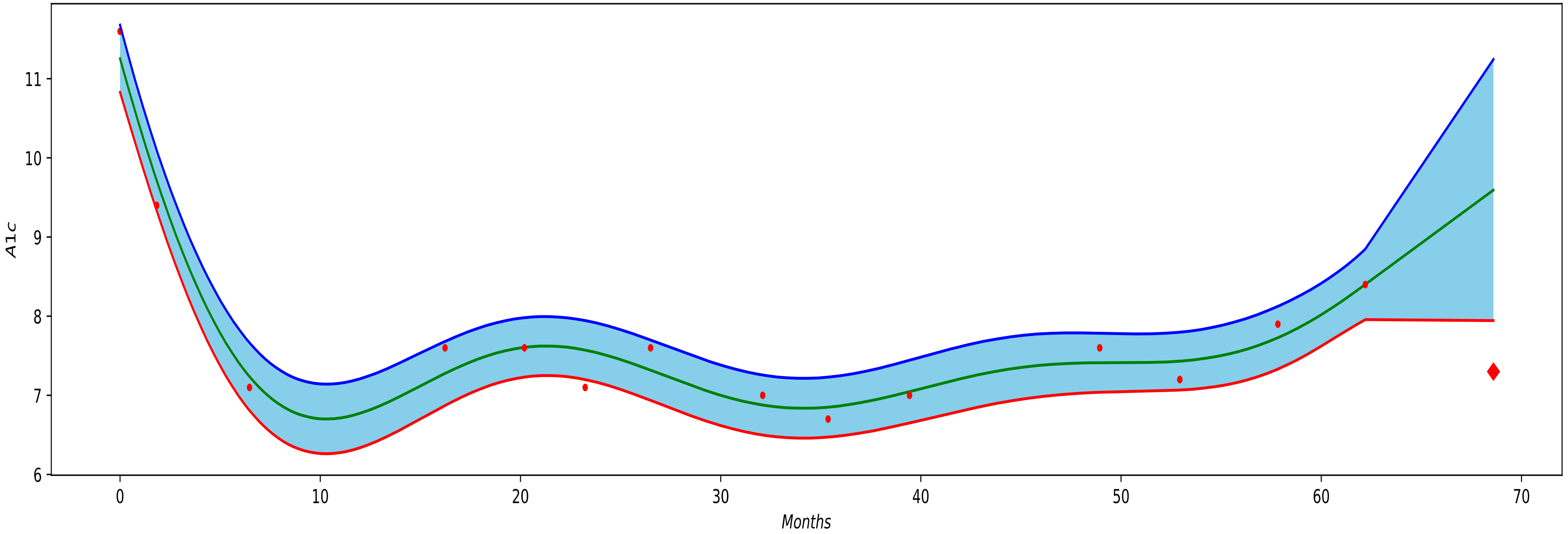
Machine Learning

- Methods characterized by the use of complex mathematical algorithms trained and optimized on large amounts of data
- Supervised learning
  - Regressions
  - Decision trees
  - Support vector machines
  - Neural networks
- Unsupervised learning
  - Clustering and association algorithms
- Semi-supervised learning
- Reinforcement learning

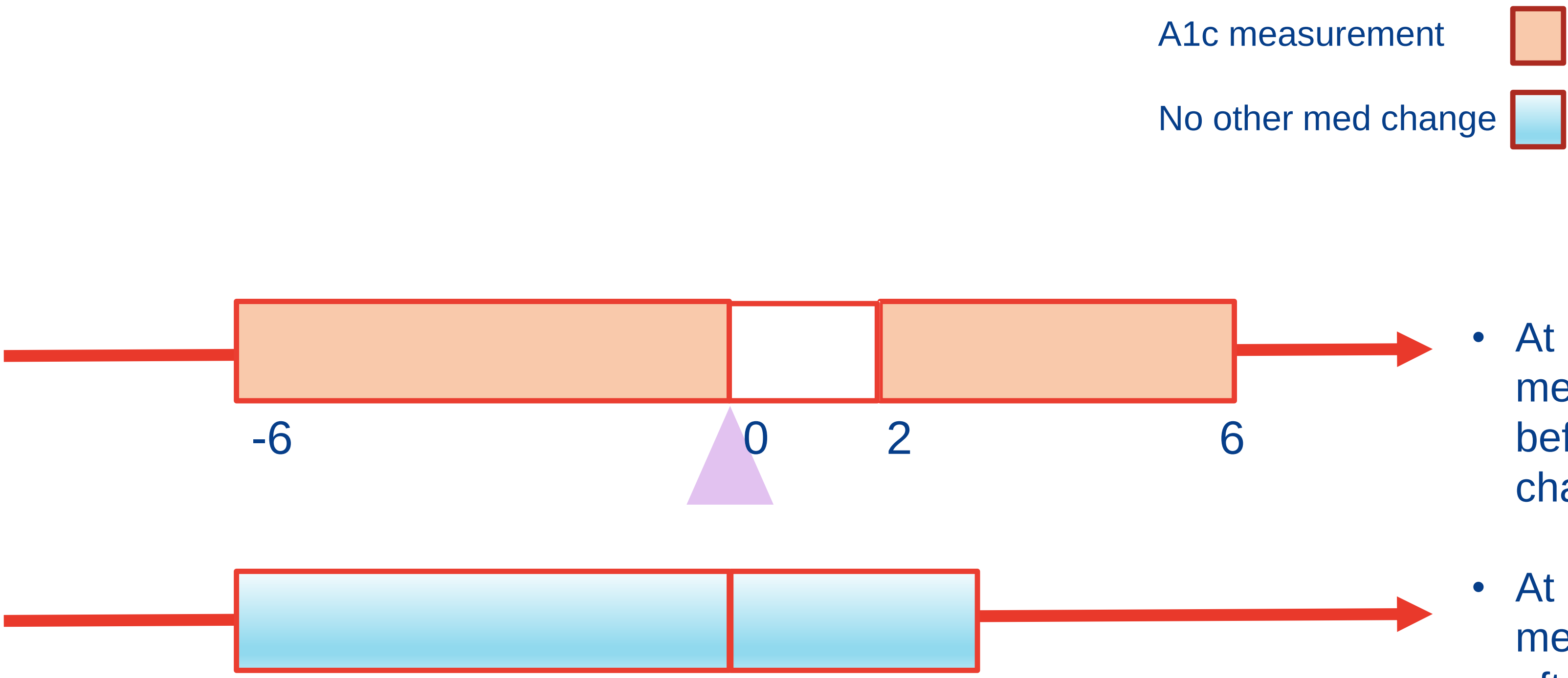
Support tool for glucose management



Prediction by HbA1c trend

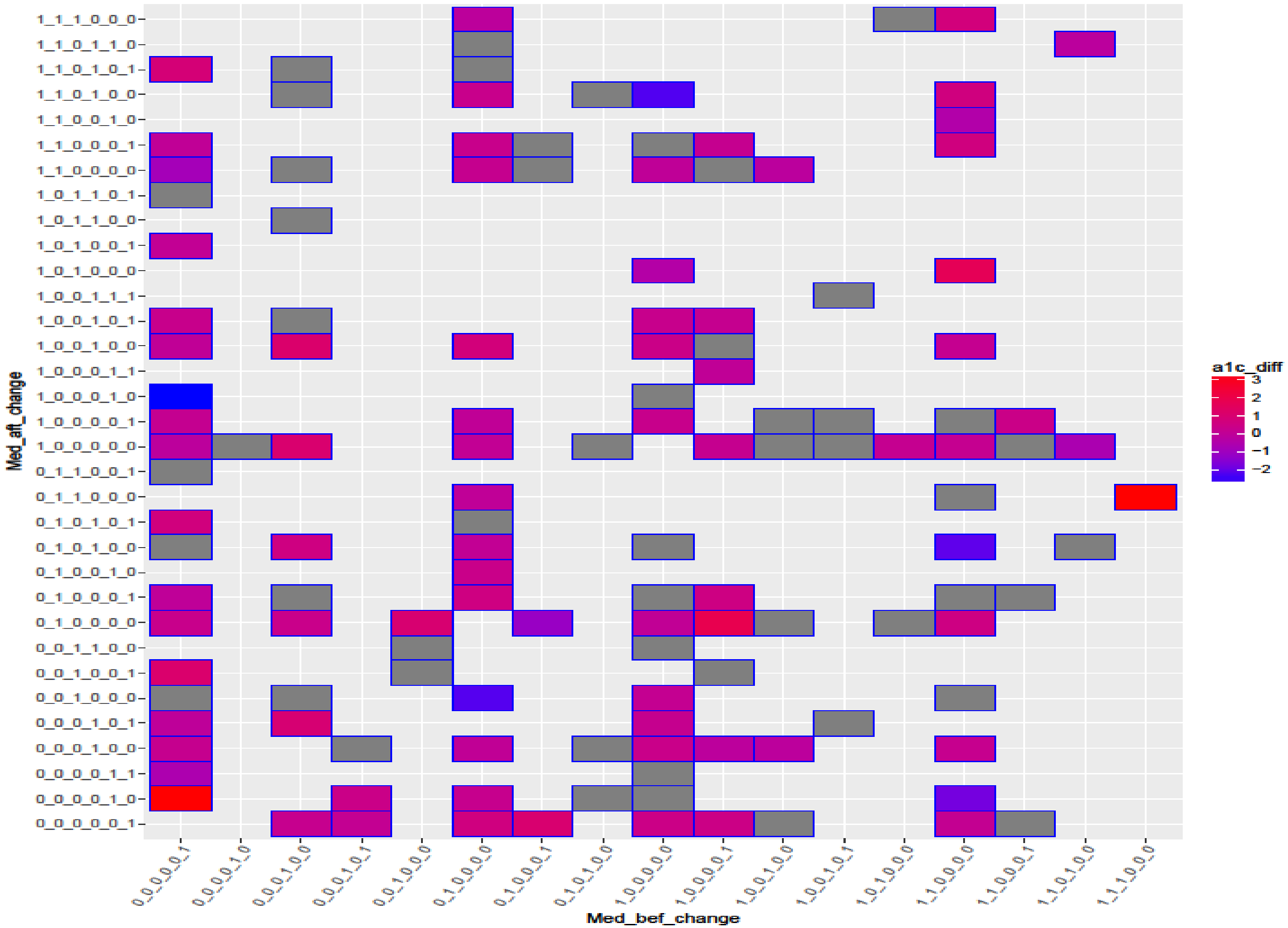


Inclusion criteria

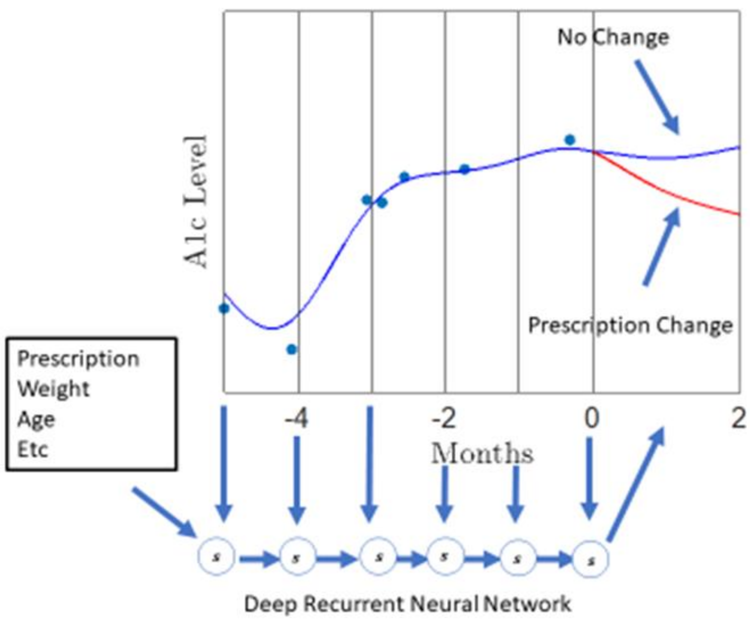


- At least two A1c measurements before the change
- At least one A1c measurement after the change

Comparing case and control HbA1cs after the medication change



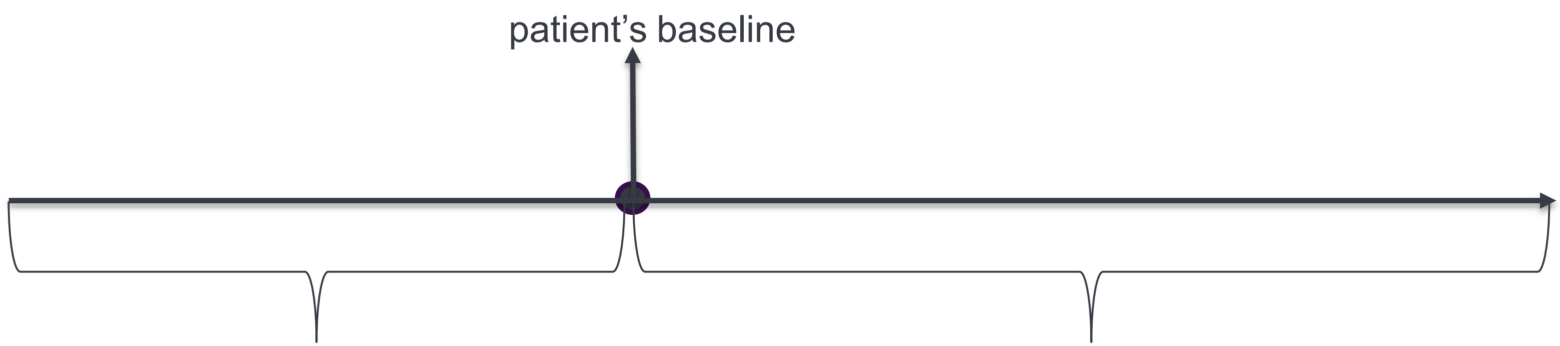
Machine Learning model



- Deep Recurrent Neural Network
  - Historical A1c values to evaluate trends
  - Incorporates covariate information (i.e. what prescriptions, height, weight, age, etc., small set of comorbidities)
  - Deep Learning on all available data to learn a representation capable of predicting future trends
  - Representation is learned on historical patients
  - Use Bayesian non-parametrics to handle non-uniform sampling and incorporate uncertainty (i.e. no measurement for 2-3 months etc.)

Medication nonadherence

- Collaborative study with Massachusetts General Hospital (MGH) sponsored by Sanofi



12 M
Variables in Prediction:
Demographics, Comorbidities, Labs

Doctor notes + the above
EHR

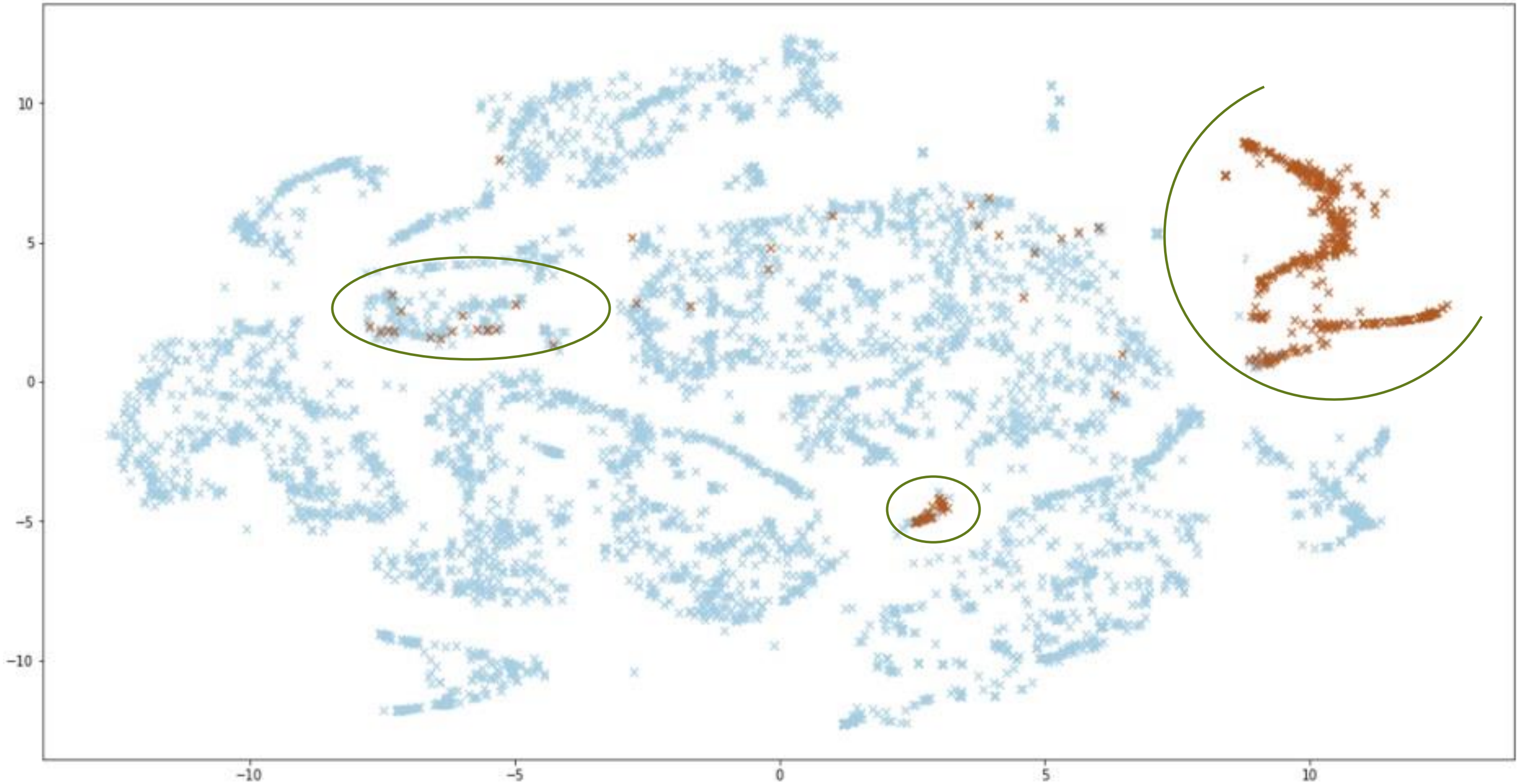
18 M
Any indication of nonadherence in the notes?

- Modeling nonadherence with doctor notes besides EHR

Non-adherence labeling scheme

Phrases referencing non-adherence		
Noncompliant	Did not take his med	Refusing to take insulin
Poor compliance	Not adherent	Refused her insulin
Poor adherence	Poor med compliance	Refused his insulin
Poorly compliant	Poor medical compliance	Refuses to take his insulin
Non adherent	Not taking insulin	Refuses to take her insulin
Not taking her med	Does not take insulin	Refused to take his insulin
Not taking his med	Decided not to take insulin	Refused to take her insulin
History of noncompliance	Refuses to use insulin	Refusing to take his insulin
Did not take her med	Refuses to take insulin	Refusing to take her insulin
Refuses med	Refused to take insulin	
Poor medication compliance	Refuse to take insulin	

Indications of medication adherence using clinical narrative



A 2-dimensional visualization of the higher dimension encoded layer by tSNE

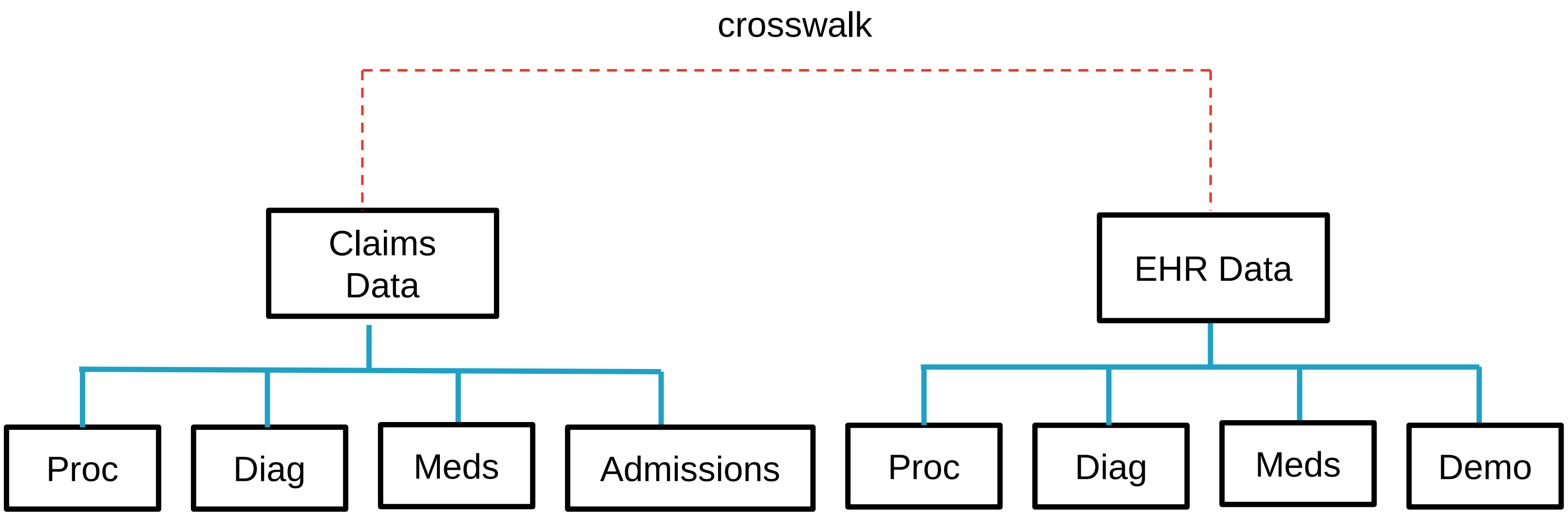
Each BLUE cross represents a note associated with a CONTROL patient

Each RED cross represents a note associated with a NON-ADHERENT patient

ML for Medicare Shared Savings Program

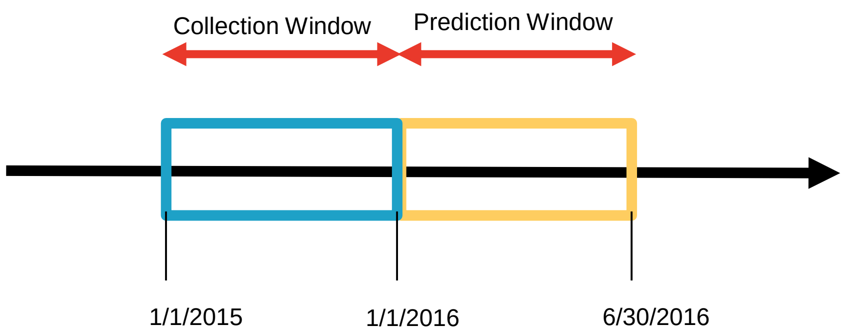
- Predict the risk of patients' admissions
- Allocate resources and provide better care for high risk patients
- Identify potential factors that contribute to higher risk of admissions

Description of data



Description of data

- Claims
  - 91456 unique patients and 1086 covariates (618 with more than 0.1% occurrences)
  - 85.94% are censored 14.06% are uncensored



- EHR
  - 79158 unique patients and 1948 covariates (883 with more than 0.1% occurrences)
  - Demographics include sex, race, employment status, marital status, emergency contact

- Combined
  - 91456 unique patients and 1507 covariates
  - 36 principal diagnosis as outcomes

Multi-layer perceptron for admissions

Principal Diagnosis	Test AUC
Schizophrenia and other psychotic disorders	0.901007
Chronic obstructive pulmonary disease and bronchiectasis	0.893717
heart failure	0.875055
Hypertension	0.857705
Mood disorders	0.842237
Respiratory failure	0.838877
Diabetes	0.836429
Anemia	0.826879
Aspiration pneumonitis	0.817202
Complications	0.813736
Secondary malignancies	0.808924
Gastrointestinal hemorrhage	0.784936
Other nervous system disorders	0.783756

Principal Diagnosis	Test AUC
Other gastrointestinal disorders	0.782207
Diseases of the urinary system	0.771511
Hepatobiliary Disorders	0.761109
Fluid and electrolyte disorders	0.758073
Bacterial infection	0.749335
Upper gastrointestinal disorders	0.749259
Rhythm Disorders	0.749036
Epilepsy	0.743410
ANY	**0.741165**
Respiratory infections	0.737710
Intestinal infection	0.733422
Symptoms	0.729875
CAD/AMI	0.727519

Deep Poisson factor model for predicting diabetes complications at one year

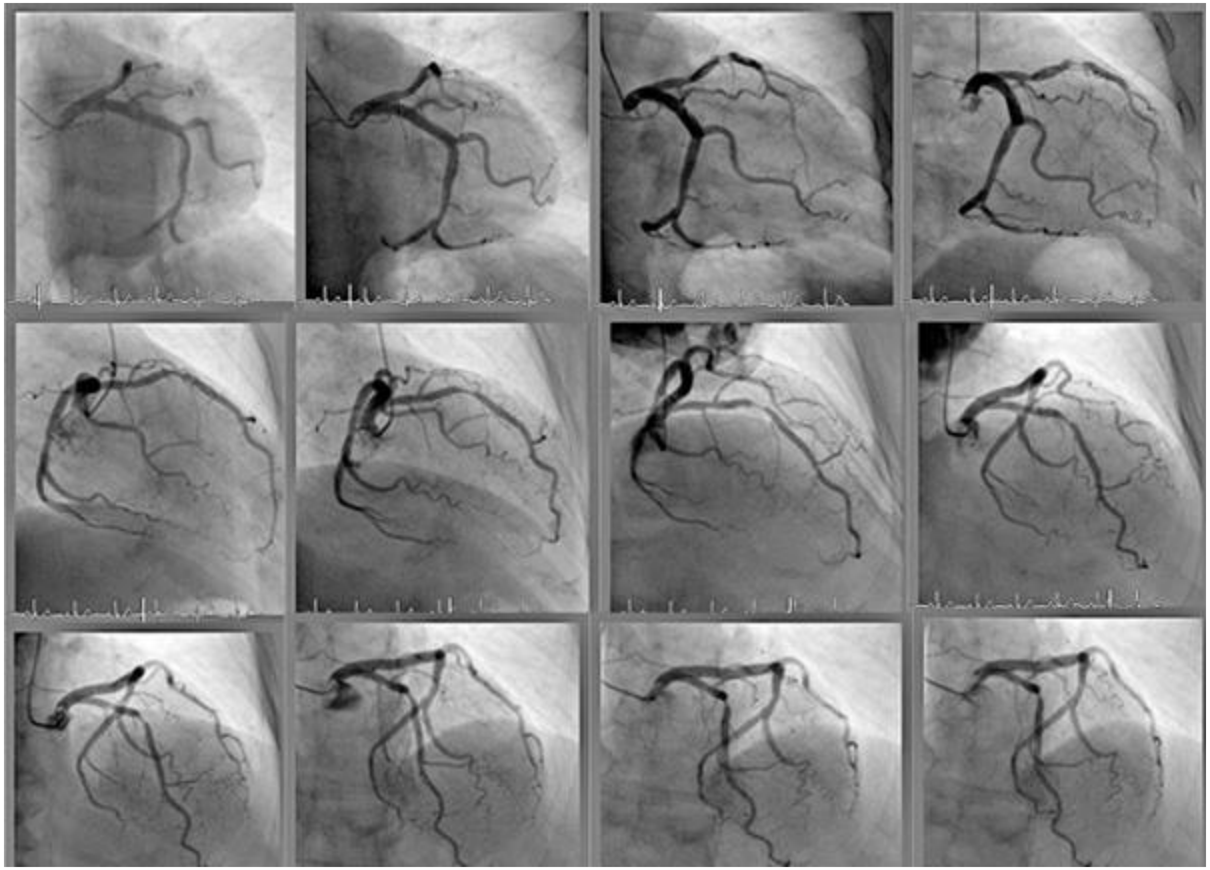
Deep Poisson	LASSO regression
Non-Linear	Linear
Search of complex model space	Need to specify functional form
Imputation via correlation structure	Explicitly recode missing predictors
One model for all outcomes	Separate model for each outcome
Requires advance computer hardware	Fast computation on standard comp
Less straight-forward means of identifying “important predictors”	Straight-forward identification of relationship of predictors to outcomes

C statistics for Deep Poisson vs. LASSO

Outcome	Event Rate	Deep Poisson	LASSO
CVD	14.1%	0.87	0.88
Infection	12.6%	0.83	0.83
Renal	11.6%	0.88	0.89
Peripheral vascular	5.0%	0.89	0.91
Cerebrovascular	3.9%	0.91	0.91
Ophthalmologic	2.2%	0.79	0.74

Puppy or muffin?

Using machine learning to advance imaging



Contact information:

Michael J. Pencina, PhD

Professor of Biostatistics and Bioinformatics, Duke University
Director of Biostatistics, Duke Clinical Research Institute
michael.pencina@duke.edu

To learn more about the DCRI’s work in advanced insights with data:

- **“The decision is in the question”:** https://www.dcri.org/our-work/analytics-and-data-science/
- **Center for Predictive Medicine:** https://www.dcri.org/our-work/analytics-and-data-science/center-predictive-medicine/
- **Program for Comparative Effectiveness Methodology:** https://dcri.org/our-work/analytics-and-data-science/cem/

To learn more about the Duke Forge and health data science:

- https://healthdatascience.duke.edu/