THE UNIVERSITY OF CHICAGO
MEDICINE & BIOLOGICAL SCIENCES

CCHE
The Chicago Center for HIV Elimination

Predictive Analytics for Retention in HIV Care

Jessica Ridgway, MD, MS; Arthi Ramachandran, PhD; Hannes Koenig, MS; Avishek Kumar, PhD; Joseph Walsh, PhD; Christina Sung; Rayid Ghani, MS; John A. Schneider, MD, MPH

Big Data/Predictive Analytics

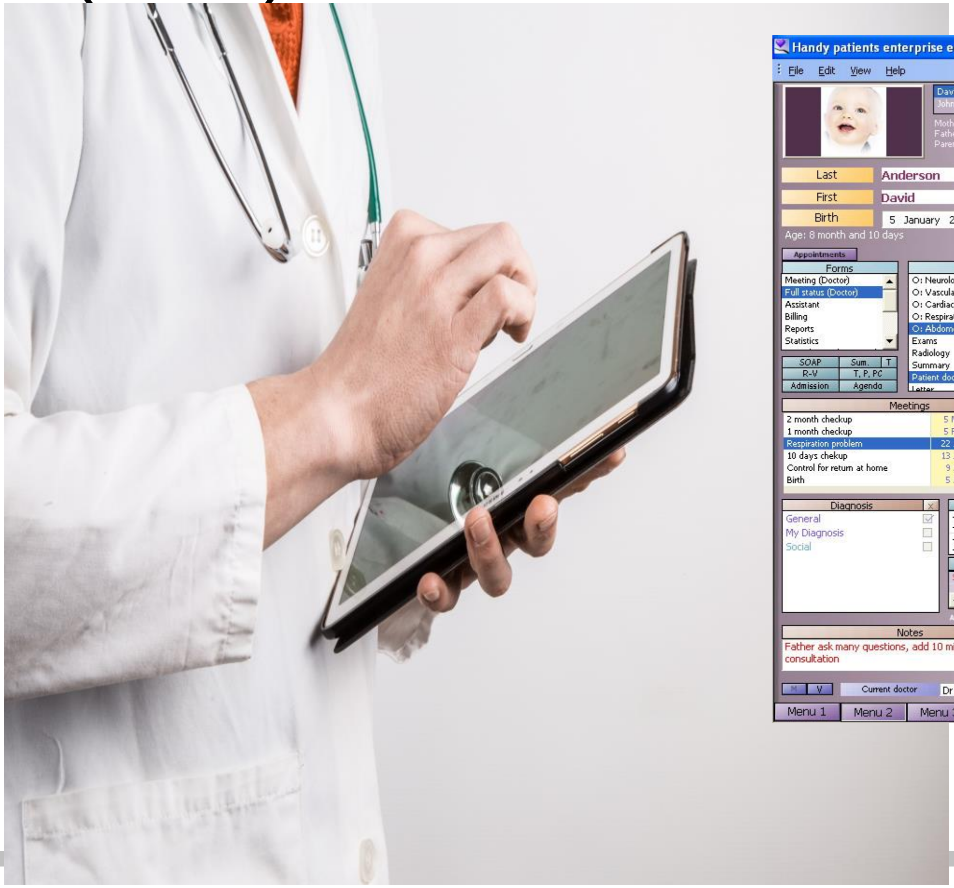
Google

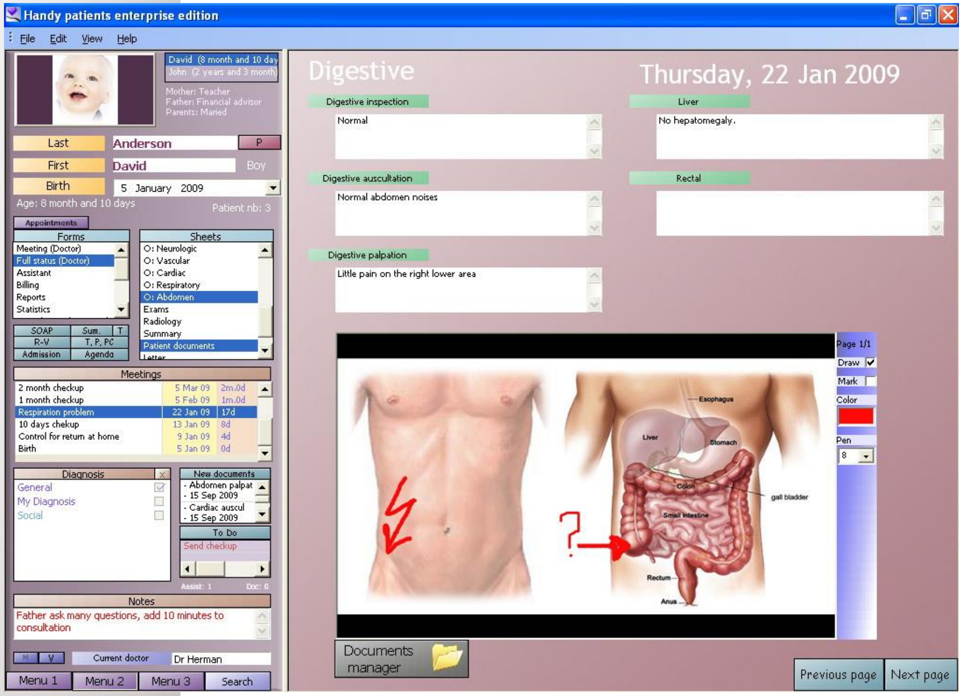
amazon

NETFLIX

Predictive Analytics in Healthcare

- Big Data source: Electronic medical records (EMR)





Predictive Analytics in Healthcare: Examples of Predicted Outcomes

- In-hospital cardiac arrest
- Readmissions
- Hospital acquired infections
- Length of stay in hospital
- Missed clinic appointments

How can predictive analytics be used to improve retention in HIV care?

- Predict each client's risk for retention in care failure *before* client falls out of care
- Real time, individualized assessment of risk
- Can be used to target retention resources for clients at greatest risk of falling out of care

Aim

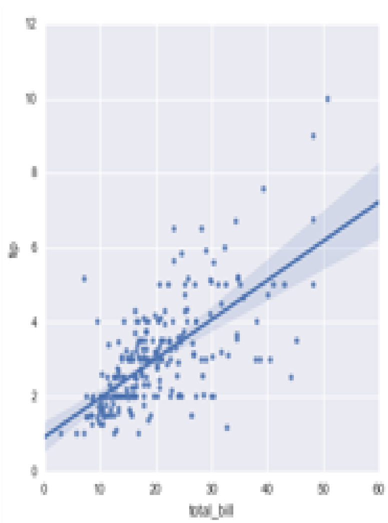
To create a predictive model of retention in care using EMR data and electronic contextual metadata, utilizing machine learning methods

What is Machine Learning?

- Derived from computer science
- Uses historical information to identify patterns or predict future events without necessarily having pre-programmed rules
- Captures hard to detect relationships in the data
  - Scalable
  - Non linear/complex models
- Goal to maximize predictive accuracy rather than interpret regression coefficients

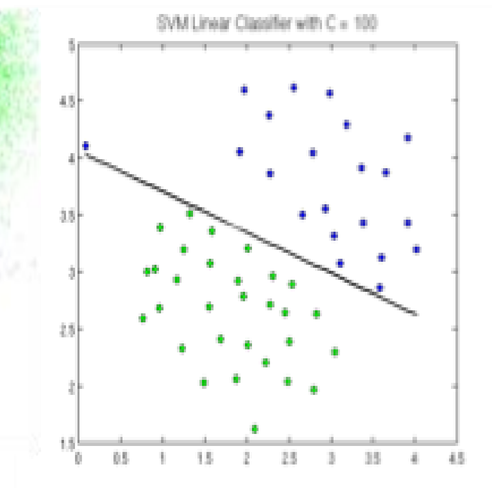
What is Machine Learning?

Most Common Machine Learning Tasks...

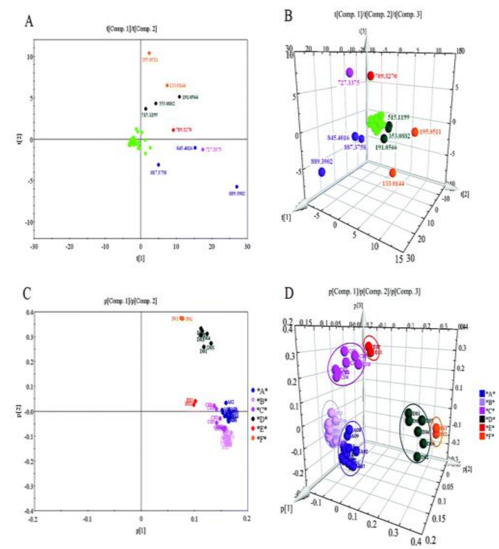


Regression
Using trends to predict outcomes

Clustering
Finding existing groups or categories



Classification
Labeling and sorting into groups



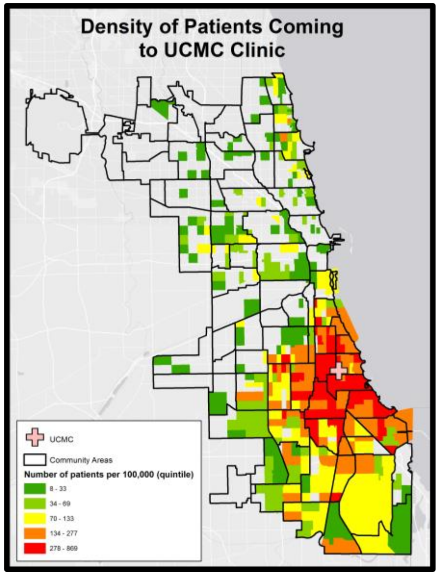
Dimension Reduction
Create a simplified abstraction of the data

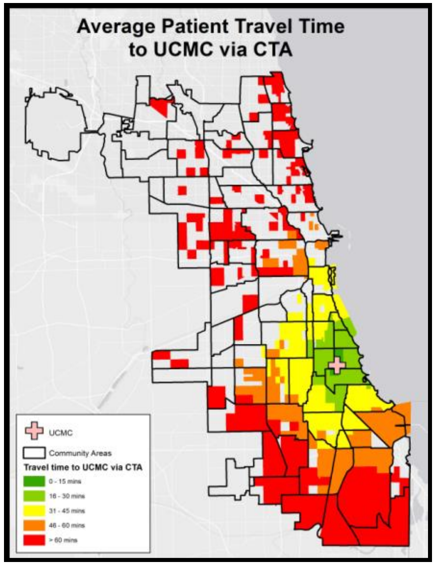
Data Source

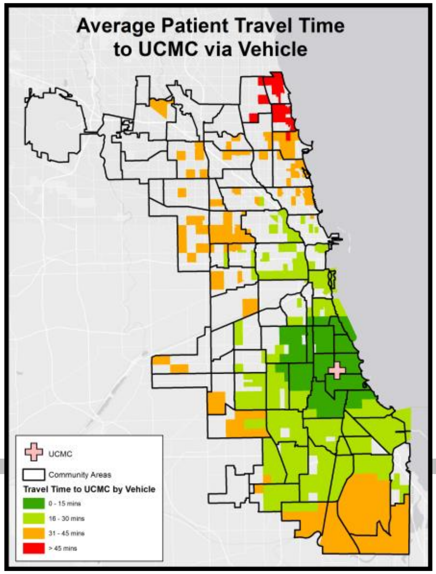
EMR data for all HIV+ patients who received care in adult ID clinic from 2008-2016:

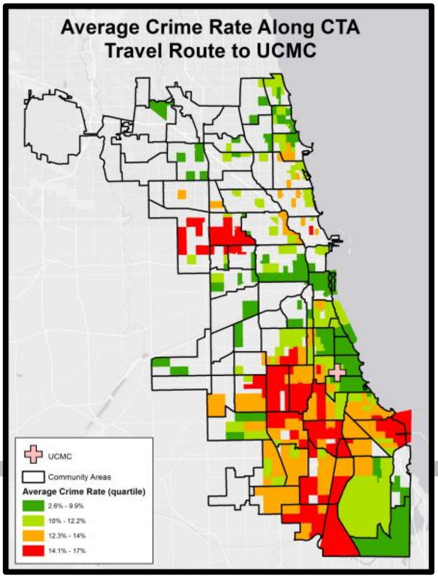
- Appointments scheduled/attended/missed/cancelled
  - Encounters in ID and other departments
- Diagnoses
  - billing codes, problem lists, past medical history
- Social history
- Laboratory values
  - CD4, viral load
- Medications
  - ART regimen, pill burden
- Demographics, Insurance

Location Based Data: Geocoded Patient Addresses









Abbreviations: UCMC, University of Chicago Medical Center; CTA, Chicago Transit Authority

Location Based Data

- Data from American Community Survey and Chicago Open Data Portal
- Characteristics of clients' neighborhoods
  - Average income level
  - Average education level
  - Racial/ethnic composition
  - Crime rates

Methods

Retention in Care Definition:

2 kept visits within 12 months > 90 days apart

Methods

- Machine Learning Methods used
  - Decision trees
  - Random forest
  - Logistic regression
  - Gradient boosting
- Validated using temporal cross-validation

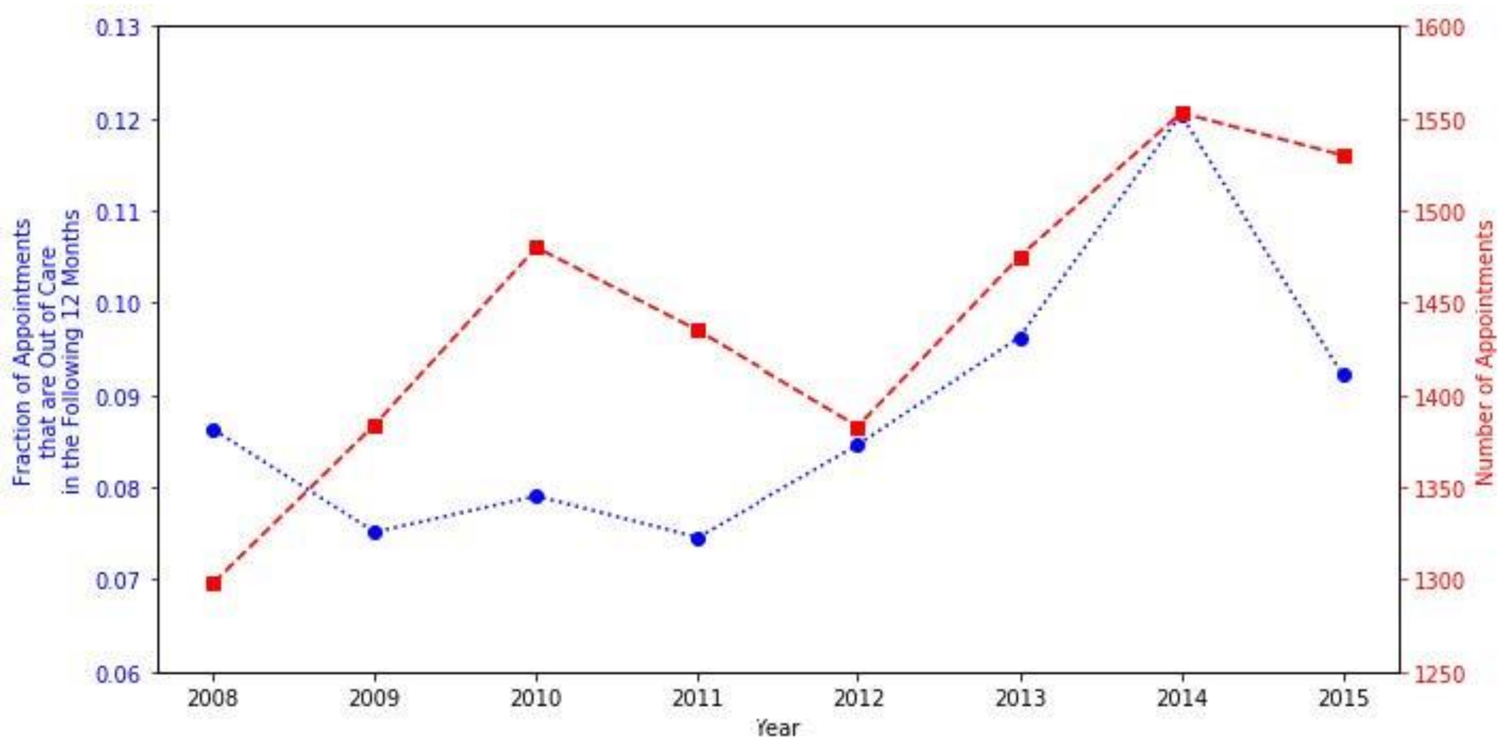
Methods

- Compared precision of each model to baseline retention rate and to a simple logistic regression model meant to simulate expert heuristics

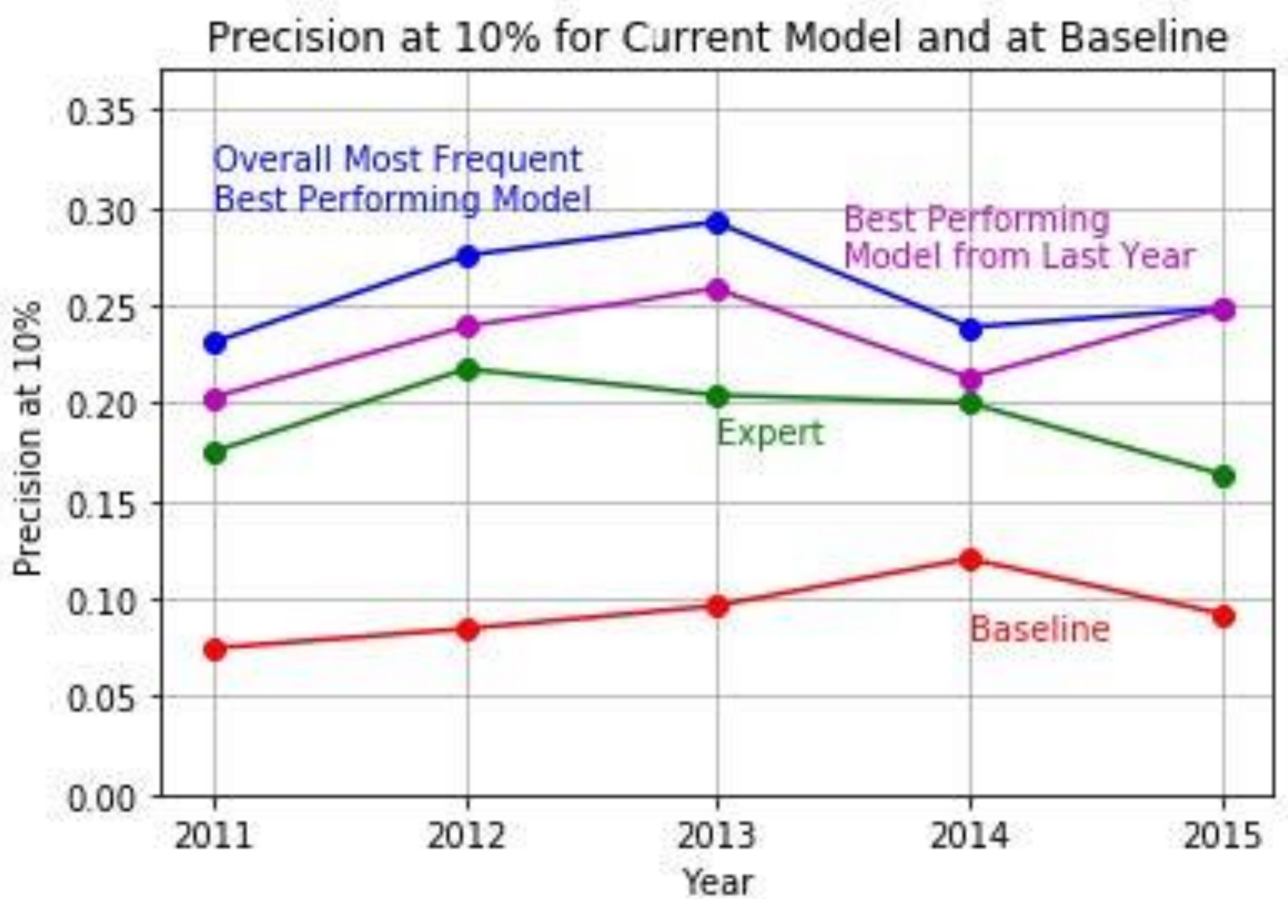
Results: Patient Demographics

Characteristics	N (%) N=713
Male sex	399 (56%)
Race African American White Other	 585 (82%) 93 (13%) 35 (5%)
Insurance Private Medicaid Medicare	 312 (44%) 309 (43%) 85 (12%)
	Mean (SD)
Age	47.3 (13.6)
# of attended appointments	19.5 (17)

Appointments per year in HIV care clinic



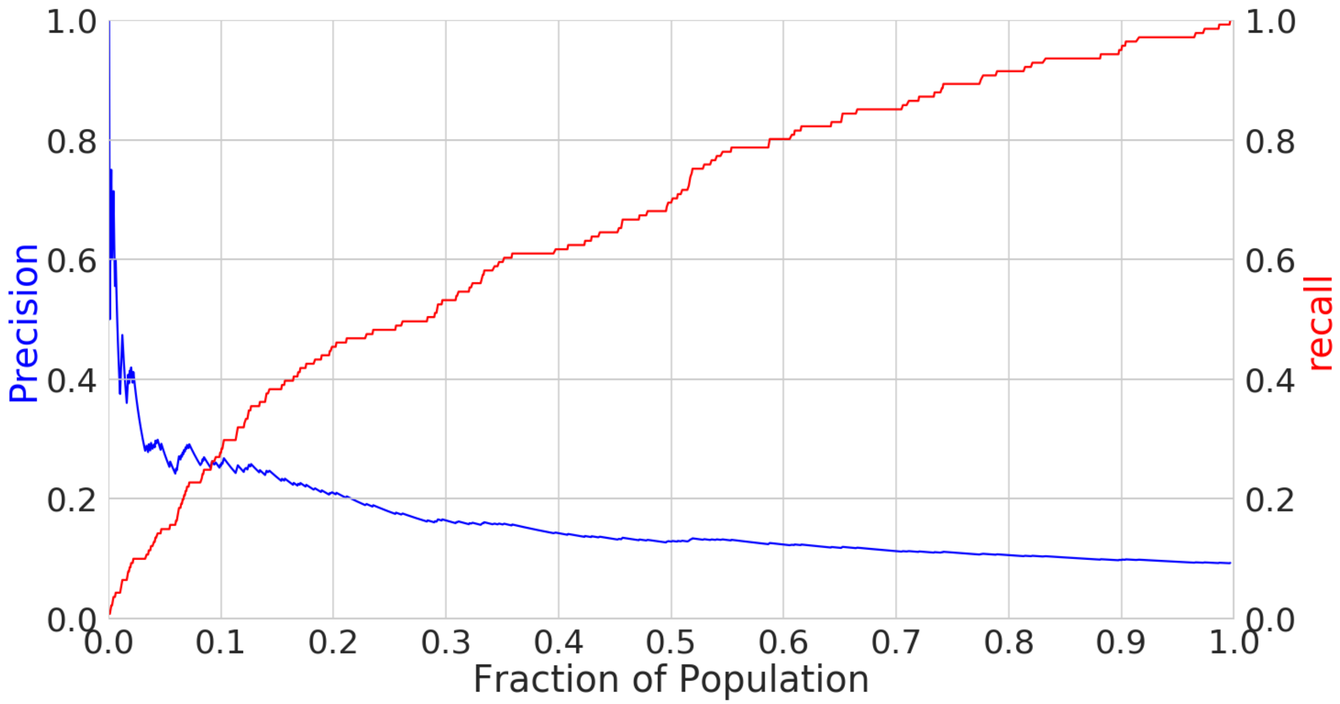
Comparison of Precision among Models



Best Performing Model: Random Forest Model

- Included 1,466 features
- Most important features for prediction of retention in care:
  - Previous ID encounters
  - CD4 count
  - Provider
  - Viral load
  - Substance use
  - Previous encounters in departments other than ID

Precision and Recall for Random Forest Model



Future Plans

- Incorporate natural language processing of text of provider and social work notes into the model
- Validate model using EMR data from CFAR Network of Integrated Systems (CNICS) research network
- Create interactive tool showing risk of retention failure in real time during clinical encounter

Acknowledgments

Center for Data Science and Public Policy

THE UNIVERSITY OF CHICAGO

Third Coast CFAR

CENTER FOR AIDS RESEARCH IN CHICAGO

Funding provided by pilot award from the Third Coast Center for AIDS Research (CFAR), an NIH funded center (P30 AI117943).