Analytics Tools to Identify Patients with Chronic Kidney Disease

Mark Sendak, Dr. Blake Cameron, and Joseph Futoma

presenting on behalf of an interdisciplinary team supported by the Duke Translational Research Institute and the Duke Institute for Health Innovation
Patient Story
Kidney Function (GFR)

Age
Untreated diabetes & high blood pressure.

Normal kidney function, but with evidence of kidney damage.

No regular medical care.
Age 48
A few more ER trips
Kidney function falling
Age 49
Kidney function now 50%
Age 50
Gets a Duke Primary Care doc
Kidney function 30%
Age 51
CKD first noted on problem list
Referred to Nephrology
Three months later presents to ER with kidney failure symptoms and “crash starts” dialysis.
Missed Opportunities:

To prevent or delay kidney failure

To prepare for kidney failure
42% starting dialysis have no prior nephrology care
Project Complexity
Extracting Data

Welcome to D.E.D.U.C.E.

Extract, Transform, and Load

Chronicles

Extract, Transform, and Load

Clarity
Storing Data

Office of Research Informatics
Virtual Machines
Accessing Data

Outside Contractor

Request guest Duke Health Enterprise account

University Student

Duke Health Enterprise VPN and multi-factor authentication access

Health System

Obtain DHTS Linux server account for Virtual Machines

ORI Virtual Machine user added
Transforming Data

Test_Name
- CREA
- CREA BLD
- CREA NT
- CREA-WB
- CREAT I-STAT
- CREAT-WB
- Creatinine
- CREATININE
- Creatinine - Labcorp
- Creatinine - LabCorp
- Creatinine (mg/mL) Jaffe
- Creatinine I-Stat
- Creatinine Whole Blood
- DUAP CREA
- DUAP CREATININE
- ICRE
- POC Creatinine
- POC-CREA

1996 2015
Population Rounding
Statistical Modeling
Challenges with EHR Data

- Generalizability, e.g. Institutional Factors, Geography
- Cohort definition / phenotyping
- Outcome / Problem definition
- Censoring, Dropout
- Variable Selection
- Missing Data
- Noisy / Incorrect Data
- Episodic, irregular, asynchronous
Typical real-world puzzle

• Problem: alignment of patient time series

• Possible solutions:
  
  • By age:
    
    • People progress through disease over different time scales, at different ages
  
  • By first time eGFR dips below threshold:
    
    • Threshold is arbitrary; can’t align new patient with no eGFR reading below threshold
Learnings

• Incredibly rewarding to see real-world impact

• Lots of interesting problems
  • High demand for people who can help solve them!

• Hot application area in statistics / machine learning

• Pressure to publish methodologically new work and take ownership of ideas
  • Better to focus on the application and solving the problem well