Health Data Science
Kickoff Event

Advances and partnerships in health data science at Duke

March 20, 2017
Partnerships in Health Data Sciences @ Duke

Larry Carin, PhD, Vice Provost for Research
Vision / Scope

• University-wide *initiative* to advance health through data science, translation and innovation

• Shape and lead the platform for 21\textsuperscript{st} century cures initiatives

• System for Evidence Generation

• Micro learning environments for learning health systems

• Platform for transformation of healthcare delivery model

• Market translation through incubation and entrepreneurship
Shaping future of health care: streamlining knowledge generation, research, innovation and translation

Build micro learning environments into hospitals and health delivery system

LHS
Duke Heart, DCI, Primary Care and Mental Health...etc.

Data Science

Translational Health

Programs to accelerate and translate research and innovations to address healthcare needs of populations

Business Incubation & Entrepreneurship

Venture market bridge for translations into business to develop the new healthcare ecosystem

Organized medicine driven by data

Big Data & Medical IoT
Capabilities and Operational View

Applied Quantitative Sciences in Health

- Project 1
- Project 2
- Project 3
- Project N

Teams

Project Team

Strategic Domains

Functional Cores

Quantitative and Data Science Scholars and Data Curators

Implementation Science and Translational Clinics

Project management and program management

Robust Data and Computing Infrastructure

Team Constitution

Methods & Quantitative expertise

Health implementation Faculty

Clinical and Data Sc. Leads

UG/Graduate Students

Methods and Operational Core
Stakeholder Engagement & Leveraging

- Education and Methods (iiD, Duke Stats, B&B, Pop Health)
- National Translation through LHS networks
- Implementation and Translation at Duke Health and local community
- Leveraging all of Duke: Clinically driven Challenges
- Leveraging all of Duke: Quantitative Science
- B&B
- Pop Health
- Math
- Statistics
- Pratt & CS
- Computing and IT infrastructure (DHTS & OIT)
- DUHS & PDC
- DIHI
- Team Science
- SoM Depts, Centers & Institutes
DCRI and the Role of Data Sciences in Health & Medicine

Eric Peterson, MD, DCRI
Facilitating Health Data Science at Duke

Michael J. Pencina, PhD
Professor of Biostatistics & Bioinformatics, DUSOM
Chief Data Scientist, DCRI
DCRI Faculty Associate Director
Health Data Science

- Content Expertise
- Methodologies
- Implementation
- Education
- Data Analysis
Machine Learning in Healthcare

Comparative Effectiveness Methodology

Methods for EHR Data

Pragmatic Clinical Trials

Predictive Modeling

Machine Learning in Healthcare

Other

METHODOLOGIES
Academic Programs Support Data Science

• Serve as a laboratory for data scientists to bring FRESH IDEAS and insights to research design

• Create opportunity for data scientists to develop new methods — “IMAGINEERS” OF DATA ANALYSIS

• Support a COLLABORATIVE APPROACH, modeling cross-pollination for the success of the program
Aligning Duke and Duke Health in Data Science – Comparative Effectiveness Research

Fan Li
Department of Statistical Science

Laine Thomas
Duke Clinical Research Institute & Department of Biostatistics & Bioinformatics

3/20/2017
Who we are?

- Fan Li: Associate Professor in the Department of Statistical Science - a top-10 department in statistics in the U.S. Expertise: causal inference
- Laine Thomas: Assistant Professor in Duke Clinical Research Institute (DCRI) and Department of Biostatistics & Bioinformatics (B&B). Expertise: causal inference and vast experience in clinical applications, e.g., cardiology
- Michael Pencina: Professor of B&B, Director of Biostatistics at DCRI
What ties us together? CER

• **Comparative effectiveness research (CER):** Direct comparison of existing health care interventions to determine which work best or worst for which patients

• Central to health care: Good CER assists consumers, clinicians, and policy makers to make informed decisions that will improve health care

• Heavily funded across many government agencies (e.g., NIH, NSF, AHRQ, PCORI) and private sectors (e.g. pharmaceuticals, medical devices)
CER – a data science problem

• CER is traditionally based on randomized experiments, but is increasingly based on observational studies, e.g. registry data, national surveys, and (maybe) electronical health records

• CER is data science problem: requires expertise in data management, statistical methods, computer programming, and close collaboration with clinicians

• known as causal inference in statistics, and program evaluation in economics
Across Campus Collaboration: the Model

- Fan Li: CER methodology provider, in constant “hunger” for data
- Laine Thomas: CER methodology provider, with vast experience in clinical applications – bridge between methodology and practice
- Matt Brennan: Cardiologist and DCRI faculty – clinical collaborator, problem provider and “consumer” of CER
- Michael Pencina: Provide research infrastructure and oversee the collaborative projects
- PhD students in B&B and Statistical Science (and staff statisticians): Get job done under supervision and conduct thesis research
- It was not always easy to start...
Program for Comparative Effectiveness Methods (CEM)

• Extending the model of collaboration across campus - CEM
• Lead by Laine Thomas and Fan Li
• Core members: faculty and students (PhD and MS in B&B, Statistical Science, etc.) across campus with interest and expertise in CER
• Focus on developing statistical/data science methodology motivated by important health care problems from Duke medicine
• Monthly meetings, active ongoing collaborations, advising students
• Products: publications (both clinical and stat methods), PhD thesis, online tools for patients, postdoc training, funding opportunities
Benefits to the whole Duke community

• Bring together the strengths of Duke Health and Duke University
• Contribute to the research infrastructure for Duke Health Data Science
• Create new research, education, and funding opportunities
• Provide a template for future across-campus collaborations
Looking forward

• Identify new methodology problems with great clinical and policy implications
• Collaborate with other methodology core (e.g., EHR program)
• Incorporate new technologies in data science (e.g., machine learning, artificial intelligence) and provide off-the-shelf research products
• Entrepreneurship
BREAK
See you soon for the panel discussion
Case Studies in Health Data Science

Moderated by Erich Huang, MD, PhD
Data Science for Surgery

Allan Kirk, MD, PhD, Department of Surgery
Katherine Heller, PhD, Statistical Science and Center for Cognitive Neuroscience
Duke PDC Outcomes Research Team (PORT)

Mark Newman, MD, Private Diagnostic Clinics
Ben Goldstein, PhD, Biostatistics & Bioinformatics
Eric Poon, MD, PhD, Duke University Health System
The Private Diagnostic Clinic (PDC) identified eight key areas of focus for its future success and development. Six of those eight goals had effective data acquisition and data analytics as a key component to driving our success.

The **PDC Outcomes Research Team (PORT)** is a cross-functional group tasked with delivering high quality data derived from various electronic sources at Duke to members of the PDC in support of PDC’s Strategic Plan.
PORT Development Process: Overview

Clinical leadership identifies problems to be solved and submits via RFA process.

IT experts and Medical Leadership selects projects and Agile lifecycle begins.

Cross-functional IT group creates useable data stored in PORT data mart.

Reports/data sets to address Use Cases (Tableau, WebI reports/dashboards)

- Knee and Hip Osteoarthritis Care, Pathways & Analysis at DUHS
- No Shows for Urology & Pediatric Subspecialties
- Low Back Pain Services and Cost at DUHS

Current (2nd Round) Use Cases
PORT Outcome: Creation of a Learning Healthcare System - Use Case for No Shows

- Identifying high impact need “Patient No Shows”
- Extract EHR data into Data mart
- Develop risk model with implementation use case in mind
- Generate daily report for clinics highlighting high risk patients
- Implement randomized prospective study to evaluate intervention

**PORT Data mart**

Additional phone call to patients with risk > 20%.

Cluster randomized cross-over trial powered to complete in 10 wks
Analytics Architecture to Support Reusability
Ready Data Sets as of 2/2017

Universes
• Beaker (Lab)
• Lab Turn Around Time (TAT)
• Pharmacy - Medication Usage Evaluation
• Stork (OB)
• Transplant
• Meaningful Use
• Clinician Efficiency in MC
• Appointments and Schedules
• HIM Coding
• Hospital Billing - Transaction Activity
• Hospital-based Episode Variation
• Professional Billing - Transaction Activity
• Professional Billing Denials and Denial Rates
• Security and Access - Access log

Datamarts
• PORT Data Mart
  • ICU Short Term Readmissions for DUHS
  • Surgical Wound Dehiscence Signal Rate
  • Appointment No Show and Late Cancellation Data (GI, Ortho, Urology, Pediatric Cardiology, Endocrine, GI, Allergy and Neurology)
  • Osteoarthritis
• Patient Location History
• Adult Immunization
• Aortic Stenosis & Heart Failure
• Transplant Finance
• Neuro ICU
• Anesthesia
• PB Denials Finance Claims
• Surgical Supplies Finance
• Research Operational Support (RMDM)
Deep-SEDI

Ricardo Henao, PhD
Assistant Research Professor, Electrical and Computer Engineering,
Duke Center for Applied Genomics and Precision Medicine

Bradi Granger, PhD, MSN
Professor, School of Nursing,
Duke Heart Center Nursing Research Program
Supporting a platform for real-time monitoring and evaluation of population health through spatially enabled data architecture and analytics

CENTRAL OBJECTIVES

• Improve healthcare delivery at the individual- and population-level
• Improve health outcomes and quality of life
• Deploy multiple levels of intervention, stratified by risk factors of individual patients
• Reduce overall healthcare costs for populations by reducing hospital and ED admissions and major procedures

THEMES

• Use of electronic health record and publicly available data to monitor the health of a population
• Use of analytics to target interventions where they are most needed
• Longitudinal monitoring through data sharing with multiple health provider partners

Deep-SEDI: Deep learning to improve morbidity and mortality prediction among patients with Type 2 diabetes.

- Deep multi-modality discriminative Poisson factor model
  - Automatic feature (sparse factor) generation
  - Predictors are count-based
  - Supervised multi-outcome prediction
  - Model covariance structure of different modalities separately
  - Multi-layer model for structured covariance modeling

Properties:
- Fully additive representation of counts
- PFA is a model of occurrences
- Variable selection via sparse activations
- Risk is monotonic with predictors
- Joint modeling of multiple outcomes

Deep-SEDI: Deep learning to improve morbidity and mortality prediction among patients with Type 2 diabetes.

- Simple data aggregation (counts) performs as well as traditional risk prediction model (UKPDS)
- Poisson factor modeling performs better than traditional sparse generalized linear models (LASSO)
- Joint outcome modeling improves overall performance
Big Data and Precision Medicine: Big Opportunities

Geoffrey S Ginsburg MD PhD
Director, Center for Applied Genomics
And Precision Medicine
Director, MEDx
Professor of Medicine
and of Biomedical Engineering
Precision Medicine

**Precision medicine** is an emerging approach for disease prevention and treatment that takes into account people's individual variations in genes, environment, and lifestyle.
The US Precision Medicine Initiative: 1 Million Person Cohort (All of US)
Factors of Risk in the Development of Coronary Heart Disease—Six-Year Follow-up Experience

Kannel WB et al.
November 1961

- High blood pressure
- Increased cholesterol
- Smoking
- Diabetes
- Family history
- Male sex

“People with this history and lab values also had... ... and responded to this treatment”
“Chronic diseases can be studied, but not by the methods of the past. If one wishes to create useful data … computer technology must be exploited.”

— Eugene Stead, MD 1969

- Led to the concept of “computerized textbook of medicine”
- Formed foundation of the Duke Databank for CV Diseases
- Spurred a generation of clinical and quantitative researchers
Step 1: Gather High Dimensional Data

Integrated Data Repository

- Sensor/ mHealth Data
- Electronic Health Records
- Exposure Data
- Omics, Imaging Data
- Geo-spatial Data
- Social Network Data

Discovery
Disease Classification
Predictive Modeling
Step 2: Make the Data Widely Accessible

17-year-old programs artificial 'brain' to diagnose breast cancer

A high school junior has created a computer brain that can diagnose breast cancer with 99 percent sensitivity.

Seventeen-year-old Brittany Wenger of Sarasota, Fla., wrote a breast cancer-diagnosing app based on an artificial neural network, basically a computer program whose structure is inspired by the way brain cells connect with one another. She won grand prize at the Google Science Fair for her invention in ceremony held in Palo Alto, Calif. last night (July 23).

Like other artificial intelligence programs, artificial neural networks ‘learn’ what it is doing by making more regular their responses, and they perform better if they get...
Step 3: Create a Learning Data Ecosystem
Individual Level Data (Infectious Disease)

Models:

- Health State
- Health to Disease
- Disease State
Individual Level Data Provides Population Insights

Psychological Language on Twitter Predicts County-Level Heart Disease Mortality

Eichstaedt, Psychological Science, 2105

Individual Data Informs Population Models: Google Flu Trends
Duke Opportunities in Data Science and Precision Medicine

• Lead in data sharing, quality and security
• Define health
• Predict health to disease transitions
• Early disease detection
• Diagnosis and sub classify disease
• Predict therapeutic response / Adverse events
• Apply these data to the health of our campus and our community