Training in Health and Biomedical Data Science at Columbia University

Noémie Elhadad, PhD
noemie.elhadad@columbia.edu
@noemieelhadad
Columbia DBMI Training Program

- 120 trainees and graduates (24 current PhD students)
- NLM T15
  - 2015: BD2K supplement on data science
  - 2017: NLM supplement on curriculum and faculty enrichment in data science
Data from Biology, Medicine, and Health

• Observational data from biology, medicine, and health are increasingly prevalent, in larger and larger amounts
  – Electronic health records, biomedical literature, self-reported and tracked health data, Internet and social media

• With the right approach, these data can
  – Help answer critical questions in a brand new way
  – Discover medical and public-health knowledge
  – Improve healthcare
  – Promote health of populations
Columbia DBMI Training Program

• Partnerships with healthcare institutions and international initiatives ➔ Laboratory for innovation for our trainees
  – NewYork-Presbyterian Hospital
  – Observational Health Data Science and Informatics (OHDSI)
  – eMERGE
Data Science at Columbia University

- Columbia Data Science Institute
  - 7 research centers, including Health Analytics
  - 200+ faculty across 9 Schools (80 new faculty)
  - General training opportunities: Certificate, Masters in Data Science

- Fertile ground for research mentorship in data science + health
  - Experts in informatics, statistics, biostatistics, computer science, applied math, etc.

- But: unmet need to train students both in the fundamentals of data science and in the health and biomedical ecosystem that generated these data and will use the product of informatics research
Training objectives for health data science at Columbia

1. Train students in computational, data-driven methods that can solve biomedical and health problems

2. Promote understanding of the socio-technical processes that shape the way biomedical and health datasets are generated and used

3. Instill in students the methodological principles of “doing” data science as part of the biomedical and health ecosystems
   – e.g., be cognizant of and proactive about reproducibility needs in biomedical data science research
Research Mentorship Objectives

1. Train to work in multi-disciplinary, data-science teams
   - Interactions with researchers and fellow trainees from across departments and schools at Columbia
   - Co-mentorships between informatics and stats/CS faculty

2. Support students to become the next generation of investigators in biomedical data sciences
   - Strong skill set in disseminating for audiences with varied backgrounds, all relevant to data and biomedical sciences.
Present Illness: 79 m w/cad, as, dm, htn, hyperlipidemia with acute onset bilateral shoulder pain occurring at rest, radiating down the arms. He sleeps sitting upright in a chair due to orthopnea.

Medications: ASA 325 mg daily, plavix 75 mg daily, atenolol 50 mg daily, isosorbide d-50 mg qam, norvasc 10 mg qpm, lasix 40 mg daily, metformin 500 mg bid.

Diagnosis: Aortic stenosis (s/p aortic valve replacement), CAD (s/p coronary artery bypass graft x 4), hypercholesterolemia, diabetes mellitus, prostate cancer (s/p prostatectomy), NIDDM, hypertension.

Hospital Course: On admission: ASA 325 mg daily, plavix 75 mg daily, atenolol 50 mg daily, isosorbide d-50 mg qam, norvasc 10 mg qpm, lasix 40 mg daily, metformin 500 mg bid.

The patient was medically managed and transferred to hospital for further care. On the second day of hospitalization, on (day 2), the patient was noted to have a temperature of 38.0°C, blood pressure 130/80 mmHg, heart rate 90 bpm, respiratory rate 20/min, and oxygen saturation 98% on room air. His laboratory results showed WBC 14,000/cu mm, RBC 3,500,000/cu mm, Hgb 11.0 g/dL, Hct 33.0%, MCV 94.0 fl, and MCH 30.0 pg.

Chest x-ray showed bilateral pleural effusions. The patient was started on nitro patch at night for persistent chest pain and heparin gtt, which was d/c'd when noted to be guaiac positive.

Catheterization: D-dimers, troponin, and other cardiology markers were repeated and were within normal limits.

The patient was started on nitro gtt for persistent chest pain and heparin gtt, which was d/c'd when noted to be guaiac positive.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.

The patient has not had fever, abdominal pain, or nausea/vomiting. He has not had sick contacts or recent travel.

He also had thoracentesis on the left and his respiratory status improved.
Bayesian formulation of deep learning in healthcare

Deep Survival Analysis

Rajesh Ranganath  
Princeton University  
Princeton, NJ 08540

Adler Perotte  
Columbia University  
New York City, NY, 10032

Noémie Elhadad  
Columbia University  
New York City, NY, 10032

David Blei  
Columbia University  
New York City, NY, 10027

Figure 1: A comparison of traditional survival analysis (top frame) and failure aligned survival analysis (bottom frame). A filled circle represents an observed event, while an empty circle represents a censored one. In the case of standard survival analysis, patients in a cohort are aligned by a starting event. In failure aligned survival analysis, patients are aligned by a failure event.
Coupling Data Mining and Laboratory Experiments to Discover Drug Interactions Causing QT Prolongation

Tal Lorberbaum, MA, Kevin J. Sampson, PhD, Jeremy B. Chang, PhD, Vivek Iyer, MD, MSE, Raymond L. Woosley, MD, PhD, Robert S. Kass, PhD, Nicholas P. Tatonetti, PhD
Understanding the role of tumor heterogeneity in GBM under therapy:
Topological data analysis in single cells

Data assimilation in diabetes

- Joining mechanistic models & empirical data
  - Glucose, insulin production, excretion, etc.
  - Estimate parameters from data
  - New: accommodate sparse, irregular, noisy data
  - Constrain the search space

Example course trajectory example for student in data science track with focus on EHR data and healthcare

Core DBMI Course

Qualitative Obj.

Domain Objective

Quantitative Obj.

IT Objective

Acculturation in Prog and Stats*

Acculturation to Medicine

Intro to Qualitative Methods

Introduction to Biomed Inform

Research Methods in Biomed Inform

Algorithms for Data Science*

Symbol Methods in Biomed Inform

Comp Methods in Biomed Inform*

Foundations of Graphical Models*

Deep Learning*

Healthcare Process Redesign

Fall Year 1

Spring Year 1

Fall Year 2

Spring Year 2

RESEARCH
Diversity of students and backgrounds: Acculturation to Programming and Statistics

• 1st-semester course (open to all DBMI students)
  – Introductory data science fundamentals
  – Computing (e.g., Linux environment, Python, Data Persistence)
  – Statistics (e.g., sampling, estimation, basics of prediction)
  – Reproducibility (e.g., Git, GitHub)

• Flipped classroom; focus on “doing”
  – Lectures/readings outside the classroom
  – Labs in the classroom with real-world, very large health datasets
  – Two instructors + 1 TA for 12 1st-year students
  – Rotating teams of 3 students for each lab
Evaluation

• **Student Feedback**
  – Formal course evaluation and direct interaction

• **DBMI Training Committee Feedback**
  – Review course evaluations, discuss feedback and the syllabi with the course instructors, and propose changes
  – Meet with elected student representatives regularly

• **External Advisory Committee Feedback**
  – Russ Altman, Ted Shortliffe, Kevin Johnson, Justin Starren
  – Senior researchers in data science: Dr. David Blei (CS and Statistics) and Dr. Shih-Fu Chang (Electrical Engineering, CS, Senior Vice Dean Eng)

• **Student Enrollment**
  – New data science courses and the overall track in data science

• **Impact on data science research within and across DBMI**
  – Number of research papers published by students enrolled in the courses
  – Number of projects and collaborations that started from a project in one of the proposed courses