Expansive Use of Data Repositories for Clinical Research

Shawn Murphy MD, Ph.D.
Chief Research Information Officer
Partners Healthcare
High Throughput Methods for supporting Translational Research

- Set of patients is selected from medical record data in a high throughput fashion

- Investigators explore phenotypes of these patients using Machine Learning tools and a translational team developed to work specifically with medical record data

- Distributed networks cross institutional boundaries for phenotype selection, public health, and hypothesis testing

- Personalized medicine is delivered into clinical care
The Data Warehouse at Partners Healthcare

Partners Enterprise

Research Patient Data Registry

Multiple Systems at Partners:
- Billing Data
- Epic Data
- Research Data (consent to contact)
- Specimen Data
- Laboratory Data
...
1) Queries for aggregate patient numbers
- Warehouse of in & outpatient clinical data
- 6.7 million Partners Healthcare patients
- 2.5 billion diagnoses, medications, genomics, procedures, laboratories, & physical findings coupled to demographic & visit data
- Authorized use by faculty status
- Clinicians can construct complex queries
- Queries cannot identify individuals, internally can produce identifiers for (2)

2) Returns detailed patient data
- Start with list of specific patients, usually from (1)
- Authorized use by IRB Protocol
- Returns contact and PCP information, demographics, providers, visits, diagnoses, medications, procedures, laboratories, microbiology, reports (discharge, LMR, operative, radiology, pathology, cardiology, pulmonary, endoscopy), and images into a Microsoft Access database and text files.
FINDING PATIENTS

Query items

Person who is using tool

Query construction

Results - broken down by number distinct of patients
One year (2014) usage of RPDR

- 3000 registered users, 655 new in 2014

- 583 teams/year gathering data for research studies

- 2634 detailed patient data sets returned to these teams in 2014, containing data of 24.7 million patient records.

- From a survey of 153 teams
  - Importance of the data received from the RPDR was evaluated in relation to the study it was supporting.
  - $94-136 million/year total research support critically dependent on RPDR from patient data received throughout life of funding.

- ~300 data marts were created to support hospital operations, representing about 80 million patient records

![Usefulness of Detailed Data](chart.png)

- Critical: 43%
- Useful: 42%
- Not Useful: 15%

106 Total Responses
The Partners Biobank provides samples (plasma, serum, and DNA) collected from consented patients.

- 84,000 patients have consented to date
- Samples are available for distribution to Partners investigators* to help identify novel Personalized Medicine opportunities that reduce cost and provide better care

*with required approval from the Partners Institutional Review Board (IRB).

Improved Clinical Care for All Patients
Biobank Integrative Genomics Strategy

Partners BioBank Samples
(Whole Blood Extracted DNA/RNA)

Genotyping
- Illumina MEGArray:
  Multi-Ethnic GWAS/Exome SNP
  Array
  Array Cost: $59/simle

Transcriptome
- Whole Transcriptome Analysis:
  RNA-seq
  Array Cost: $40-50/sample

Epigenome Profiling
- Methylation Analysis:
  HumanM450K Array
  Array Cost: $150/sample

Genome/Transcriptome Analysis: ~$100/sample

Genome/Transcriptome/Epigenome Analysis: ~$260/sample
Detailed data is gathered for request and distributed.

Data is gathered from RPDR and other Partners sources.

Files include Small Database.

Output files placed in special directory.
Results: QTc interval and medication use

Mean QTc (ms) for medications:

- Selective serotonin reuptake inhibitor or methadone
  - Citalopram (mg)*
  - Escitalopram (mg)*
  - Fluoxetine (mg)
  - Paroxetine (mg)
  - Sertraline (mg)
  - Methadone (mg)*

- Other antidepressant or methadone
  - Amitriptyline (mg)*
  - Bupropion (mg)*
  - Duloxetine (mg)
  - Mirtazapine (mg)
  - Nortriptyline (mg)
  - Venlafaxine (mg)
  - Methadone (mg)*

* Dose a significant predictor of QTc in fully adjusted linear models at \( \alpha = 0.05 \)
† QTc at specified dose is significantly different from that at prior dose in fully adjusted linear models at \( \alpha = 0.05 \)

Mean (SD) corrected QT (QTc) interval recorded on electrocardiogram 14–90 days after prescription of antidepressant or methadone, by drug dose
Find Normal Brain MRI’s of Children

Number of patients who had a brain MRI scan at a particular age in months from 0 to 6 years (A) and in weeks from 0 to 4 months (B)

Providing Clinically Relevant Brain MRI ADC Maps of Normal Children Ages 0-6

Process to find appropriate clinical imaging studies

- N=100,000
  - Brain MRI in MGH
- N = 2,871
  - Scanned 2006-2013 with ADC maps in Siemens 3T scanner
  - 0-6 years old at the time of scan
  - Radiological reports suggest no abnormality
- N = 1,648
  - ADC maps found that are not corrupted
- N = 705
  - ADC maps and clinical cases re-examined & confirmed to be normal by a neuro-radiologist (Dr. Grant) and a neonatologist (Dr. Bates)

Medical Records Query using i2b2 finds Normal Children with Clinical Brain MRI Images. NIBIB funded mi2b2 Workbench allows special audited access for investigator to Hospital Clinical Imaging Studies of Patients specified by Institutional Review Board (IRB) approval. 
https://www.i2b2.org
http://mi2b2help.partners.org
Atlases provide a visual guide for Radiology Decision Support, such as determining Perinatal Hypoxic Ischemic Encephalopathy

ADC map from 4 infants: Each statistically compared to age matched atlas yields visual guide to pathology

Quantitative analysis tools + large data sets = Great insights for practicing doctors
Participate in Networks

- Partners HealthCare System
- Boston Children’s Hospital
- BIDMC
- Boston Health Net (BMC and Community Health Centers)
- Columbia U. Medical Center and New York Presbyterian Hospital
- University of California, Davis
- Washington University in St. Louis
- Wake Forest Baptist Medical Center
- Morehouse/Grady/RCMI
- U Texas Health Science Center/Houston
Accrual for Clinical Trials (ACT) - Workflow

**Phase One**
Query SHRINE network for aggregate patient counts

**Phase Two**
Review and select patients locally in i2b2

- Generate table of patients
- Your i2b2
- SHRINE Web Client
- I2b2 Web Client
- Review/export list of patients
The "emerge network" is an initiative focused on the electronic medical records and genomics of individual patients. The process involves genotyping 25,000 individuals, leading to the identification of 100 high-priority genes with discovery, penetrance, and pleiotropy. The outcomes include various measures such as MD visits, labs/cholesterol levels, health care costs, and family screening.
Congestive Heart Failure

- Affects 2% of the adult population
- Risk of death first year after diagnosis: 35%
- Inpatient hospital costs in 2011: $10.5B which is a small fraction of all heart failure related care
Early Detection of Worsening or Improving Anemia

**Background and Methods**
- Anemia is one of the strongest predictors of morbidity and mortality in CHF.
- Increasing or decreasing HGB is a further strong predictor, but there is no good way to determine whether a patient’s HGB is on its way up or down (Circulation. 2005;112:1121-1127)

**Results and Conclusions**
- A novel mathematical model of the RBC lifecycle enables estimation of patient-specific rates of RBC maturation and turnover from a routine CBC.

**Applications**
1. CHF patients most likely to have decreasing HGB may benefit from altered treatment or longer hospitalization to avoid readmission.
2. CHF patients most likely to have increasing HGB may be responding well to treatment and benefit from earlier discharge or maintenance of current therapy.

Dynamic Model of the RBC Lifecycle
- Quantify Maturation and Clearance Rates
Creatinine Prediction: Hypothetical Application

- Hypothetical analysis of creatinine times series where possible treatments are introduced into the model

- The model hypothetically provide a future trajectory conditioned on each treatment
Population Based Predictive Analytics to Support Improved Decision Making

Exploring Integration of MGH Path renal predictive model
Virtual Patient Steward watches Mary for each chronic Dz and stratifies:

- Mary

Followed for:
- Prevention
- Diabetes Type II
- Hypertension
- Heart Failure

Virtual Patient Steward watches Mary for each chronic Dz and stratifies:

- HF
- DM2
- HTN

AI
Tribute to…

- Big Data Commons Team
  - Christopher Herrick
  - Brent Richter
  - Mariah Mitchel
  - Michael Mendis
  - Martin Rees
  - Alyssa Goodson
  - Lori Phillips
  - Edana Merchan
  - Eloy Toscano
  - Wayne Chan

- AI Phenotyping Team
  - Vivian Gainer
  - Victor Castro
  - Kavi Waghlikar
  - Nich Wattanasin
  - Scott Weiss
  - Natalie Bouton

- HIP Team
  - Sandy Aronson
  - Calum MacRae
  - Mike Oats
  - Layne Ainsworth
I2b2, SHRINE, and SMART Information and Software on the Web

i2b2 Homepage (https://www.i2b2.org)
i2b2 Software (https://www.i2b2.org/software)
i2b2 Community Site (https://community.i2b2.org)
SMART Platforms Homepage (http://smarthealthit.org)
THANK YOU