

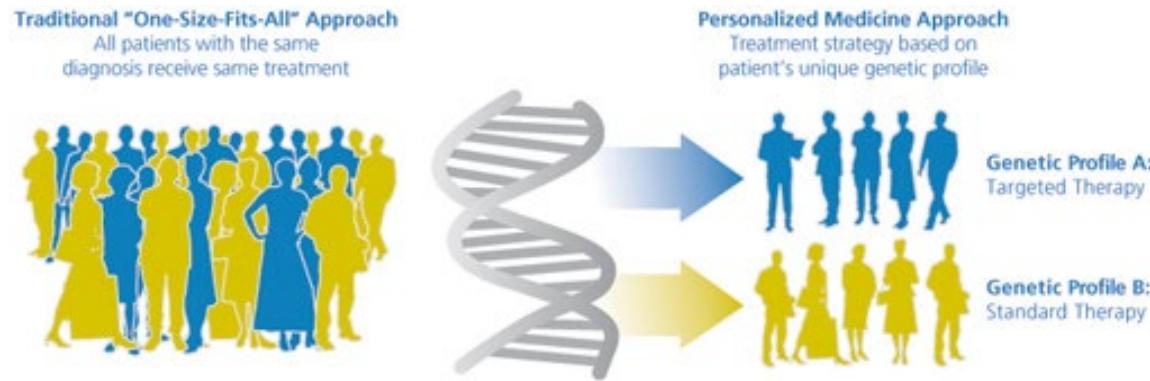
Instrumenting the Health Care Enterprise for Discovery in the Course of Clinical Care

Shawn Murphy MD, Ph.D.

Chief Research Information Officer

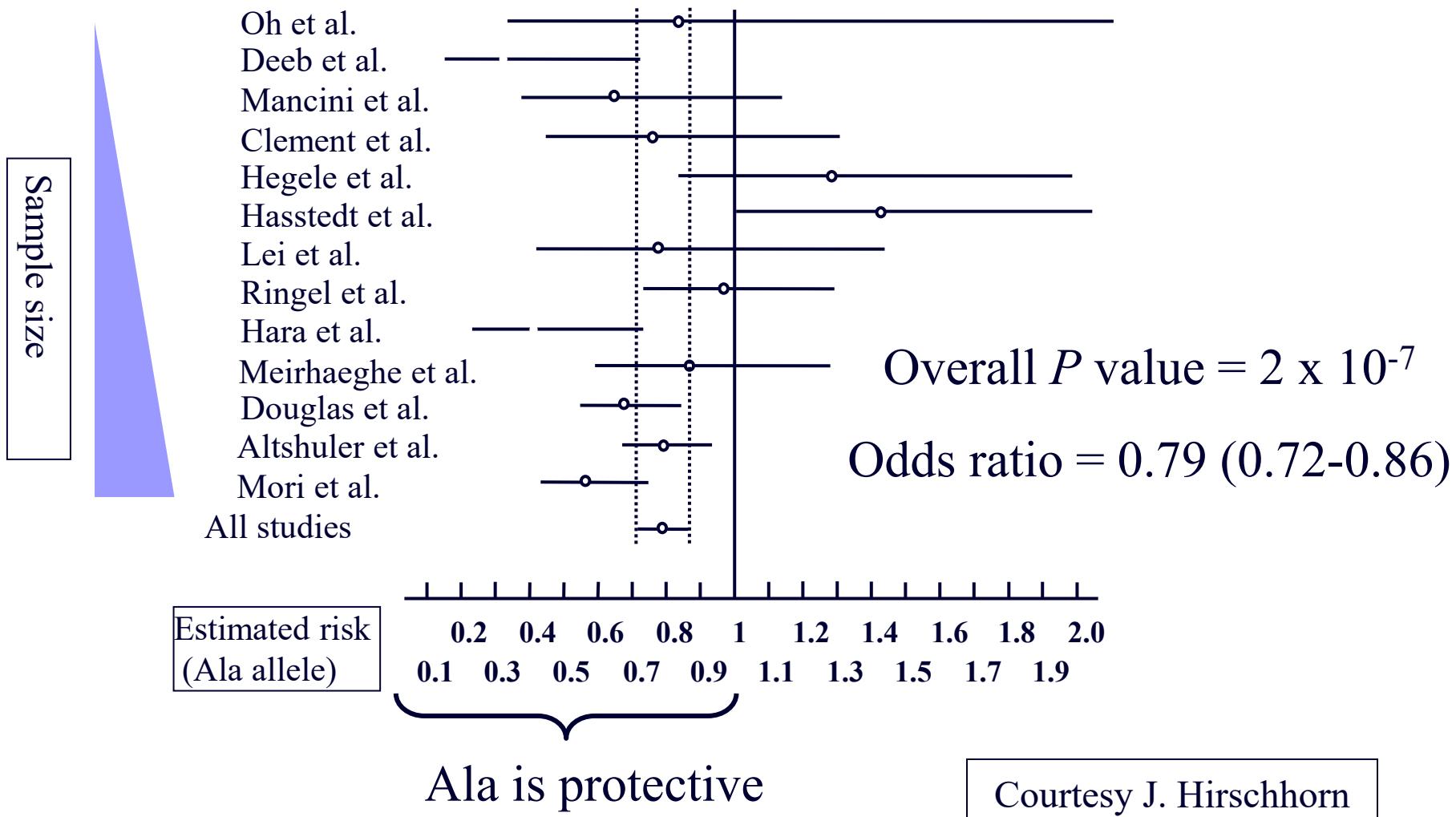
***Harvard Medical School / Mass General
Brigham***

Personalized Medicine and Genomic technology are critical to managing populations



- Managing a population involves improving health outcomes of the group as a whole by identifying, monitoring and addressing health needs of individuals through:
 - Subpopulation stratification
 - Targeted, evidence-based treatment protocols
 - Predictive analytics

Example: PPAR γ Pro12Ala and Diabetes



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
- Digital medicine is delivered into clinical care

Data problems that make working with Electronic Healthcare Data to conduct research difficult

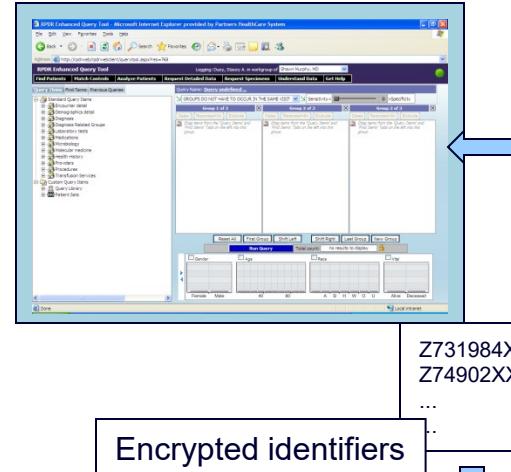
- 1) There are significant risks of a data breach which will result in very large fines and loss of confidence in the hospitals where the breach occurred.
- 2) The data are not collected for research purposes, and therefore the data can be poorly structured with significant omissions, biases, and inaccuracies.

Research Patient Data Registry (RPDR) at Mass General Brigham to find patient cohorts and distribute data

1) Queries for aggregate patient numbers

- Warehouse of in & outpatient clinical data
- 6.7 million Mass General Brigham patients
- 2.6 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)

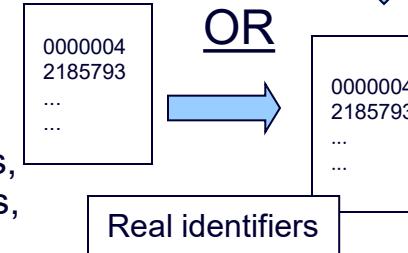
Query construction in web tool



De-
identified
Data
Warehouse

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.



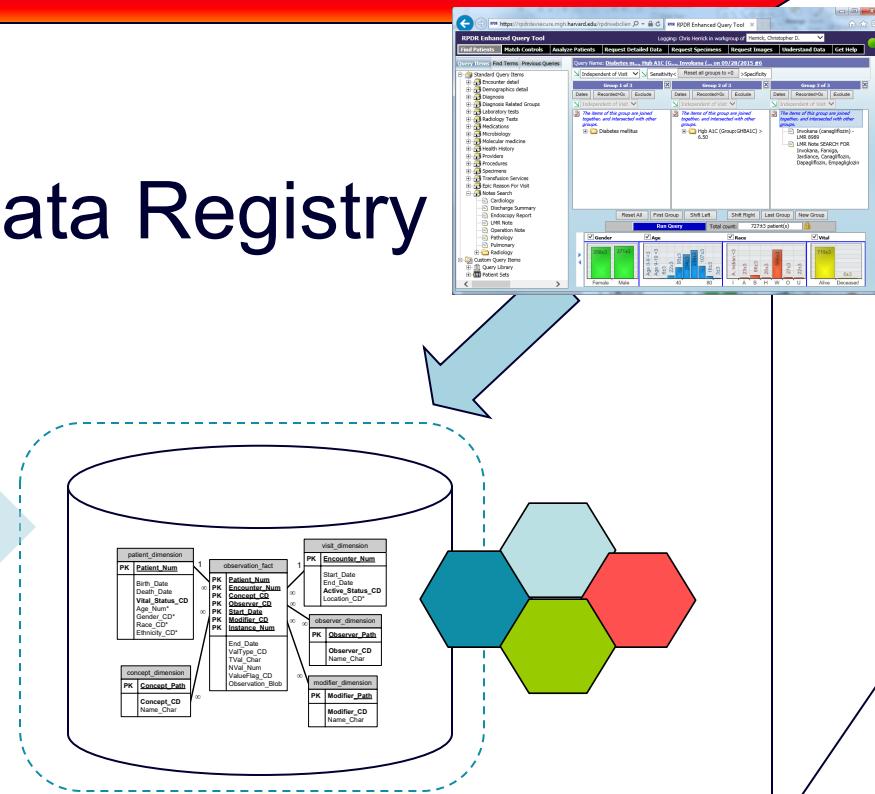
Test ID	Test Description	Result	Result Text	Anomalous	Flag Reference	Unit Reference Range
SQ-PTT	Superior APTT	29.8			980	22-1-36.1
SQ-PTT	APTT	37.0			980	22-1-36.1
SQ-PTT	PTT	46.4			980	22-1-36.1
SQ-PTT	APTT	43.1	MODERATELY H		980	22-1-36.1
SQ-PTT	APTT	26.7			980	22-1-36.1
SQ-PTT	PTT	23.7			980	22-1-36.1
SQ-PTT	APTT	24.7			980	22-1-36.1
SQ-PTT	PTT	25.4			980	22-1-36.1
SQ-PTT	APTT	24.1			980	22-1-36.1
SQ-PTT	APTT	24.0			980	22-1-36.1
SQ-PTT	APTT	24.7			980	22-1-36.1
SQ-PTT	APTT	24.3			980	22-1-36.1
SQ-PTT	APTT	34.5			980	22-1-36.1
SQ-PTT	PTT	40.0		H	980	22-1-36.1
SQ-PTT	APTT	45.0		H	980	22-1-36.1
SQ-PTT	Superior APTT	55.2	*** Note: New n. H		980	22-1-36.1
SQ-PTT	APTT	50.0		H	980	22-1-36.1
SQ-PTT	Superior APTT	34.3			980	22-1-36.1
SQ-PTT	APTT	34.0			980	22-1-36.1
SQ-PTT	PTT	22.6			980	22-1-34.1
SQ-PTT	APTT	37.4		H	980	22-1-34.1
SQ-PTT	PTT	27.1	SLT HEMOLYS	H	980	22-1-34.1
SQ-PTT	APTT	36.1		H	980	22-1-34.1
SQ-PTT	APTT	36.4	MODERATE H	H	980	22-1-34.1

The Data Warehouse at Mass General Brigham

Mass General Brigham Enterprise Scope

Research Patient Data Registry

Multiple Systems at MGB:
Billing Data
Epic Data
Research Data (consent to contact)
Specimen Data
Laboratory Data
...



FINDING PATIENTS

RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

Query items

Person who is using tool

Query construction

Results - broken down by number distinct of patients

The screenshot shows the RPDR Enhanced Query Tool interface. At the top, there's a toolbar with icons for Tools and Help, and a menu bar with File, Home, Search, Favorites, etc. Below the toolbar is a navigation bar with Address (http://rpdrweb/rpdrwebclient/querytool.aspx?res=768), Logging (Duey, Stacey A. in workgroup of Shawn Murphy, MD), and a green status indicator.

The main area is titled "RPDR Enhanced Query Tool" and contains several tabs: Find Patients, Match Controls, Analyze Patients, Request Detailed Data, Request Specimens, Understand Data, and Get Help. The "Query Items" tab is selected, showing a tree view of query items categorized under Standard Query Items (e.g., Encounter detail, Demographics detail, Diagnosis, Diagnosis Related Groups, Laboratory tests, Medications, Microbiology, Molecular medicine, Health History, Providers, Procedures, Transfusion Services) and Custom Query Items (e.g., Query Library, Patient Sets).

The central workspace is titled "Query Name: Query undefined ..." and features three groups for constructing queries:

- Group 1 of 3:** Contains a checkbox for "Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group." Below it are buttons for Dates, Recorded>0x, and Exclude.
- Group 2 of 3:** Contains a checkbox for "Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group." Below it are buttons for Dates, Recorded>0x, and Exclude.
- Group 3 of 3:** Contains a checkbox for "Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group." Below it are buttons for Dates, Recorded>0x, and Exclude.

At the bottom of the workspace are buttons for Reset All, Run Query, and New Group. The "Run Query" button is highlighted in blue. Below these buttons are four filter sections: Gender (Female, Male), Age (40, 80), Race (A, B, H, W, O, U), and Vital (Alive, Deceased). The "Run Query" button has a tooltip: "GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT".

The bottom of the interface includes a "Done" button and a "Local intranet" link.

RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help



Address <http://rpdrweb/rpdrwebclient/querytool.aspx?res=768>

RPDR Enhanced Query Tool

Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD

Find Patients Match Controls Analyze Patients Request Detailed Data Request Specimens Understand Data Get Help

Query Items Find Terms Previous Queries

Standard Query Items

- + Encounter detail
- + Demographics detail
- + Diagnosis
 - Circulatory system
 - + Acute Rheumatic fever
 - + Arterial vascular disease
 - Cardiac problem-Oncall
 - Cardiac risk factors-Oncall
 - Cardiac risk stratification-Oncall
 - + Cerebrovascular disease
 - + Chronic Rheumatic heart disease
 - + Disease of capillaries
 - + Diseases of pulmonary circulation
 - + Hypertensive disease
 - Ischemia-Oncall
 - Ischemic heart disease
 - + Acute myocardial infarction
 - + Angina pectoris
 - Ischemic heart disease-Oncall
 - Old myocardial infarction
 - + Other acute and subacute forms of ischemic
 - + Other forms of chronic ischemic heart disease
 - + Other forms of heart disease
 - Vascular problem-Oncall
 - + Venous and lymphatic disease
 - Conditions in the perinatal period
 - Congenital anomalies
 - Digestive system
 - Endocrine disorders
 - Events of pregnancy

Query Name: Jsut Diagnos AMI

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT Sensitivity < Reset all groups to >0 >Specificity

Group 1 of 3

Dates Recorded>0x Exclude

One or more items recorded

+ Acute myocardial infarction

Group 2 of 3

Dates Recorded>0x Exclude

Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.

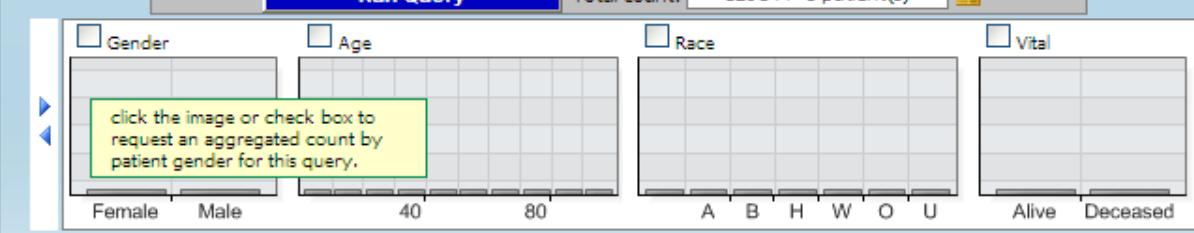
Group 3 of 3

Dates Recorded>0x Exclude

Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.

Reset All First Group Shift Left Shift Right Last Group New Group

Run Query Total count: 120144±3 patient(s)



RPDR Enhanced Query Tool - Microsoft Internet Explorer provided by Partners HealthCare System

File Edit View Favorites Tools Help



Address <http://rpdrweb/rpdrwebclient/querytool.aspx?res=768>

RPDR Enhanced Query Tool

Logging: Duey, Stacey A. in workgroup of Shawn Murphy, MD

Find Patients Match Controls Analyze Patients Request Detailed Data Request Specimens Understand Data Get Help

Query Items Find Terms Previous Queries

Search For:

Containing ck-mb index

Find

All Categories

Search Items
CK-MB Index (Group:CKMBRI)
CK-MB INDEX (Test:sc400,4452)

Query Name: Acute myocardia..., CK-MB Index (Gr... on 01/24/2011 #2

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT

Sensitivity <

Reset all groups to >0

>Specificity

Group 1 of 3

Dates Recorded>0x Exclude
 One or more items recorded
+ Acute myocardial infarction

Group 2 of 3

Dates Recorded>0x Exclude
 One or more items recorded
+ CK-MB Index (Group:CKMBRI) > 3.5

Group 3 of 3

Dates Recorded>0x Exclude
 Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.

Reset All

First Group

Shift Left

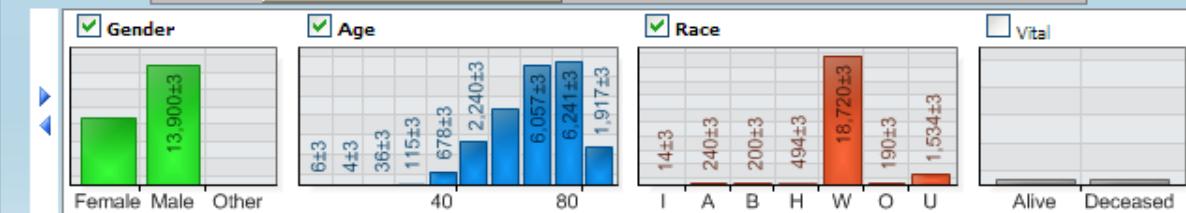
Shift Right

Last Group

New Group

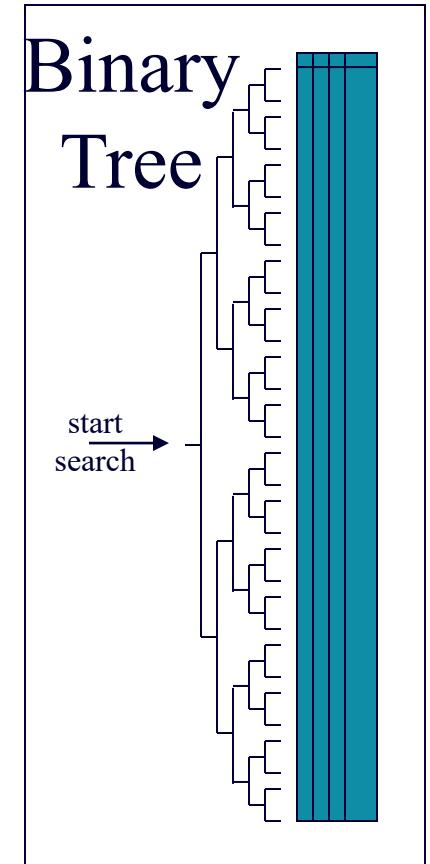
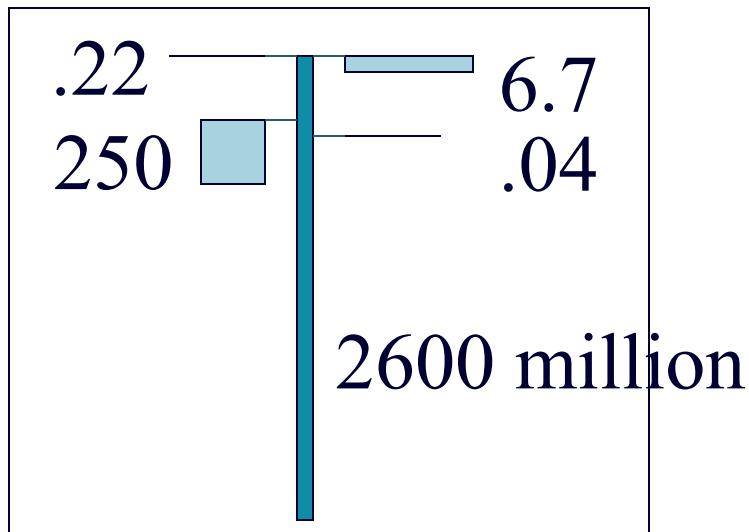
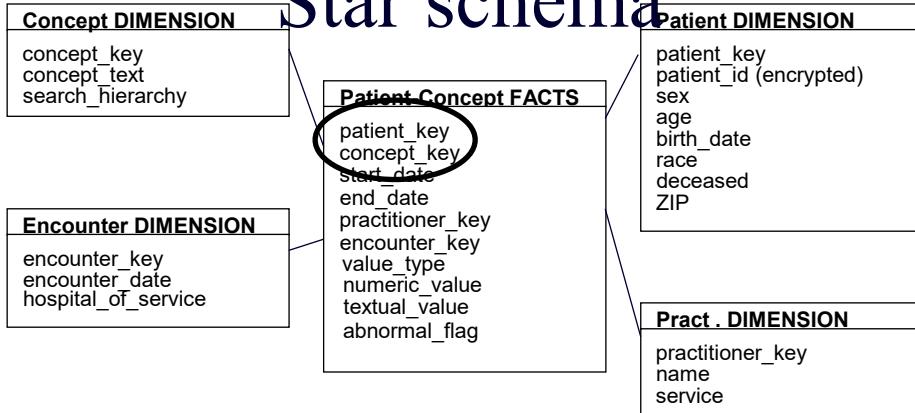
Run Query

Total count: 21647±3 patient(s)



Theory of Kimball translated to Healthcare Data

Star schema



RPDR Detailed Data Request Wizard -- Web Page Dialog



Welcome to the RPDR Data Request Wizards

The RPDR is a HIPAA compliant system, which returns aggregate patient information via a Query Tool, based on user-defined criteria. With proper IRB approval, RPDR users can:

- use their previously queried patient set
- or import their own approved set of Medical Record Numbers

to request detailed or identified patient clinical data. These wizards are included in the RPDR for human research investigators to request identified patient data from their respective Partners sites.

You are now launching a wizard in order to request identified patient data.

Your request must conform and comply with the allowances of your Partners sponsored IRB human studies protocol. This responsibility rests entirely on the faculty sponsor who is requesting the identified data or who is approving the request of identified data from a workgroup member. It is very important that the correct IRB protocol number be chosen for each request of protected health information.

This information is protected under the Partners Privacy and Confidentiality Policy and provided with approval by the Human Research Committee only for the use specified in your protocol number. It may not be used for any other purpose without specific approval by the Human Research Committee. It may not be distributed to any individual not specifically authorized under that approval. The data must be managed in a manner that complies with HIPAA Security Regulations.

I accept responsibility for the data returned by this query.

[Accept](#) [Cancel](#)

Partners Healthcare System HIPAA Compliance

Additional HIPAA information for the research community is available from these links, sponsored by Partners and the Human Research Council (PHRC).

[HIPAA and the Privacy Rule](#)

[HIPAA Central](#)

RPDR Detailed Data Request Wizard -- Web Page Dialog

RPDR DETAILED DATA REQUEST WIZARD

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

Select protocol number(s)

Partners IRB (required): mgh-demo-1

Title: RPDR protocol - Demonstration IRB number for Dr. Murphy

Status: Active

Newton Wellesley Hospital IRB: NWH Demo 1

Title: test

Status: Active

Spaulding Rehabilitation Hospital IRB:

Options for returned set of patients:

Create a static set of patients from this query that can be used in other RPDR queries

Rerun the base query shown above to obtain a fresh set of patients

Help

< Back

STEP 3

Next >

Cancel

RPDR Detailed Data Request Wizard -- Web Page Dialog

RPDR DETAILED DATA REQUEST WIZARD

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

Please select if you would like a HIPAA-defined (deidentified)
limited data set or an identified data set

What's a limited data set?

Limited Data Set

- The files that result from this request will be available in a protected file share with no special encryption.

Identified Data Set

- The text files that result from this request will be encrypted and the Microsoft Access file will be password protected. In order to access the data, a password will be provided.

Help

< Back

STEP 8

Next >

Cancel

RPDR Detailed Data Request Wizard -- Web Page Dialog

RPDR DETAILED DATA REQUEST WIZARD

Using IRB#mgh-demo-1 (found in the RPDR Identified database) to obtain data from the RPDR
You are logged in as Murphy, Shawn N. in workgroup Shawn Murphy, MD

Select the types of data that should be returned from the RPDR

Only data allowed by your protocol should be chosen

(Identified data sets will always return a set of identified patient medical numbers)

Detail Data Items

- + Demographic Data
- + Identifying Patient Information - not available for Limited Data Sets
- + LMR (Longitudinal Medical Record)
- + Medications, Diagnoses and Procedures
- + Medications, Diagnoses and Procedures from Billing Data - only visits where query criteria occur all in the same visit
- + Patient Clinical Reports- not available for Limited Data Sets
 - Cardiology Reports
 - Discharge Summaries
 - Endoscopy Reports
 - Microbiology Data
 - Operative Notes
 - Pathology Reports
 - Pulmonary Reports
 - Radiology Reports
 - Transfusion Data, Blood Bank Data

Help

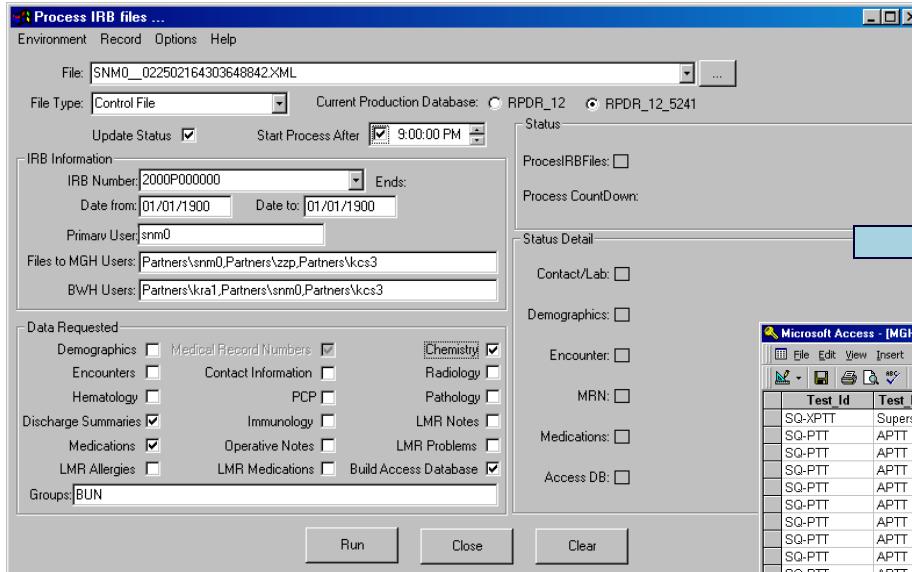
< Back

STEP 9

Next >

Cancel

Detailed data is gathered for request and distributed



Output files placed
in special directory

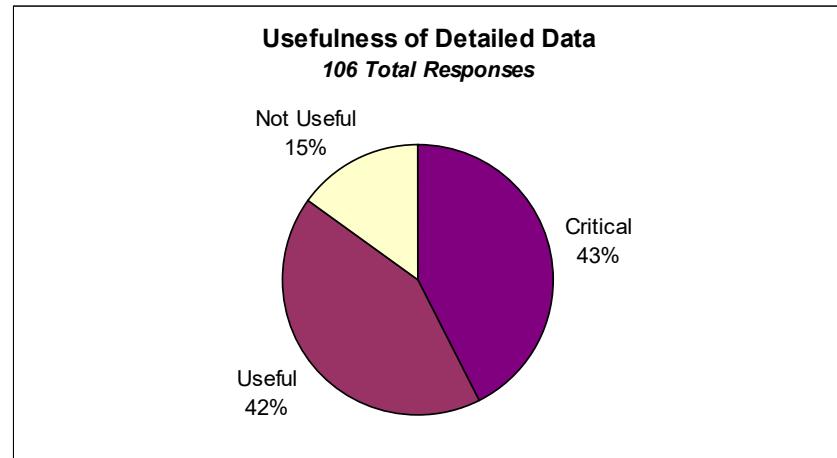
Test_Id	Test_Description	Result	Result_Text	Abnormal_Flag	Reference_Unt	Reference_Range
SQ-XPTT	Superstat APTT	29.8			sec	22.1-35.1
SQ-PTT	APTT	32.8			sec	22.1-35.1
SQ-PTT	APTT	37.8	H		sec	22.1-35.1
SQ-PTT	APTT	46.4			sec	22.1-35.1
SQ-PTT	APTT	43.1	MODERATELY H		sec	22.1-35.1
SQ-PTT	APTT	25.7			sec	22.1-35.1
SQ-PTT	APTT	23.7			sec	22.1-35.1
SQ-PTT	APTT	25.4			sec	22.1-35.1
SQ-PTT	APTT	24.7			sec	22.1-35.1
SQ-PTT	APTT	24.0			sec	22.1-35.1
SQ-PTT	APTT	24.7			sec	22.1-35.1
SQ-XPTT	Superstat APTT	31.3			sec	22.1-35.1
SQ-PTT	APTT	34.5			sec	22.1-35.1
SQ-PTT	APTT	40.0	H		sec	22.1-35.1
SQ-PTT	APTT	45.0	H		sec	22.1-35.1
SQ-XPTT	Superstat APTT	55.2	*** Note: New n	H	sec	22.1-35.1
SQ-PTT	APTT	33.6			sec	22.1-35.1
SQ-XPTT	Superstat APTT	34.3			sec	22.1-35.1
SQ-PTT	APTT	37.9	H		sec	22.1-35.1
SQ-PTT	APTT	22.6			sec	22.1-34.1
SQ-PTT	APTT	37.4	H		sec	22.1-34.1
SQ-PTT	APTT	37.2	SLT HEMOLYS	H	sec	22.1-34.1
SQ-PTT	APTT	35.1	H		sec	22.1-34.1
SQ-PTT	APTT	36.4	MODERATE HE H		sec	22.1-34.1

Data is gathered from RPDR
and other MGB sources

Files include Small Database

One years usage of RPDR

- 4526 registered users, 1113 new in just 2019
- 834 teams/year gathering data for research studies
- 4472 detailed patient data sets returned to these teams in 2019, 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.
 - Calculated over 4 years (FY15-FY19) the total agreement amounts were \$2.27 Billion, making per year consumption critically dependent on RPDR \$244 Million.



Rapid investigation of QTc prolongation

■ FDA warning 2011 for Celexa

Safety Announcement:

[8-24-2011] "should no longer be used at doses greater than 40 mg per day because it can cause abnormal changes in the electrical activity of the heart."

■ But, did NOT include Lexapro (which is active ingredient of Celexa [s-enantiomer])

■ Shown to be true with RPDR-derived data set with >38,000 EKGs obtained within 14 – 90 day window after medication initiated

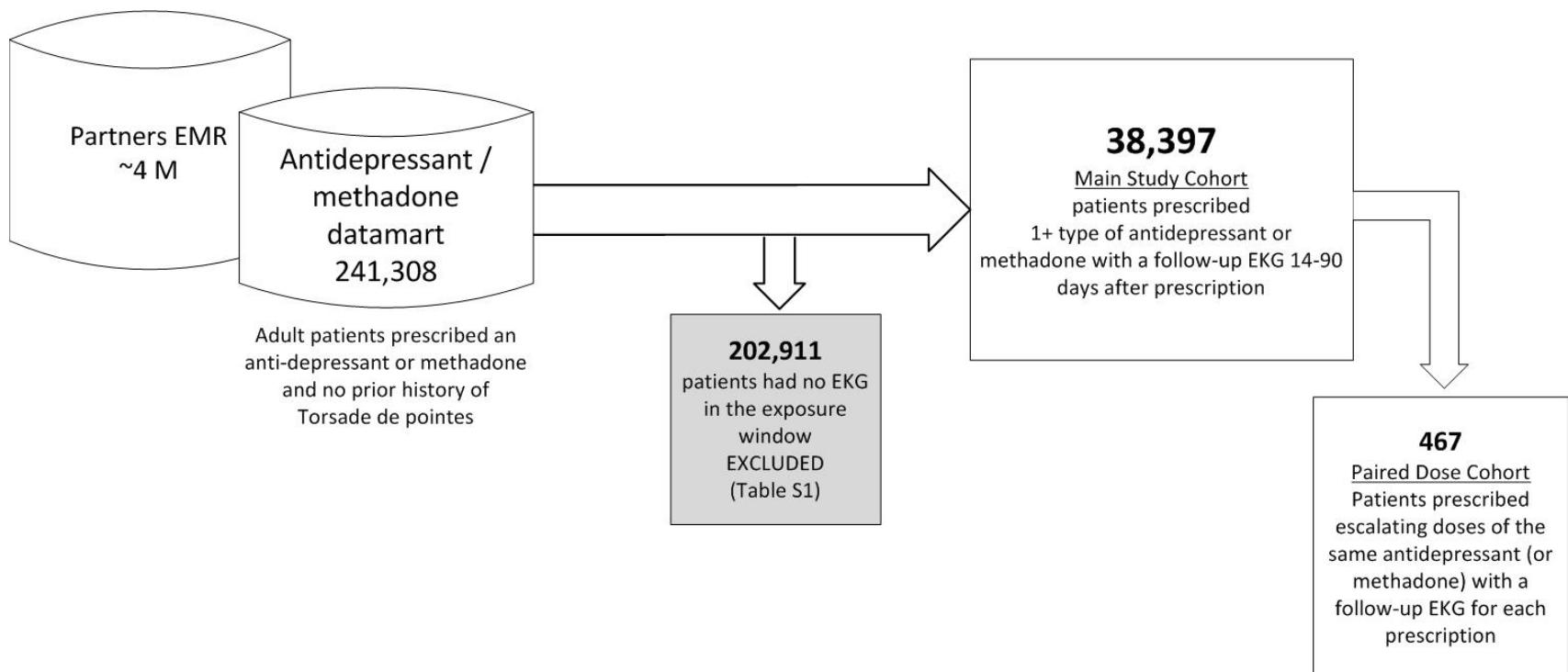
Anti-depressant	Adjusted model†	
	prolongation n	p-value
SSRI		
Citalopram (Celexa)	2.85	0.004
Escitalopram (Lexapro)	3.80	< 0.001
Fluoxetine (Prozac)	1.44	0.150
Paroxetine (Paxil)	0.07	0.943
Sertraline (Zoloft)	0.87	0.383
Other anti-depressants		
Amitriptyline	4.10	< 0.001
Bupropion	-2.15	0.032
Duloxetine	0.60	0.547
Mirtazapine	-1.46	0.145
Nortriptyline	1.23	0.219
Venlafaxine	1.15	0.251
previously known prolonger		
Methadone	5.32	< 0.001

† Adjusted for age, gender, race, type of insurance, history of major depression, history of myocardial infarction and Charlson comorbidity score

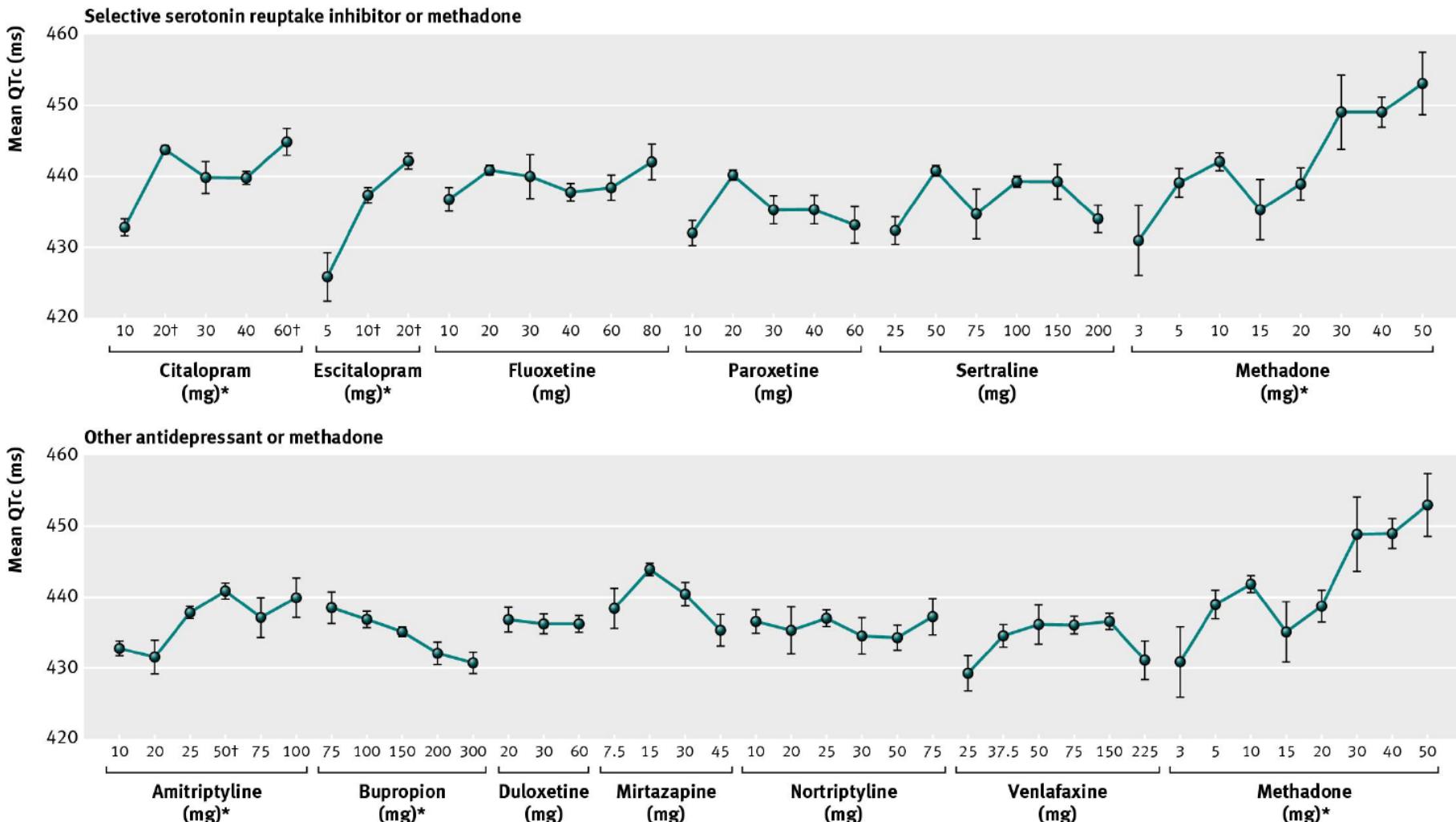
Roy Perlis MD, MSc and team

Relevant Cohorts of Patients are Gathered through RPDR and Detailed Data Obtained

- Medication use by individual patients over time
- Patient EKG QTc values at various time points



Results: QTc interval and medication use



* 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

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
- Digital medicine is delivered into clinical care

RPDR Evolved into international “Informatics for Integrating Biology and the Bedside (i2b2)” sponsored by the National Institutes of Health, what is it?

- Software for explicitly organizing and transforming person-oriented clinical data to a way that is optimized for clinical genomics research
 - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
 - Software is built in a modular pattern that allows additions without disturbing core parts
 - Available as open source at <https://www.i2b2.org>

I2b2 Community Software distributed as open source

community.i2b2.org/wiki/

i2b2 Community Wiki Search Log in

Home Documentation Get Software Community Projects Community Events Working Groups Support

i2b2 Community Wiki

Quick Launch

i2b2 Web Client Demo

i2b2 Github

Pages

- i2b2 Community Projects
- i2b2 Documentation
- i2b2 Software
 - i2b2 Hive
 - i2b2 Software Downloads Links

Recently updated

- i2b2 Documentation updated Feb 07, 2020
- Welcome to the i2b2 Community Wiki updated Jan 02, 2020
- i2b2 Software updated Jan 02, 2020
- i2b2 Software updated Dec 27, 2019
- i2b2 Software Downloads Links updated Dec 27, 2019

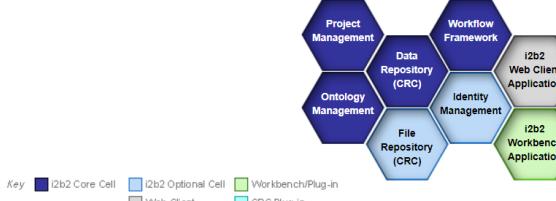
Show More

Dashboard

Welcome to the i2b2 Community Wiki

i2b2 is an open-source clinical data warehousing and analytics research platform used at over 250 locations worldwide. i2b2 enables sharing, integration, standardization, and analysis of heterogeneous data from healthcare and research.

- The i2b2 Community is a life-sciences-focused open-source, open-data community. This wiki is the central place for the i2b2 Community to communicate and share projects with other users. Here you will find the latest information on the i2b2 Software, what others in the community are doing, and find resources to help answer any questions you may have about the i2b2.
- i2b2 is part of the [i2b2 transSMART Foundation](#), which brings together an NIH-funded enterprise clinical research platform (i2b2) and pharma-developed software for translational research studies (transSMART).

The diagram illustrates the core components of the i2b2 system. It features a central hexagonal node labeled "Data Repository (CRC)" connected to five surrounding hexagonal nodes: "Project Management" (top), "Workflow Framework" (top-right), "Identity Management" (bottom-right), "File Repository (CRC)" (bottom), and "Ontology Management" (left). Additionally, there are three separate hexagonal nodes: "i2b2 Web Client Application" (top-right), "i2b2 Workbench Application" (bottom-right), and "i2b2 Hive" (bottom-right). A legend below the diagram defines the colors and symbols: dark blue for "i2b2 Core Cell", light blue for "i2b2 Optional Cell", green for "Workbench/Plug-in", and cyan for "CRC Plug-in".

 Older, unsupported add-ons are on the Archived Optional Components page.

Documentation	Get Software			Announcements
Getting Started <ul style="list-style-type: none">Installation GuideUpgrade GuideVideos and Tutorials	Software	Version		What's New in i2b2?
	i2b2 Core Server Source Code	1.7.12 (December, 2019)	Download	i2b2 Release 1.7.12 now available! <ul style="list-style-type: none">Easier i2b2 installRedesigned FindTermsREDCAP import to i2b2ACT OntologySupport for OKTA and NTLM2 User Authenticationi2b2 Workbench download for Windows 64-bit platform... and much more!
For Developers <ul style="list-style-type: none">Server-side MessagingServer ArchitectureServer-side DesignWeb Client DesignRelease Notes	i2b2 Core Data Source Code	1.7.12 (December, 2019)	Download	
	i2b2 Web Client	1.7.12 (December, 2019)	Download	
For End Users	i2b2 Documentation	1.7.12 (December, 2019)	Release Notes Upgrade Guide	Important links: <ul style="list-style-type: none">i2b2 transSMART Working GroupsNew Project: i2b2 on OMOP

I2b2 Implementations

CTSA's

- Boston University
- Case Western Reserve University (*including Cleveland Clinic*)
- Children's National Medical Center (GWU), Washington D.C.
- Duke University
- Emory University (*including Morehouse School of Medicine and Georgia Tech*)
- Harvard University (*including Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, Children's Hospital Boston, Dana Farber Cancer Center, Joslin Diabetes Center, Massachusetts General Hospital*)
- Medical University of South Carolina
- Medical College of Wisconsin
- Oregon Health & Science University
- Penn State Milton S. Hershey Medical Center
- Tufts University
- University of Alabama at Birmingham
- University of Arkansas for Medical Sciences
- University of California Davis
- University of California, Irvine
- University of California, Los Angeles*
- University of California, San Diego*
- University of California San Francisco
- University of Chicago
- University of Cincinnati (*including Cincinnati Children's Hospital Medical Center*)
- University of Colorado Denver (*including Children's Hospital Colorado*)
- University of Florida
- University of Kansas Medical Center
- University of Kentucky Research Foundation
- University of Massachusetts Medical School, Worcester
- University of Michigan
- University of Pennsylvania (*including Children's Hospital of Philadelphia*)
- University of Pittsburgh (*including their Cancer Institute*)
- University of Rochester School of Medicine and Dentistry
- University of Texas Health Sciences Center at Houston
- University of Texas Health Sciences Center at San Antonio
- University of Texas Medical Branch (Galveston)
- University of Texas Southwestern Medical Center at Dallas
- University of Utah
- University of Washington
- University of Wisconsin - Madison (*including Marshfield Clinic*)
- Virginia Commonwealth University
- Weill Cornell Medical College

Academic Health Centers (does not include AHCs that are part of a CTSA):

- Arizona State University
- City of Hope, Los Angeles
- Georgia Health Sciences University, Augusta
- Hartford Hospital, CN
- HealthShare Montana
- Massachusetts Veterans Epidemiology Research and Information Center (MAVERICK), Boston
- Nemours
- Phoenix Children's Hospital
- Regenstrief Institute
- Thomas Jefferson University
- University of Connecticut Health Center
- University of Missouri School of Medicine
- University of Tennessee Health Sciences Center
- Wake Forest University Baptist Medical Center

HMOs:

- Group Health Cooperative
- Kaiser Permanente

International:

- Georges Pompidou Hospital, Paris, France
- Hospital of the Free University of Brussels, Belgium
- Inserm U936, Rennes, France
- Institute for Data Technology and Informatics (IDI), NTNU, Norway
- Institute for Molecular Medicine Finland (FIMM)
- Karolinska Institute, Sweden
- Landspitali University Hospital, Reykjavik, Iceland
- Tokyo Medical and Dental University, Japan
- University of Bordeaux Segalen, France
- University of Erlangen-Nuremberg, Germany
- University of Goettingen, Goettingen, Germany
- University of Leicester and Hospitals, England (Biomed. Res. Informatics Ctr. for Clin. Sci)
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

Companies:

- Johnson and Johnson (TransMART)
- GE Healthcare Clinical Data Services

i2b2 Software adapts through new plugins

The screenshot shows a Windows Internet Explorer window displaying the 'Dashboard - i2b2 Wiki' page. The URL in the address bar is <https://community.i2b2.org/wiki/dashboard.action>. The page lists several 'Related Project' entries, each with a globe icon, a link, a brief description, and two small icons (a magnifying glass and a document). The projects listed are:

- [i2b2 Sponsored Project - i2b2 Web Client](#)
The i2b2 Web Client is one of several core projects that are directly sponsored by the i2b2 team.
- [i2b2 Sponsored Project - NCBO Ontology Tools](#)
Tools to extract and organize ontologies from the NCBO, organized by Lori Phillips
- [Related Project - Clinical Trender](#)
The Clinical Trender aims to allow researchers to track and visualize certain clinical variables related to a selected p...
- [Related Project - CRC Tester](#)
A Workbench plugin that tests the CRC web services by Mike Mendis
- [Related Project - Crimson](#)
A project to make specimens available through i2b2 infrastructure, organized by Lynn Bry
- [Related Project - ExportXLS](#)
The i2b2 Web Client Plugin that tabulates patient data & applicable specified concepts; and facilitates export to spread...
- [Related Project - FACE caGrid CQL2 Data Source](#)
A caGrid/TRIAD data service that runs i2b2 queries via the RESTful interface to CRC
- [Related Project - Familial, Associational, & Incidental Relationships \(FAIR\) Initiative.](#)
A collection of DBA tools and webclient plugins to facilitate the identification of related concepts amongst related pat...
- [Related Project - Federated Query Simulations](#)
Simulations of federated query tools that return aggregate counts, such as SHRINE, by Griffin Weber

At the bottom of the browser window, the status bar shows 'Internet | Protected Mode: Off' and a zoom level of '100%'. The browser's title bar reads 'Dashboard - i2b2 Wiki - Windows Internet Explorer'.

Use NLP to extract the relevant features from the set of patient notes.

Programmer's File Editor - [050210_1629\MiniDem1.txt]

SOCIAL HISTORY: The patient is married with four grown daughters, uses tobacco has wine with dinner. **Smoker**

PRINCIPAL DIAGNOSIS: LEFT LOWER LOBE PNEUMONIA

SECONDARY DIAGNOSIS: 1. CHF 2. HEART FAILURE

HISTORY OF PRESENT ILLNESS: SOCIAL HISTORY: The patient is a nonsmoker. No alcohol. **Non-Smoker**

PAST MEDICAL HISTORY: (1) Hip Fracture. (2) Bronchiectasis.

BRIEF RESUME OF HOSPITAL COURSE: 63 yo woman with COPD, 50 pack-yr tobacco (quit 3 wks ago), **Past Smoker**

ALLERGIES: (1) Aspirin. (2) Ciprofloxacin. (3) Penicillin.

SOCIAL HISTORY: The patient lives alone and denies tobacco or alcohol use. **Unclear smoking history** ???

PATIENT COMMUNICATION: Temperature 97.2, pulse 60, respirations 20, blood pressure 160/63, oxygen saturation 95% on room air. HEENT: Normocephalic and atraumatic. Pupils equal and reactive.

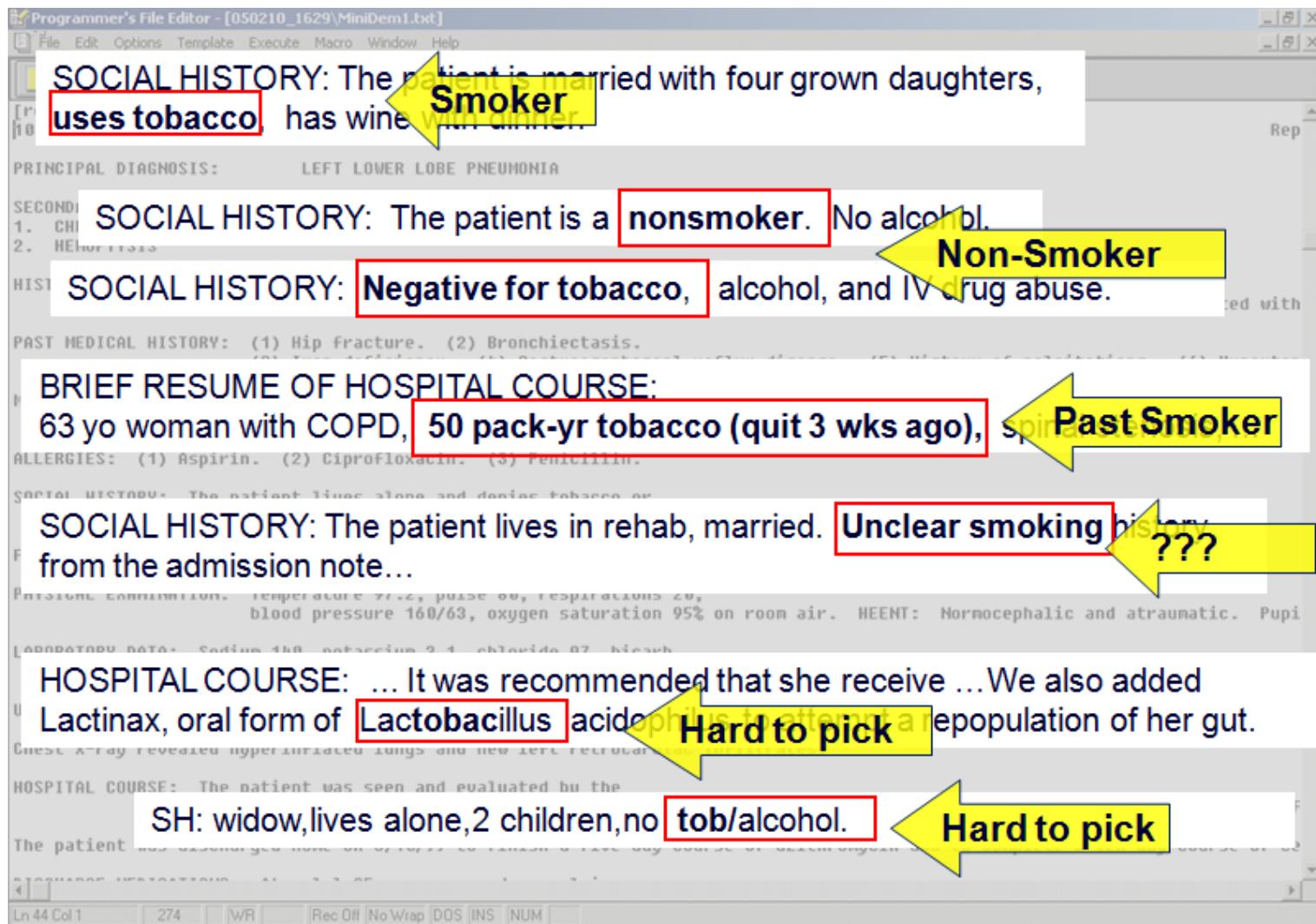
LABORATORY DATA: Sodium 140, potassium 3.4, chloride 97, bicarbonate 24

HOSPITAL COURSE: ... It was recommended that she receive ... We also added Lactinax, oral form of **Lactobacillus acidophilus** to attempt a repopulation of her gut. **Hard to pick**

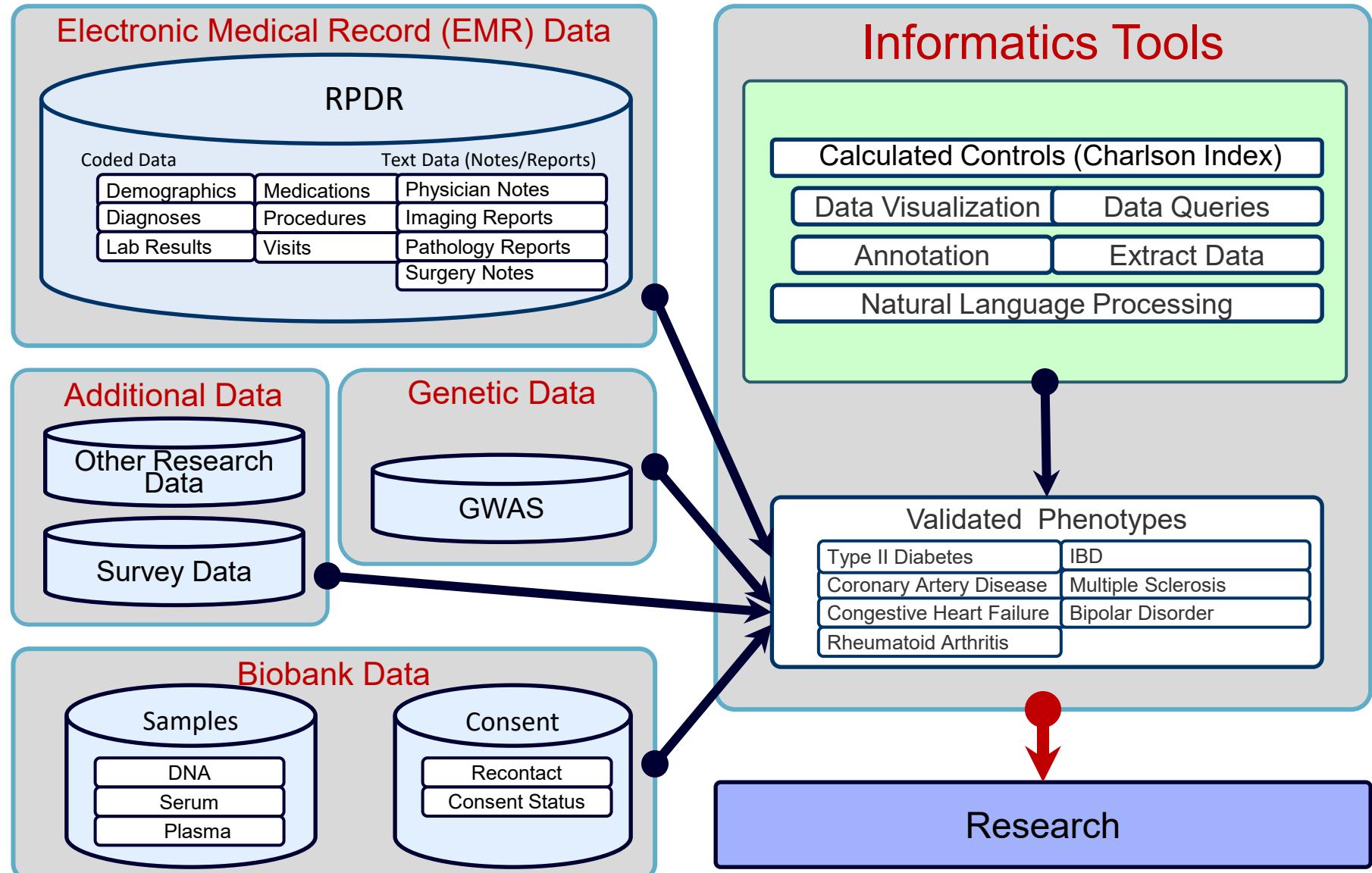
HOSPITAL COURSE: The patient was seen and evaluated by the physician. The patient ... **Hard to pick**

SH: widow, lives alone, 2 children, no tobacco/alcohol.

Ln 44 Col 1 274 WR Rec Off No Wrap DOS INS NUM

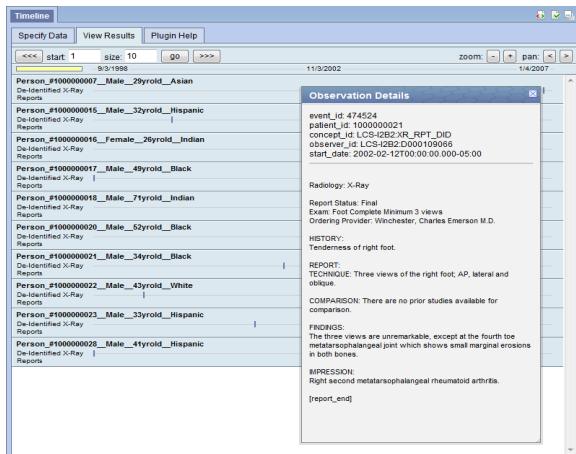


Data Integration in Big Data Commons

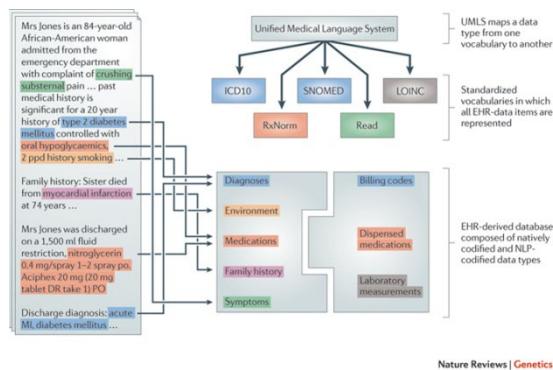


Curating a Disease Algorithm

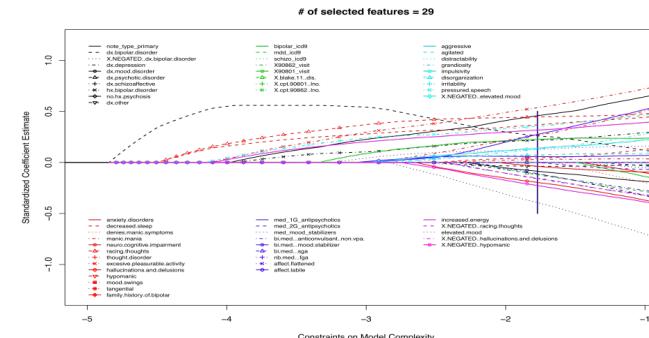
1. Create a gold standard training set.



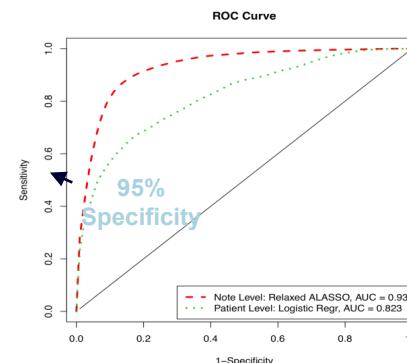
2. Create a comprehensive list of features from patient's electronic data that describe the disease of interest



3. Develop the classification algorithm. Using the data analysis file and the training set from step 1, assess the frequency of each variable. Remove variables with low prevalence. Apply adaptive LASSO penalized logistic regression to identify highly predictive variables for the algorithm



4. **Apply the algorithm to all subjects** in the superset and assign each subject a probability of having the phenotype



Biobank Portal | Curated Diseases

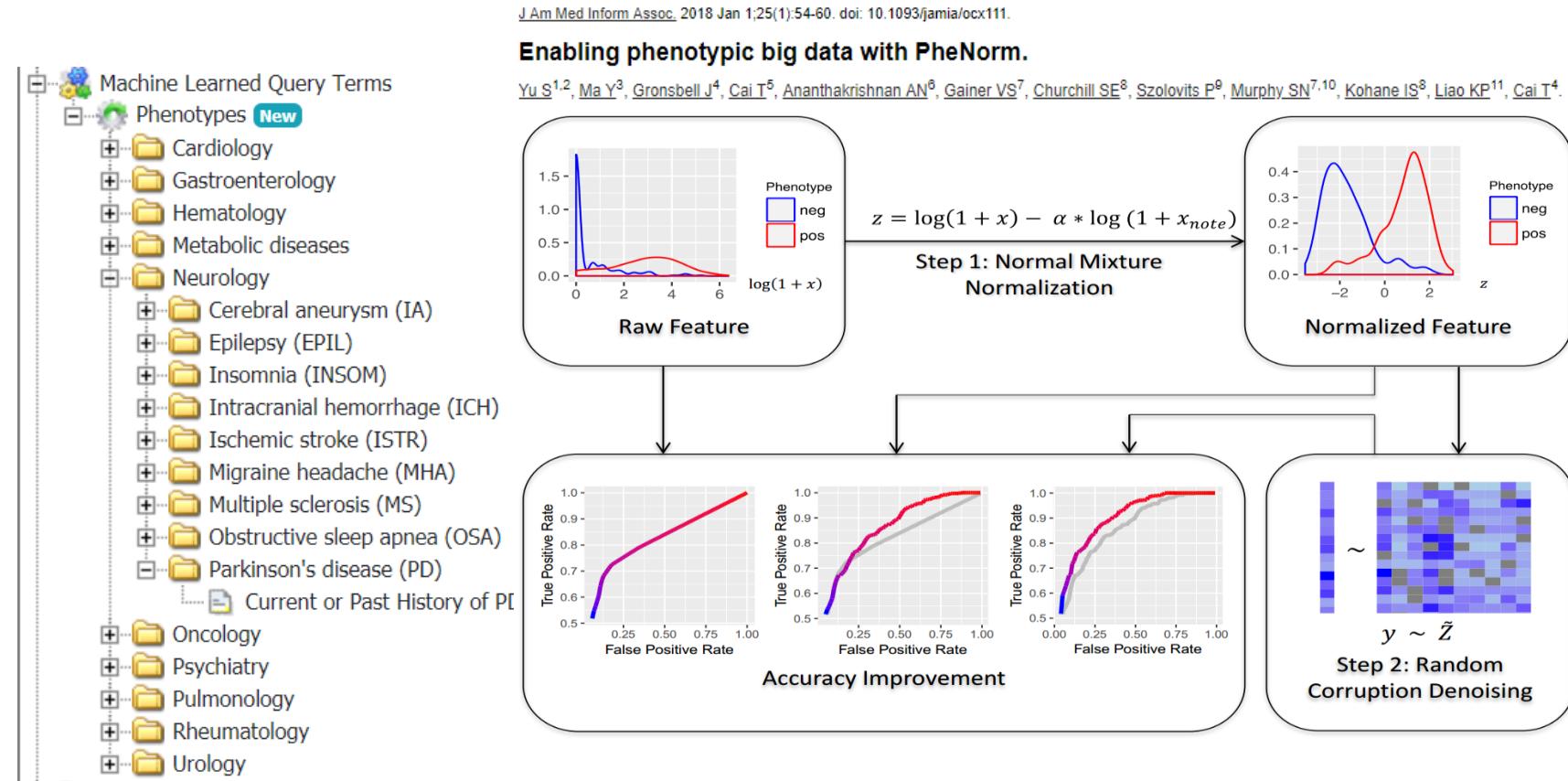
Validated Phenotype	Count*	Predictive Positive Value
Bipolar Disease	71	89%
Congestive Heart Failure	387	90%
Coronary Artery Disease	2,420	97%
Crohn's Disease	453	90%
Multiple Sclerosis	94	90%
Rheumatoid Arthritis	550	90%
Type 2 Diabetes Mellitus	1,887	97%
Ulcerative Colitis	330	90%

Healthy Controls based on Charlson Index	Count**
0 – 10-year survival probability is >98.3%	2,206
1 – 10-year survival probability is >95.87%	4,343
2 – 10-year survival probability is >90.15%	6,545

* Based on 15,880 patients

** Based on 21,300 patients

Automated Learning Algorithms enabled in RPDR such as PheNorm Algorithm



Machine Learned Phenotypes

- Abdominal hernia
- Acute bronchitis and bronchiolitis
- Acute pancreatitis
- Alcoholism
- Alzheimer's disease
- Aortic aneurysm
- Aplastic anemia
- Atrial fibrillation
- Atrioventricular block
- Autism spectrum disorders
- Basal cell carcinoma
- Bipolar Disease
- Bladder cancer
- Brain cancer
- Breast cancer
- Cerebral aneurysm
- Cholelithiasis
- Chronic pancreatitis
- Chronic sinusitis
- Coronary atherosclerosis
- Crohn's disease
- Deep vein thrombosis
- Depression
- Diverticulosis and diverticulitis
- Eating disorder
- Epilepsy
- Gastroesophageal reflux disease
- Gout
- Heart valve disorders
- Hyperlipidemia
- Hyperparathyroidism
- Hypertension
- Hypothyroidism
- Insomnia
- Intracranial hemorrhage
- Ischemic stroke
- Leukemia
- Lung cancer
- Melanoma
- Migraine headache
- Multiple sclerosis
- Myocardial infarction
- Neutropenia
- Non-Hodgkin lymphoma
- Obesity
- Obsessive compulsive disorder
- Obstructive sleep apnea
- Ovarian cancer
- Pancreatic cancer
- Parkinson's disease
- Peripheral vascular disease
- Pneumonia
- Polycystic ovaries
- Prostate cancer
- Pulmonary heart disease
- Renal cancer
- Renal failure
- Schizophrenia
- Substance addiction
- Suicidal ideation
- Suicide attempt or self-inflicted injury
- Thyroid cancer
- Tobacco use disorder
- Type 1 diabetes
- Type 2 diabetes
- Ulcerative colitis
- Urinary calculus
- Uterine cancer

Phenotype Automation: Phenotype Quality Dashboard

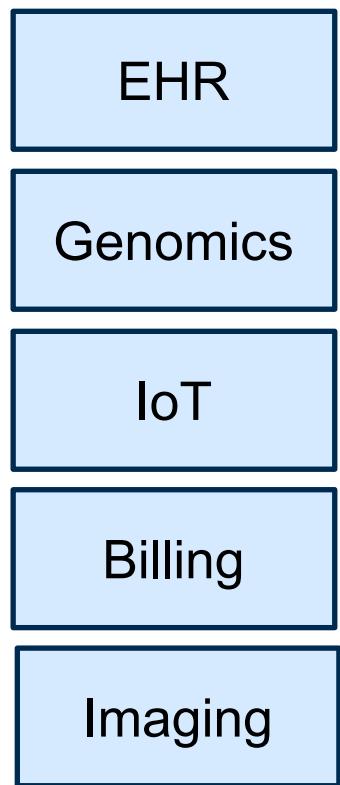
 **Computable Phenotype Dashboard**

	category	PheWAS_code	abbr	PheWAS_name	model	ICD_PPV	ICD_AUC	AUC	PPV	TPR
1	ONC	PheWAS:189.21	BLCA	Bladder cancer	PheNorm_ICD	0.80	0.903	1.000	1.00	0.42
2	ONC	PheWAS:204	LEUK	Leukemia	PheNorm_ICD	0.73	1.000	1.000	1.00	0.91
3	PSYCH	PheWAS:297.1	SI	Suicidal ideation	PheNorm_ICDNLP	0.93	0.786	1.000	1.00	0.43
4	PSYCH	PheWAS:305.2	EATD	Eating disorder	PheNorm_ICDNLP	0.53	0.482	1.000	1.00	1.00
5	NEURO	PheWAS:327.4	INSOM	Insomnia	PheNorm_ICDNLP	0.93	0.821	1.000	1.00	0.50
6	CARDIO	PheWAS:452.2	DVT	Deep vein thrombosis	PheNorm_ICDNLP	0.87	0.692	1.000	1.00	1.00
7	NEURO	PheWAS:817	CONC	Concussion	PheNorm_NLP	0.73	0.682	1.000	1.00	1.00
8	METAB	PheWAS:250.1	T1DM	Type 1 diabetes	PheNorm_ICD	0.17	0.882	0.982	0.982	0.982
9	ONC	PheWAS:184.11	OVCA	Ovarian cancer	PheNorm_ICDNLP	0.60	0.926	0.926	0.926	0.926
10	ONC	PheWAS:182	UTCA	Uterine cancer	PheNorm_ICD	0.50	0.867	0.900	0.900	0.900
11	GI	PheWAS:555.1	CD	Crohn's disease	PheNorm_mean	0.54	0.961	0.961	0.961	0.961

Phenotypes New

- Cardiology
 - Aortic aneurysm (AA)
 - Atrial fibrillation (AFIB)
 - Atrioventricular block (AVB)
 - Coronary atherosclerosis (CAD)
 - Deep vein thrombosis (DVT)
 - Heart valve disorders (HVD)
 - Hypertension (HTN)
 - Myocardial infarction (MI)
 - Peripheral vascular disease (PWD)
 - Pulmonary heart disease (PHD)

Combined with Generative AI can produce Digital Twin of Patient



Digital Twin – Abigail Test

TEST, ABIGAIL 106894405 (PHS) 04/01/1970 (50 y) F
RC: N/A

Search this patient

RECORDS NOTES ALL | ✓

2018-01-12 DR Orders Only
2018-01-09 DR Telephone
2018-01-03 DR Telephone
2017-11-10 DR Orders Only
2017-10-19 DR Orders Only
239 more

STUDIES ALL | ✓

2018-08-31 DR Letter (Out)
2018-08-18 DR Letter (Out)
2015-07-31 DR Telephone

RADIOLOGY ALL | ✓
No records found

MICROBIOLOGY ALL CX | ✓
2016-07-27 MIC SPUTUM
2016-07-27 MIC URINE
2016-07-12 MIC URINE

PATHOLOGY ALL | ✓
No records found

LABS RECENT | ✓
No records found

ALLERGIES
2018-11-26 ALGY PENICILLINS

MEDICATIONS CURRENT | ✓
2020-04-18 MED trazODone (DESYREL) 50 MG tab
2019-09-10 MED methylphenidate HCl (RITALIN LA) ...
2019-09-10 MED methylphenidate HCl (RITALIN) 10 ...
2018-04-11 MED phenylephrine (NEO-SYNEPHRIN)
2015-11-08 MED phenylephrine (NEO-SYNEPHRIN)
2 more

Last updated: Jun 19, 2020 12:41:22 PM

Covid-19

COVID STATUS	RISK FACTORS	SEVERITY				
SARS-CoV-2 PCR...	Height Weight BMI Smoker Healthcare worker Housing Pregnant Recent travel Asthma COPD Diabetes Heart failure High cholesterol Hypertension MI Immunodeficiency Autoimmune disease	ACE inhibitors ARBs Beta blockers Bisphosphonates CCBs Chemotherapy Statins Steroids NSAIDS Thiazide diuretics Transplant Transplant meds Warfarin Flu vaccine BCG vaccine	Pulse Respiratory rate Oxygen Therapy High-flow oxygen CPK ECMO IL-6 SOB/dyspnea Sputum Muscle aches/rigidity Cough Nausea/vomiting/d... Congestive heart failur... Abnormal CXR Abnormal Chest CT	ER-Urgent admit Hypercapnia Pharyngitis Muscle aches/rigidity Mechanical Ventilator... Pneumonia Procalcitonin Lactic acid Potassium ECG Troponin Legionella Strep pneumonia Mycoplasma	Influenza A/B (Posi... RSV (Positive) CPK IL-6 Procalcitonin Lactic acid Potassium ECG Troponin Legionella Strep pneumonia Mycoplasma	ALB AST ALT Bilirubin BNP BUN Cholesterol Creatinine eGFR Hematocrit LDL Lymphocytes Neutrophils Prothrombin Troponin HBV (serology) HCV (serology) HIV

MANAGEMENT

- Cefepime/ceftirax...
- Azithromycin
- Vancomycin
- Oseeltamivir
- Remdesivir
- Hydroxychloroquine
- Darunavir/ribociclib
- Lopinavir/ritonavir
- Interferon beta B1
- Tocilizumab

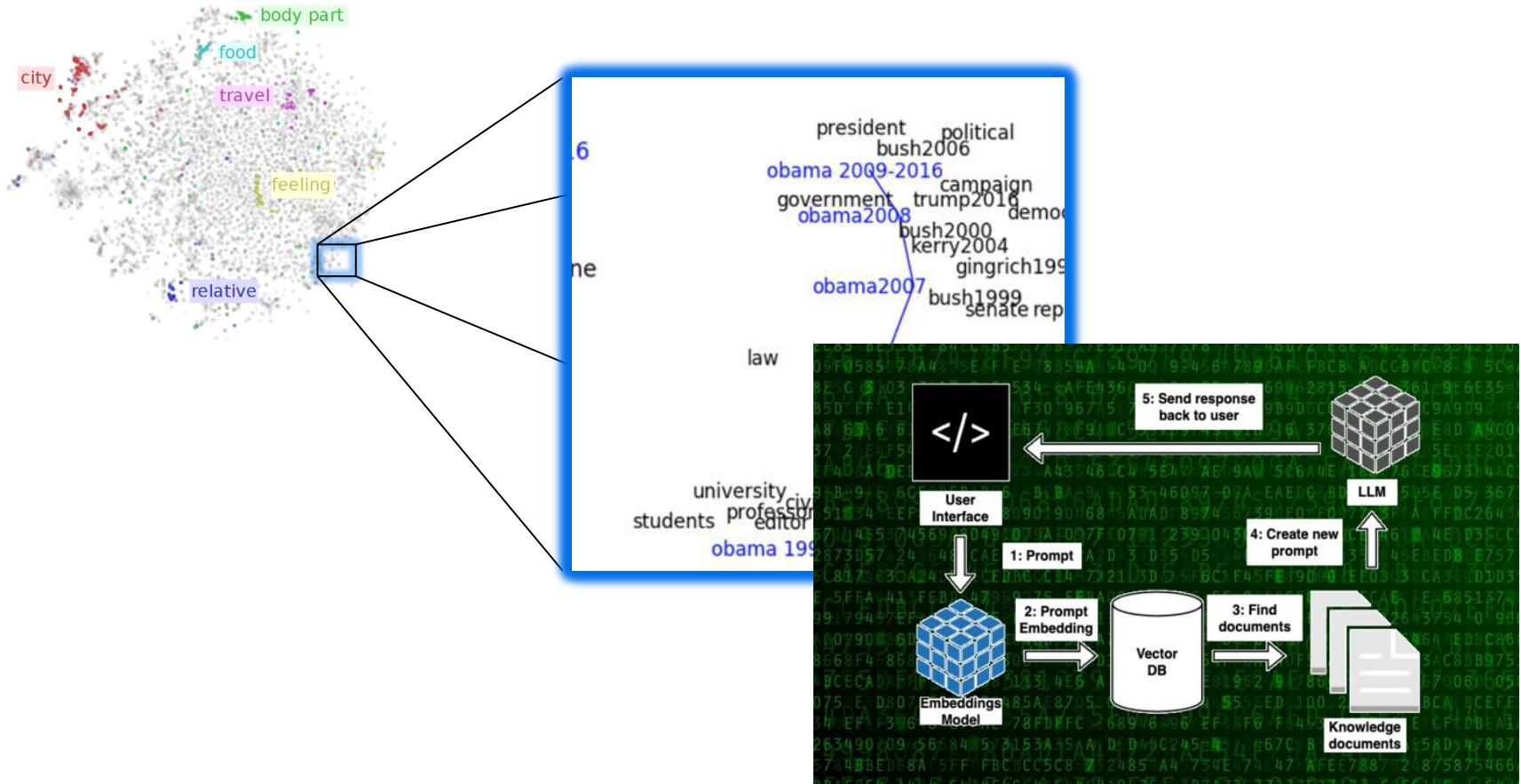
Study Specific Conditions and Common Data Elements

Core Conditions

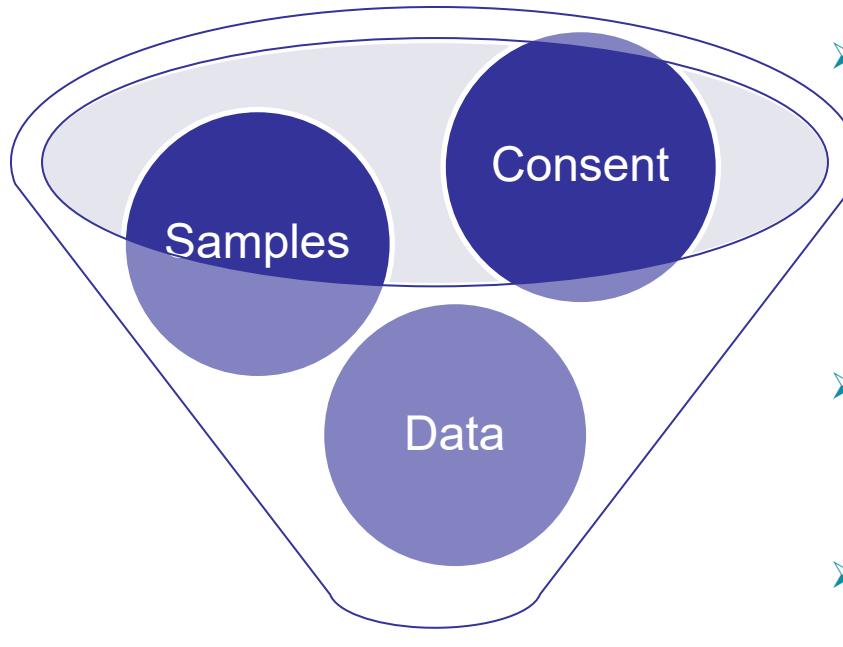
CV	PULM	GI	GU	ID	NEURO	HEME	OTHER
AAA/Dxn	COPD	Divericulosis	AKI	Abscess	Dementia	Anemia	Agitation
AFib/Flutter	DVT	Hemoptysis	CKD	Bacteremia	MS	BIO	Capacity
CAD	ILD	Ketoacidosis	Infec.	Endocrinopathy	Depression	Coag	Depression
Chest Pain	OSA	Na Disorder	HIV/AIDS	Neuropathy	MM	Hypercoag	EBOH
CHF	PE	Pulm HTN	Pneumonia	PRES	Myelodysplasia	TTP/HUS	Drug Abuse
Conduction Dz	PVD	Pulm Nodules	Soft Tissue Inf	Seizures			Noncompliance
MI	Valve Disease		Ulcers	Stroke/TIA			Psych
PTV	VT/VF		Varices	UTI	XRT		Smoking

DATA UPDATED - CLICK HERE TO UPDATE AGAIN

LMM Enhanced interaction with Patient Representation



The MGB Biobank



- The MGB Biobank provides samples (plasma, serum, and DNA) collected from consented patients.
- 140,000 patients have consented to date
- Samples are available for distribution to MGB investigators* to help identify novel Personalized Medicine opportunities that reduce cost and provide better care

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

Improved Clinical Care for All Patients

Genotype Data

The screenshot shows the i2b2 Query Tool interface. On the left, a tree view of terms under 'Biobank Genomics' includes 'dbSNP rs Identifier - 10097' and 'Gene - 10097'. A red circle highlights 'Gene - 10097', and a red arrow points from it to the 'Query Tool' window. The 'Query Tool' window shows a 'Temporal Constraint' section with 'Group 1' containing 'Gene - 10097'. Below it is a yellow box with the text 'drop a term on here'. At the bottom, a 'Search by Gene' dialog is open, showing a dropdown menu with 'APO' selected. A red arrow points from the 'Gene Name*' dropdown to the list of genes below. The list includes: APOA1, APOA1BP, APOA2, APOA4, APOA5, APOB, APOBEC1, APOBEC2, APOBEC3A, APOBEC3A_B, APOBEC3B.

The screenshot shows the i2b2 Query Tool interface. On the left, a tree view of terms under 'Biobank Genomics' includes 'dbSNP rs Identifier - 10097'. A red circle highlights 'dbSNP rs Identifier - 10097', and a red arrow points from it to the 'Query Tool' window. The 'Query Tool' window shows a 'Temporal Constraint' section with 'Group 1' containing 'dbSNP rs Identifier - 1009'. Below it is a yellow box with the text 'drop a term on here'. At the bottom, a 'Search by dbSNP rs Identifier' dialog is open, showing a dropdown menu with 'rs1234' selected. A red arrow points from the 'rs identifier*' dropdown to the list of variants below. The list includes: rs12340033 | C to G, rs12340061 | G to A, rs12340067 | C to T, rs12340088 | T to G, rs12340105 | A to C, rs12340107 | G to T, rs12340117 | G to A, rs12340120 | G to A, rs12340129 | A to G, rs12340149 | G to A, rs12340158 | T to A.

[36](https://community.i2b2.org/wiki/display/IGD>Loading+Genomic+VCF+Files+into+i2b2</p></div><div data-bbox=)

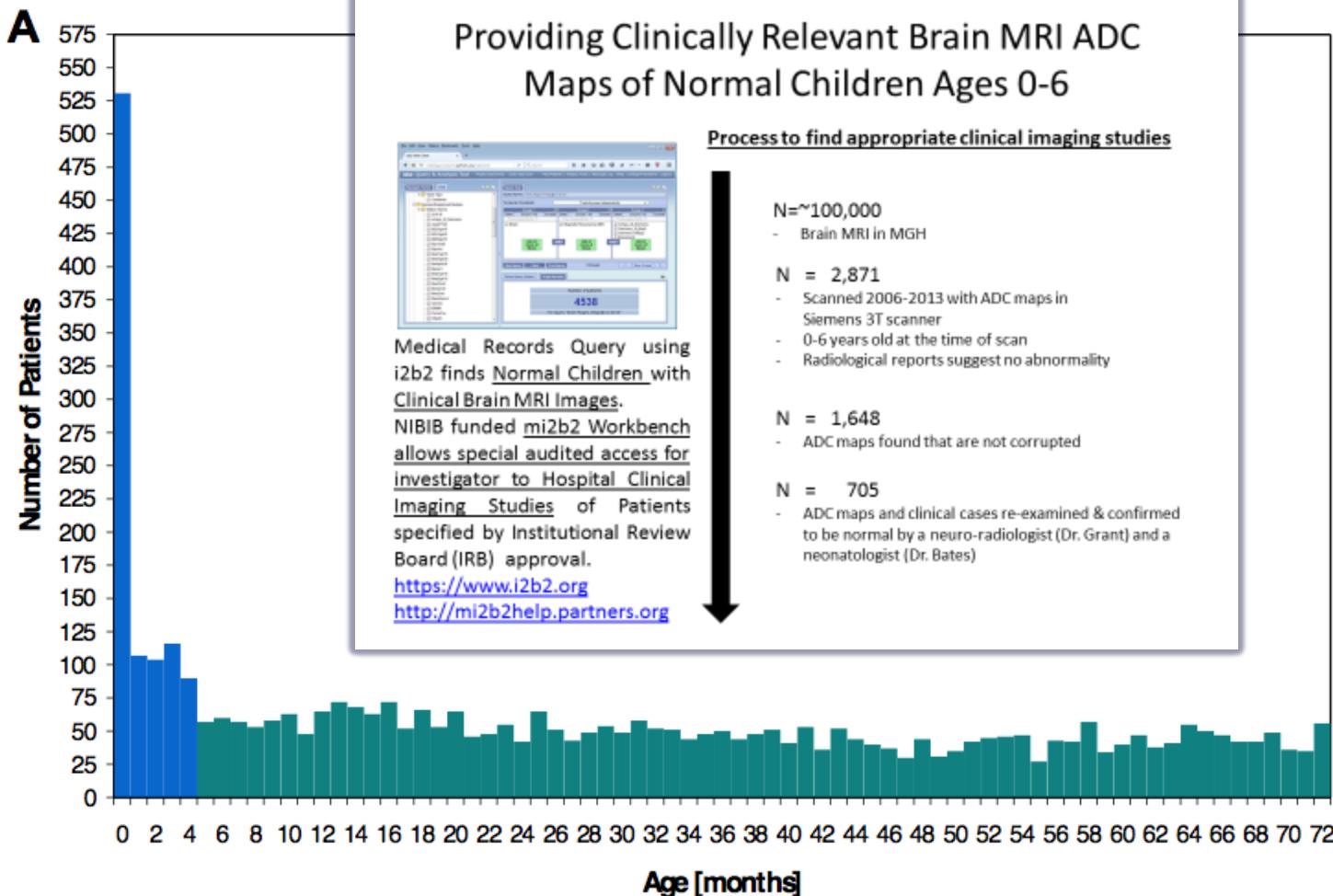
High Quality Phenotypes for Research Studies

The screenshot shows the Partners Biobank Portal Query Tool window. The left sidebar contains a tree view of "Navigate Terms" with categories like Biobank Consent Information, Biobank Demographics, Biobank Genomics, Biobank Health Information Survey, Biobank Sample Types, Curated Disease Populations, and various diseases and conditions. The main area is titled "Query Tool" and shows a query named "Prima-CHF --Gene@14:22:40". The temporal constraint is set to "Treat all groups independently". The query is structured into three groups:

- Group 1:** Dates, Occurs > 0x, Exclude. Constraint: Treat Independently. Condition: Primary dilated cardiomyopathy - 4002. Description: one or more of these.
- Group 2:** Dates, Occurs > 0x, Exclude. Constraint: Treat Independently. Condition: CHF - current or past history (PPV 0.90) - 700. Description: one or more of these.
- Group 3:** Dates, Occurs > 0x, Exclude. Constraint: Treat Independently. Condition: Gene [contains "TTN AND Homozygous AND (Frameshift OR missense OR nonsense OR start_loss OR stop_loss)"]. Description: one or more of these.

Below the groups, there are buttons for "Run Query", "Clear", and "3 Groups". At the bottom, there are links for "Show Query Status", "Graph Results" (which is selected), "Query Report", and "Download Results". A summary section displays "Number of patients" as 70 for the query "Prima-CHF --Gene@14:22:40".

Find Normal Brain MRI's of Children

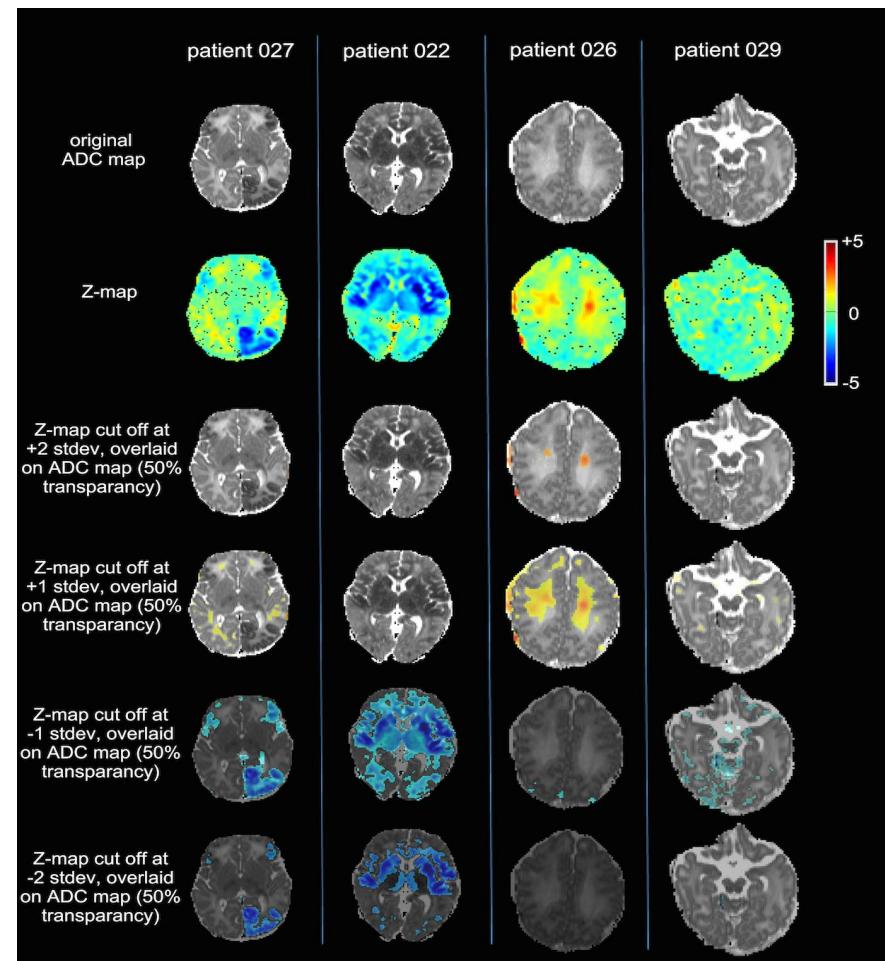


Number of patients who had a brain MRI scan at a particular age in months from 0 to 6 years (A) and in weeks from 0 to 4 months (B)

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

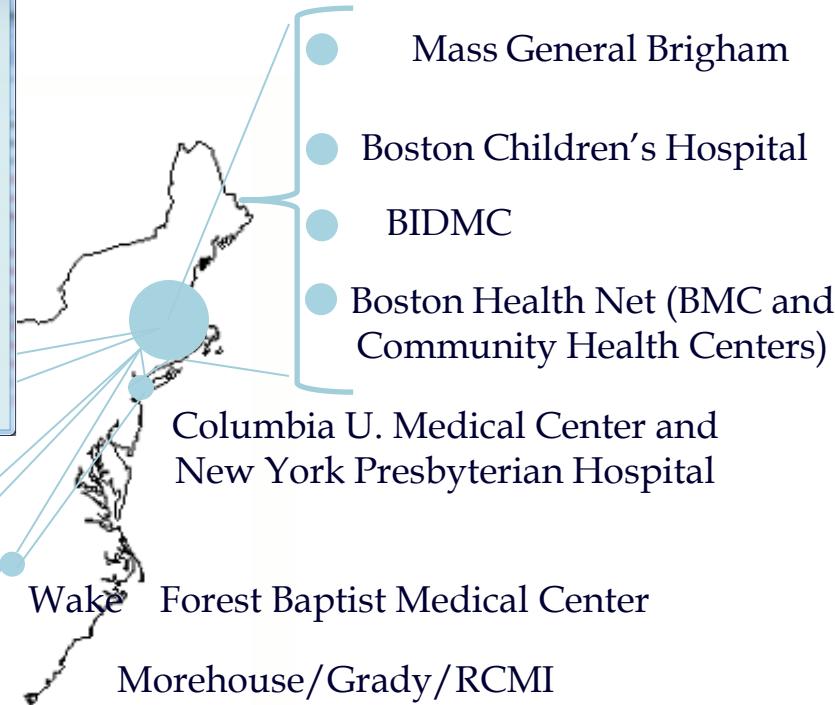
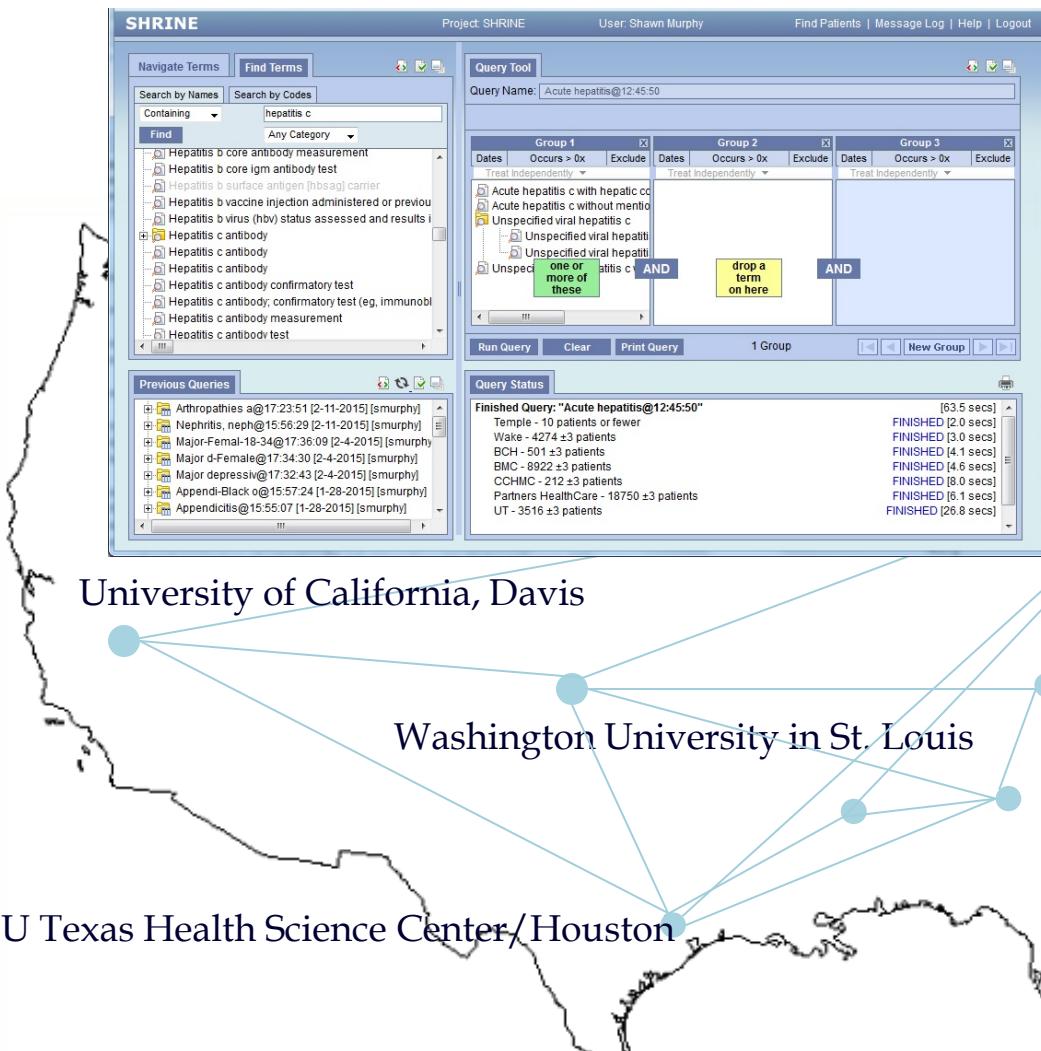
Quantitative analysis tools + large data sets = Great insights for practicing doctors



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
- Digital medicine is delivered into clinical care

Federated Queries

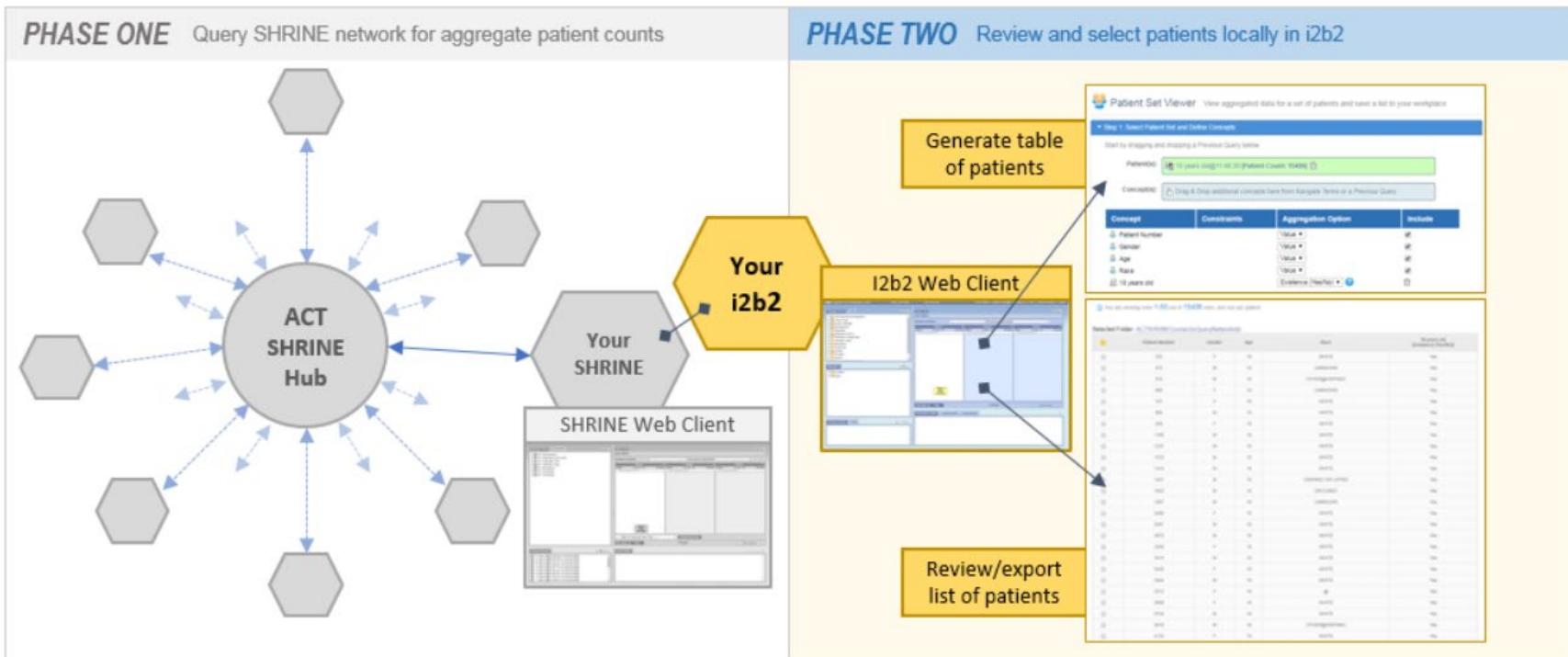


Drive Pragmatic Clinical Studies at Sites

The screenshot displays the i2b2 Query & Analysis Tool interface, specifically the ACTFullDemo version. The interface is divided into several sections:

- Left Sidebar (Workplace):** Contains a tree view of medical categories like Cardiovascular agents, Central nervous system agents, and various analgesics.
- Patient Set Viewer (Main Panel):** Shows a grid of patient records with columns for ID, Gender, Age, Race, and two status columns (Yes and No). A large blue arrow points from this grid towards the analysis chart on the right.
- Patient Data Analysis (Right Panel):** This section includes:
 - Vitals:** Options for Weight, Blood Pressure, etc., with "Weight, Blood Pressure" currently selected.
 - Medications:** Options for Metropol, Spironolactone, Valsartan, etc., with "Metropol, Spironolactone, Valsartan" selected.
 - Labs:** Options for Creatinine, Potassium, eGFR, etc., with "Creatinine, Potassium, eGFR" selected.
- Events and Visit Summary:** Shows visit counts for different locations: ER Visits (14), ICU Visits (2), Floor Visits (56), and Clinic Visits (33).
- Time Series Data:** Displays trends for various vital signs and medications over time (01/07/17 to 01/20/17).

Clinical Studies Workflow



emerge network

ELECTRONIC MEDICAL RECORDS AND GENOMICS



2500



Sequencing
100 high-
priority genes

25,000 Network-wide

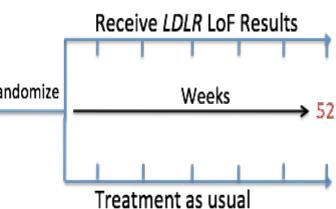


25,000

Discovery



Penetrance and Pleiotropy

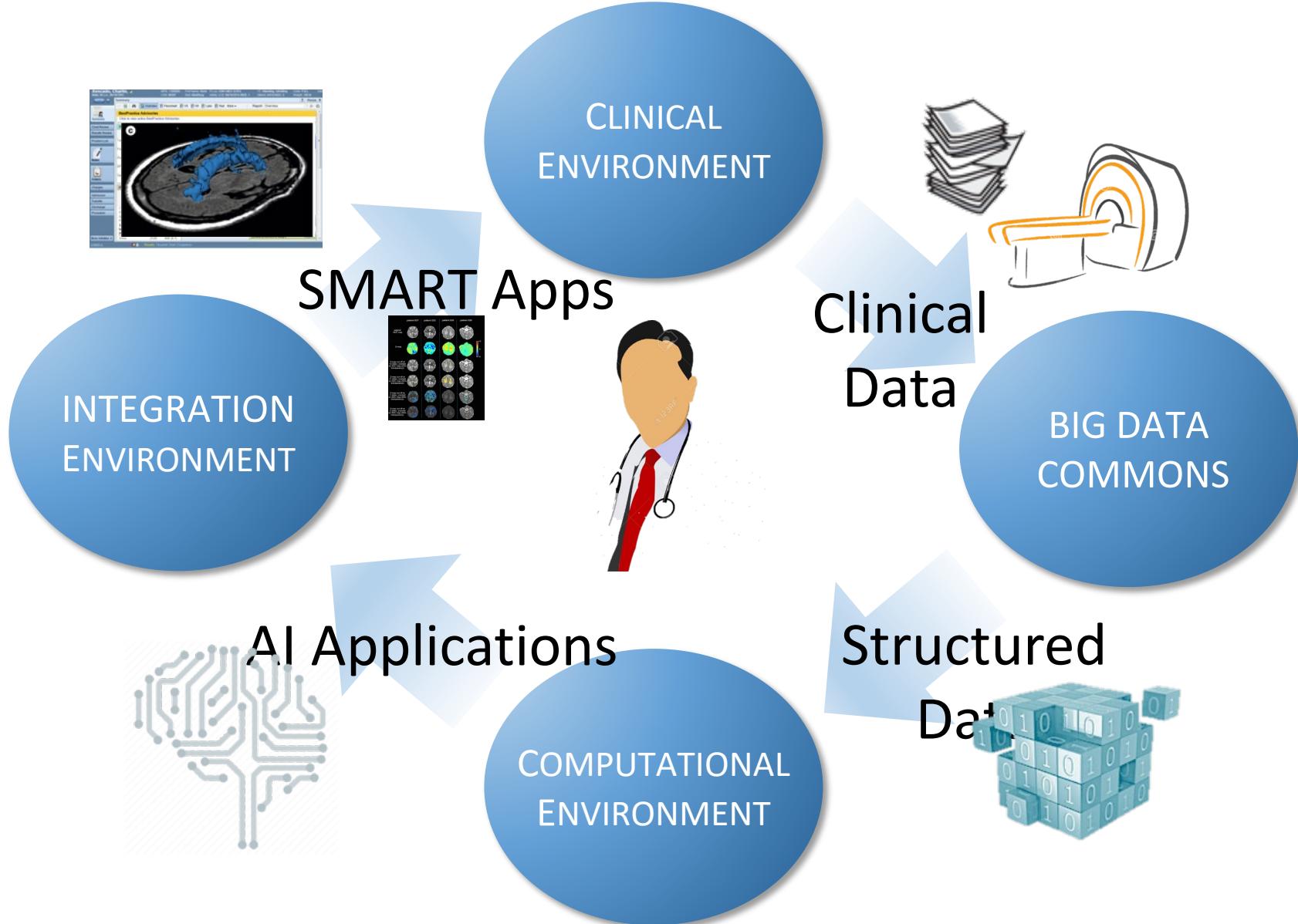


Genotyping

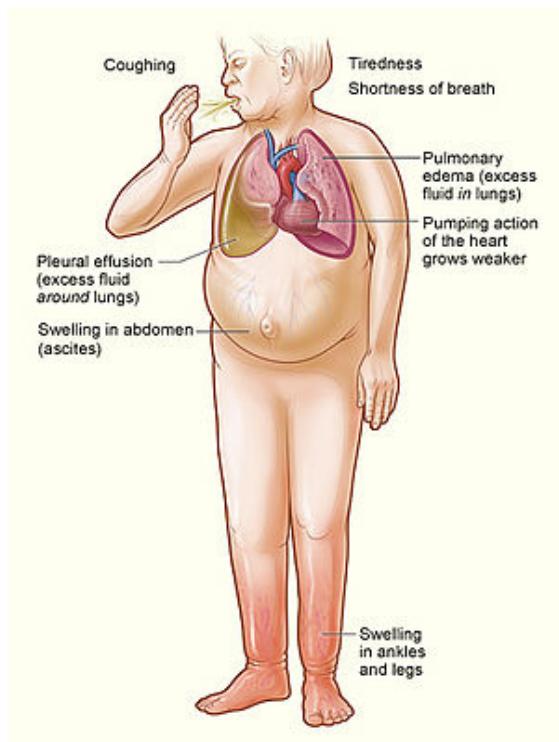
High Throughput Methods for supporting Translational Research

- 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
- Digital medicine is delivered into clinical care

CONTROLLING CYCLE OF LEARNING HEALTHCARE SYSTEM



Congestive Heart Failure



- Affects 2% of the adult population
- Risk of death first year after diagnosis: 35%
- In patient 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)

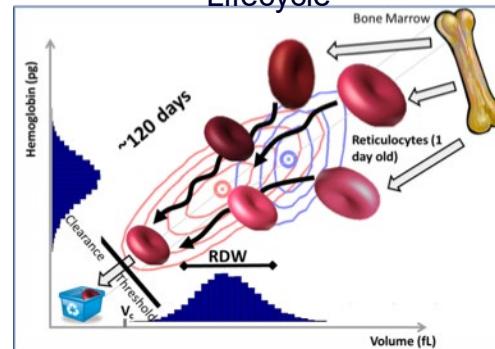
Results and Conclusions

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

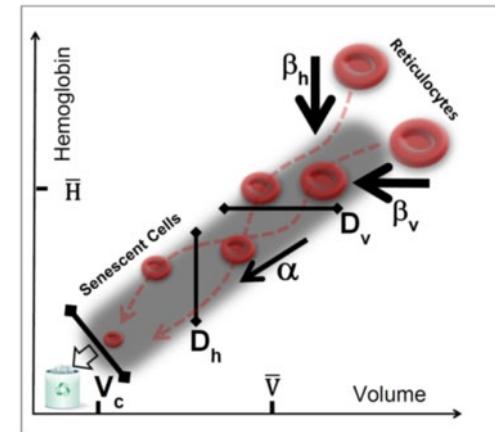
Applications

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

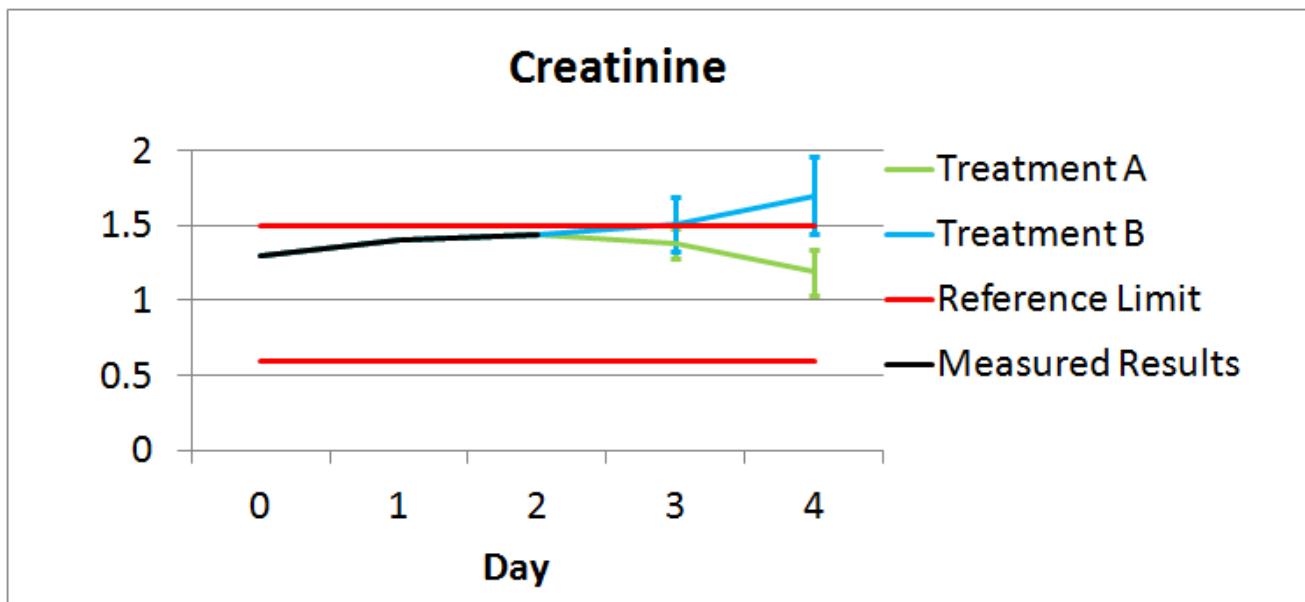
Dynamic Model of the RBC Lifecycle



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

Heart Failure Physiology Tool

John Doe

Date of Birth: January 1, 1940
MRN: 1234567890

Actions Under Consideration

Prescribe 30mg / day HCT [remove](#)

Discharge Patient [remove](#)

[Add Action](#)

Predicted Events (With/Without Actions)

Length of Stay: 1 days -> 0 days

30 Day Readmission: 30% -> 50%

Population Based Predictive Analytics to Support Improved Decision Making

Longitudinal Data

[current admission](#) [2 year](#) [5 year](#) [10 year](#) [lifetime](#) [custom](#)

ED Visits



Admissions



Patient Rank

89

80

70

65

[Add Data Stream](#)

Notes

Creatinine

[set thresholds](#)

O₂ Saturation

[set thresholds](#)

Heart Rate

[set thresholds](#)

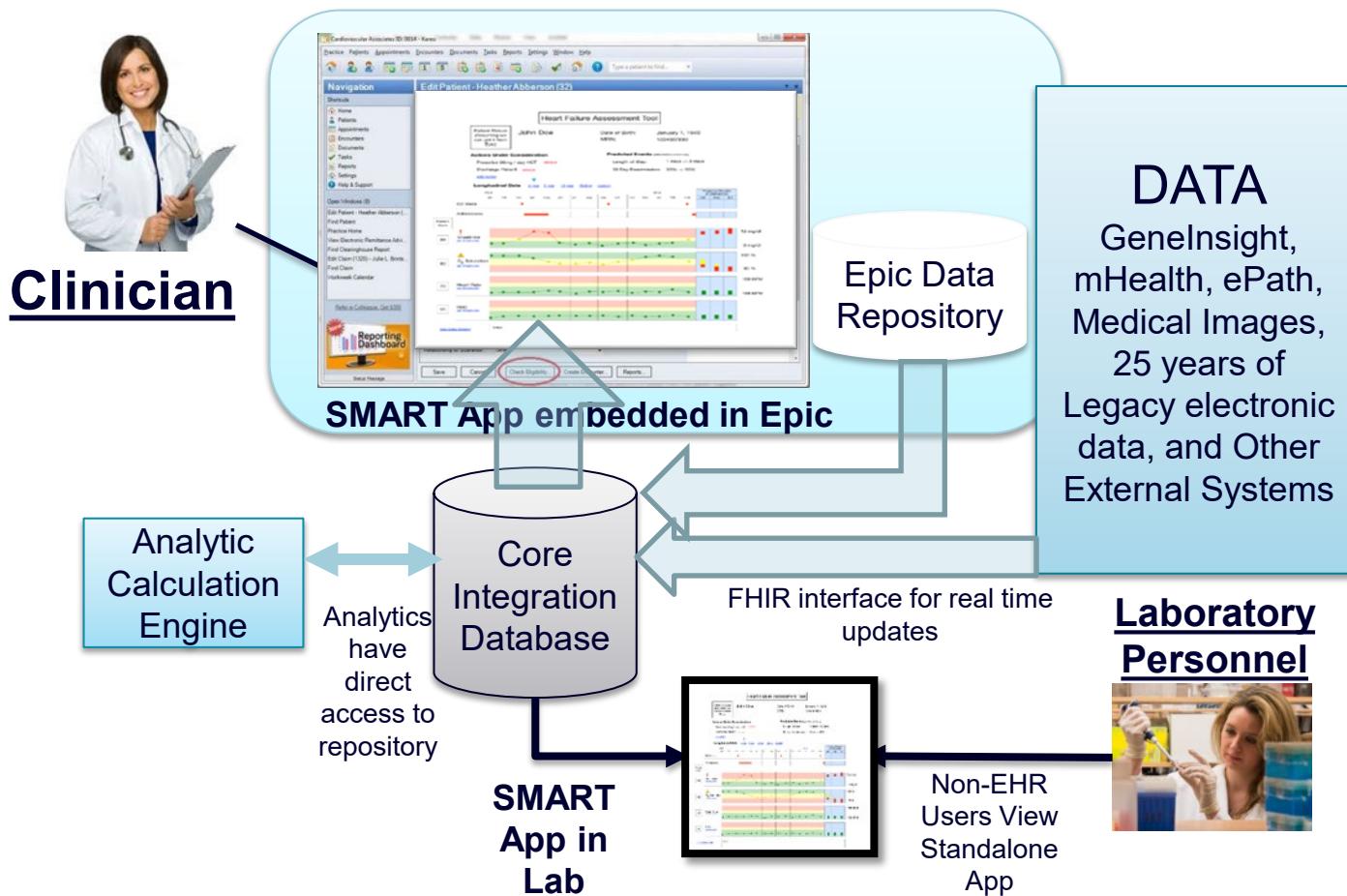
RBC

[set thresholds](#)

Exploring Integration of MGH Path renal predictive model

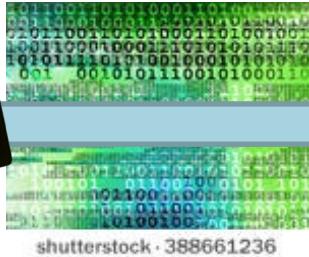
Population Based Predictive Analytics to Support Improved Decision Making

Bringing Big Data into Clinical Care with Open App Development



Transforming Care in the Digital Age

Digital and IoT devices continuously output Patient Data

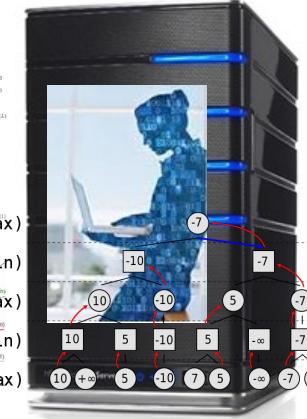
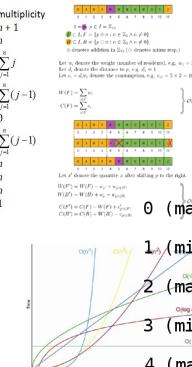


```

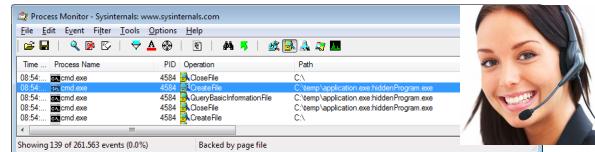
1. for ( i = 1; i <= a1.length; i++ )
  1.1. j = i
  1.2. while ( a1[i] != a2[j] )
    if ( j >= a1.length)
      1.3. return false
    j++
  1.4. a2[j] = a2[i]
  1.5. a2[i] = tmp
2. return true

```

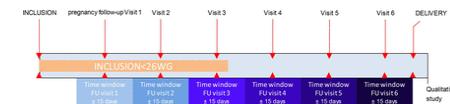
cost multiplicity
 c_1 $n+1$
 c_2 n
 c_3 $\sum_{j=1}^n j$
 c_4 $\sum_{j=1}^n (j-1)$
 c_5 $\sum_{j=1}^n \left(\frac{1}{c_2 j} + \sum_{i=1}^{j-1} \frac{1}{c_2 i} \right)$
 c_6 n
 c_7 n
 c_8 n
 c_9 n
 c_{10} 1



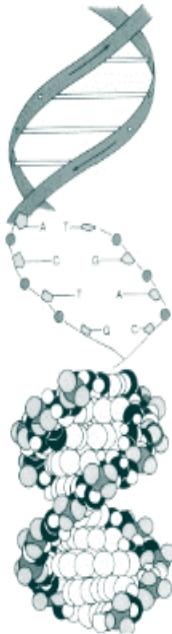
Digital Twin of patient enables continuous assessment of patient with Real Time Algorithms



Navigator Model dramatically increases Frequency and Convenience for Patient Communication



System drives Pragmatic Clinical Trials Leading to Continuous Process Improvement



i2b2 and SMART Software

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>)