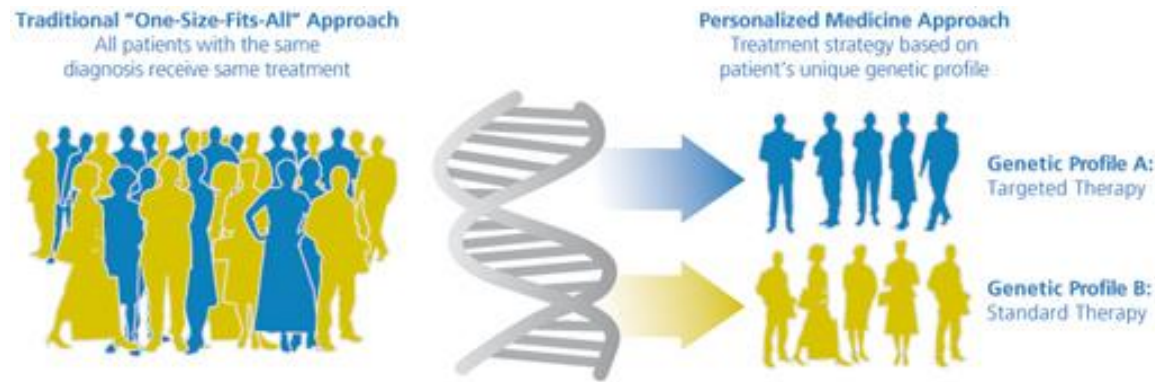


# Instrumenting the Health Care Enterprise for Discovery in the Course of Clinical Care

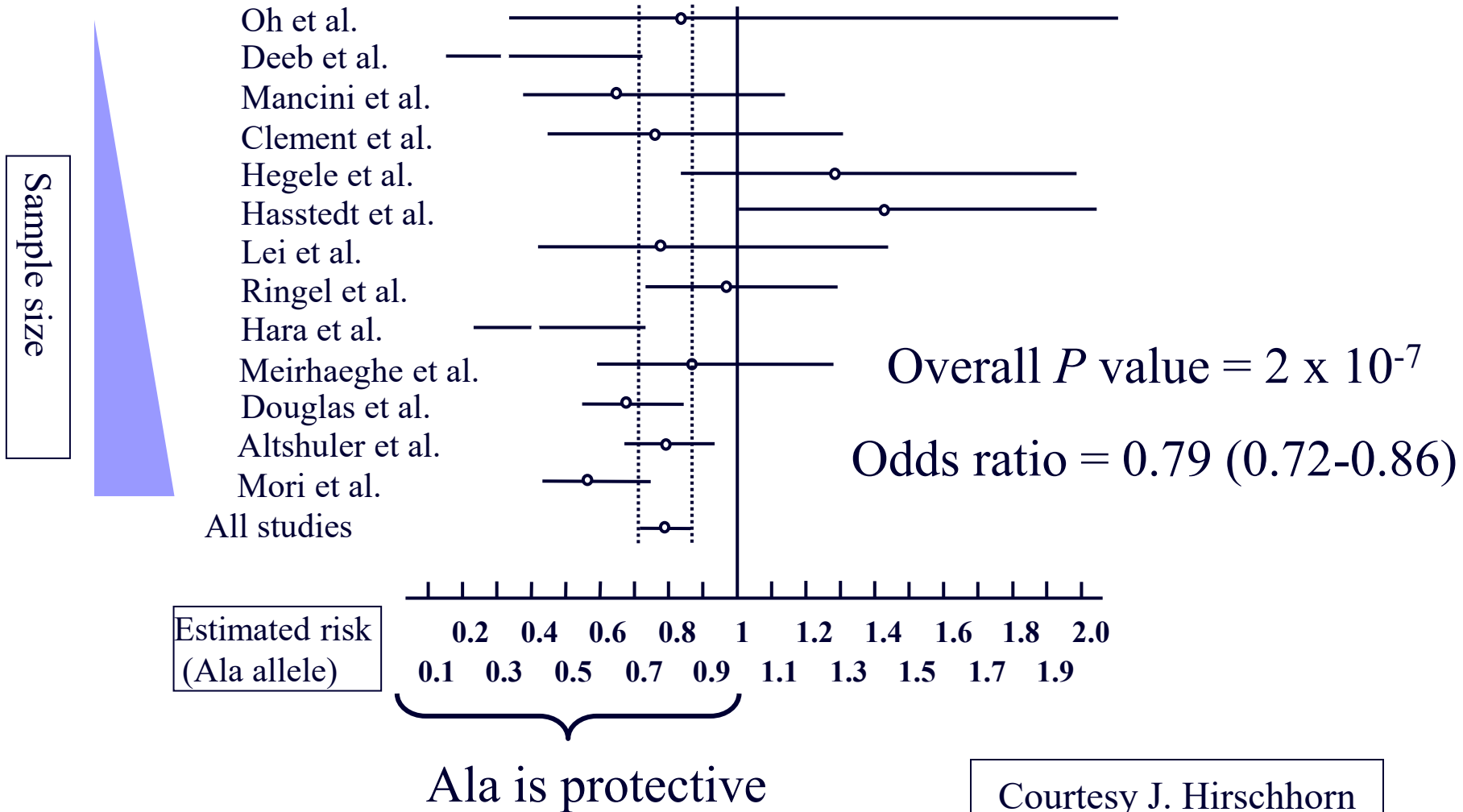
***Shawn Murphy MD, Ph.D.***  
***Chief Research Information Officer***  
***Harvard Medical School / Mass General  
Brigham***

# 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 $\gamma$ 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 through Digital Twin

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

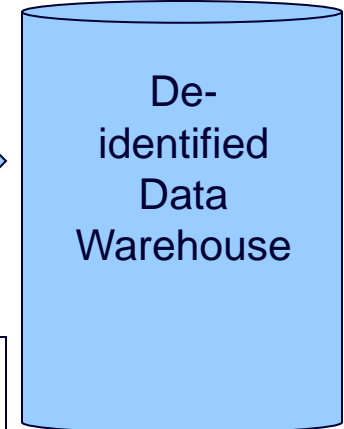
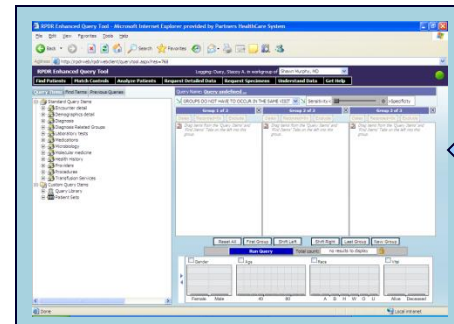
- 1) There are significant risks of a data breach which will result in very large fines and loss of confidence in the hospitals where the breach occurred.
- 2) The data are not collected for research purposes, and therefore the data can be poorly structured with significant omissions, biases, and inaccuracies.

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

## 1) Queries for aggregate patient numbers

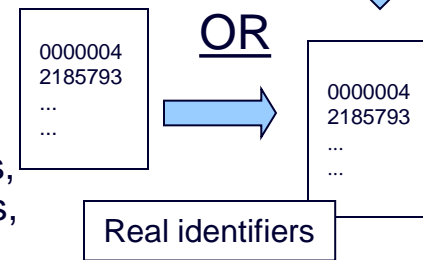
- Warehouse of in & outpatient clinical data
- 6.7 million Mass General Brigham patients
- 2.6 billion diagnoses, medications, genomics, procedures, laboratories, & physical findings coupled to demographic & visit data
- Authorized use by faculty status
- Clinicians can construct complex queries
- Queries cannot identify individuals, internally can produce identifiers for (2)

### Query construction in web tool



## 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     | Abnormal Flag | Reference | Unit | Reference Range |
|---------|------------------|--------|-----------------|---------------|-----------|------|-----------------|
| SO-PTT  | Supravital APTT  | 32.8   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 32.8   |                 | H             |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 46.4   |                 | H             |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 43.1   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 26.7   | MODERATELY H    |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 23.7   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 24.4   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 24.7   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 34.0   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 24.7   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | Supravital APTT  | 31.3   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 34.5   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 46.0   |                 | H             |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 46.0   |                 | H             |           | sec  | 22-35.1         |
| SO-PTT  | Supravital APTT  | 50.2   | Rate Slow H     |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 33.6   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | Supravital APTT  | 34.3   |                 |               |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 37.9   |                 | H             |           | sec  | 22-35.1         |
| SO-PTT  | APTT             | 22.6   |                 |               |           | sec  | 22-34.1         |
| SO-PTT  | APTT             | 37.4   |                 | H             |           | sec  | 22-34.1         |
| SO-PTT  | APTT             | 37.2   | SLT HEMOLYSIS H |               |           | sec  | 22-34.1         |
| SO-PTT  | APTT             | 38.1   |                 | H             |           | sec  | 22-34.1         |
| SO-PTT  | APTT             | 38.4   | MODERATE HE H   |               |           | sec  | 22-34.1         |

# FINDING PATIENTS

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

Query items

Person who is using tool

The screenshot displays the RPDR Enhanced Query Tool interface. The top navigation bar includes buttons for 'Find Patient', 'Match Controls', 'Analyze Patients', 'Request Detailed Data', 'Request Specimens', 'Understand Data', and 'Get Help'. The 'Query Items' tab is active, showing a tree view of categories such as 'Standard Query Items' (Encounter detail, Demographics detail, Diagnosis, etc.) and 'Custom Query Items'. The main workspace is titled 'Query Name: Query undefined ...' and contains three groups for query construction. Each group has a 'Dates' field, a 'Recorded > 0x' checkbox, and an 'Exclude' button. Below the groups are buttons for 'Reset All', 'Add Group', 'Delete Group', 'New Group', and a prominent 'Run Query' button. The status bar shows 'Total count: no results to display'. At the bottom, there are four summary tables: 'Gender' (Female, Male), 'Age' (40, 80), 'Race' (A, B, H, W, O, U), and 'Vital' (Alive, Deceased).

Query construction

Results - broken down by number distinct of patients

- 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: Isut Diagnos AMI

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

| Group 1 of 3  | Group 2 of 3  | Group 3 of 3  |       |             |         |       |             |         |
|---|---|---|-------|-------------|---------|-------|-------------|---------|
| Dates   | Recorded>0x   | Exclude   | Dates | Recorded>0x | Exclude | Dates | Recorded>0x | Exclude |
| <input checked="" type="checkbox"/> One or more items recorded<br><input checked="" type="checkbox"/> Acute myocardial infarction | <input type="checkbox"/> Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group. | <input type="checkbox"/> 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)

Gender  Age  Race  Vital

click the image or check box to request an aggregated count by patient gender for this query.

|        |      |    |    |   |   |   |   |   |   |       |          |
|--------|------|----|----|---|---|---|---|---|---|-------|----------|
| Female | Male | 40 | 80 | A | B | H | W | O | U | Alive | Deceased |
|--------|------|----|----|---|---|---|---|---|---|-------|----------|



Query Items Find Terms Previous Queries

Search For:

Containing

All Categories

Search Items

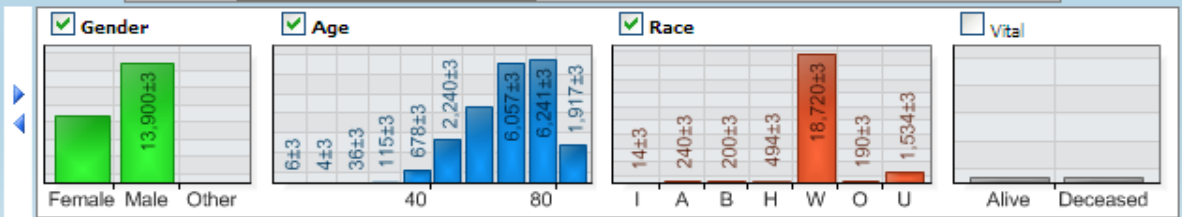
- CK-MB Index (Group:CKMBRI)
- CK-MB INDEX (Test:sc400.4452)

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

GROUPS DO NOT HAVE TO OCCUR IN THE SAME VISIT

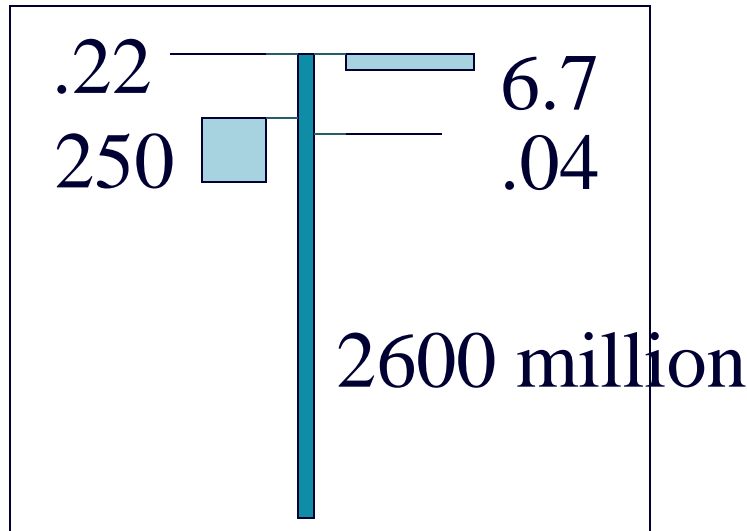
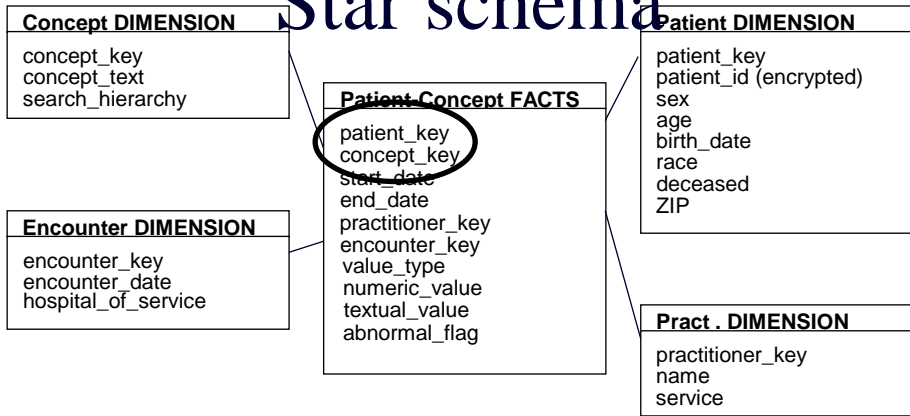
| Group 1 of 3  | Group 2 of 3  | Group 3 of 3  |
|---|---|---|
| <input type="button" value="Dates"/> <input type="button" value="Recorded &gt;0x"/> <input type="button" value="Exclude"/> <p><i>One or more items recorded</i></p> <ul style="list-style-type: none"> <li>Acute myocardial infarction</li> </ul> | <input type="button" value="Dates"/> <input type="button" value="Recorded &gt;0x"/> <input type="button" value="Exclude"/> <p><i>One or more items recorded</i></p> <ul style="list-style-type: none"> <li>CK-MB Index (Group:CKMBRI) &gt; 3.5</li> </ul> | <input type="button" value="Dates"/> <input type="button" value="Recorded &gt;0x"/> <input type="button" value="Exclude"/> <p><i>Drag items from the 'Query Items' and 'Find Items' Tabs on the left into this group.</i></p> |

Total count: 21647±3 patient(s)

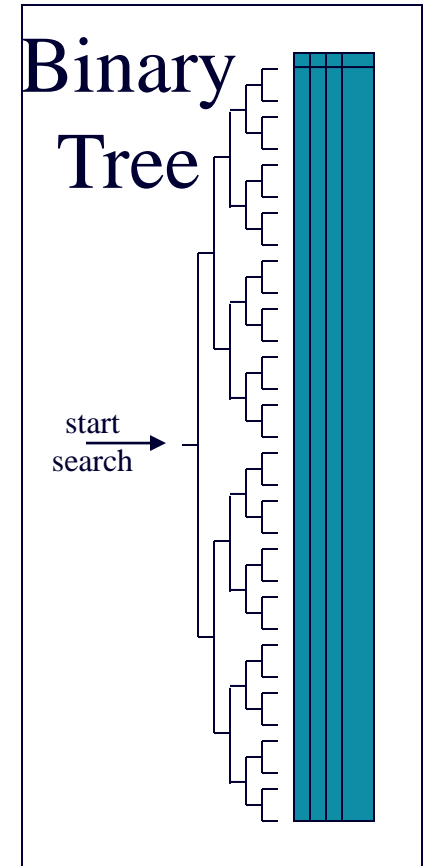


# Theory of Kimball translated to Healthcare Data

## Star schema



## Binary Tree





### 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):

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

Newton Wellesley Hospital IRB:

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

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

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

Environment Record Options Help

File: SNM0\_022502164303648842.XML

File Type: Control File Current Production Database: RPDR\_12 RPDR\_12\_5241

Update Status  Start Process After 9:00:00 PM

IRB Information

IRB Number: 2000P000000 Ends:

Date from: 01/01/1900 Date to: 01/01/1900

Primary User: smn0

Files to MGH Users: Partners\smn0,Partners\zvp,Partners\kcs3

BWH Users: Partners\kra1,Partners\smn0,Partners\kcs3

Data Requested

Demographics  Medical Record Numbers  Chemistry

Encounters  Contact Information  Radiology

Hematology  PCP  Pathology

Discharge Summaries  Immunology  LMR Notes

Medications  Operative Notes  LMR Problems

LMR Allergies  LMR Medications  Build Access Database

Groups: BUN

Run Close Clear

Output files placed in special directory

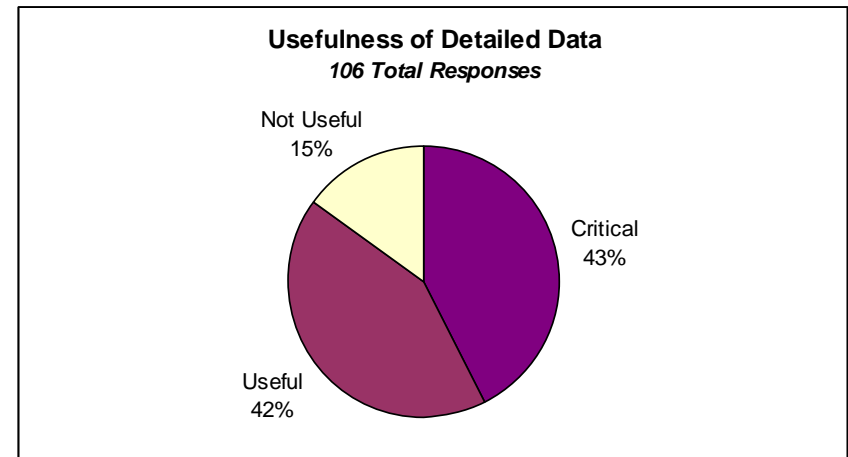
Data is gathered from RPDR and other MGB sources

| Test Id | Test Description | Result | Result Text     | Abnormal Flag | Reference Uni | 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   |                 | H             | 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       |

Files include Small Database

# One year's 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†  |         |
|----------------------------|------------------|---------|
|                            | prolongatio<br>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 |

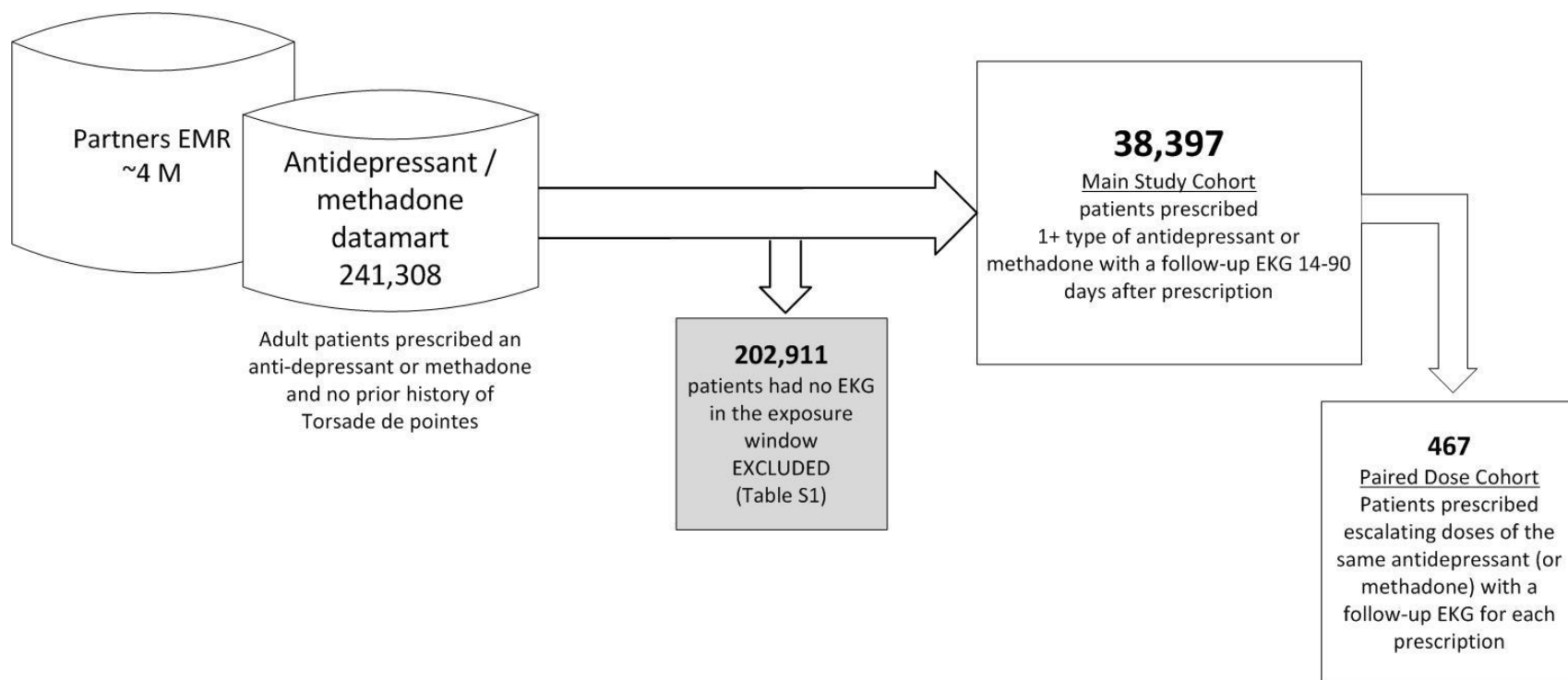
† Adjusted for age, gender, race, type of insurance, history of major depression, history of myocardial infarction and Charlson comorbidity score

Roy Perlis MD, MSc and team

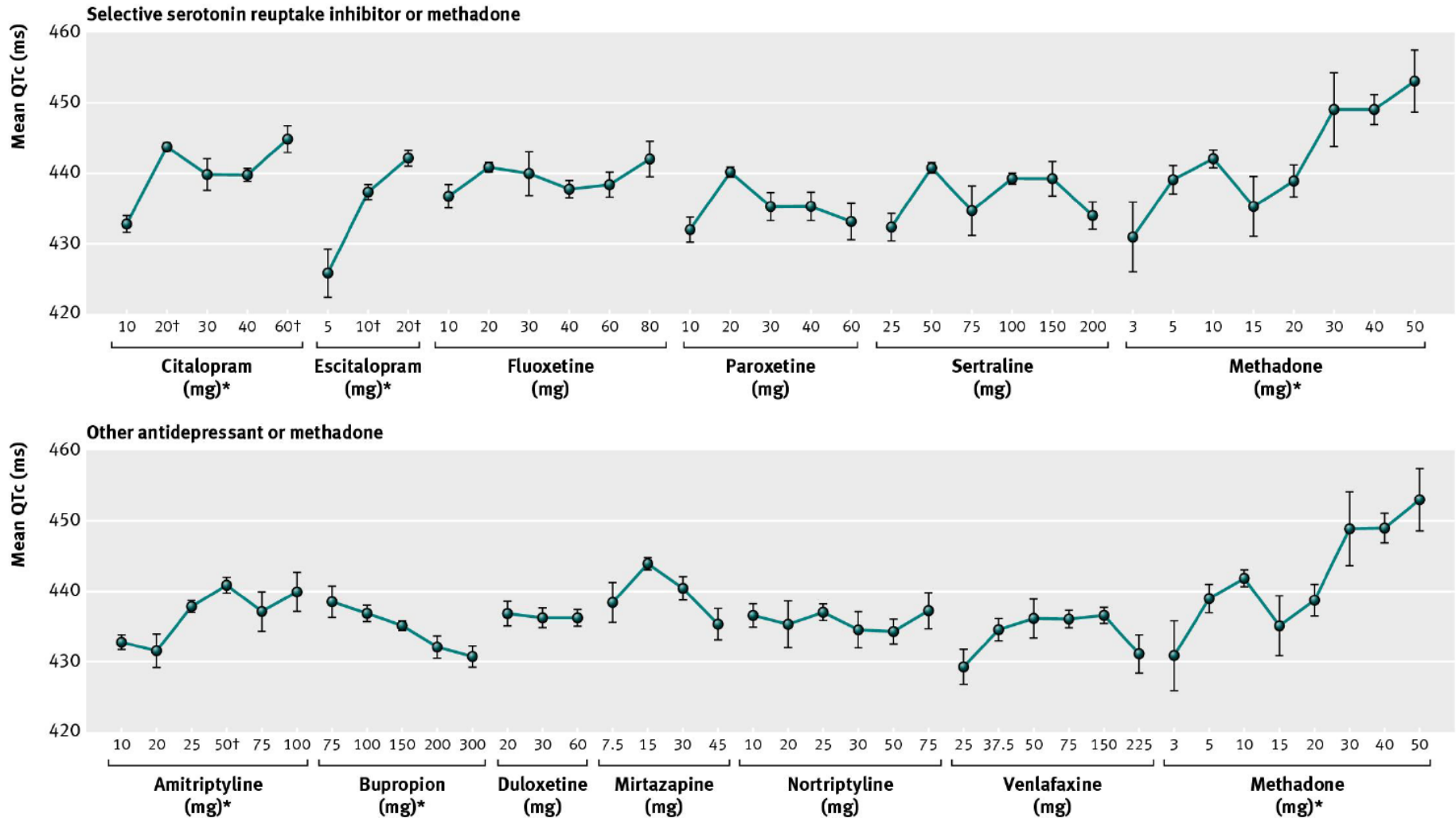


# Relevant Cohorts of Patients are Gathered through RPDR and Detailed Data Obtained

- 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 through Digital Twin

## **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

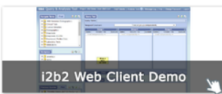
community.i2b2.org/wiki/

i2b2 Community Wiki


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

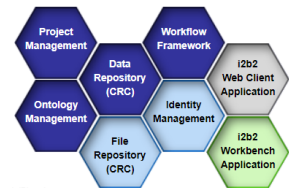
- i2b2 Documentation updated Feb 07, 2020
- Welcome to the i2b2 Community Wiki updated Jan 02, 2020
- i2b2 Software updated Jan 02, 2020
- i2b2 Software updated Dec 27, 2019
- i2b2 Software Downloads Links updated Dec 27, 2019

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 heterogenous 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).



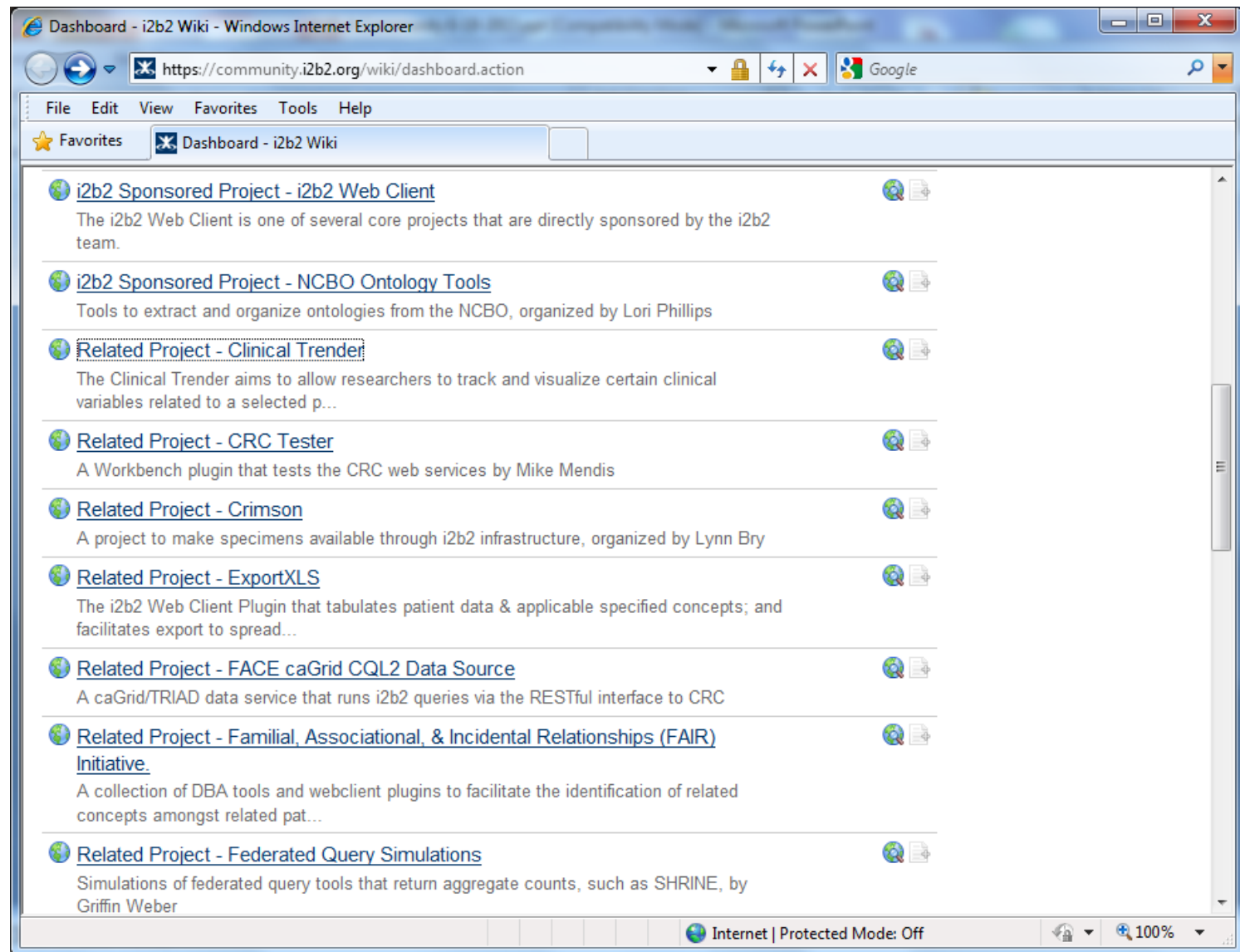
Key: i2b2 Core Cell, i2b2 Optional Cell, Workbench/Plug-in, Web Client, CRC Plug-in



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

| Documentation   | Get Software  | Announcements  |         |  |                              |                         |                          |                            |                         |                          |                 |                         |                          |                    |                         |  |   |
|---|---|--|---------|--|------------------------------|-------------------------|--------------------------|----------------------------|-------------------------|--------------------------|-----------------|-------------------------|--------------------------|--------------------|-------------------------|--|---|
| <p><b>Getting Started</b></p> <ul style="list-style-type: none"> <li>Installation Guide</li> <li>Upgrade Guide</li> <li>Videos and Tutorials</li> </ul> <p><b>For Developers</b></p> <ul style="list-style-type: none"> <li>Server-side Messaging</li> <li>Server Architecture</li> <li>Server-side Design</li> <li>Web Client Design</li> <li>Release Notes</li> </ul> <p><b>For End Users</b></p> | <table border="1"> <thead> <tr> <th>Software</th> <th>Version</th> <th></th> </tr> </thead> <tbody> <tr> <td>i2b2 Core Server Source Code</td> <td>1.7.12 (December, 2019)</td> <td><a href="#">Download</a></td> </tr> <tr> <td>i2b2 Core Data Source Code</td> <td>1.7.12 (December, 2019)</td> <td><a href="#">Download</a></td> </tr> <tr> <td>i2b2 Web Client</td> <td>1.7.12 (December, 2019)</td> <td><a href="#">Download</a></td> </tr> <tr> <td>i2b2 Documentation</td> <td>1.7.12 (December, 2019)</td> <td><a href="#">Release Notes</a><br/><a href="#">Upgrade Guide</a></td> </tr> </tbody> </table> | Software   | Version |  | i2b2 Core Server Source Code | 1.7.12 (December, 2019) | <a href="#">Download</a> | i2b2 Core Data Source Code | 1.7.12 (December, 2019) | <a href="#">Download</a> | i2b2 Web Client | 1.7.12 (December, 2019) | <a href="#">Download</a> | i2b2 Documentation | 1.7.12 (December, 2019) | <a href="#">Release Notes</a><br><a href="#">Upgrade Guide</a> | <p><b>What's New in i2b2?</b></p> <ul style="list-style-type: none"> <li><b>i2b2 Release 1.7.12</b> now available!                     <ul style="list-style-type: none"> <li>Easier i2b2 install</li> <li>Redesigned FindTerms</li> <li>REDCAP import to i2b2</li> <li>ACT Ontology</li> <li>Support for OKTA and NTLM2 User Authentication</li> <li>i2b2 Workbench download for Windows 64-bit platform</li> <li>... and much more!</li> </ul> </li> </ul> <p><b>Important links:</b></p> <ul style="list-style-type: none"> <li>i2b2 transSMART Working Groups</li> <li>New Project: i2b2 on OMOP</li> </ul> |
| Software  | Version   |  |         |  |                              |                         |                          |                            |                         |                          |                 |                         |                          |                    |                         |  |   |
| i2b2 Core Server Source Code  | 1.7.12 (December, 2019)   | <a href="#">Download</a>                                       |         |  |                              |                         |                          |                            |                         |                          |                 |                         |                          |                    |                         |  |   |
| i2b2 Core Data Source Code  | 1.7.12 (December, 2019)   | <a href="#">Download</a>                                       |         |  |                              |                         |                          |                            |                         |                          |                 |                         |                          |                    |                         |  |   |
| i2b2 Web Client   | 1.7.12 (December, 2019)   | <a href="#">Download</a>                                       |         |  |                              |                         |                          |                            |                         |                          |                 |                         |                          |                    |                         |  |   |
| i2b2 Documentation  | 1.7.12 (December, 2019)   | <a href="#">Release Notes</a><br><a href="#">Upgrade Guide</a> |         |  |                              |                         |                          |                            |                         |                          |                 |                         |                          |                    |                         |  |   |

# i2b2 Software adapts through new plugins



# Genotype Data

**Query Tool**

Query Name: \_\_\_\_\_

Temporal Constraint

Group 1

| Dates                 | Occurs > 0x | Exclude | Da |
|-----------------------|-------------|---------|----|
| Treat Independently ▾ |             |         |    |
| Gene - 10097          |             |         |    |

drop a term on here

**Search by Gene**

Use the gene name box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first characters.

Gene Name\*: APO

Please note the zygosity to query for patients without a particular variant (for example, APOA2 to query for patients without APOA2).

Zygosity\*: \_\_\_\_\_

Consequence: \_\_\_\_\_

APOA1  
APOA1BP  
APOA2  
APOA4  
APOA5  
APOB  
APOBEC1  
APOBEC2  
APOBEC3A  
APOBEC3A\_B  
APOBEC3B

OK Cancel

**Query Tool**

Query Name: \_\_\_\_\_

Temporal Constraint

Group 1

| Dates                       | Occurs > 0x | Exclude | Da |
|-----------------------------|-------------|---------|----|
| Treat Independently ▾       |             |         |    |
| dbSNP rs Identifier - 10097 |             |         |    |

drop a term on here

**Search by dbSNP rs Identifier**

Use the rs identifier box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first three numbers.

rs identifier\*: rs1234

Please note the zygosity to query for patients without a particular variant (for example, rs12340033 to query for patients without rs12340033). After you start typing (nucleotide on the right) to query for patients without a particular variant (nucleotide on the left) to the alternate nucleotide.

Zygosity\*: \_\_\_\_\_

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

OK Cancel

drop a term on here

<https://community.i2b2.org/wiki/display/IGD>Loading+Genomic+VCF+Files+into+i2b2>

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

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

File Edit Options Template Execute Macro Window Help

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

PRINCIPAL DIAGNOSIS: LEFT LOWER LOBE PNEUMONIA

SECONDARY DIAGNOSES:

1. CHRONIC BRONCHITIS  
2. HEART FAILURE

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

HISTORY OF PRESENT ILLNESS: **Negative for tobacco**, alcohol, and IV drug abuse.

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 ???**

PHYSICAL EXAMINATION: Temperature 97.2, pulse 66, respirations 20, blood pressure 160/63, oxygen saturation 95% on room air. HEENT: Normocephalic and atraumatic. Pupils equal and reactive to light.

LABORATORY DATA: Sodium 148, potassium 2.4, chloride 87, bicarbonate 24, glucose 108, creatinine 1.2, BUN 18, Hgb 12.5, Hct 38, WBC 12,000, platelets 250,000.

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**

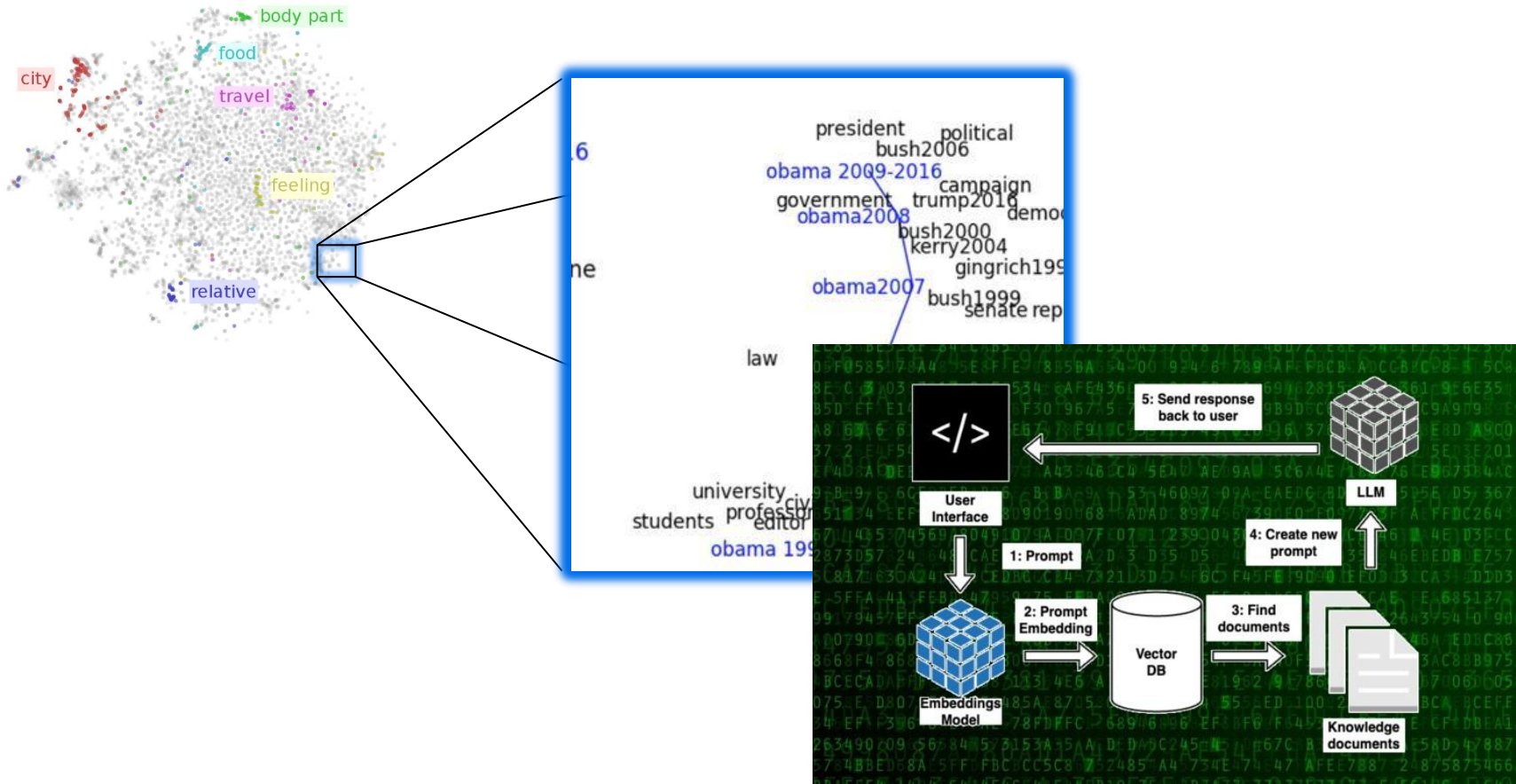
HOSPITAL COURSE: The patient was seen and evaluated by the physician on 10/10/77. She was discharged home on 10/10/77 to return to a 14-day course of treatment.

The patient is a widow, lives alone, 2 children, no **tob/alcohol**. **Hard to pick**

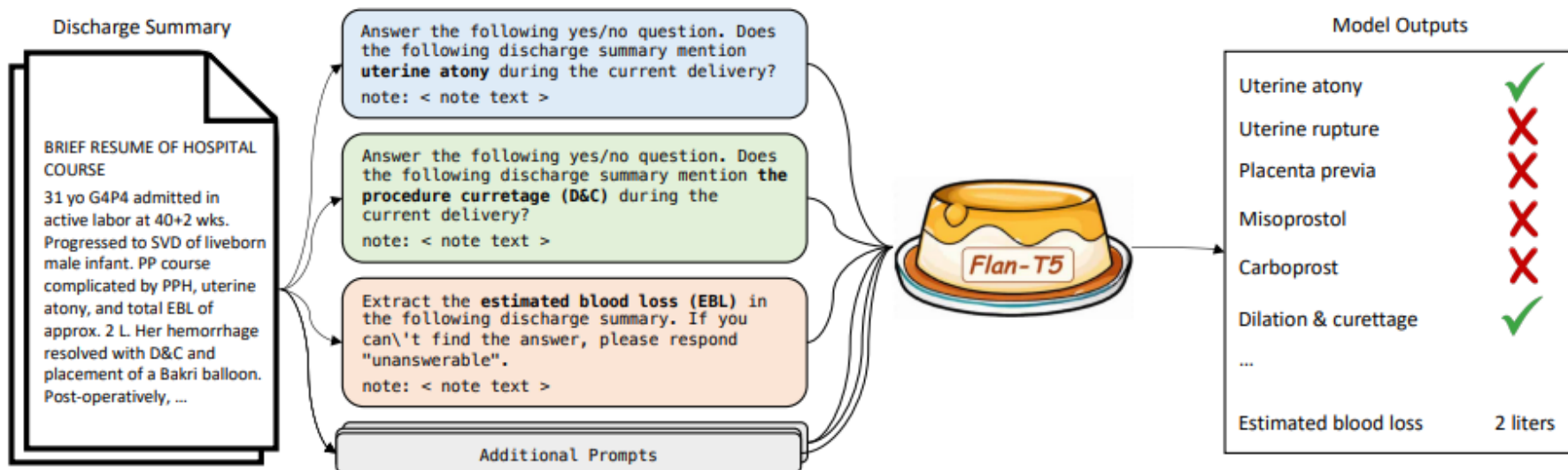
Ln 44 Col 1 | 274 | |WR| |Rec Off|No Wrap|DOS|INS|NUM



# LMM Enhanced interaction with Patient Representation

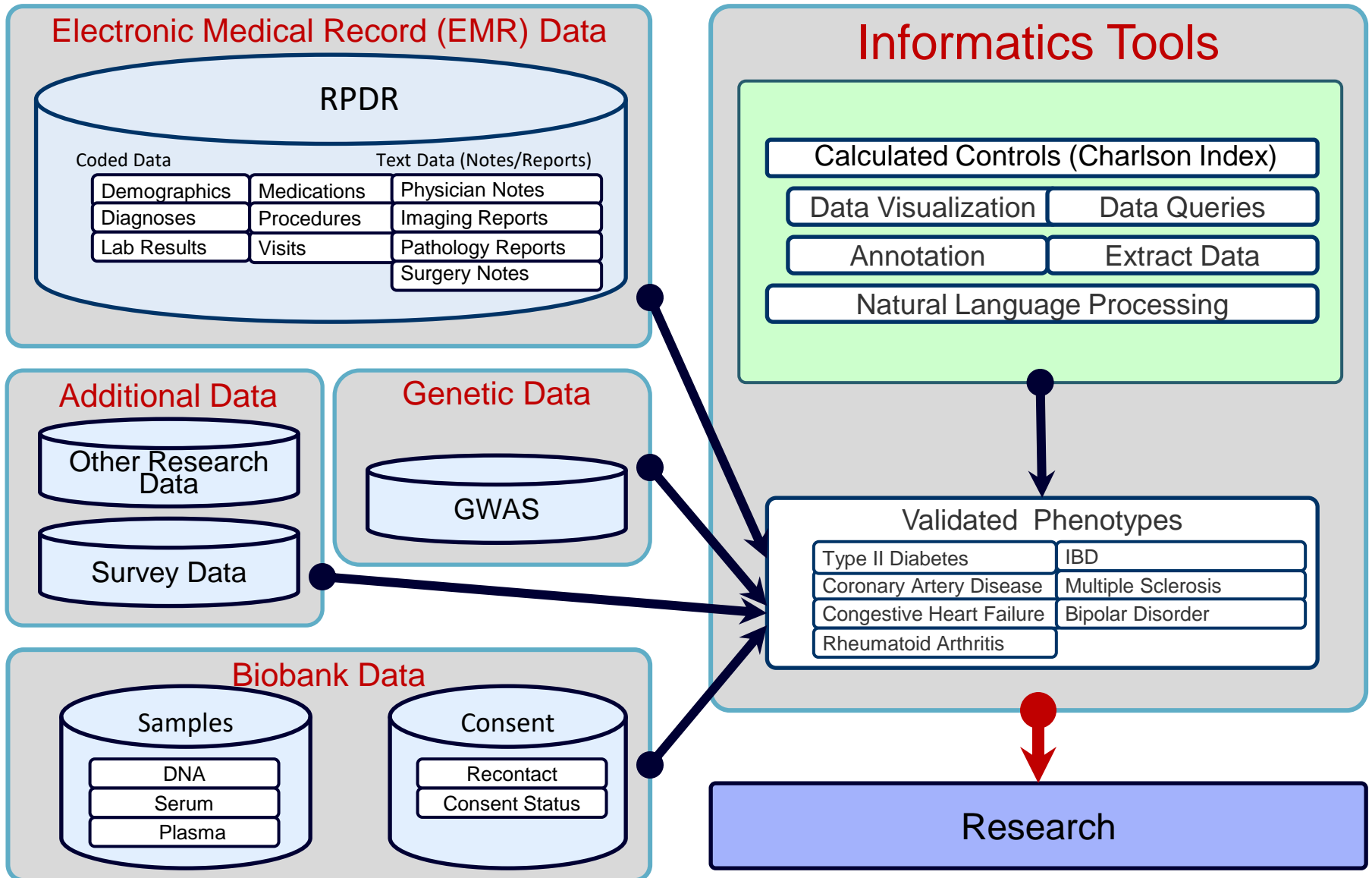


# Medical conditions supported by description in chart



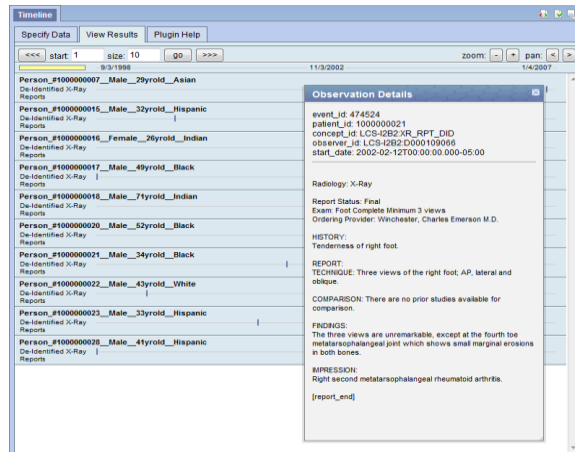
Emily Alsentzer et al  
Zero-shot Interpretable Phenotyping of Postpartum Hemorrhage Using Large Language Models  
medRxiv preprint doi:  
<https://doi.org/10.1101/2023.05.31.23290753>

# Data Integration in Big Data Commons

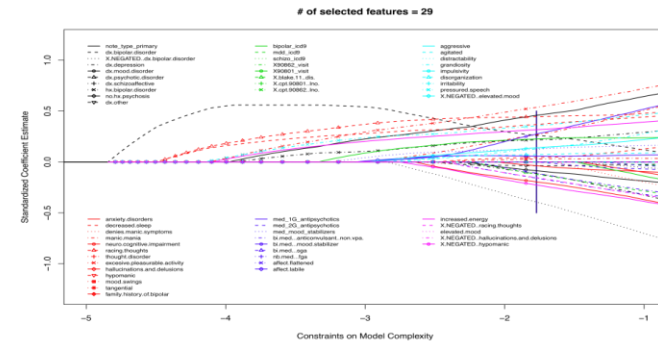


# Curating a Disease Algorithm

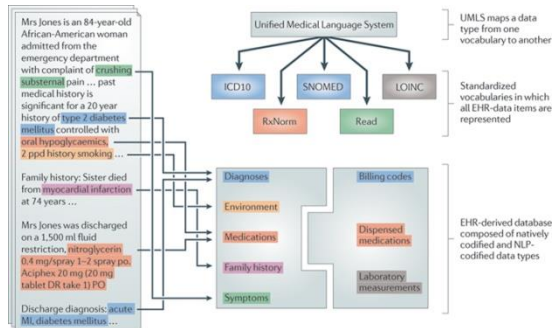
## 1. Create a gold standard training set.



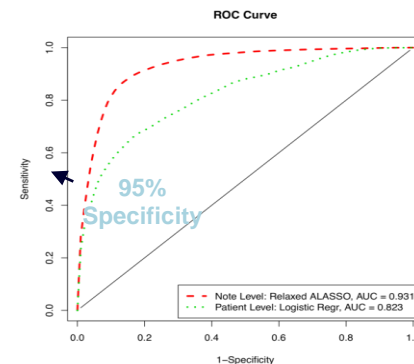
## 3. Develop the classification algorithm. Using the data analysis file and the training set from step 1, assess the frequency of each variable. Remove variables with low prevalence. Apply adaptive LASSO penalized logistic regression to identify highly predictive variables for the algorithm



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



## 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

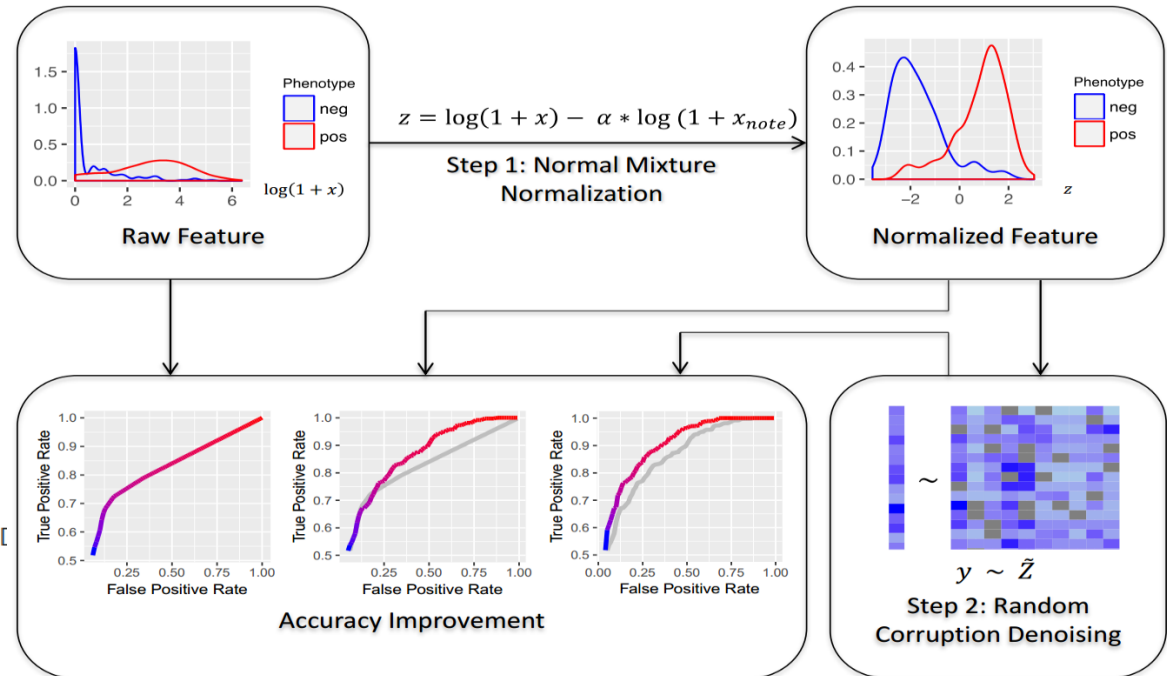
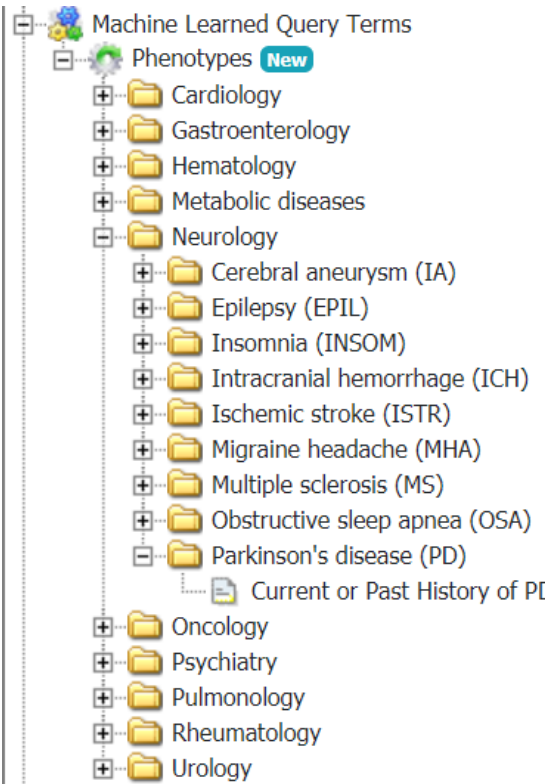
\*\* Based on 21,300 patients

# Automated Learning Algorithms enabled in RPDR such as PheNorm Algorithm

J Am Med Inform Assoc. 2018 Jan 1;25(1):54-60. doi: 10.1093/jamia/ocx111.

## Enabling phenotypic big data with PheNorm.

Yu S<sup>1,2</sup>, Ma Y<sup>3</sup>, Gronsbell J<sup>4</sup>, Cai T<sup>5</sup>, Ananthakrishnan AN<sup>6</sup>, Gainer VS<sup>7</sup>, Churchill SE<sup>8</sup>, Szolovits P<sup>9</sup>, Murphy SN<sup>7,10</sup>, Kohane IS<sup>9</sup>, Liao KP<sup>11</sup>, Cai T<sup>4</sup>.

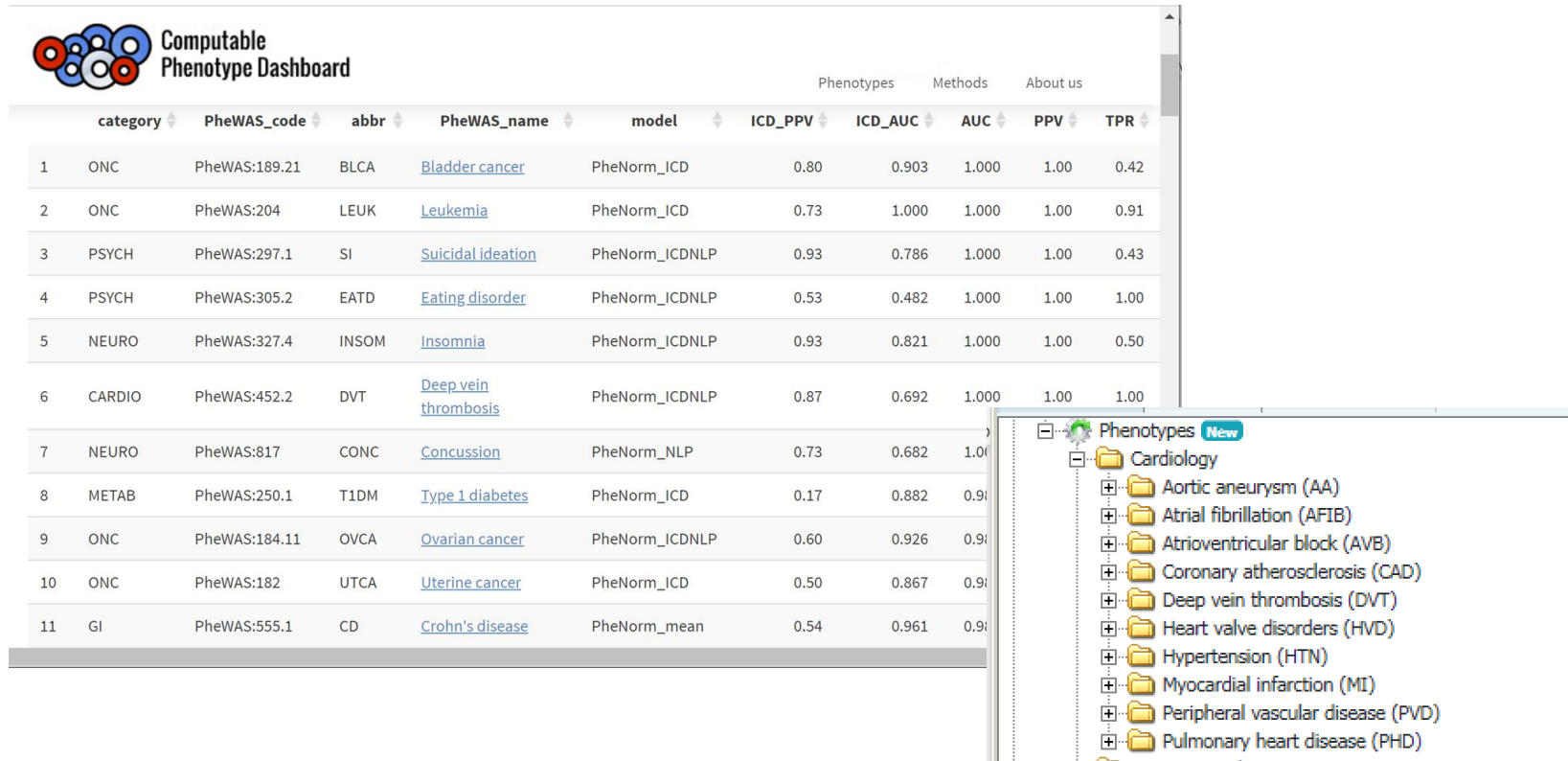


# Machine Learned Phenotypes

---

- 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



The dashboard features a logo with colored circles and the text "Computable Phenotype Dashboard". It includes navigation tabs for "Phenotypes", "Methods", and "About us". The main table displays 11 phenotypes with their respective quality metrics. A sidebar on the right shows a tree view of phenotypes under the "Cardiology" category.

| category | PheWAS_code | abbr          | PheWAS_name | model                                | ICD_PPV        | ICD_AUC | AUC   | PPV   | TPR  |      |
|----------|-------------|---------------|-------------|--------------------------------------|----------------|---------|-------|-------|------|------|
| 1        | ONC         | PheWAS:189.21 | BLCA        | <a href="#">Bladder cancer</a>       | PheNorm_ICD    | 0.80    | 0.903 | 1.000 | 1.00 | 0.42 |
| 2        | ONC         | PheWAS:204    | LEUK        | <a href="#">Leukemia</a>             | PheNorm_ICD    | 0.73    | 1.000 | 1.000 | 1.00 | 0.91 |
| 3        | PSYCH       | PheWAS:297.1  | SI          | <a href="#">Suicidal ideation</a>    | PheNorm_ICDNLP | 0.93    | 0.786 | 1.000 | 1.00 | 0.43 |
| 4        | PSYCH       | PheWAS:305.2  | EATD        | <a href="#">Eating disorder</a>      | PheNorm_ICDNLP | 0.53    | 0.482 | 1.000 | 1.00 | 1.00 |
| 5        | NEURO       | PheWAS:327.4  | INSOM       | <a href="#">Insomnia</a>             | PheNorm_ICDNLP | 0.93    | 0.821 | 1.000 | 1.00 | 0.50 |
| 6        | CARDIO      | PheWAS:452.2  | DVT         | <a href="#">Deep vein thrombosis</a> | PheNorm_ICDNLP | 0.87    | 0.692 | 1.000 | 1.00 | 1.00 |
| 7        | NEURO       | PheWAS:817    | CONC        | <a href="#">Concussion</a>           | PheNorm_NLP    | 0.73    | 0.682 | 1.000 | 1.00 | 0.42 |
| 8        | METAB       | PheWAS:250.1  | T1DM        | <a href="#">Type 1 diabetes</a>      | PheNorm_ICD    | 0.17    | 0.882 | 0.926 | 0.99 | 0.42 |
| 9        | ONC         | PheWAS:184.11 | OVCA        | <a href="#">Ovarian cancer</a>       | PheNorm_ICDNLP | 0.60    | 0.926 | 0.926 | 0.99 | 0.42 |
| 10       | ONC         | PheWAS:182    | UTCA        | <a href="#">Uterine cancer</a>       | PheNorm_ICD    | 0.50    | 0.867 | 0.926 | 0.99 | 0.42 |
| 11       | GI          | PheWAS:555.1  | CD          | <a href="#">Crohn's disease</a>      | PheNorm_mean   | 0.54    | 0.961 | 0.926 | 0.99 | 0.42 |

**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 (PVD)
  - Pulmonary heart disease (PHD)



# High Quality Phenotypes for Research Studies

The screenshot displays the Partners Biobank Portal interface. On the left is a 'Navigate Terms' sidebar with a tree view of categories including Biobank Consent Information, Demographics, Genomics, Health Information Survey, Sample Types, and Curated Disease Populations. The 'Curated Disease Populations' section is expanded to show 'Congestive Heart Failure (CHF)', which includes sub-phenotypes like 'CHF - current or past history (PPV 0.90) - 700' and 'CHF - no history (NPV 0.99) - 36024'. Other diseases listed include Asthma, Bipolar Disorder, Breast Cancer, COPD, Crohn's Disease, Depression, Epilepsy, Gout, Hypertension, Multiple Sclerosis, Obesity, Rheumatoid Arthritis, Schizophrenia, and Type 1/2 Diabetes Mellitus.

The main 'Query Tool' area 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                               |             |         | Group 2  |             |         | Group 3  |             |         |
|---------------------------------------|-------------|---------|--|-------------|---------|--|-------------|---------|
| Dates                                 | Occurs > 0x | Exclude | Dates  | Occurs > 0x | Exclude | Dates  | Occurs > 0x | Exclude |
| Treat Independently                   |             |         | Treat Independently                            |             |         | Treat Independently  |             |         |
| Primary dilated cardiomyopathy - 4002 |             |         | CHF - current or past history (PPV 0.90) - 700 |             |         | Gene [contains "TTN AND Homozygous AND (Frameshift OR missense OR nonsense OR start_loss OR stop_loss)"] |             |         |
| one or more of these                  |             |         | AND  |             |         | one or more of these   |             |         |
|                                       |             |         | AND  |             |         | one or more of these   |             |         |

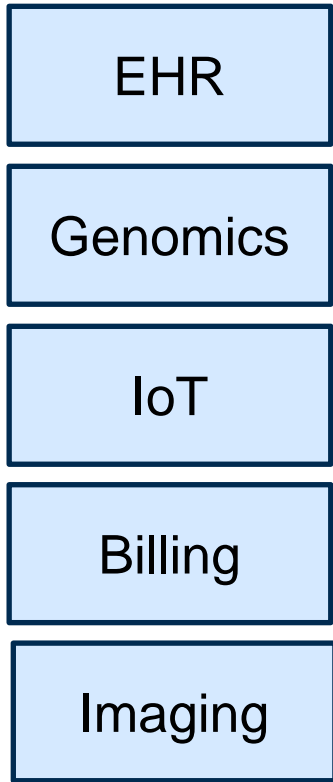
Buttons for 'Run Query', 'Clear', and 'New Group' are visible. Below the query tool, tabs for 'Show Query Status', 'Graph Results', 'Query Report', and 'Download Results' are present. The 'Graph Results' tab is active, showing a summary box:

Number of patients

# 70

For Query "Prima-CHF --Gene@14:22:40"

# 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

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

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

RADIOLOGY

No records found

MICROBIOLOGY

2016-07-27 MIC SPUTUM  
2016-07-27 MIC URINE  
2016-07-12 MIC URINE

PATHOLOGY

No records found

LABS

No records found

ALLERGIES

2018-11-26 ALGY PENICILLINS

COVID 40

| COVID STATUS       | RISK FACTORS       | SEVERITY           | MANAGEMENT           |
|--------------------|--------------------|--------------------|----------------------|
| SARS-COV-2 PCR...  | Height             | ACE inhibitors     | Cefepime/ceftria...  |
| SARS-COV-2 PCR...  | Weight             | ARBs               | Azithromycin         |
| SARS-COV-2 IgM ... | BMI                | Beta blockers      | Vancomycin           |
| SARS-COV-2 IgG ... | Smoker             | Bisphosphonates    | Oseltamivir          |
| COVID precautions  | Healthcare worker  | CCBs               | Remdesivir           |
| Known contact      | Housing            | Cough              | Hydroxychloroquine   |
|                    | Pregnant           | Chemotherapy       | Darunavir/cobicistat |
|                    | Recent travel      | Statins            | Lopinavir/ritonavir  |
|                    | Asthma             | Steroids           | Interferon beta B1   |
|                    | COPD               | NSAIDs             | Tocilizumab          |
|                    | Diabetes           | Thiazide diuretics |                      |
|                    | Heart failure      | Transplant meds    |                      |
|                    | High cholesterol   | Warfarin           |                      |
|                    | Hypertension       | Flu vaccine        |                      |
|                    | MI                 | BCG vaccine        |                      |
|                    | Immunodeficiency   |                    |                      |
|                    | Autoimmune disease |                    |                      |

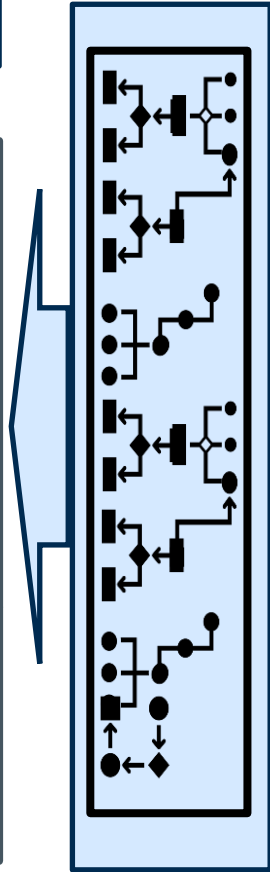
Study Specific Conditions and Common Data Elements

Recent Past Medical History

| CV            | PULM         | GI             | GU          | ID              | NEURO          | HEME           | OTHER         |
|---------------|--------------|----------------|-------------|-----------------|----------------|----------------|---------------|
| AAA/Dxsm      | COPD         | Divericulosis  | AKI         | Abscess         | AMS            | Anemia         | Agitation     |
| AFib/Flutter  | DVT          | ESLD           | CRD         | Bacteremia      | Dementia       | DIC            | Capacity      |
| CAD           | Hemophysis   | GERD           | Ematuria    | Cholecystitis   | Encephalopathy | HIT            | Depression    |
| Chest Pain    | ILD          | GI Bleeding    | Kidney Dis  | Endo Dis        | Hypercoag      | HT             | EOH           |
| CHF           | OSA          | Hepatitis      | Na Disorder | HIV/AIDS        | Neuropathy     | MM             | Drug Abuse    |
| Conduction Dz | PE           | IBD            | RRT         | Pneumonia       | PRES           | Myelodysplasia | Noncompliance |
| MI            | Pulm HTN     | Intes. Ischem. | Stones      | Soft Tissue Inf | Seizures       | TTP/HUS        | Psych         |
| PVD           | Pulm Nodules | Ulcers         | UTI         | TB              | Stroke/TIA     | XRT            | Smoking       |
| Valve Disease |              | Varices        |             | UTI             |                |                | Suicide       |
| VT/VF         |              |                |             |                 |                |                |               |

Core Conditions

DATA UPDATED: 01/16/2020 12:41:22 PM

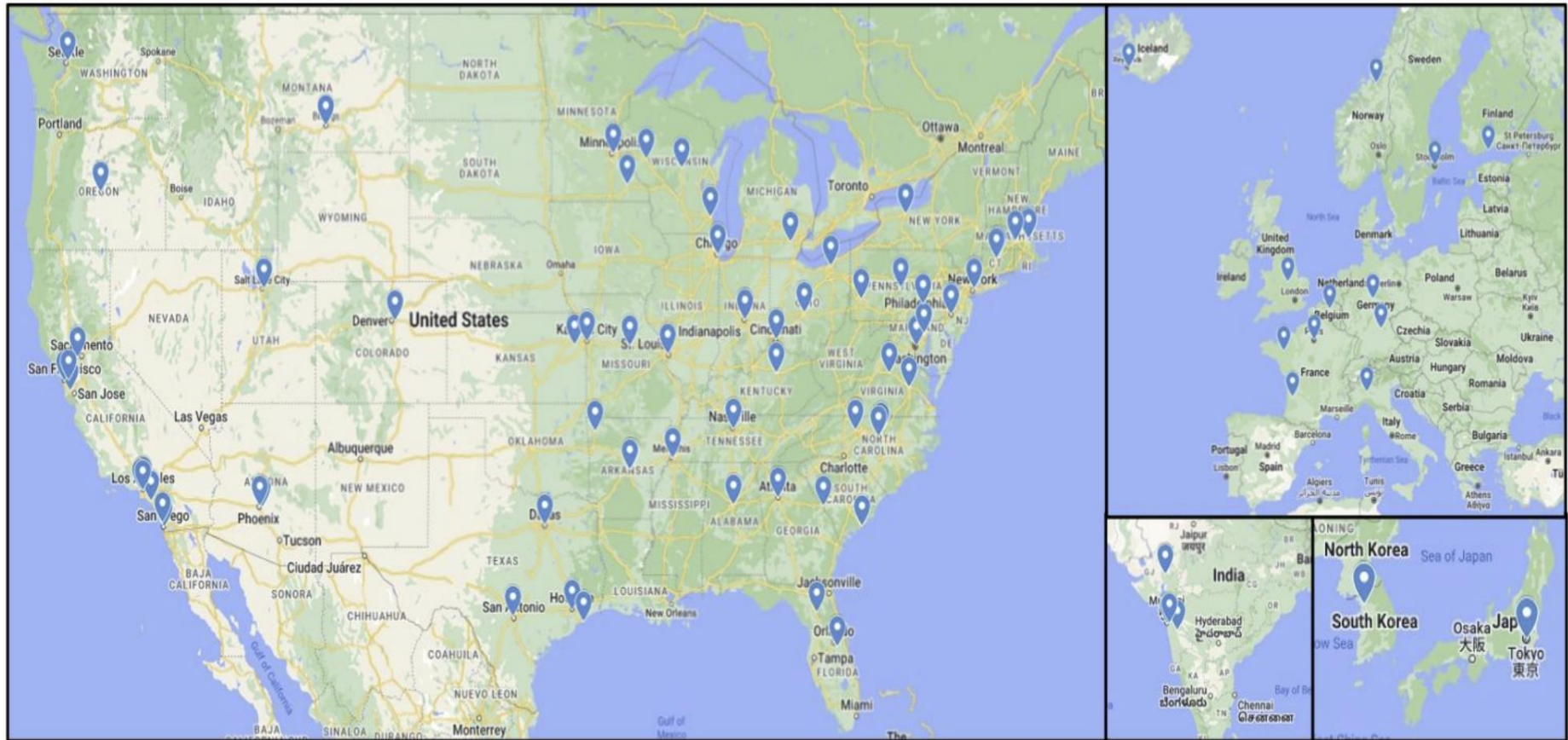


# High Throughput Methods for supporting Translational Research

- 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 through Digital Twin

# I2b2 Implementations

>250 across the USA and Internationally, some illustrated below:



# Federated Queries

**SHRINE** Project: SHRINE User: Shawn Murphy Find Patients | Message Log | Help | Logout

**Query Tool**

Query Name: Acute hepatitis@12:45:50

Search by Names Search by Codes

Containing hepatitis c

Find Any Category

- Hepatitis b core antibody measurement
- Hepatitis b core igm antibody test
- Hepatitis b surface antigen [hbsag] carrier
- Hepatitis b vaccine injection administered or previous
- Hepatitis b virus (hbv) status assessed and results i
- Hepatitis c antibody
- Hepatitis c antibody
- Hepatitis c antibody confirmatory test
- Hepatitis c antibody; confirmatory test (eg, immunobi
- Hepatitis c antibody measurement
- Hepatitis c antibody test

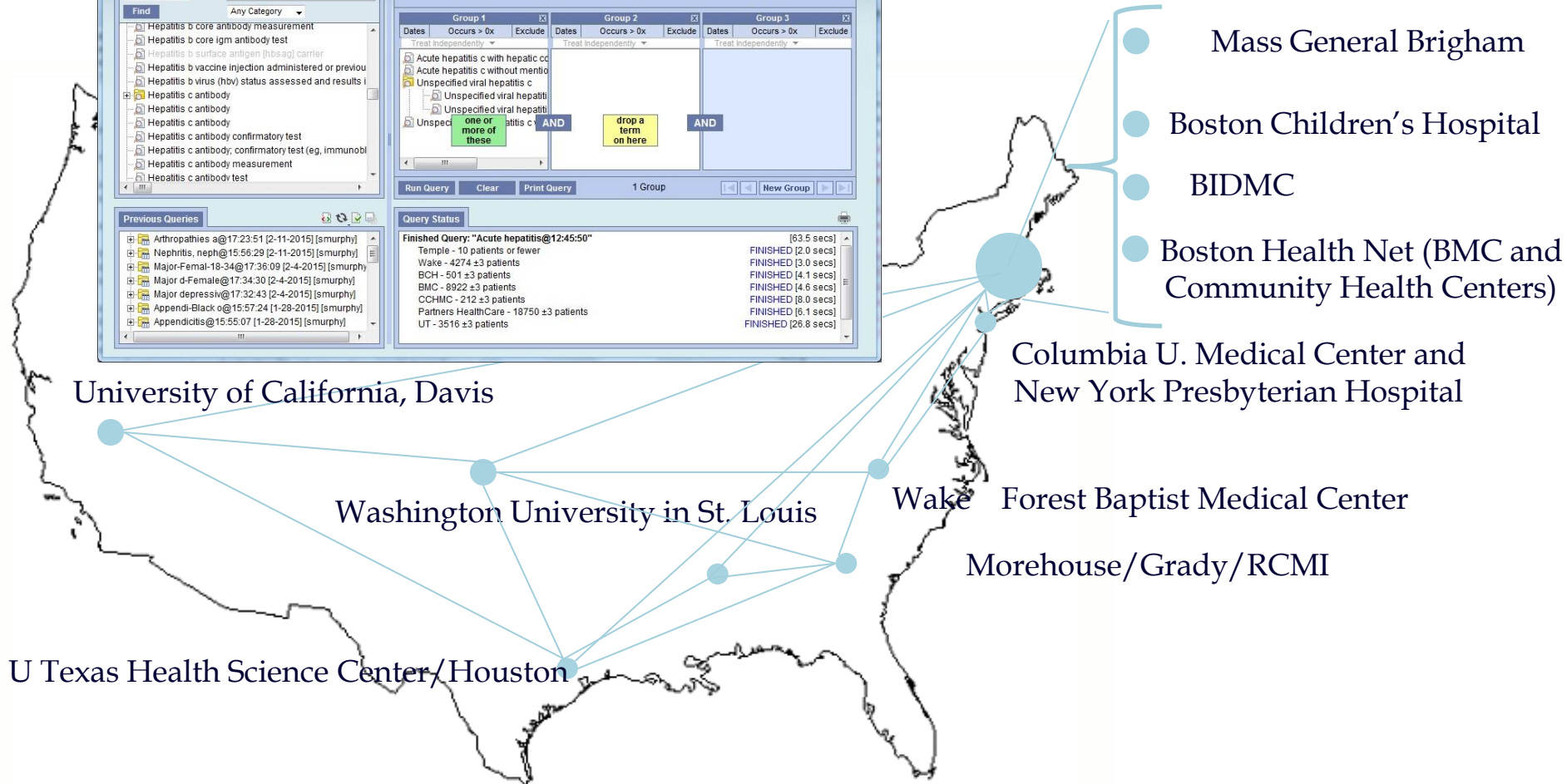
Previous Queries

- Arthropathies a@17:23:51 [2-11-2015] [smurphy]
- Nephritis, neph@15:56:29 [2-11-2015] [smurphy]
- Major-Femal-18-34@17:36:09 [2-4-2015] [smurphy]
- Major d-Female@17:34:30 [2-4-2015] [smurphy]
- Major depressio@17:32:43 [2-4-2015] [smurphy]
- Appendi-Black o@15:57:24 [1-28-2015] [smurphy]
- Appendicitis@15:55:07 [1-28-2015] [smurphy]

**Query Status**

Finished Query: "Acute hepatitis@12:45:50" [63.5 secs]

|   |                      |
|---|----------------------|
| Temple - 10 patients or fewer           | FINISHED [2.0 secs]  |
| Wake - 4274 ±3 patients                 | FINISHED [3.0 secs]  |
| BCH - 501 ±3 patients                   | FINISHED [4.1 secs]  |
| BMC - 8922 ±3 patients                  | FINISHED [4.6 secs]  |
| CC-HMC - 212 ±3 patients                | FINISHED [8.0 secs]  |
| Partners HealthCare - 18750 ±3 patients | FINISHED [6.1 secs]  |
| UT - 3516 ±3 patients                   | FINISHED [26.8 secs] |



# Drive Pragmatic Clinical Studies

The screenshot displays the i2b2 Query & Analysis Tool interface. The browser address bar shows the URL `i2b2act.dipr.partners.org/ACTFullDemo/#`. The main interface is divided into several sections:

- Navigate Terms:** A tree view on the left showing medical categories such as Cardiovascular agents, Central nervous system agents, Analgesics, Anorexiant, Anticonvulsants, Antiemetic/antivertigo agents, Antiparkinson agents, General anesthetics, and Miscellaneous central nervous system agents.
- Workplace:** A section below the terms list showing a project named 'cronjob' and a 'Patients List' with IDs: 160803, 24339, 60677, and 91326.
- Previous Queries:** A list of recent queries, including 'Parkins-Tobacco@12:12:29 [4-12-2017] [cronjob]' and others.
- Patient Set Viewer:** A central table with columns for patient ID, gender, age, race, and medication. The table contains 15 rows of patient data. The row for patient ID 160803 is highlighted in blue, and a large blue arrow points from this row towards the analysis section.
- Patient Data Analysis:** A section on the right for configuring data analysis. It includes dropdown menus for 'Vitals' (Weight, Blood Pressure), 'Medications' (Metoprolol, Spironolactone, Valsartan), and 'Labs' (Creatinine, Potassium, eGFR). It also shows a timeline for 'Viewing 2 weeks starting from 01/07/17' with counts for ER Visits (14), ICU Visits (2), Floor Visits (56), and Clinic Visits (33). Below the timeline is a grid showing data points for various metrics over time.

# RECOVER Study Data Harmonization



En Español



[HOME](#) | [WHAT IS LONG COVID?](#) | [RESEARCH](#) ▾ | [NEWS & EVENTS](#) ▾ | [ABOUT THE INITIATIVE](#) ▾

## RECOVER: Researching COVID to Enhance Recovery

The National Institutes of Health (NIH) created the RECOVER Initiative to learn about the long-term effects of COVID.

The goal of RECOVER is to rapidly improve our understanding of and ability to predict, treat, and prevent PASC (post-acute sequelae of SARS-CoV-2), including Long COVID.

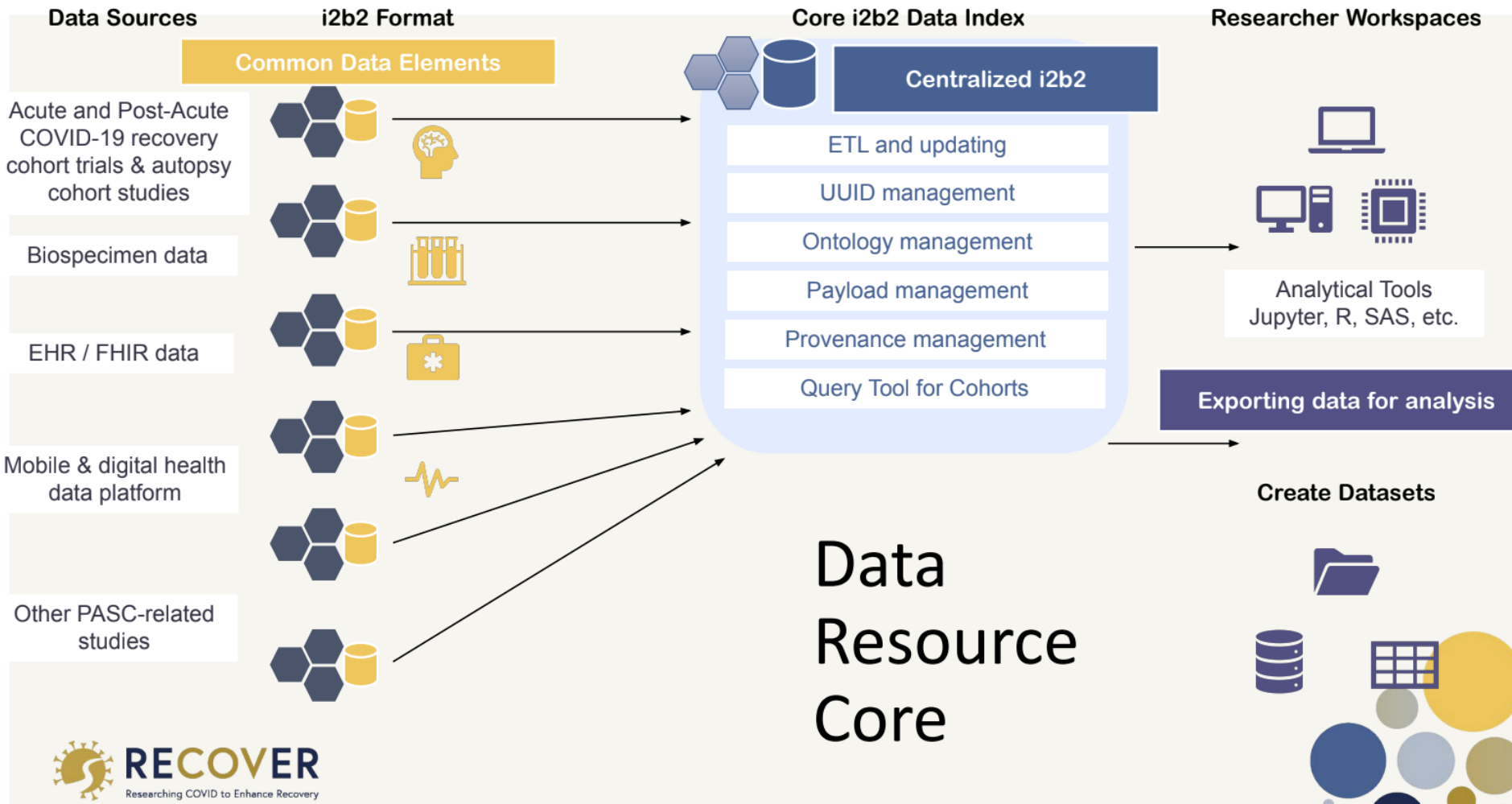
[LEARN MORE ABOUT LONG COVID](#)



<https://recovercovid.org>



# Data harmonized within i2b2 star schema





# Concepts in database available in harmonized ontology

## Ontology-driven normalization of source data

**Adult REDCap**  
*V. Castro, R. Metta, B. Benoit, et al.*

**Autopsy REDCap**  
*V. Castro, R. Metta, B. Benoit, et al.*

**Bio-specimens**  
*A. Cagan, P. Svendsen, et al.*

**ENACT OMOP**  
*M. Morris, S. Visweswaran, et al.*

**Sensor Data**  
*G. Weber, A. Cagan, et al.*

**RECOVER i2b2 Query & Analysis Tool** RECOVER i2b2 (Stage)  
Researching COVID to Enhance Recovery

Terms Info

- RECOVER Adult (2023.06)
  - + Enrollment and Consent Tracking
  - + Medications
  - + Participant Surveys
  - + Tier 1 Clinical Assessments
  - + Tier 2 Procedures
- + RECOVER Autopsy (2023.06)
- RECOVER Biospecimens
  - + Adult
  - + Autopsy
  - + Pediatric
- RECOVER Demographics
  - + Age at enrollment
  - + Enrollment protocol
  - + Enrollment site
  - + Sex at birth
  - + Vital status
- + RECOVER EHR - Diagnoses ICD10-ICD9 (2023.08)
- + RECOVER EHR - Medications VA Classes (2023.08)
- + RECOVER Pediatric (2023.06)
- + RECOVER Pediatric Caregiver (2023.06)

Find Patients Analysis Tools

Find Patients  with  with

drag a concept here to define inclus

Name:

Query Status

# Includes study participant survey data

## Adult survey metadata (Castro, Metta, Benoit, et al.)

Adult  
REDCap

V. Castro, R.  
Metta, B. Benoit,  
et al.

The screenshot displays the RECOVER i2b2 Query & Analysis Tool interface. The main content area shows a hierarchical tree view of survey metadata under the heading "RECOVER Adult (2023.06)". The tree includes the following categories and items:

- RECOVER Adult (2023.06)
  - Enrollment and Consent Tracking
  - Medications
  - Participant Surveys
    - Alcohol and Tobacco (Enrollment)
    - Alcohol and Tobacco (Followup)
    - Comorbidities
    - COVID Treatment (Enrollment)
    - Demographics Survey (Enrollment)
    - Disability (Enrollment)
    - Long COVID Treatment Trial
    - New COVID Infection
      - Did the participant get infected with COVID between their previous event and this event?
        - Did the participant get infected with COVID between their previous event and this event? (0, No)
        - Did the participant get infected with COVID between their previous event and this event? (1, Yes)
      - Date of COVID diagnosis:
      - How was the participant diagnosed with COVID?
        - How was the participant diagnosed with COVID? (1, Diagnosed by positive PCR test)
        - How was the participant diagnosed with COVID? (2, Diagnosed by positive antigen result (rapid test))
        - How was the participant diagnosed with COVID? (3, Diagnosed by positive antibody result (blood test))
        - How was the participant diagnosed with COVID? (4, Diagnosed by a positive result, but not sure which test)
        - How was the participant diagnosed with COVID? (5, Diagnosed by a doctor based on symptoms)
        - How was the participant diagnosed with COVID? (6, Self diagnosed)
    - PASC Symptoms

The interface also features a search bar with a magnifying glass icon and a dropdown menu set to "Any Category". On the right side, there is a "Find Patients" panel with a search input field and a "Find Patient" button. Below this, there is a "Name:" label and an input field. At the bottom right, there is a "Query Status" panel.

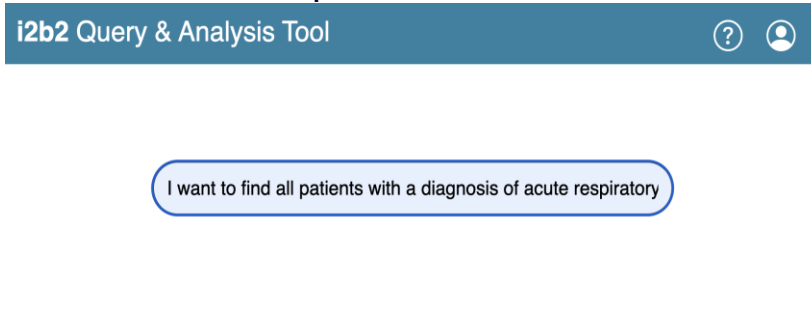
# New i2b2 Query Tool to be released:

The screenshot displays the i2b2 Query & Analysis Tool interface. On the left, a list of queries is shown, including "Diabetes mellitus-Disorders of parathyroid gland@10:15:45 [9-22-2023] [demo]", "Query 8 - Timeline Notes [9-22-2023] [demo]", and "Query 4- Female-Age Temporal - Endocrine Hormones [9-22-2023] [demo]". The right panel shows the "Find Patients" window for the selected query. It includes options for "with", "without", and "when" (selected), and a list of events: "Event 1: Endocrine, nutritional and metabolic diseases (e00-e89)" and "Event 2: Hormones". A relationship is defined: "The start of the first occurrence of Event 1 occurs before the start of the first occurrence of Event 2". The "Name" field shows "(t) Query 4- Female-Age Temporal - Endocrine Hormones". The "Query Status" section indicates "Finished Query: (t) Query 4- Female-Age Temporal - Endocrine Hormones" with a compute time of 0.4 seconds. A large orange box displays the "Number of patients" as 6.

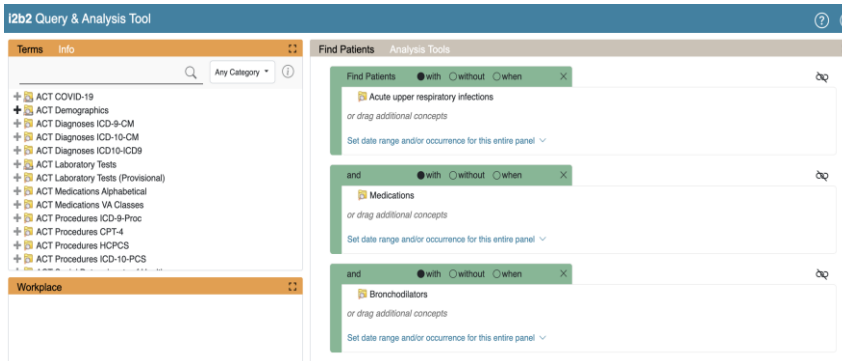
<https://i2b2transmart.org/2023-i2b2-symposium/2023-symposium-recordings-slides/>

# I2B2 AI

User asks a question



Result rendered in web client



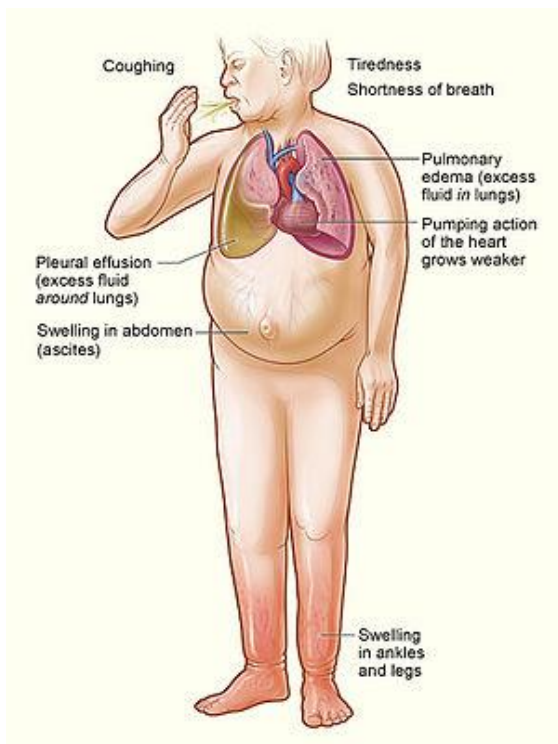
AI returns response in i2b2 format



# 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 through Digital Twin

# 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*)

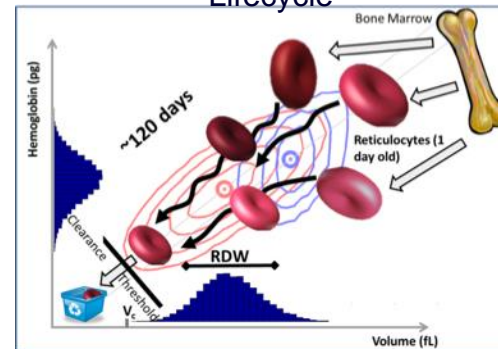
## Results and Conclusions

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

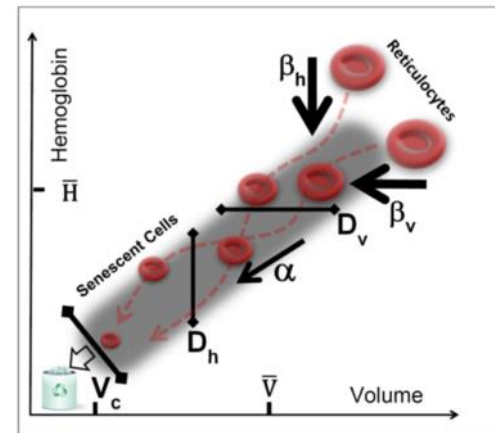
## Applications

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.

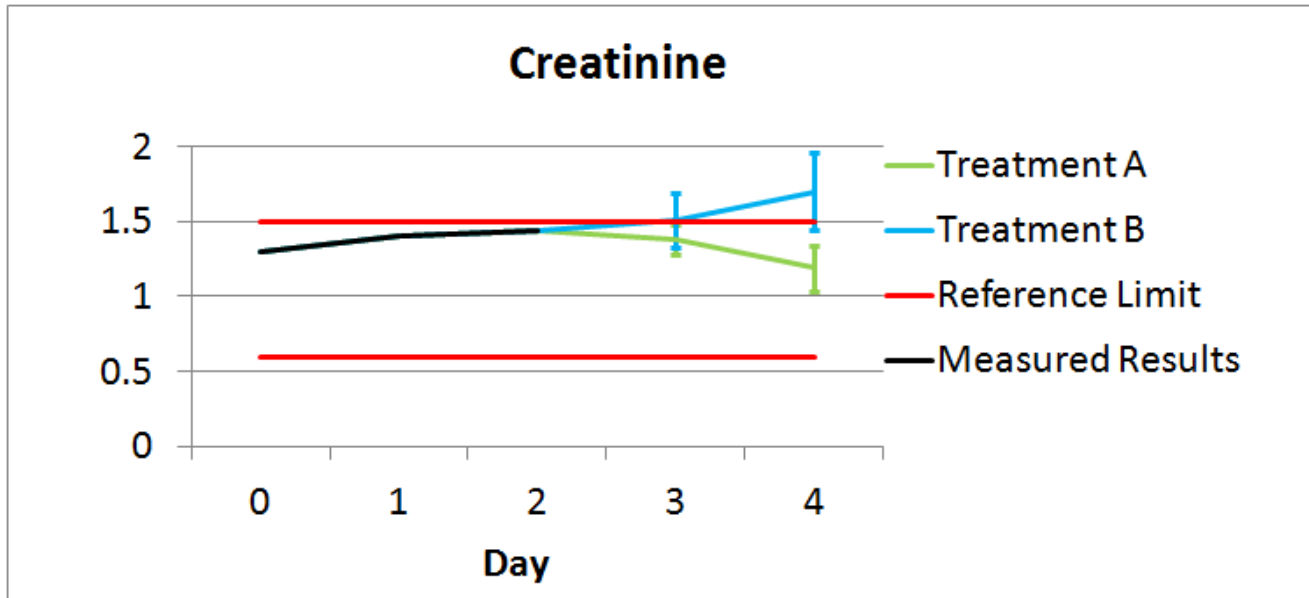
## Dynamic Model of the RBC Lifecycle



## 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



# 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](#)



Exploring  
Integration of  
MGH Path renal  
predictive model

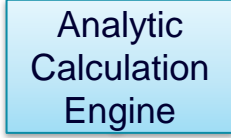
# Bringing Big Data into Clinical Care with Open App Development



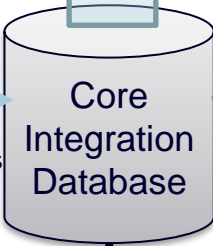
**Clinician**



**DATA**  
 GeneInsight,  
 mHealth, ePath,  
 Medical Images,  
 25 years of  
 Legacy electronic  
 data, and Other  
 External Systems



Analytics have direct access to repository

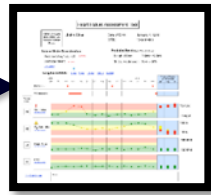


FHIR interface for real time updates

**Laboratory Personnel**



**SMART App in Lab**



**Non-EHR Users View Standalone App**

# Transforming Care in the Digital Age

Digital and IoT devices continuously output Patient Data



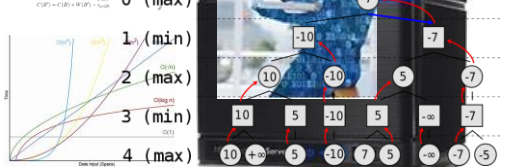
```

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



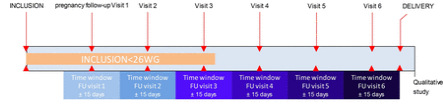
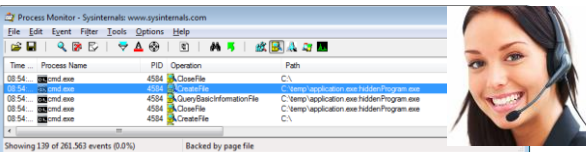
shutterstock · 388661236

cost  
 $n + 1$   
 $n$   
 $n$   
 $\sum_{i=1}^n \sum_{j=1}^n$   
 $\sum_{i=1}^n (j-1)$   
 $0$   
 $\sum_{i=1}^n (j-1)$   
 $n$   
 $n$   
 $n$   
 $n$   
 $n$   
 $n$   
 $n$   
 $1$



# PATIENT

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

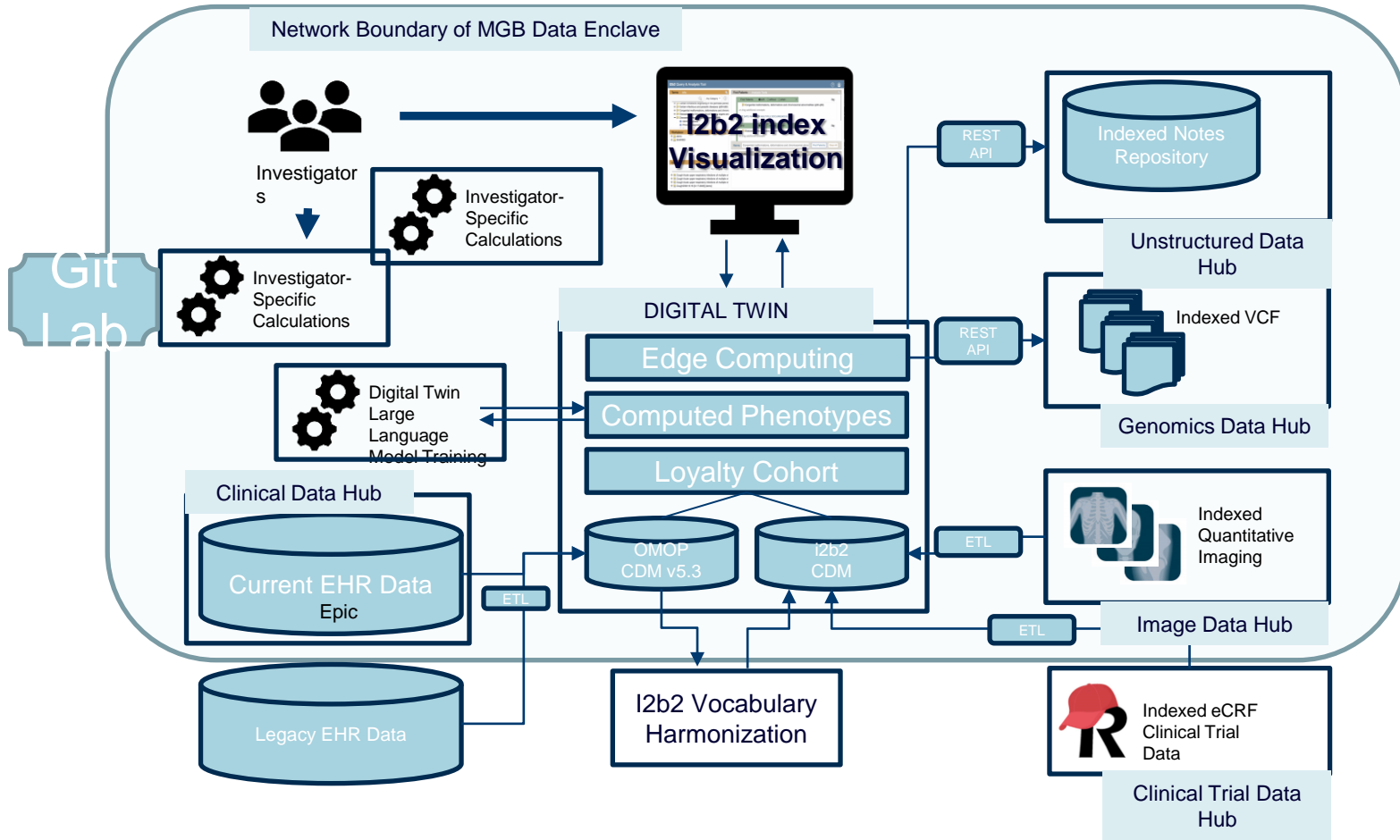


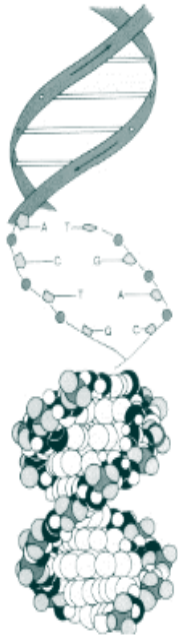
Navigator Model dramatically increases Frequency and Convenience for Patient Communication

System drives Pragmatic Clinical Trials Leading to Continuous Process Improvement



# MGB Data Enclave Overview





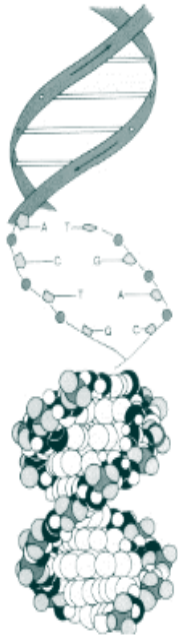
# I2b2 tranSMART Software

*i2b2 Homepage (<https://www.i2b2.org>)*

*i2b2 Software (<https://www.i2b2.org/software>)*

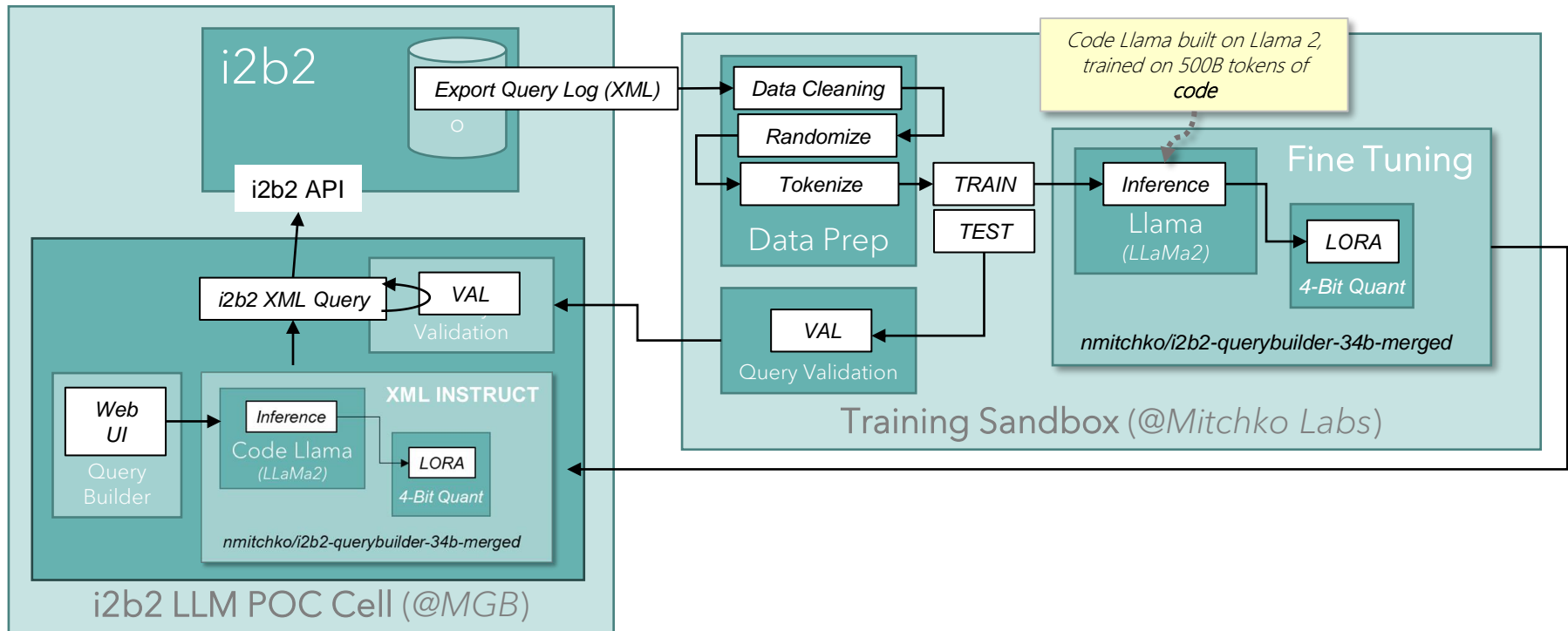
*i2b2 Community Site (<https://community.i2b2.org>)*

*<https://i2b2transmart.org/2023-i2b2-symposium/2023-symposium-recordings-slides/>*



# APPENDIX

# AI-ENABLED QUERY BUILDER: (I.E. INSTRUCTION-TUNED POC)



# I2B2 TRAINING ENVIRONMENT

## Training Requirements

GPUs (48GB of VRAM or More)  
Patience  
Cooling

## Software

Qlora – State of the art mixed precision training library  
Transformers, Pytorch, Accelerate – ML Libraries

## Experience:

Nicholai Mitchko background in ML and other model finetunes

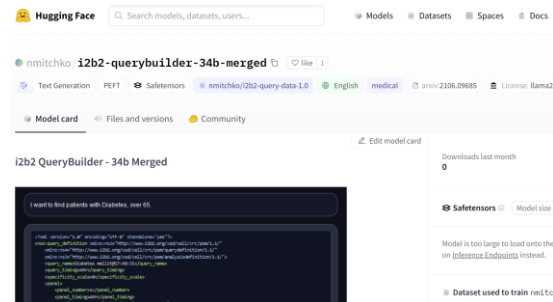
## Training Run

3x Nvidia RTX 6000  
3 days time  
271 Quadrillion Floating Point Operations  
10.05 epochs (runs through the dataset)

## All Open source

Found at

<https://huggingface.co/nmitchko>



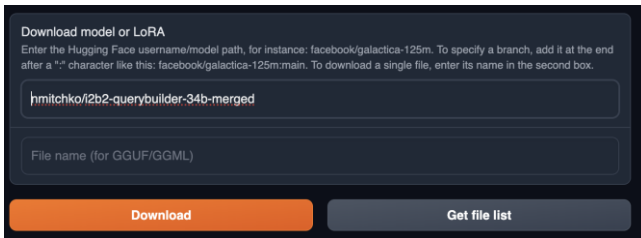


# I2B2 AI DEVELOPMENT SANDBOX

WebUI oobabooga

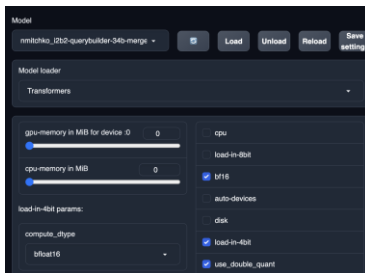
<https://github.com/oobabooga/text-generation-webui>

## Download LLM



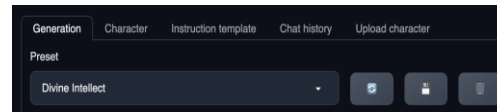
hmitchko/i2b2-querybuilder-34b-merged

## Load Model



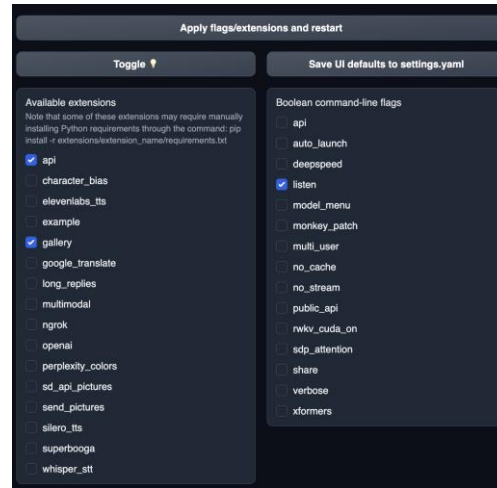
Compute\_dtype: bfloat16  
Check: bf16  
Check: load-in-4bit  
Check: use\_double\_quant

## Parameters



## Divine Intellect

## Enable API



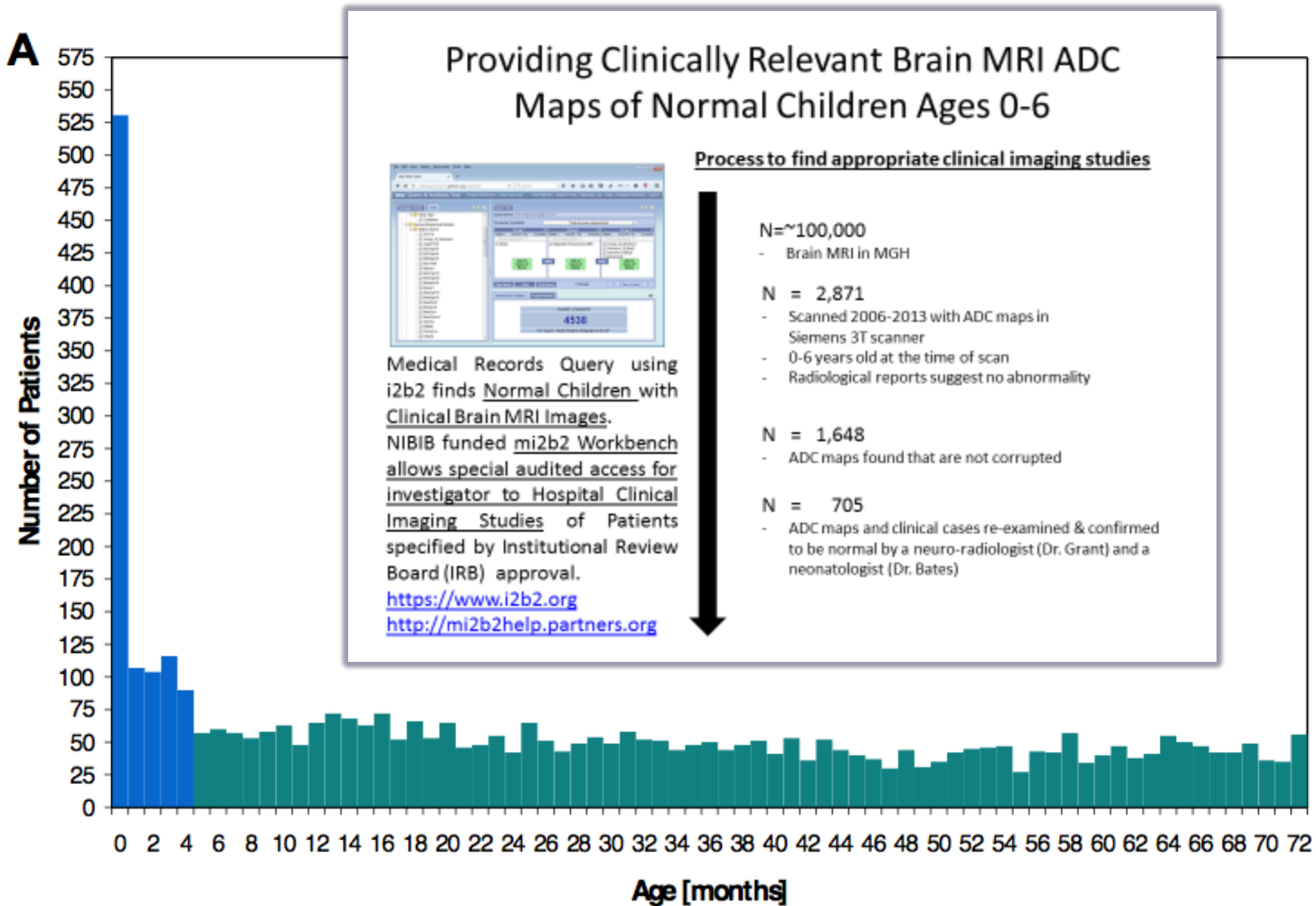
## I2b2 AI Cell

<https://github.com/i2b2/i2b2-ai>

## I2b2 Web Client

[https://github.com/i2b2/i2b2-webclient-prototype/tree/i2b2\\_ai](https://github.com/i2b2/i2b2-webclient-prototype/tree/i2b2_ai)

# 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**

