Opportunities and Limitations of Big Data to Address Diversity

Shawn Murphy MD, Ph.D.
The Research Data Warehouse at Partners Healthcare

Research Patient Data Registry

Multiple Systems at Partners:
- Billing Data
- Epic Data
- Research Data (consent to contact)
- Specimen Data
- Laboratory Data

...
Research Patient Data Registry (RPDR) at Partners Healthcare to find patient cohorts for clinical research

1) Queries for aggregate patient numbers
   - Warehouse of in & outpatient clinical data
   - 6.5 million Partners Healthcare patients
   - 2.2 billion diagnoses, medications, procedures, laboratories, & physical findings coupled to demographic & visit data
   - Authorized use by faculty status
   - Clinicians can construct complex queries
   - Queries cannot identify individuals, internally can produce identifiers for (2)

2) Returns detailed patient data
   - Start with list of specific patients, usually from (1)
   - Authorized use by IRB Protocol
   - Returns contact and PCP information, demographics, providers, visits, diagnoses, medications, procedures, laboratories, microbiology, reports (discharge, LMR, operative, radiology, pathology, cardiology, pulmonary, endoscopy), and images into a Microsoft Access database and text files.
2014’s usage of RPDR

- > 5100 registered users over 12 years, 655 new in 2014
- 583 teams/year gathering data for research studies
- 2634 detailed patient data sets returned to these teams in 2014, containing data of 24.7 million patient records.
- From a survey of 153 teams
  - Importance of the data received from the RPDR was evaluated in relation to the study it was supporting.
  - $94-136 million total research support critically dependent on RPDR from patient data received throughout life of funding.
- ~300 data marts were created to support hospital operations, representing about 80 million patient records
Security and Patient Confidentiality of Step 1

- All patients at Partners are added when seen in “Affiliated Covered Entity”
  - HIPAA notification that their data may be used for research and operations upon registration.

- RPDR data is anonymized at the Query Tool.
  - Aggregated numbers are obfuscated to prevent identification of individuals; automatic lock out occurs if pattern suggests identification of an individual is being attempted.

A Security Architecture for Query Tools used to Access Large Biomedical Databases
Shawn N. Murphy, MD, Ph.D. and Henry C. Chuah, MD, M.S.
Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA.

- Queries done in Query Tool available for review by RPDR team, a user lock out will specifically direct a review.

- Concept of “established medical investigator” is promoted by classification as a faculty member or designated list from IRB.
Security and Patient Confidentiality of Step 2

- Only studies approved by designated Directors of Operation at Hospitals or the Institutional Review Board (IRB) are allowed to receive detailed data.

- Queries may be set up by a workgroup member, but faculty sponsor on Approval / IRB protocol must directly approve all queries that return identified data.

- Special controls exist when distributing data regarding HIV antibody and antigen test results, substance abuse rehab programs, and genetic data, due to specific state and federal laws.
FINDING PATIENTS

Query items

Person who is using tool

Query construction

Results - broken down by number distinct of patients
Please supply a name for this query:

Jsut Diagnos AMI

- Create a patient set from this query

Please choose one of these two timing models ...

Either (1) Some item from each group must have occurred in the same visit; or (2) Some item from each group may have occurred at any visit. For example, if one wanted to find patients whose diabetes was treated at the MGH while an inpatient, choose the timing model where items occurred at the same visit.

- Same Visit

<table>
<thead>
<tr>
<th>Patient</th>
<th>Date</th>
<th>Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000004</td>
<td>12/15/1999</td>
<td>MGH</td>
<td>Inpatient</td>
<td>Diabetes</td>
<td></td>
</tr>
</tbody>
</table>

- Any Visit

<table>
<thead>
<tr>
<th>Patient</th>
<th>Date</th>
<th>Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000004</td>
<td>2/28/1999</td>
<td>BWH</td>
<td>Inpatient</td>
<td>Seizure</td>
<td></td>
</tr>
<tr>
<td>00000004</td>
<td>10/05/1999</td>
<td>MGH</td>
<td>Outpatient</td>
<td>Atib</td>
<td></td>
</tr>
<tr>
<td>00000004</td>
<td>12/15/1999</td>
<td>BWH</td>
<td>Outpatient</td>
<td>Diabetes</td>
<td></td>
</tr>
</tbody>
</table>
Please complete...

Please supply a name for this query:

AMI and CK-MB > 3.5

- [ ] Create a patient set from this query
- [ ] Run the query as background job

Please choose one of these two timing models...

- Same Visit
- Any Visit

Either (1) Some item from each group must have occurred in the same visit; or (2) Some item from each group may have occurred at any visit. For example, if one wanted to find patients whose diabetes was treated at the MGH while an inpatient, choose the timing model where items occurred at the same visit.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Date</th>
<th>Groups</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000004</td>
<td>12/15/1999</td>
<td>MGH</td>
<td>Inpatient</td>
<td>Diabetes</td>
<td></td>
</tr>
<tr>
<td>00000004</td>
<td>2/28/1999</td>
<td>BWH</td>
<td>Inpatient</td>
<td>Seizure</td>
<td></td>
</tr>
<tr>
<td>00000004</td>
<td>10/05/1999</td>
<td>MGH</td>
<td>Outpatient</td>
<td>Alib</td>
<td></td>
</tr>
<tr>
<td>00000004</td>
<td>12/15/1999</td>
<td>BWH</td>
<td>Outpatient</td>
<td>Diabetes</td>
<td></td>
</tr>
</tbody>
</table>
MATCHING PATIENTS

Previous query items

Control set construction

Case set construction

Estimate set size and run program
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
Select the sites from which you would like to receive data

(Please note that different sites and Institutional Review Boards (IRBs) may have different policies regarding obtaining patient data. Detail of the policies can be read here)

- Massachusetts General Hospital (MGH)
- Brigham and Women’s Hospital (BWH)
- Newton Wellesley Hospital (NWH)
- Spaulding Rehabilitation Hospital (SRH)
- Faulkner Hospital (FH)
- North Shore Medical Center (NSMC)
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: NIWH Demo 1
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

Select others who require access to this data

Selected faculty sponsor(s):
Shawn Murphy, MD

Available
- Peng, Zhaoping Z.
- Nalichowski, Ruth

Chosen
- Gainer, Vivian S.

Add ->
Add All ->>
<- Remove
<- Remove All
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

☐ 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.
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)
Please select the laboratory tests that should be returned

(Drag the lab items from the left panel and drop into the panel below)

Selected Data Items

- CK-MB Index (Group:CKMBRI) (Loinc:2158-4)
RPDR DISCLAIMER: All requests, such as this one, are reviewed by the IRB to assure compliance with the written protocol. If the data requested is not authorized by the protocol, those involved with its retrieval face disciplinary action including loss of federal funding, loss of employment, and/or criminal prosecution.

[Read Full IRB Policy]

Enter IRB Password: 
Data is gathered from RPDR and other Partners sources.

Output files placed in special directory.

Files include a Microsoft Access Database.
Creating **Enterprise Wide** Query and Analysis System for Big Data

Integrates disparate islands of patient data (clinical and research data) onto a Common Big Data Platform
The Partners Biobank provides samples (plasma, serum, and DNA) collected from consented patients.

- 31,000 patients have consented to date!
- Samples are available for distribution to Partners investigators* to help identify novel Personalized Medicine opportunities that reduce cost and provide better care

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

Research Discoveries

Improved Clinical Care for All Patients
Biobank Integrative Genomics Strategy

Partners BioBank Samples
(Whole Blood Extracted DNA/RNA)

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

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

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

Genome/Transcriptome Analysis: ~$100/sample

Genome/Transcriptome/Epigenome Analysis: ~$260/sample
Curating a Disease Algorithm

1. Create a gold standard training set.

2. Create a comprehensive list of features (concepts/variables) that describe the phenotype 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

<table>
<thead>
<tr>
<th>Validated Phenotype</th>
<th>Count*</th>
<th>Predictive Positive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar Disease</td>
<td>71</td>
<td>89%</td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>387</td>
<td>90%</td>
</tr>
<tr>
<td>Coronary Artery Disease</td>
<td>2,420</td>
<td>97%</td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td>453</td>
<td>90%</td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>94</td>
<td>90%</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>550</td>
<td>90%</td>
</tr>
<tr>
<td>Type 2 Diabetes Mellitus</td>
<td>1,887</td>
<td>97%</td>
</tr>
<tr>
<td>Ulcerative Colitis</td>
<td>330</td>
<td>90%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Healthy Controls based on Charlson Index</th>
<th>Count**</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 10-year survival probability is &gt;98.3%</td>
<td>2,206</td>
</tr>
<tr>
<td>1 – 10-year survival probability is &gt;95.87%</td>
<td>4,343</td>
</tr>
<tr>
<td>2 – 10-year survival probability is &gt;90.15%</td>
<td>6,545</td>
</tr>
</tbody>
</table>

* Based on 15,880 patients
** Based on 21,300 patients
Step 1: Run a query in the Biobank Portal

Two criteria:

a. The patient is a smoker (from the Health Information Survey)

b. The patient agrees to be re-contacted
You can download the de-identified data for this query as an Excel/CSV file.
### Patient Set Viewer

**Setting up Parameters**

1. Drop a **Patient Set** or a **Previous Query** into the Patient Set box. The ontology terms (Concepts) used in the Previous Query will automatically populate the Concepts box.
2. Additional Concepts can be dropped into the same Concepts box by dragging them from the Navigate Terms panel or the Workplace panel to the left.
3. Change the default options by clicking on the Options box to suit your display needs.
4. Click on Create Table to start getting the Patients and the Concepts you entered.

**Patient Set:**

| Diabetes mellitus@12:09:29 [10-5-2015] [demo] | Size = 22093 |

**Concept(s):**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Count</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes mellitus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family history [Contains: diabetes]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGB A1C (LOINC:4548-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGB A1C (LOINC:4548-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGB A1C (LOINC:4548-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Options:**

- **Starting with the 1st Patient, retrieve 100 Patients**

  - **Table Size:** 100
  - **Max = 200, Min = 1**

- **Starting with the 1st Patient**

  - **Create Table**
Download table and select patients

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age</th>
<th>Race</th>
<th>Diabetic Mellitus Count</th>
<th>Family History (Contains: Diabetes Count)</th>
<th>HGB A1C (LOINC:4540-4) min</th>
<th>HGB A1C (LOINC:4540-4) max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000000003</td>
<td>Male</td>
<td>41</td>
<td>asian</td>
<td>12</td>
<td>54</td>
<td>4.1</td>
<td>6.9</td>
</tr>
<tr>
<td>1000000008</td>
<td>Male</td>
<td>23</td>
<td>hispanic</td>
<td>3</td>
<td>12</td>
<td>4.2</td>
<td>6.7</td>
</tr>
<tr>
<td>100000011</td>
<td>Female</td>
<td>60</td>
<td>white</td>
<td>21</td>
<td>21</td>
<td>4.6</td>
<td>7.1</td>
</tr>
<tr>
<td>100000013</td>
<td>Female</td>
<td>83</td>
<td>black</td>
<td>34</td>
<td>9</td>
<td>4.7</td>
<td>5.9</td>
</tr>
<tr>
<td>100000015</td>
<td>Male</td>
<td>32</td>
<td>hispanic</td>
<td>1</td>
<td>10</td>
<td>4.4</td>
<td>6.3</td>
</tr>
<tr>
<td>100000020</td>
<td>Male</td>
<td>25</td>
<td>hispanic</td>
<td>3</td>
<td>25</td>
<td>5.1</td>
<td>6.8</td>
</tr>
<tr>
<td>100000043</td>
<td>Male</td>
<td>61</td>
<td>hispanic</td>
<td>14</td>
<td>21</td>
<td>4.9</td>
<td>6.5</td>
</tr>
<tr>
<td>100000045</td>
<td>Female</td>
<td>89</td>
<td>asian</td>
<td>37</td>
<td>15</td>
<td>4.9</td>
<td>6.6</td>
</tr>
<tr>
<td>100000054</td>
<td>Female</td>
<td>66</td>
<td>asian</td>
<td>43</td>
<td>6</td>
<td>4.8</td>
<td>6.6</td>
</tr>
<tr>
<td>100000066</td>
<td>Female</td>
<td>22</td>
<td>black</td>
<td>32</td>
<td>4</td>
<td>5.2</td>
<td>7.0</td>
</tr>
<tr>
<td>100000070</td>
<td>Male</td>
<td>38</td>
<td>black</td>
<td>43</td>
<td>38</td>
<td>5.5</td>
<td>6.5</td>
</tr>
<tr>
<td>100000077</td>
<td>Male</td>
<td>13</td>
<td>black</td>
<td>12</td>
<td>8</td>
<td>4.4</td>
<td>6.7</td>
</tr>
<tr>
<td>100000080</td>
<td>Male</td>
<td>33</td>
<td>Indian</td>
<td>32</td>
<td>13</td>
<td>4.2</td>
<td>6.5</td>
</tr>
<tr>
<td>100000083</td>
<td>Female</td>
<td>65</td>
<td>black</td>
<td>33</td>
<td>4</td>
<td>4.7</td>
<td>6.6</td>
</tr>
<tr>
<td>100000087</td>
<td>Male</td>
<td>55</td>
<td>asian</td>
<td>36</td>
<td>6</td>
<td>4.8</td>
<td>6.9</td>
</tr>
<tr>
<td>100000096</td>
<td>Female</td>
<td>76</td>
<td>black</td>
<td>54</td>
<td>17</td>
<td>4.9</td>
<td>6.5</td>
</tr>
<tr>
<td>100000099</td>
<td>Male</td>
<td>36</td>
<td>black</td>
<td>32</td>
<td>0</td>
<td>5.2</td>
<td>6.6</td>
</tr>
<tr>
<td>100000108</td>
<td>Male</td>
<td>66</td>
<td>asian</td>
<td>43</td>
<td>1</td>
<td>5.3</td>
<td>6.7</td>
</tr>
<tr>
<td>100000109</td>
<td>Male</td>
<td>63</td>
<td>hispanic</td>
<td>31</td>
<td>4</td>
<td>4.9</td>
<td>5.9</td>
</tr>
<tr>
<td>100000112</td>
<td>Male</td>
<td>17</td>
<td>black</td>
<td>17</td>
<td>0</td>
<td>4.7</td>
<td>6.8</td>
</tr>
<tr>
<td>100000119</td>
<td>Female</td>
<td>72</td>
<td>white</td>
<td>43</td>
<td>32</td>
<td>4.4</td>
<td>5.9</td>
</tr>
<tr>
<td>100000133</td>
<td>Female</td>
<td>16</td>
<td>asian</td>
<td>20</td>
<td>25</td>
<td>4.9</td>
<td>6.6</td>
</tr>
</tbody>
</table>
Partners Biobank Portal – Request Genetic Data

Biobank Portal Genomics Request

The Biobank contains subjects who have consented to make their genomic data available for research. To request genomic data, please fill out the form below.

Once you have submitted this form, you will be contacted by the Biobank team to complete your request. Please note that no genomic data will be sent out without a discussion with the Biobank team.

For more assistance on making a genomic data request, please refer to the Biobank Wiki section How to Make Genomic Requests Using the Biobank Portal or contact us.

Step 1. Patient Selection:

Choose Your Patients:
- Enter a list of Biobank Subject IDs
- Select a Previous Query

Your Patient List:
Type or paste a list of Biobank Subject IDs, one per line. You can find Subject IDs by downloading the de-identified data set. Do not enter MRNs.

10021040
10022384
10018524
10018243
10018028
10020983

Step 2. Details of Request:

Request Type:
- De-Identified
- Identified

Contact Name:
Wattanasin, Nich

Contact E-mail:
nwattanasin@partners.org
Distributed Query System
Workflow at the sites to find patients for a clinical trial:

- After a query is run across the “SHRINE” network, the query is automatically saved at every site.

- The query saved at each site is transformed into a patient set.

- The patient set is studied and sorted for the specific patients eligible for the Clinical Trial.
Run Query Using SHRINE
I2b2 Workbench works directly on new database created from patient set Locally.

- At the Local site, one will:
  - View a patient list summary with matched criteria
  - Review a patient’s health data one at a time
  - Designate suitable patients
  - Export to Excel / CSV format
### Review de-identified Patients

#### i2b2 Workbench for i2b2 Demo (Oracle)

#### Select Patients | Select Concepts | Render Tables

<table>
<thead>
<tr>
<th>SMART</th>
<th>Patient ID</th>
<th>PSet #</th>
<th>Patient Name</th>
<th>Gender</th>
<th>Race</th>
<th>Date of Birth</th>
<th>Age</th>
<th>Obesity</th>
<th>Acute Myocardial Infarction</th>
<th>HMG-CoA reductase inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>100000002</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>white</td>
<td>1966-08-29T...</td>
<td>40</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000003</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>asian</td>
<td>1969-03-07T...</td>
<td>38</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000004</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>black</td>
<td>1976-08-13T...</td>
<td>30</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000005</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>hispanic</td>
<td>1973-06-25T...</td>
<td>33</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000006</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>hispanic</td>
<td>1981-08-05T...</td>
<td>25</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000007</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>asian</td>
<td>1981-07-10T...</td>
<td>25</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000008</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>hispanic</td>
<td>1986-12-16T...</td>
<td>20</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000009</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>hispanic</td>
<td>1989-01-11T...</td>
<td>18</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000010</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>hispanic</td>
<td>1976-11-23T...</td>
<td>30</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000011</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>white</td>
<td>2049-10-19T...</td>
<td>57</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000012</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>black</td>
<td>1991-03-29T...</td>
<td>16</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000013</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>black</td>
<td>2027-02-27T...</td>
<td>80</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000014</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>white</td>
<td>1956-05-22T...</td>
<td>50</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000015</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>M</td>
<td>hispanic</td>
<td>1978-09-11T...</td>
<td>28</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>0.1</td>
<td>100000016</td>
<td>1-1</td>
<td>xxxxx, xxxxx</td>
<td>F</td>
<td>indian</td>
<td>1983-11-19T...</td>
<td>23</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Review identified Patients
A “Criteria Matcher” App can help assess candidate patients using the clinical trial metadata.
Using Big Data to Find Similar People

To Help Predict Outcomes and Support Medical Decisions

To Learn About what Differences are Important for Predicting Disease

To Understand the Disease Traits Caused by a Gene Variant

To Help Interpret Features in Medical Images and Tissues
Deciding to Perform a Genetic Test
Inherited Heart Disease - Cardiomyopathy

Population Background Driven Analysis
MRN-00000 Mother Goose 1/1/1850

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All Repository Patients</th>
<th>Test Positive Patients</th>
<th>Proband Values</th>
<th>Similar Patient Population</th>
<th>Results</th>
<th>Pretest Probability Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV Wall Thickness</td>
<td><img src="chart.png" alt="Chart" /></td>
<td>11 mm</td>
<td></td>
<td></td>
<td>&lt; 11 mm</td>
<td>x</td>
</tr>
<tr>
<td>LV Dimensions</td>
<td><img src="chart.png" alt="Chart" /></td>
<td>42 mm</td>
<td></td>
<td></td>
<td>&gt; 15 mm</td>
<td>-</td>
</tr>
<tr>
<td>GLA Activity</td>
<td><img src="chart.png" alt="Chart" /></td>
<td></td>
<td></td>
<td></td>
<td>&gt; 20% Activity</td>
<td>- 6%</td>
</tr>
<tr>
<td>Latest EKG Results</td>
<td><img src="chart.png" alt="Chart" /></td>
<td>Indeterminant</td>
<td></td>
<td></td>
<td>&lt; 5% Activity</td>
<td>+ 70%</td>
</tr>
</tbody>
</table>

Hypertrophic Cardiomyopathy (HCM) Panel
Calculated Positive Pretest Probability: 11%
Health Innovation Platform Architecture
Treating IHD Patient with Precision Medicine

Cardiomyopathy Patient

Clinical Data Collected

Targeted Therapy

Biopsy

Routine Pathology

Molecular Pathology

Partners Big Data Commons
SMART Apps link i2b2 to the EMR

- Substitutable Medical Application and Reusable Technology – Commissioned form the Office of the National Coordinator
- Allows Big Data from i2b2 to integrate Clinical “Apps” into the Epic EMR.
- Paradigm is similar to Mobile Apps with a proposed standard interface using FHIR sponsored by the National Argonaut EMR Project.
New network storage in the cloud

On-Premise

- Archive
- Backup
- Files

CloudArray
Virtual or Physical Appliance

- Volume-Based Dynamic Cache
- In-Cloud Snapshots
- CloudArray Recovery
- Encryption
- Bandwidth Scheduler
- Data Reduction

iSCSI / NAS

Cloud API
delivered as a Service

Self-Service in 2016
Presenting the researcher desktop

Robust desktop capabilities for the Partners researcher

Analysis Pipeline

Public Data Sets

Database Schemas

ML Cluster Analysis

Image Analysis
Enabling Innovation to reach into EMR

SMART Apps run in EMR

Got Statins?

BP Centiles

Big Data Analytics performed with High Performance Computing

Partners Big Data Commons

SMART-Enabled DW - I2b2/RPDR

SMART-Enabled Document/Pivotal

SMART-Enabled EMR - Epic
Tribute to…

- **RPDR/I2b2 Core Team**
  - Shawn Murphy
  - Christopher Herrick
  - Mariah Mitchell
  - Vivian Gainer
  - Alyssa Goodson
  - Martin Rees
  - Charles Wang
  - Laurie Bogosian
  - Stacey Duey
  - Andrew Cagan
  - Michael Mendis
  - Lori Phillips
  - Janice Donahoe
  - Nich Wattanasin
  - David Wang

- **Biobank Team**
  - Natalie Boutin
  - Scott Weiss
  - Vivian Gainer

- **HPC and Cloud Team**
  - Brent Richter
  - Jon Jackson
  - Alan Harris

- **Genomics Innovation Team**
  - Sandy Aronson
  - Heidi Rehm
  - Calum MacRea
I2b2 and SMART Information and Software on the Web

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